Federal Reserve Bank of Minneapolis
Research Department Staff Report 101

December 1985

A METHOD FOR ESTIMATING THE TIMING INTERVAL IN A LINEAR ECONOMETRIC MODEL, WITH AN APPLICATION TO TAYLOR'S MODEL OF STAGGERED CONTRACTS

Lawrence J. Christiano*

Federal Reserve Bank of Minneapolis

Abstract

This paper describes and implements a procedure for estimating the timing interval in any linear econometric model. The procedure is applied to Taylor's model of staggered contracts using annual averaged price and output data. The fit of the version of Taylor's model with serially uncorrelated disturbances improves as the timing interval of the model is reduced.

*This paper is forthcoming in the Journal of Economic Dynamics and Control. Earlier versions were presented to the Money and Banking Workshop at the University of Chicago and the summer 1984 North American meetings of the Econometric Society. I have benefited from conversations with Robert Lucas, Michael Mussa, Edward Prescott, Laurence Weiss, and Arnold Zellner. I am particularly grateful for the comments and suggestions of Lars Hansen, John Huizinga, and Boyan Jovanovic. The detailed comments of a referee have substantially improved the exposition. I am also grateful to Kathleen Rolfe for her tireless, invaluable editorial assistance.

1. Introduction

This paper describes and implements a procedure for estimating the timing interval in any linear econometric model. This section begins by explaining why such a procedure may be of interest. The second part of this section explains why Taylor's (1980a,b) model was chosen for the application. An outline of the paper appears in the third part of this section.

1.1. Why treat the timing interval as unknown?

Why an econometric procedure which treats a model's timing interval as a free parameter is of interest may not be obvious. To explain, I need some simplifying notation. Let N index the length of the timing interval in a theoretical model, with N = 1 indicating a model in which the timing interval coincides with the data sampling interval; N = 2, a model in which the timing interval is half the data sampling interval; and so on. In a fully specified model, N is an endogenous variable, chosen by the agents whose decisions are being modeled. Explicitly modeling the endogeneity of N would produce a relation of the form $N = f(\phi)$, where f is some function and ϕ is the vector of free parameters of the model. [In the general case--not considered here--N varies with time and the values of the state variables. See Barro (1970) and Garber (1977) for examples.] This paper supplies a method for estimating N without restrictions. The method is useful whether or not f is available.

The function f is not available in most empirical applications. Without an f function, theory provides no guidance about the value of N, so that it is an unknown parameter (subject only to $N > 0$). One way of proceeding in this context is to treat N like other parameters whose values are unknown and estimate it. An advantage of this approach is that when a cross section of time series is available, unrestricted estimates of N and ϕ can be used to yield information about the properties of f which would be a useful

guide when theorizing about f. To see this, suppose that time series i yields estimates $\hat{\phi}_i$ and $\hat{N}_i$ and that there are H time series (i = 1, ..., H). The assumption that $N_i = f(\phi_i)$ for i = 1, ..., H indicates the nature of f even when the theory says nothing about it. The procedure in this paper is also useful as a diagnostic tool when an econometric estimate of N is otherwise not needed. If the standard route of fixing N = 1 is taken, my procedure can be used to check whether estimation results are sensitive to that specification.

When f is available, restricted estimates of N and ϕ--denoted $\hat{N}^r$, $\hat{\phi}^r$--can be obtained by imposing the restriction $\hat{N}^r = f(\hat{\phi}^r)$ during estimation. A useful test of the model is to then compare $\hat{N}^r$ with $\hat{N}^u$, an estimate of N obtained without imposing $N = f(\phi)$. The computation of $\hat{N}^u$ requires the procedure outlined in this paper. (Much of my discussion about identification and the calculation of the likelihood function would also be relevant for the problem of computing $\hat{N}^r$.)

There is no basis for the nearly universal practice of setting N = 1 and not checking the results for sensitivity to N. Such a basis could exist only if permitting N to be a free parameter were computationally intractable or if hypothesis tests and estimates of the remaining parameters were known in general to be insensitive to the choice of N. Both of these cases may be ruled out. This paper demonstrates concretely that allowing N to be free is computationally tractable. The paper also adds an example to the work which has already shown, both theoretically and empirically, that empirical results can be sensitive to the choice of model timing interval [Christiano (1984,1985), Mundlak (1961), Zellner (1968), and Zellner and Montmarquette (1971)].

1.2. Why study the staggered contracts model?

I apply my procedure to Taylor's staggered contracts model because it is ideal for illustrating how aggregation over time affects model estima-

tion and fit as N increases. The simple (benchmark) version of Taylor's model, in which error terms are white noise, predicts one-way Granger causality from the price level to output. However, the work of Sims (1971) and Geweke (1978) has shown that the sampled representation of a one-way Granger-causal system with $N > 1$ will in general display bidirectional Granger causality. This and other considerations mentioned below suggest that raising N above the value of 1--which is what Taylor sets it to--is likely to improve model fit. Taylor's strategy for improving the model's fit is to introduce serial correlation in the error terms.

In Taylor's model, the degree of overlap in wage contracts determines the timing of the model, so that N simultaneously indexes the timing interval and the degree of contract overlap. Besides freeing up N, I vary the contract length, n, which Taylor (1980a) fixes a priori at a value of 2. (The unit of measurement for n is the sampling interval, which is one year.) Thus, I attempt to improve the benchmark model and its fit by freeing up two parameters, n and N, held fixed by Taylor (1980a). This is an attractive strategy because the parameters n and N lend themselves more easily to economic interpretation than do Taylor's (1980a) serial correlation parameters. For instance, freeing up n and N and repeating the cross-country analysis of Taylor (1980a) would permit a study of the empirical relationships among the contract length, the degree of contract overlap, and the parameters of government policy. The policy implications of Taylor's model depend critically on the contract length and degree of overlap not varying with policy. Under the invariance assumption, the model implies a policy tradeoff between output and price stability, which sets the model off sharply from the rational expectations models of Barro (1976), Lucas (1973), and Sargent and Wallace (1975).

1.3. Outline

The paper has five more sections. Section 2 presents Taylor's model. Sections 3 and 4 describe the proposed methodology for estimating the timing interval of an arbitrary linear econometric model. Section 3 describes how to compute the likelihood function when $N > 1$ and the data are averages over the sampling interval. Section 4 discusses the aliasing identification problem that arises when estimating ϕ with $N > 1$. Section 5 applies the methodology to Taylor's model. Section 6 summarizes the conclusions.

2. Taylor's model

In Taylor (1980a), a model of aggregate output and price dynamics is estimated using annual averaged data for the economies of the United States, Canada, and eight European countries. The model specifies that half the population of workers negotiate two-year wage contracts each year. This model, once closed by an aggregate demand equation, restricts the vector time series representation of annual aggregate output and price data.

The timing interval in Taylor's model is determined by the length of the period separating contract negotiations. A natural way to modify this timing is to preserve the specification that two-year wage contracts are negotiated, but to have contracts come up for renewal more often. For example, a fourth of the population could negotiate two-year wage contracts every half year or an eighth could negotiate every quarter. The basic economics underlying these specifications are the same. However, the specifications have different implications for the time series representation of the annual averaged data; some may fit the data better than others. In addition to permitting the model timing interval to vary, I consider contract periods longer than two years.

A formal description of Taylor's model follows. I adopt the normalization that the length of the sampling interval (the period) is 1. (In the application, the sampling interval is one year.)

To begin, define

$n \equiv$ the number of periods covered by a contract (an integer greater than 1)

$N \equiv$ a positive integer such that n-period wage contracts are negotiated every $1/N$ periods by $1/nN$ of the population of workers

$x(t) \equiv$ the log of the wage contract negotiated at the beginning of period t to expire at the end of period $t + n - 1/N$, in deviation from trend

$p(t) \equiv$ the log of the aggregate price level in the interval $[t, t + 1/N)$, in deviation from trend

$y(t) \equiv$ the log of aggregate output in the interval $[t, t + 1/N)$, in deviation from trend

$\varepsilon(t) \equiv$ a shock to $x(t)$

$v(t) \equiv$ an aggregate demand shock.

The t index ranges over $0, \pm 1/N, \pm 2/N, \pm 3/N, \ldots$. Taylor (1980a) sets $n = 2$, $N = 1$.

Workers negotiating wage contracts at the beginning of period t are assumed to be interested in two factors: their wage relative to other wages and the current and prospective levels of output relative to trend. In particular, the wage-setting process is modeled as follows:

$$x(t) = \sum_{s=1}^{nN-1} b_s \hat{x}(t + \tfrac{s}{N}) + \sum_{s=1}^{nN-1} b_s x(t - \tfrac{s}{N}) + \frac{\gamma}{nN} \sum_{s=0}^{nN-1} \hat{y}(t + \tfrac{s}{N}) + \varepsilon(t) \qquad (1)$$

where a caret (^) over a variable indicates its expectation, conditioned on information dated t - 1/N and earlier, and

$$b_s = \frac{nN - s}{nN(nN-1)} \qquad s = 1, 2, \ldots, nN - 1 \qquad (2)$$

so that $\sum_{s=1}^{nN-1} b_s = 1/2$. That is, b_s is the fraction of a period over which an n-year wage contract negotiated at the beginning of t + s/N or t - s/N overlaps with the contract negotiated at the beginning of period t.

Fig. 1 may help visualize the dating in the model. It depicts the time covered by two adjacent n-period contracts. The upper bar indicates the interval over which an n-period contract negotiated at time t applies. The lower bar indicates the interval over which the next n-period contract, negotiated at time t + 1/N, applies.

The remaining equations of the model are[1/]

$$y(t) = -\beta \hat{p}(t) + v(t) \qquad 0 < \beta < 1 \qquad \text{(aggregate demand)} \qquad (3)$$

$$p(t) = \frac{1}{nN} \sum_{i=0}^{nN-1} x(t - \tfrac{i}{N}) \qquad \text{(price markup)} \qquad (4)$$

$$v(t) = \eta(t) + \theta\eta(t - \tfrac{1}{N}) \qquad |\theta| < 1 \qquad \text{(demand shock)} \qquad (5)$$

$$\varepsilon(t) = u(t) + \delta u(t - \tfrac{1}{N}) \qquad |\delta| < 1 \qquad \text{(wage shock)} \qquad (6)$$

where $\{\eta(j); j=0, \pm 1/N, \pm 2/N, \ldots\}$ and $\{u(j); j=0, \pm 1/N, \pm 2/N, \ldots\}$ are serially independent, normally distributed random variables with

$$E\left(\begin{matrix} \eta(j) \\ u(j) \end{matrix} \right) \left(\eta(j + \tfrac{\tau}{N}) u(j + \tfrac{\tau}{N}) \right) = \begin{cases} V, & \tau = 0 \\ \underset{\sim}{0}, & \tau \neq 0. \end{cases} \qquad (7)$$

The arguments in Taylor (1980a,b) can be used to show that the reduced form $Y(t) = (y(t), p(t))$ process corresponding to (1)-(7) is

$$A(B)Y(t) = C(B)w(t) \tag{8}$$

where $B \equiv L^{1/N}$ (where L is the unit interval time shift operator; that is, $Lz_t = z_{t-1}$), A(B) is a scalar polynomial in B:

$$A(B) = 1 - a_1 B - \ldots - a_{(nN-1)} B^{nN-1} \tag{9}$$

and C(B) is a 2 x 2 polynomial matrix:

$$C(B) = I + C_1 B + \ldots + C_{nN} B^{nN}$$

$$= \begin{bmatrix} C_{11}(B) & C_{12}(B) \\ C_{21}(B) & C_{22}(B) \end{bmatrix} . \tag{10}$$

The object A(B) is the unique polynomial with zeros outside the unit circle and lead term equal to unity that satisfies

$$\lambda A(B)A(B^{-1}) = b(B) \tag{11}$$

where

$$b(B) = \sum_{s=-(nN-1)}^{nN-1} b_s B^s \tag{12a}$$

$$b_0 \equiv - \frac{nN + \gamma\beta}{nN - \gamma\beta(nN-1)} \tag{12b}$$

$$b_s = b_{-s} \tag{12c}$$

and b_s for positive values of s is defined in (2). Also, in (11),

$$\lambda = \frac{-b_{(nN-1)}}{a_{(nN-1)}} . \tag{13}$$

In (10),

$$C_{11}(B) = -\beta C_{21}(B) + A(B)(1+\theta B) \tag{14a}$$

$$C_{21}(B) = \frac{k_1}{nN} \sum_{s=1}^{nN} B^s \tag{14b}$$

$$C_{12}(B) = \beta[A(B) - C_{22}(B)] \tag{14c}$$

$$C_{22}(B) = (1+k_2 B) \sum_{s=0}^{nN-1} B^s. \tag{14d}$$

The disturbance in (8) is

$$w(t) = \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix} = \begin{pmatrix} \eta(t) \\ \frac{1}{nN} u(t) \end{pmatrix}. \tag{15}$$

Finally,

$$k_1 = \frac{-1}{\lambda}\left[\frac{\theta\gamma}{nN - \gamma\beta(nN-1)}\right] \tag{16a}$$

$$k_2 = \frac{-1}{\lambda}\left[\frac{nN\delta}{nN - \gamma\beta(nN-1)}\right]. \tag{16b}$$

Using (14a)-(14d),

$$\det C(B) = A(B)\left(\frac{1-B^{nN}}{1-B}\right)\left[1 + \left(k_2 + \theta - \frac{\beta k_1}{nN}\right)B + \theta k_2 B^2\right]. \tag{17}$$

Evidently the zeros of det C(B) are those of the three polynomials to the right of (17). The roots of A(B) lie outside the unit circle. The second polynomial to the right of the equality in (17) has nN - 1 roots lying on the unit circle. These are

$$\cos\left(\frac{2\pi}{nN} j\right) + i \sin\left(\frac{2\pi}{nN} j\right) \qquad j = 1, 2, \ldots, nN - 1$$

where i is the imaginary number. The location of the zeros of the third polynomial to the right of (17) is uncertain. For example, when N = 2, n = 2,

$\theta = .8$, $\delta = -.2$, $\beta = .5$, and $\gamma = .2$, the zeros of this polynomial are 4.309 and -1.257. When $\delta = .9$, the roots are -1.217 and -.989.

To summarize, the structure in (1)-(7) implies the covariance stationary representation (8), with a moving average (MA) part whose determinant has $nN - 1$ roots on the unit circle and possibly some inside as well. (The condition $\theta = \delta = 0$ guarantees that no roots lie inside the unit circle.)

Note from (8) that the theory restricts the time series representation of $Y(t) = (y(t),p(t))$ sampled every $1/N$ of a period. The data I want to analyze, however, are period averages sampled once a period. That is, I have no observations on $\{Y(t);t=0,\pm 1/N,\pm 2/N,\ldots\}$. Instead, I have observations on $\{\overline{Y}(t);t=0,\pm 1,\pm 2,\ldots\}$, where

$$\overline{Y}(t) = \frac{1}{N}\sum_{i=0}^{N-1} Y(t + \frac{i}{N}).$$

For example, when $N = 12$, I interpret the 1960 observation of the gross national product as the geometric average of the (unavailable) 12 monthly observations. In section 3, I show how to calculate the likelihood function conditional on this feature of the data and for a given set of values for the structural parameters. These are β, γ, δ, θ, and the three elements of V in (7). For convenience, stack these in the vector ϕ, where

$$\phi = (\beta, \gamma, \delta, \theta, V_{11}, V_{22}, V_{12}). \tag{18}$$

The vector ϕ is an element of Φ, which is the subset of R^7 such that $0 < \beta < 1$, $\gamma > 0$, $|\delta| < 1$, $|\theta| < 1$, $V_{11} > 0$, $V_{22} > 0$, $V_{11}V_{22} - V_{12}^2 > 0$, and the factorization (11) exists.

Section 4 is concerned with the problem of identifying ϕ for given values of n and N. Readers less interested in the technical details and eager for empirical results can skip sections 3 and 4.

3. The likelihood function

Here I describe a method for computing the exact log-likelihood of a set of data, $\overline{Y}(1), \ldots, \overline{Y}(T)$, given values for the structural parameters, ϕ, and for n and N. The strategy is related to that proposed by Hansen and Sargent (1980) for estimating the parameters of continuous time systems from sampled data. An alternative strategy for computing the likelihood function in this context uses the Kalman filter [Ansley and Kohn (1983)].

If the exact normal likelihood function is to be evaluated, then

$$R_{\overline{y}}(\tau,\phi) \equiv E\overline{Y}(t)\overline{Y}(t-\tau)^T \qquad \tau = 0, 1, \ldots, T-1 \tag{19}$$

is required. These are computed by first obtaining $R_y(\tau,\phi) \equiv EY(t)Y(t-\tau)^T$ from (8). Denote the covariance generating function of $\{Y(t);t=0,\pm 1/N,\pm 2/N, \ldots\}$ by

$$S_y(s) = \sum_{i=-\infty}^{+\infty} R_y(\tfrac{i}{N},\phi)s^i \tag{20}$$

where s is the complex variable defined over a region which includes the unit circle. It is a well known result in the analysis of covariance stationary time series that

$$S_y(s) = \frac{C(s)VC(s^{-1})^T}{A(s)A(s^{-1})}. \tag{21}$$

R_y may be found by expanding (21) and matching the result with (20).

Write

$$A(s) = (1-\rho_1 s)(1-\rho_2 s)\ldots(1-\rho_m s) \tag{22}$$

where $m \equiv nN - 1 > 0$. Comparison with (9) indicates that ρ_i for $i = 1, \ldots, m$ are the reciprocals of the roots of the polynomial defined there. Also, $|\rho_i| < 1$ for all i. Note from (10) that the order of C(s) is m + 1. The partial fractions expansion of (21) (after some algebra) is

$$S_y(s) = \sum_{j=1}^{m} \frac{W_j}{1 - \rho_j s} + \sum_{j=1}^{m} \frac{W_j^T}{1 - \rho_j s^{-1}} + Ks + K^T s^{-1} + \tilde{K} - \sum_{j=1}^{m} W_j^T \qquad (23a)$$

where

$$K = \frac{-1}{a_m} C_{m+1} V \qquad (23b)$$

$$\tilde{K} = -\left[C_{m+1} V C_1^T + C_m V - K(a_m a_1 - a_{m-1})\right]\frac{1}{a_m} \qquad a_0 \equiv -1 \qquad (23c)$$

$$W_j = \frac{C(\rho_j^{-1}) V C(\rho_j)^T}{(1-\rho_j^2) \prod_{\substack{i=1 \\ i \neq j}}^{m} [1 + \rho_i^2 - \rho_i(\rho_j + \rho_j^{-1})]}. \qquad (23d)$$

It may be verified that if

$$R_y(0,\phi) = \tilde{K} + \sum_{j=1}^{m} W_j \qquad (24a)$$

$$R_y(\tfrac{1}{N},\phi) = K + \sum_{j=1}^{m} W_j \rho_j \qquad (24b)$$

$$R_y(\tau,\phi) = \sum_{j=1}^{m} W_j \rho_j^{N\tau} \qquad \tau = \tfrac{2}{N}, \tfrac{3}{N}, \ldots \qquad (24c)$$

$$R_y(\tau,\phi) = R_y^T(-\tau,\phi) \qquad \tau < 0 \qquad (24d)$$

then (20) and (23a) are satisfied so that (24a)-(24d) define the covariance function corresponding to the representation in (8).[2/]

To get eq. (19), note that

$$R_{\underline{y}}(\tau,\phi) \equiv E\overline{Y}(t)\overline{Y}(t-\tau)^T$$

$$\equiv E\left[\frac{1}{N} \sum_{j=0}^{N-1} Y(t + \tfrac{j}{N})\right]\left[\frac{1}{N} \sum_{k=0}^{N-1} Y(t + \tfrac{k}{N} - \tau)^T\right]$$

$$= \frac{1}{N^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} R_y(\tfrac{j-k}{N} + \tau). \qquad (25)$$

Substituting (24a)-(24d) into (25) gives (after some algebra)

$$R_{\underline{y}}(0,\phi) = \frac{1}{N}\tilde{K} + \frac{N-1}{N^2}(K+K^T) + \sum_{k=1}^{m}[(1-\rho_k^N)(1-\rho_k^{-N})]^{-1}\{\tilde{W}_k[1-\rho_k^N+N(1-\rho_k^{-1})] + \tilde{W}_k^T[1-\rho_k^N-N(1-\rho_k)]\} \quad (26a)$$

$$R_{\underline{y}}(1,\phi) = \frac{1}{N^2}K + \sum_{k=1}^{m}\tilde{W}_k\rho_k^N \quad (= \sum_{k=1}^{m}W_k\rho_k \text{ if } N=1) \quad (26b)$$

$$R_{\underline{y}}(\tau,\phi) = \sum_{k=1}^{m}\tilde{W}_k\rho_k^{N\tau} \qquad \tau = 2, 3, 4, \ldots. \quad (26c)$$

Here

$$\tilde{W}_k = W_k\frac{(1-\rho_k^N)(1-\rho_k^{-N})}{N^2(1-\rho_k)(1-\rho_k^{-1})} \qquad k = 1, 2, \ldots, m. \quad (26d)$$

Define the following block-Toeplitz matrix:

$$\Omega(\phi) = \begin{bmatrix} R_{\underline{y}}(0,\phi) & R_{\underline{y}}^T(1,\phi) & \cdots & R_{\underline{y}}^T(T-1,\phi) \\ R_{\underline{y}}(1,\phi) & R_{\underline{y}}(0,\phi) & \cdots & R_{\underline{y}}^T(T-2,\phi) \\ \vdots & \vdots & & \vdots \\ R_{\underline{y}}(T-1,\phi) & R_{\underline{y}}(T-2,\phi) & \cdots & R_{\underline{y}}(0,\phi) \end{bmatrix}. \quad (27)$$

Let

$$\overline{Y} = \begin{pmatrix} \overline{Y}(1) \\ \vdots \\ \overline{Y}(T) \end{pmatrix}.$$

Then the log-likelihood of $\overline{Y}(1), \ldots, \overline{Y}(T)$ for a given value of $\phi \in \Phi$, $n \geq 1$, and $N \geq 1$ is

$$\mathfrak{L}(\overline{Y};\phi,n,N) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log\det\Omega(\phi) - \frac{1}{2}\overline{Y}^{T}\Omega^{-1}(\phi)\overline{Y}. \qquad (28)$$

Dunsmuir and Hannan (1976) show that maximizing (28) with respect to ϕ delivers a consistent estimator of ϕ when n and N are known. Dunsmuir and Hannan's consistency results apply here since they permit unit roots in the determinant of the MA representation. (Their results assume identification, a topic discussed in section 4.) Dunsmuir and Hannan's asymptotic distribution results do not apply here, since for these results they rule out unit MA roots. The results of Pham-Dinh (1978) suggest that my maximum likelihood estimator of ϕ is asymptotically normally distributed because I impose the unit MA roots exactly during estimation. Pham-Dinh proves the asymptotic normality of the maximum likelihood estimator in the scalar ARMA case when the MA has a unit root and the restriction is imposed exactly during estimation. The results of Sargan and Bhargava (1983) suggest that the assumption that the unit root is imposed exactly is critical. They show that the maximum likelihood estimator of the MA coefficient of a scalar, zero mean, MA(1) representation is not asymptotically normally distributed if the MA has a root at plus or minus unity.

To conclude this section, my estimator for n, N, ϕ, denoted $\hat{n}$, $\hat{N}$, $\hat{\phi}$, is defined by

$$\hat{n}, \hat{N}, \hat{\phi} = \underset{n\geq 2, N\geq 1, \phi\in\Phi}{\operatorname{argmax}} \mathfrak{L}(\overline{Y};\phi,n,N).$$

When n and N are fixed and known a priori, the problem of estimating ϕ closely matches the standard estimation problem in time series, and much is known about the asymptotic distribution properties of $\hat{\phi}$ (as summarized above). However, in this paper, n and particularly N are treated as unknown and to be estimated. The sampling theory for this case is left for future research. Despite this, when reporting the results in section 5, I use the distribution

theory appropriate when n and N are known. In particular, for given n and N, I estimate the variance-covariance matrix of $\hat{\phi}$, cov $(\hat{\phi})$, by this formula:

$$\hat{\text{cov}}\ (\hat{\phi}) = -\left[\left.\frac{\partial^2 f(\overline{Y};\phi,n,N)}{\partial\phi\partial\phi^T}\right|_{\phi=\hat{\phi}}\right]^{-1}. \tag{29}$$

Under the normality assumption on w(t), and assuming n and N are fixed and known, $\hat{\text{cov}}$ converges strongly to cov.

4. The aliasing identification problem

Estimating a model with a timing interval finer than the data sampling interval (with N > 1) involves an identification problem with regard to ϕ that is not present in standard applications (when the two intervals are assumed to coincide). Section 4.1 describes this aliasing identification problem and compares it with the standard, or classical, identification problem. Four examples illustrate these concepts and motivate the more rigorous material in sections 4.2 and 4.3.

Section 4.2 presents a set of sufficient conditions that guarantee aliasing identification when the data-generating mechanism is (8) and the data are temporally averaged (Theorem 1). One of the examples in section 4.1 and a new example show that these conditions are not necessary for aliasing identification, but they suggest that a set of necessary and sufficient conditions for as general a class of models as (8) will be hard to work out. [This parallels the findings of Hansen and Sargent (1983) in the continuous time context.] The virtue of my Theorem 1 is that its conditions are relatively easy to check, either analytically or computationally. For example, I show analytically (in Proposition 2) that Taylor's model with n = 2 and N = 2 satisfies the conditions. Moreover, I describe a computational algorithm for checking the conditions in the more general case, when $n \geq 2$ and $N > 2$.

Section 4.3 describes sufficient conditions for aliasing identification when the available data are sampled point-in-time (Theorem 2). Some remarks on the role of data averaging in aiding aliasing identification are offered there.

4.1. The problem defined

There is a parameter vector ϕ, belonging to a space of admissible points Φ, which gives rise to polynomials $A(\cdot;\phi)$ and $C(\cdot;\phi)$ and to $V(\phi)$, where

$$A(B;\phi)Y(t) = C(B;\phi)w(t) \tag{30}$$

$$Ew(t)w(t - \tfrac{\tau}{N})^T = \begin{cases} V(\phi), & \tau = 0 \\ \underset{\sim}{0}, & \tau \neq 0 \end{cases}$$

and $w(t)$ is defined in (15). Here $B \equiv L^{1/N}$, L is the unit shift operator, A is an m^{th} order scalar polynomial in B, and C is an $(m+1)^{th}$ order matrix polynomial in B. Throughout this section, I assume $m > 0$ and $N > 1$. Eq. (30) is (8) reproduced here for convenience, using a notation that emphasizes the dependence of A, C, and V on $\phi \in \Phi$, the structural parameters.

Standard asymptotic consistency and distribution results require that the structural parameters of the model be identified at the true parameter values [Dunsmuir and Hannan (1976), Kohn (1979)]. When observations on $\{Y(t);t=0,\pm1/N,\pm2/N,\ldots\}$ are available, the definition of identification is standard:

Classical identification condition. Suppose (i) $\phi_0 \in \Phi$ and (ii) for any $\phi' \in \Phi$, $\phi_0 \neq \phi'$ a value of $\tau \in \{0,\pm1/N,\pm2/N,\ldots\}$ exists such that $R_y(\tau,\phi_0) \neq R_y(\tau,\phi')$. Then the model is identified in the classical sense at ϕ_0.

When only observations on $\{\bar{Y}(t);t=0,\pm1,\pm2,\ldots\}$ (or $\{Y(t);t=0,\pm1,\pm2,\ldots\}$) are available, the classical identification condition is not sufficient to guarantee identification; a stronger condition is required:

Aliasing identification condition. The model is identified in the aliasing sense at $\phi_0 \in \Phi$ if, for any $\phi' \in \Phi$, $\phi_0 \neq \phi'$,

(i) $R_{\bar{y}}(\tau,\phi_0) \neq R_{\bar{y}}(\tau,\phi')$ for some $\tau \in \{0,1,2,\ldots\}$ when observations on $\{\bar{Y}(t);t=0,\pm1,\pm2,\ldots\}$ are available.

(ii) $R_y(\tau,\phi_0) \neq R_y(\tau,\phi')$ for some $\tau \in \{0,1,2,\ldots\}$ when observations on $\{Y(t);t=0,\pm1,\pm2,\ldots\}$ are available.

To establish the aliasing identification condition when the available data are averages, the mapping from $\phi_0 \in \Phi$ to $\{R_{\bar{y}}(\tau,\phi_0);\tau=0,1,2,\ldots\}$ must be shown to have a unique inverse in Φ. If the available data are point-in-time, then the mapping from $\phi_0 \in \Phi$ to $\{R_y(\tau,\phi_0);\tau=0,1,2,\ldots\}$ must be shown to have a unique inverse in Φ.

The following four examples illustrate the above definitions and motivate the material in sections 4.2 and 4.3. In Example 1, classical identification obtains almost everywhere in Φ, but aliasing identification does not. Examples 2 and 3 modify Example 1 so that aliasing identification does obtain. Example 2 illustrates how the restrictions of economic theory can overcome the aliasing identification problem, while Example 3 suggests the problem may be less severe when the available data are averages. In Examples 1-3, the underlying data-generating mechanism is assumed to be AR(1). Example 4 shows that this assumption alone substantially reduces the dimension of the aliasing identification problem by drastically restricting the between-sample covariances. The result is that relatively little is left to be accomplished by the restrictions of economic theory or averaging. [This point, which is very general, is basically the one made by Hansen and Sargent (1983).]

Example 1: Sampling point-in-time from a scalar AR(1). Suppose $N = 2$, $m = 1$, all variables are scalars, $A(B;\phi) = 1 - \rho B$, $|\rho| < 1$, $C(B;\phi) = 1$, and $V(\phi) = \sigma^2 > 0$. Here $\phi = (\rho,\sigma^2)$. Evidently $\Phi = \{x=(x_1,x_2)\in R^2: |x_1|<1, x_2>0\}$, $W = \sigma^2/(1-\rho^2)$, and $K = \tilde{K} = 0$. The covariance function for this process is, by (24a)-(24d), $R_y(\tau,\phi) = W\rho^{2|\tau|}$ for $\tau = 0, \pm 1/2, \pm 1, \ldots$. While the classical identification condition is satisfied for all $\phi \in \Phi$, aliasing identification does not obtain if only sampled point-in-time observations on $Y(t)$ are available and $\rho \neq 0$. In particular, given $\{R_y(\tau,\phi);\tau=0,\pm 1/2,\pm 1,\ldots\}$ for some $\phi = (\rho,\sigma^2) \in \Phi$, there is always another element $\phi' = (-\rho,\sigma^2) \in \Phi$ that implies the same covariance function when sampled at the integers. That is, while $R_y(\tau,\phi) \neq R_y(\tau,\phi')$ for $\tau = \pm 1/2, \pm 3/2, \ldots$, nevertheless, $R_y(\tau,\phi) = R_y(\tau,\phi')$ for $\tau = 0, \pm 1, \pm 2, \pm 3, \ldots$. The problem is that, although ρ^2 can be deduced from $\{R_y(\tau,\phi);\tau=0,1,2,\ldots\}$, the equation $\rho = (\rho^2)^{1/2}$ has two solutions. (In section 4.2, this set of solutions is denoted $D(\phi) = \{\rho,-\rho\}$.)

Example 2: Potential for economic theory to guarantee aliasing identification. One way to modify the setup in Example 1 so that aliasing identification obtains globally is to use economic theory to restrict things so that only one element in $D(\phi)$ is admissible, that is, corresponds to a $\phi \in \Phi$. For example, this would be true if theory required that ρ be nonnegative; that is, $\Phi = \{x=(x_1,x_2)\in R^2: 0\leq x_1<1, x_2>0\}$. (Section 4.2 highlights this strategy.)

Example 3: Sampling averages from a scalar AR(1). Another way to modify the setup in Example 1 so that aliasing identification obtains globally is to assume that the available data are not point-in-time but rather temporally averaged. Then, although both elements in $D(\phi)$ are admissible, only one may be consistent with the sampled covariance function. To see this, note that

(26a)-(26d) imply that $R_{\bar{y}}(0,\phi) = 2\tilde{W}/(1+\rho^{-1})$ and $R_{\bar{y}}(\tau,\phi) = \tilde{W}\rho^{2\tau}$ for $\tau = 1, 2, \ldots$, where $\tilde{W}$ is defined in (26d) and $\bar{Y}(t) = (1/2)[Y(t) + Y(t+(1/2))]$. Then

$$\tilde{W} = \frac{[R_{\bar{y}}(1,\phi)]^2}{R_{\bar{y}}(2,\phi)}$$

$$\rho = \frac{R_{\bar{y}}(0,\phi)}{2\tilde{W} - R_{\bar{y}}(0,\phi)}.$$

Here both elements in $D(\phi)$ are consistent with $R_{\bar{y}}(\tau,\phi)$ for $\tau = 1, 2, 3, \ldots$. But one of those elements is inconsistent with $R_{\bar{y}}(0,\phi)$. (Some further comments on the role of averaging appear in section 4.3.)

Example 4: Role of AR(1) assumption in determining dimension of aliasing problem in Example 1. The aliasing identification problem in Example 1 is of small dimension in that it requires discrimination between only two observationally distinct models (that is, $\{\rho,\sigma^2\}$ and $\{-\rho,\sigma^2\}$). Thus, relatively small changes in the setup produce examples in which aliasing identification does occur. (See Examples 2 and 3.) This small dimension is due to the considerable structure placed on the between-sample covariances by the AR(1) assumption on $\{Y(t);t=0,\pm1/2,\pm1,\ldots\}$. To see this, note that $R_y(\tau) = W\rho^{2|\tau|}$ for $\tau = 0, \pm1, \pm2, \ldots$ is consistent with any half-interval covariance function from the following uncountably infinite set:

$$R_y(\tau) = \begin{cases} W\rho^{2|\tau|}, & \tau = 0, \pm1, \pm2, \pm3, \ldots \\ g(\tau), & \tau = \pm1/2, \pm3/2, \ldots. \end{cases}$$

Here, $g(\tau)$ is an arbitrary function constrained only by the requirement that the implied spectral density of $\{Y(t);t=0,\pm1/2,\pm1,\ldots\}$ be positive for almost all frequencies $\omega \in (-2\pi,2\pi)$. For example,

$$g(\tau) = \begin{cases} W\rho + k, & \tau = \pm 1/2 \\ W\rho^{2|\tau|}, & \tau = \pm 3/2, \pm 5/2, \ldots \end{cases}$$

will do, where $|k| < (1/2)\sigma^2/(1+|\rho|)^2$. Note that this change in the specification of the underlying data-generating mechanism enormously increases the dimension of the aliasing identification problem without changing the model's observable implications. Thus, in designing a setup that avoids the aliasing identification problem, untestable assumptions about the structure of the underlying data-generating mechanism play at least as important a role as does averaging or the prior restrictions of economic theory. [Again, this point is Hansen and Sargent's (1983).]

4.2. Sufficient conditions with averages

Assuming classical identification and some other mild conditions, the only potential source of aliasing underidentification when the data are time averaged and the model is (30) is the fact that the equation $z = x^{1/N}$--where x is a (possibly complex) scalar--has N solutions. [Phillips (1973) showed this in a different context.] If $\bar{z}$ is any solution to this equation, then the entire set of solutions is given by $\bar{z} \exp[(2\pi i/N)J]$, where $J \in \{1,2,\ldots,N\}$. (Readers not interested in the technical details should skip to Theorem 1.)

Recall that the mapping from $\phi \in \Phi$ to $R_{\underline{y}}(\cdot,\phi)$ takes four steps. First, for a given $\phi \in \Phi$, a set of parameters $A(\cdot;\phi)$, $C(\cdot;\phi)$, $V(\phi)$ is obtained from (30). Then the roots of $A(\cdot;\phi)$--$\rho_k = \rho_k(\phi)$ for $k = 1, \ldots, m$--are computed from (22). Next, the ρ_k's, A, C, and V are used to compute $K = K(\phi)$, $\tilde{K} = \tilde{K}(\phi)$, and $W_k = W_k(\phi)$ for $k = 1, \ldots, m$ from (23a). Finally, $R_{\underline{y}}(\cdot,\phi)$ is computed from (26a)-(26d) using $\{K,\tilde{K},W_k,\rho_k;k=1,\ldots,m\}$. The aliasing identifi-

cation condition is established if these steps can be reversed uniquely, so that ϕ is recovered from $\underline{R}_y(\cdot,\phi)$.

The first step in recovering ϕ from $\underline{R}_y(\cdot,\phi)$ is to obtain $\tilde{W}_k$ and ρ_k^N for $k = 1, \ldots, m$ using (26c). Proposition 1 shows that this can be accomplished under weak conditions.

Proposition 1. If $\phi \in \Phi$ and

(i) $W_k + W_k^T \neq 0$ for $k = 1, \ldots, m$ and

(ii) $\rho_1(\phi), \ldots, \rho_m(\phi)$ are distinct and nonzero,

then

(iii) relations (26c) permit computing a unique set $\tilde{W}_k$, ρ_k^N for $k = 1, \ldots, m$ using $\underline{R}_y(\tau,\phi)$ for $\tau = 2, 3, \ldots, 2m + 1$.

(The proofs of Proposition 1, 2, and 3 are in Appendix A.)

With the $\tilde{W}_k$'s and ρ_k^N's in hand, K is easily obtained from (26b). But to proceed further, the ρ_k's are required, and many sets of ρ_k's are consistent with the ρ_k^N's already computed. The entire class of sets is

$$D(\phi) = \{\tilde{\rho}=(\tilde{\rho}_1,\ldots,\tilde{\rho}_m)\in R^m: \tilde{\rho}_k=\rho_k(\phi)\exp\left[\frac{2\pi i}{N}J_k\right],k=1,\ldots,m;$$

$$J_k\in(1,2,\ldots,N),k=1,\ldots,m\}.$$

Evidently $D(\phi)$ contains N^m elements. (For an illustration, see Example 1.)

Definition. The admissible subset of $D(\phi)$ is the set of elements $\tilde{\rho} \in D(\phi)$ such that $\tilde{\rho}_k = \tilde{\rho}_k(\phi)$ for some $\phi \in \Phi$.

If the admissible subset of $D(\phi)$ is a singleton, then that point must be $[\rho_1(\phi),\ldots,\rho_m(\phi)]$. (See Example 2 for an illustration.) Proposition 2 shows

that when n = 2 and N = 2, the restrictions, (11), of Taylor's model guarantee that the admissible subset of $D(\phi)$ is a singleton for any $\phi \in \Phi$.

Proposition 2. If

(i) N = 2, n = 2 and

(ii) the mapping from ϕ to ρ_k for k = 1, ..., m is defined by (22) and (11), then

(iii) the admissible subset of $D(\phi)$ for any $\phi \in \Phi$ is a singleton.

When the admissible subset of $D(\phi)$ is a singleton, then a unique set of ρ_k's can be inferred from ρ_k^N for k = 1, ..., m. With the ρ_k's, W_k for k = 1, ..., m can be computed from (26d) and $\tilde{K}$ from (26a).

The final step in the inverse mapping is to infer a unique element in Φ from $\{K,\tilde{K},W_k,\rho_k;k=1,2,\ldots,m\}$. Proposition 3 shows that this is possible if the classical identification condition is satisfied.

Proposition 3. If

(i) the classical identification condition is satisfied at some $\phi_0 \in \Phi$ and

(ii) $\phi' \in \Phi$ and $\phi_0 \neq \phi'$,

then

(iii) $K(\phi_0) \neq K(\phi')$ or $\tilde{K}(\phi_0) \neq \tilde{K}(\phi')$ or $W_k(\phi_0) \neq W_k(\phi')$ or $\rho_k(\phi_0) \neq \rho_k(\phi')$ for some k.

To summarize:

Theorem 1. If

(i) the observed data are averages over the sampling interval

(ii) the data are generated by (30) for some $\phi_0 \in \Phi$

(iii) the classical identification condition is satisfied

(iv) $W_k + W_k^T \neq 0$ for $k = 1, \ldots, m$

(v) $\rho_1(\phi_0), \ldots, \rho_m(\phi_0)$ are distinct and nonzero and

(vi) the admissible subset of $D(\phi_0)$ is a singleton,

then

(vii) the aliasing identification condition is satisfied at ϕ_0.

Proof. To establish the theorem, I only need show that $\phi_0 \in \Phi$ can be recovered uniquely from $\{R_{\underline{y}}(\tau,\phi_0);\tau=0,1,2,\ldots\}$. Conditions (i) and (ii) imply that these are given by (26a)-(26d). Conditions (iv) and (v) imply, by Proposition 1 and eq. (26b), that K, $\tilde{W}_k$, and ρ_k^N for $k = 1, \ldots, m$ may be inferred uniquely from $\{R_{\underline{y}}(\tau,\phi_0);\tau=1,2,\ldots\}$. Condition (vi) implies that ρ_k for $k = 1, \ldots, m$ may be inferred uniquely from ρ_k^N for $k = 1, \ldots, m$. Given the ρ_k's, the $\tilde{W}_k$'s, and K, eqs. (26a) and (26d) imply unique $\tilde{K}$ and W_k's. Finally, condition (iii) and Proposition 3 guarantee that a unique $\phi_0 \in \Phi$ may be inferred from $\{K,\tilde{K},W_k,\rho_k;k=1,\ldots,m\}$. Q.E.D.

A few notes about the conditions of Theorem 1: Condition (i) is discussed in section 4.3. Condition (ii) plays an important role in guaranteeing identification because it places strong restrictions on the between-sample covariances, $R_{\underline{y}}(\tau,\phi)$ for $\tau \notin \{0,\pm1,\pm2,\ldots\}$. If these could be arbitrary (subject only to $\{\overline{Y}(t)\}$ having a nonnegative spectral density), then there would be little hope for aliasing identification. (See Example 4.) Condition (iii) is satisfied in Taylor's model and must be verified in all estimation problems, including those with N = 1. Condition (iv) can probably, by complicating the proof, be replaced by the weaker conditions $V \neq 0$, $C(\rho_k) \neq 0$, and $C(\rho_k^{-1}) \neq 0$ for $k = 1, \ldots, m$ (that is, $W_k \neq 0$). The condition that $C(\rho_k) \neq 0$ simply rules out a redundantly parameterized model, while $C(\rho_k^{-1}) \neq 0$ would be guaranteed if (30) were required to be an invertible

representation. Condition (v) can presumably be relaxed to allow for multiple roots. Then, however, the partial fractions expansion formula in (23a) would no longer be correct, so neither would the covariance formulas in (26a)-(26d). Examples 3 and 5 (the latter appears below) establish that condition (vi) is also not necessary for the theorem. In the examples, (i)-(v) hold and (vi) does not, yet aliasing identification obtains, for slightly different reasons. In Example 3, $\{R_{\underline{y}}(\tau,\phi);\tau=0,1,2,\ldots\}$ implies a unique set $\{K,\tilde{K},W_k,\rho_k;k=1,\ldots,m\}$. In Example 5, $\{R_{\underline{y}}(\tau,\phi);\tau=0,1,2,\ldots\}$ implies several sets $\{K,\tilde{K},W_k,\rho_k;k=1,\ldots,m\}$; however, only one of these corresponds to a $\phi \in \Phi$. [Example 5 is similar to examples described in Hansen and Sargent (1983).]

Example 5: Sampling averages from a scalar ARMA(1,1). Suppose $N = 2$, $m = 1$, all variables are scalars, $A(B;\phi) = 1 - \rho B$, $|\rho| < 1$, $C(B;\phi) = 1 + \theta B$, $|\theta| < 1$, and $V(\phi) = \sigma^2 > 0$. Here $\phi = (\rho,\theta,\sigma^2)$, $\Phi = \{x=(x_1,x_2,x_3)\in R^3: |x_1|<1, |x_2|<1, x_3>0\}$, $W = \sigma^2(1+\theta\rho^{-1})(1+\theta\rho)/(1-\rho^2)$, $\tilde{W} = .25W(1+\rho)(1+\rho^{-1})$, $\tilde{K} = -\theta\sigma^2/\rho$, and $K = 0$. Also, $R_{\underline{y}}(0,\phi) = (\tilde{K}/2) + 2\tilde{W}/(1+\rho^{-1})$, and $R_{\underline{y}}(\tau,\phi) = \tilde{W}\rho^{2\tau}$ for $\tau = 1, 2, \ldots$. Both elements of $D(\phi)$ are admissible for all $\phi \in \Phi$ since ρ has no sign restriction.[3/]

Let $\phi_0 = (\rho_0,\theta_0,\sigma_0^2)$ denote some particular element in Φ, with ρ_0, $\theta_0 \neq 0$. The quantities $\tilde{W}$ and ρ^2 are uniquely identified from $R_{\underline{y}}(\tau,\phi_0)$ for $\tau = 1, 2$. Denote these by $\tilde{W}_0 = R_{\underline{y}}(1,\phi_0)^2/R_{\underline{y}}(2,\phi_0)$ and $\rho_0^2 = R_{\underline{y}}(2,\phi_0)/R_{\underline{y}}(1,\phi_0)$. Two possible values of ρ are consistent with the ρ_0^2 relation: ρ_0 and $-\rho_0$. Each of these produces different values for $\tilde{K}$ and W from (26a) and (26d), respectively. Denote the $\{\tilde{K},W\}$ that corresponds to ρ_0 and $-\rho_0$, respectively, by $\{\tilde{K}_0,W_0\}$ and $\{\tilde{K}_0',W_0'\}$. In particular,

$$\tilde{K}_0 = 2\underline{R}_y(0,\phi_0) - \frac{4\tilde{W}_0}{1 + \rho_0^{-1}} \tag{31}$$

$$\tilde{K}_0' = 2\underline{R}_y(0,\phi_0) - \frac{4\tilde{W}_0}{1 - \rho_0^{-1}} \tag{32}$$

$$W_0 = \frac{4\tilde{W}_0}{(1+\rho_0)(1+\rho_0^{-1})} \tag{33}$$

$$W_0' = \frac{4\tilde{W}_0}{(1-\rho_0)(1-\rho_0^{-1})}. \tag{34}$$

Thus, $\underline{R}_y(\tau,\phi_0)$ for $\tau = 0, 1, 2, \ldots$ maps into two distinct sets--$\{\tilde{K}_0, W_0, \rho_0\}$ and $\{\tilde{K}_0', W_0', -\rho_0\}$--for every $\phi_0 \in \Phi$. By construction, $\{\tilde{K}_0, W_0, \rho_0\}$ corresponds to $\phi_0 \in \Phi$. That no other $\phi \in \Phi$ generates $\{\tilde{K}_0, W_0, \rho_0\}$ is implied by the fact that classical identification obtains in this example. If $\{\tilde{K}_0', W_0', -\rho_0\}$ corresponds to no $\phi \in \Phi$, then aliasing identification obtains at ϕ_0, despite the failure of condition (vi) of Theorem 1. In fact, there are such elements, $\phi_0 \in \Phi$. To see this, compute all possible ϕ's that correspond to $\{\tilde{K}_0', W_0', -\rho_0\}$. All share the properties that $\rho = -\rho_0$ and that θ and σ^2 satisfy

$$\sigma^2 = \frac{\rho_0 \tilde{K}_0'}{\theta} \tag{35}$$

$$\tilde{W}_0 = \frac{(1-\theta\rho_0^{-1})(1-\theta\rho_0)(1-\rho_0^{-1})\sigma^2}{4(1+\rho_0)}. \tag{36}$$

Substituting σ^2 from (35) into (36) and rearranging reveals that values of θ that solve this system also solve

$$\theta^2 - \delta\theta + 1 = 0 \tag{37}$$

where $\delta(\phi_0) = -4(1+\rho_0)\tilde{W}_0/[(1-\rho_0)\tilde{K}_0'] + (\rho_0+\rho_0^{-1})$. If $|\delta(\phi_0)| < 2$, then (37) has no real solution, in which case the model is identified at $\phi_0 \in \Phi$. Moreover, the set $\phi_0 \in \Phi$ such that $|\delta(\phi_0)| < 2$ is not empty. For example, $\phi_0 =$

(.2,.8,1.) implies $\delta = .58$. Since δ is a continuous function of ϕ, there is a subset $\Lambda \subset \Phi$ of positive measure such that the model is identified.

I provide no proof that Taylor's model is identified in the aliasing sense when n or N or both are greater than 2. However, the preceding discussion does indicate a computational algorithm for checking identification in this more general case. Theorem 1 suggests that the only potential source of underidentification arises if the admissible subset of $D(\phi)$ is not a singleton. Whether or not this is true can easily be determined because $D(\phi)$ has a finite number of elements; its admissible subset can thus be computed. (Such an algorithm would mimic the strategy used to prove Proposition 2.)

4.3. The role of averaging

Condition (i) of Theorem 1, that the data are averages, is crucial to the theorem. If the condition is dropped and the data are point-in-time, then the theorem is false. To see this, suppose $C_m \neq 0$ so that $K \neq 0$. Then the W_j's and ρ_j's can be recovered uniquely using (24c). Eq. (24a) would then permit recovering $\tilde{K}$, but K could not in general be recovered. Yet if $K = 0$ (that is, $C_{m+1} = 0$), then Theorem 1 remains valid when the data are sampled point-in-time. For concreteness, this is stated formally:

Theorem 2. If

(i) the observed data are sampled point-in-time and $C_{m+1}(\phi_0) \equiv 0$ and

(ii) conditions (ii)-(vi) of Theorem 1 are satisfied,

then

(iii) the aliasing identification condition is satisfied at ϕ_0.

The preceding discussion understates the importance of averaging for aliasing identification by suggesting that averaging only helps when $C_{m+1} \neq 0$.

In particular, recall from Examples 1 and 3 that in the scalar AR(1) case with $N = 2$, where $C_{m+1} = 0$, the model parameters are identifiable from averaged data but not from point-in-time data. Another example is provided by the scalar ARMA(1,1) case with $N = 2$. Example 5 shows that, on a subset of the parameter space, the parameters of this model can be identified from the covariogram of the sampled, averaged data. Yet the parameters are under-identified at almost all points in Φ when the data are point-in-time.[4/] These examples suggest the possibility that the covariogram of data generated by (8) contain (weakly) more information about ϕ when the data are averaged than when they are point-in-time. This conjecture deserves further study.

Finally, here I have only studied the identifiability of ϕ when $N > 1$ but known. Of greater relevance is the joint identifiability of N and ϕ, a more difficult problem beyond the scope of this paper. [The problem is explored in the context of a simple example in Christiano (1985).]

5. Empirical results

This section reports the results of restricted and unrestricted estimation of the model in section 2. Section 5.1 discusses the unrestricted time series properties of the U.S. output and price data. An important feature of the data is that the null hypothesis that output fails to Granger-cause the price level can be rejected. The only way Taylor's model with $N = 1$ and $n > 1$ can be reconciled with this fact is to permit serial correlation in the residuals, that is, to allow $\delta, \theta \neq 0$. Section 5.2 argues that allowing $N > 1$ should improve the fit of the $\delta = \theta = 0$ version of the model. This argument is supported by section 5.3, which presents the results of maximum likelihood estimation. However, the section also shows that Taylor's (1980a) strategy of setting $N = 1$ and freeing up θ and δ produces an even better model fit. Section 5.4 explores the reasons for this by examining impulse response functions.

5.1. The data

The data are linearly detrended annual observations of the log of output (y_t) and the price level (p_t) for the United States in the period 1954-76 and appear in Appendix B. A convenient way to summarize the stochastic dynamics of the data is to fit an unrestricted vector autoregression (VAR). These are the results:

$$\begin{pmatrix} y_t \\ p_t \end{pmatrix} = \begin{bmatrix} .417 & -.431 \\ .673 & 2.04 \end{bmatrix} \begin{pmatrix} y_{t-1} \\ p_{t-1} \end{pmatrix} + \begin{bmatrix} -.040 & .197 \\ -.371 & -.961 \end{bmatrix} \begin{pmatrix} y_{t-2} \\ p_{t-2} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \qquad (38)$$

Two of the roots of the determinant of the AR part are $1.03 \pm .23i$, which are close to 1, suggesting near-nonstationarity.[5/]

The MA representation (ψ) of the VAR in (38) is in table 1. Several features stand out there. First, note the persistent effect of a shock to the price level. The effect of a price shock on the price level (ψ_{22}) builds to a peak after five years before subsiding. Second, note the strong positive cross effect of an output shock on the price level (ψ_{21}). This is statistically significant, according to a test of the null hypothesis that output does not enter the price equation in the VAR in eq. (38). The test statistic is 76.79, which is drawn from an asymptotic chi-square distribution with two degrees of freedom under the null hypothesis that output does not Granger-cause prices. Third, note the negative and persistent effect of the price shock on output (ψ_{12}). Fourth, note the nonzero effect of an output shock on output at long lags and the switch in sign between lag 1 and 2 (ψ_{11}).

5.2. Reasons for expecting improved model fit when N > 1

According to (8), the MA representation implied by Taylor's model is $Y(t) = \Gamma(B)w(t)$, where $\Gamma(B) \equiv [A(B)]^{-1}C(B)$ and

$$\Gamma(B) = \begin{pmatrix} \Gamma_{11}(B) & \Gamma_{12}(B) \\ \Gamma_{21}(B) & \Gamma_{22}(B) \end{pmatrix}.$$

By (14a)-(14d),

$$\Gamma_{11}(B) = -\beta \frac{k_1}{nN} \frac{B(1-B^{nN})}{(1-B)A(B)} + (1+\theta B) \tag{39a}$$

$$\Gamma_{21}(B) = \frac{k_1}{nN} \frac{B(1-B^{nN})}{(1-B)A(B)} \tag{39b}$$

$$\Gamma_{12}(B) = \beta\left[1 - \frac{(1+k_2B)(1-B^{nN})}{(1-B)A(B)}\right] \tag{39c}$$

$$\Gamma_{22}(B) = \frac{(1+k_2B)(1-B^{nN})}{(1-B)A(B)}. \tag{39d}$$

The expressions k_1 and k_2 are defined in (16a) and (16b) and reproduced here for convenience:

$$k_1 = -\frac{1}{\lambda}\left[\frac{\theta\gamma}{nN - \gamma\beta(nN-1)}\right] \tag{16a}$$

$$k_2 = -\frac{1}{\lambda}\left[\frac{nN\delta}{nN - \gamma\beta(nN-1)}\right]. \tag{16b}$$

Here $\lambda < 0$ is defined in (13) and is functionally independent of θ and δ.

The Γ's are the MA representation of the model at the fine time interval. What is useful for diagnostic purposes is the MA representation implied by Γ for the sampled, averaged data. Denote this $\tilde{\psi}(L) = \tilde{\psi}_0 + \tilde{\psi}_1 L + \tilde{\psi}_2 L^2 + \ldots$. For a direct comparison with $\psi(L)$, impose the normalization $\tilde{\psi}_0 \equiv I$. Formally, for given $\phi \in \Phi$, $\tilde{\psi}(L)$ is defined by

$$\tilde{\psi}(z)\tilde{V}\tilde{\psi}(z^{-1}) = \sum_{\tau=-\infty}^{+\infty} R_{\underline{y}}(\tau,\phi) z^{\tau} \tag{40}$$

where $R_{\underline{y}}$ is defined in (26a)-(26d) and $\tilde{V}$ is a positive definite matrix that is not of interest here. The indicated factorization is known to exist and be unique because $S_y(i\omega)$ in (21) is positive definite at almost all $\omega \in (-\infty,+\infty)$. [See Hannan (1970).] The algorithm described in Rozanov (1967, ch. 1, sec. 10) was used in my calculations. Of course, in the special case N = 1, $\tilde{\psi}(L) = \Gamma(L)$.[6/]

Three distinct sets of notation for MA representations have been introduced. These are summarized in table 2. The similarity between $\tilde{\psi}$ and ψ is one criterion I use to evaluate the quality of model fit.

Taylor (1980a) studies only the version of the model with n = 2, N = 1. I take this as a starting point and define the model with $\delta = \theta = 0$, n = 2, N = 1 as the benchmark model. The elements of $\tilde{\psi}(\cdot)$ [$\equiv \Gamma(\cdot)$, since N = 1] for this model are reported in table 3.

Comparing table 3 with table 1 reveals that the benchmark model misses some of the stochastic dynamics in the data. The signs in the elements of $\tilde{\psi}_{12}$ and $\tilde{\psi}_{22}$ correspond to those of ψ_{12} and ψ_{22}, respectively. However, the peak effect in $\tilde{\psi}_{12}$ and $\tilde{\psi}_{22}$ is reached quickly, at lag 1, while that in ψ_{12} and ψ_{22} is reached later. Moreover, the benchmark model predicts no cross effects from an output shock to prices ($\tilde{\psi}_{21} = 0$), an implication strongly contradicted by the data. Also contradicted is the prediction of the benchmark model that the impact of an output shock on output ($\tilde{\psi}_{11}$) is felt immediately, without lag.

Given this evidence, it is not surprising that Taylor (1980a) reports a rejection of the benchmark model against an unrestricted representation for $(y(t),p(t))$. Taylor moves away from the benchmark model by letting δ and θ be nonzero. The effects of this on $\tilde{\psi}(\cdot)$ are described in table 4. Comparing tables 1, 3, and 4 shows that allowing $\delta, \theta > 0$ moves Taylor's model

closer to the relevant features of the data, as summarized in table 1. However, as Taylor points out, moving from the model of table 3 to that of table 4 amounts to concluding that the theory is seriously incomplete, because much of the dynamics underlying the model in table 4 is delivered by unexplained dynamics in shock terms. Furthermore, Taylor reports that his modified model is rejected.

Another way to move away from the benchmark model is to increase the value of N. This preserves the model's simplicity and has a more straightforward economic interpretation than Taylor's approach. Moreover, such a move should produce results closer to the dynamics in table 1 than the benchmark model's are.

When $\delta = \theta = 0$, the theory predicts that $\Gamma_{21}(L^{1/N}) = 0$. However, as Sims (1971) and Geweke (1978) have shown, when a process with a 2,1 element in its pure MA representation equal to zero is sampled, that element in the sampled MA representation will be nonzero, unless the process is a first-order bivariate VAR. Thus, modifying the benchmark model by setting $N > 1$ creates a potential for reconciling the model with the nonzero ψ_{21} elements in table 1. In a similar way, temporal aggregation effects can place nonzero elements at lags greater than 1 in the 1,1 elements of the MA part of the sampled representation.

Another result of setting $N > 1$ is that the peak value in the $\tilde{\psi}_{22}$ weights can occur at a lag length greater than 1. To see this, note that the order of the autoregressive part on the detrended log of the price level is $nN - 1$. As N increases, this increases, and in the (1/N)-period representation the coefficients of $\Gamma(\cdot)$ may peak later. On the other hand, aggregating over time works in the opposite direction, moving the peak of the MA coefficients of the sampled representation to shorter lags. Therefore, the net effect on

the location of the peak value of $\tilde{\psi}_{22}$ of increasing N is ambiguous. However, as is shown below, the effect is positive in empirically relevant portions of the parameter space.

One characteristic that the $N > 1$, $\delta = \theta = 0$ version of the model shares with the benchmark version is that

$$\tilde{\psi}_{12} = -\beta\tilde{\psi}_{22}, \quad \tilde{\psi}_{11} = -\beta\tilde{\psi}_{21} \tag{41}$$

at lags greater than 0. This stems from the form of the aggregate demand equation (3), which implies that $E_{t-(1/N)}y(t) = -\beta E_{t-(1/N)}p(t)$, which implies that $E_{t-1}\bar{y}(t) = -\beta E_{t-1}\bar{p}(t)$. [Here $\bar{x}(t)$ denotes $(1/N)\sum_{i=0}^{N-1}x(t + (i/N))$.]

5.3. Maximum likelihood estimation

Now I examine evidence on whether or not setting $N > 1$ produces a model fit better than the benchmark model's. I also compare this strategy to Taylor's (1980a) strategy of permitting $\delta, \theta \neq 0$.

Table 5 reports the maximized values of the likelihood function for alternative values of n and N. These numbers are graphed in fig. 2. Clearly, increasing the value of N above 1 improves the quality of model fit, as section 5.2 predicts. The extent of the improvement depends heavily on the length of the contract period. With two-year contracts ($n = 2$), almost all the improvement occurs when N increases from 1 to 2. With three- and four-year contracts, improvement continues as N increases beyond 2. In fact, the results suggest that the optimal value of N lies beyond $N = 10$ for three- and four-year contracts. (My estimation routine breaks down when $nN > 30$, so I have no direct evidence on this.) Table 5 and fig. 2 also indicate that the three-year specification fits the data better than the two-year specification Taylor (1980a) studied. (I have not investigated whether the difference is significant.)

Table 6 presents the parameter estimates corresponding to some of the results in table 5.[7] Several features stand out. One is that the estimate of β is insensitive to the choice of n and N. This is probably a result of (41), according to which β is identified as the ratio of $-\tilde{\psi}_{12}$ to $\tilde{\psi}_{22}$ or $-\tilde{\psi}_{11}$ to $\tilde{\psi}_{21}$ at lags greater than 1, regardless of the value of n or N. The estimated value of γ, in contrast, is insensitive to the choice of N only when n = 2. Also, both the estimated innovation variances (V_{11} and V_{22}) and the estimated correlation between the innovations increase with N. I cannot explain this.

For comparison, table 7 presents the results of estimation when δ and θ are permitted to be free and N = 1. These results agree with table 5's in suggesting that the model with three-year labor contracts fits the data better than the two-year specification. Also, a comparison of tables 5 and 7 suggests that freeing δ and θ improves model fit more than increasing N above 1. Of course, this does not contradict the finding that increasing N does improve the fit of the benchmark model.

Since Taylor reports a rejection of the n = 2, N = 1, δ, $\theta \neq 0$ model, the above comparison suggests that my no serial correlation, free n, N model is also rejected. An open question is whether or not the six-parameter model (free serial correlation, free n, N) would be rejected.

5.4. The impulse response functions

The reasons for the results of section 5.3 can be further investigated by studying the impulse response functions graphed in figs. 3-6. The objective is to discover which features of the serial- and cross-correlation structure of the data are not explained by the model. The focus is on two questions: Why does increasing the value of N in the benchmark model improve

model fit? And why does the N = 1 (δ, $\theta \neq 0$) version of the model fit better than the N > 1 ($\delta = \theta = 0$) model?

Figs. 3 and 4 show the dynamic response of prices to innovations in prices and output, respectively; figs. 5 and 6, the dynamic response of output to innovations in output and prices, respectively. The same normalization is applied to all models, namely, that the lead matrix in the impulse response function is the identity matrix.

Begin by focusing on fig. 3a, for which n = 2. The curve labeled "N = 1" describes the response of the price level to a price innovation in the benchmark model. The curve labeled "unrestricted" describes the same relation for the unrestricted model and is a graph of the last column in table 1. Note that increasing the value of N above 1 moves the model closer to the unrestricted model in two respects. The peak in the impulse response function shifts from lag 1 to lag 2 (as discussed in section 5.2). Also, the impulse response coefficients themselves become larger. The curve labeled "N = 1 (δ, $\theta \neq 0$)" in fig. 3a describes the price response to price innovations in the N = 1 model estimated allowing δ and θ to vary freely.

In terms of their ability to track ψ_{22}, the N > 1 ($\delta = \theta = 0$) and N = 1 (δ, $\theta \neq 0$) models seem to do about equally well (or poorly). They both are improvements over the benchmark model. They both understate the magnitude of the unrestricted impulse response function and peak too soon. Figs. 3b and 3c, for which n = 3 and n = 4, tell essentially the same story as fig. 3a.

That the N = 1 curve hugs the horizontal axis in figs. 4a-c reflects the benchmark model's implication that output fails to Granger-cause the price level. This is where most of the improvement in model fit resulting from raising N should occur. In fact, the impulse response functions for N = 4, 7, and 10 are closer to the unrestricted impulse response functions than are

those for the benchmark model (N = 1). However, the higher N functions fall short of the unrestricted functions in two respects: they are much smaller in magnitude and peak earlier.

The impulse response functions implied by the N = 1 (δ, $\theta \neq 0$) model in fig. 4 appear not to match the unrestricted response functions as well as do those implied by N = 4, 7, and 10; their magnitude is far too small. Thus, the N > 1 ($\delta = \theta = 0$) model dominates the N = 1 (δ, $\theta \neq 0$) model in terms of its ability to track ψ_{21}.

Shifting to output responses, look first at fig. 6. These output responses to price shocks are similar to fig. 3's price responses to such shocks. Increasing the value of N increases the magnitude of the impulse response functions and, when n = 2, shifts the peak to a higher lag. Both these factors move the model closer to the unrestricted model, although not much. Also, the N = 1 (δ, $\theta \neq 0$) model generally does about as well as the N > 1 ($\delta = \theta = 0$) model. When n = 3, it does a bit better.

The responses of output to output shocks (fig. 5) also suggest that the benchmark model is seriously deficient. In the benchmark model (N = 1), an innovation in output has no lagged effect on output. The (unrestricted) data suggest, however, that an unexpected jump in output this year is typically associated with a jump in output next year and a fall in subsequent years. Increasing the value of N is unlikely to help the benchmark model reproduce this complicated pattern. This is because, as (41) indicates, $\tilde{\psi}_{11}$ and $\tilde{\psi}_{21}$ have either the same or the opposite sign at all lags greater than zero. Consequently, the model can match ψ_{11} and ψ_{21} either at lag 1 or at lags greater than 1, but not both. Fig. 5 shows that the estimation criterion has chosen the second option. Increasing N above 1 moves the benchmark model

toward the unrestricted representation for lags greater than 1, but away from it for lag 1.

Allowing $\theta > 0$ but keeping $N = 1$ has a greater chance of reproducing the unrestricted function from output shocks to output. This is because, when $\theta > 0$, an output innovation affects future output both directly and indirectly, through prices. The direct effect is that a one-unit positive shock hitting output this year drives output up by θ next year. The indirect effect is evident in (1). By increasing expectations about next period's output, an innovation to output this period increases next period's wage if $\gamma > 0$. Because of the backward-looking nature of the contract model in (1), this leads to a jump in every future wage as well. The jump in wages translates into a jump in prices by (4), and that jump pushes output down by (3), starting next period. Next period's positive direct effect should thus be partially offset by the negative indirect effect as long as γ is not too big. Only the indirect effects should exist in periods after the next.

In sum, with $\theta > 0$ and $\gamma > 0$ but not too big, the $N = 1$ version of the model can reproduce an innovation response pattern from output to output similar to that found in the data.

Altogether, the impulse response functions offer several conclusions. The $N > 1$ model fits the data better than the benchmark model for all the reasons in section 5.2. In particular, increasing N above 1 eliminates the implication that output not Granger-cause prices. Also, increasing the value of N increases the serial persistence in the model, raising the magnitude of impulse response coefficients and shifting their peak to later lags. A shortcoming of the $N > 1$ model is that it does not go far enough in these directions. In addition, the model fails to adequately treat the dynamic response of output to output shocks. This seems to be the reason that the N =

1 (δ, $\theta \neq 0$) model fits the data better than the $N > 1$ ($\delta = \theta = 0$) model. The former model captures the switch in sign in the impulse response from output to output. The importance of this factor in determining the relative value of the two models' likelihoods is consistent with the fact that the estimates of θ in table 7 are from five to ten times their standard errors.

6. Conclusion

This paper shows that by increasing the degree of overlap in contract wages, the fit of the benchmark version of Taylor's model is improved. The improvement is due to aggregation over time. As the degree of overlap in the model increases, the model timing interval shrinks and the model loses the implication of a one-way Granger-causal relation between annual averages of prices and output. Since the data display two-way Granger causality, the result is an improved model fit.

The empirical example in this paper--together with those in Christiano (1984,1985), Mundlak (1961), Zellner (1968), and Zellner and Montmarquette (1971)--illustrates the potential sensitivity of empirical results to timing specification. The approach here is a practical procedure for estimating the timing interval in a model, thereby avoiding the pitfalls of timing misspecification.

Notes

1/ Eq. (3) is derived from the quantity theory equation: $p(t) + y(t) = m(t) + v(t)$, where $v(t)$ is velocity and is restricted to have a first-order moving average representation. Government monetary policy is assumed to obey the feedback rule: $m(t) = (1-\beta)p(t)$. Substituting this into the quantity equation gives $m(t) = -\beta p(t) + v(t)$. Finally, $p(t)$ is replaced by $\hat{p}(t)$ to reflect the idea that money supply decisions feed back on $\hat{p}(t)$, not $p(t)$. [Strictly speaking, this requires that $p(t)$ also be in the equation for $m(t)$. See Taylor (1980a, n. 9).]

2/ Inspection of (23a) reveals that in fact 2^{2m} functions $R_y(\cdot,\phi)$ solve (20) and (21). These are obtained by noting that, for each j, $W_j/(1-\rho_j s)$ has two expansions, one in positive powers of s and the other in negative. The unique function $R_y(\cdot,\phi)$ which solves (20) and (21) and which has the property that the expansion (20) converges for s in a region that includes the unit circle is the one in (24a)-(24d). This is the only element in the set of 2^{2m} expansions which can be the covariance function of a covariance stationary process. The others either are nonsymmetric or do not go to zero as $\tau \to \infty$.

3/ Note that when only point-in-time data are available, this model is underidentified in the aliasing sense for all $\phi \in \Phi$ except those with $\rho = \theta = 0$. This is because (ρ,θ,σ^2) and $(-\rho,-\theta,\sigma^2)$ imply identical values of W and $\tilde{K}$.

4/ See note 3.

5/ The coefficients in (38) were estimated equation by equation by OLS. Let $\sigma_i^2 \equiv$ estimated Eu_{it}^2 for i = 1, 2, and $\rho \equiv \sigma_{12}/(\sigma_1\sigma_2)$, where $\sigma_{12} \equiv$ estimated $Eu_{1t}u_{2t}$. The results are $\sigma_1^2 = 2.783$, $\sigma_2^2 = .885$, and $\rho = -.117$. The

roots of the determinant of the AR part of (38) are 2.554 ± 1.230i and 1.033 ± .228i. My estimation results differ somewhat from Taylor's (1980a, p. 115). The difference may be due to data revisions, although I have not checked this formally.

6/ The computer programs required to compute $\tilde{\psi}$ are elaborate and are available from the author on request.

7/ The results for values of N between 1 and 10 that are not in table 6 are primarily simple interpolations between results for the values listed. The estimate of γ has two exceptions. For n = 3 it rises sharply from .01 to .86 between N = 9 and N = 10, and for n = 4 it doubles from .12 to .24 between N = 6 and N = 7.

Fig. 1. Two adjacent contracts.

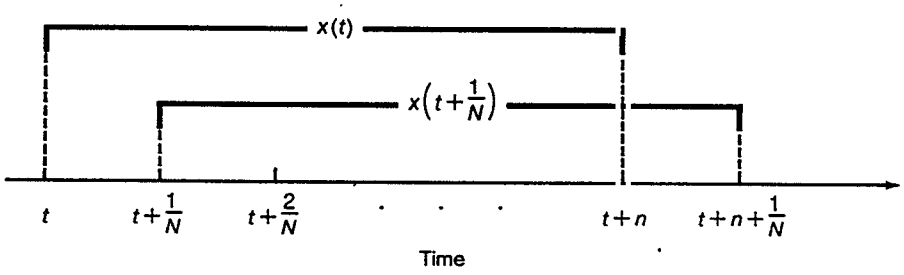

Fig. 2. Maximized log-likelihood function for various values of the contract length n and the timing interval index N. (Graphs of data in table 5.)

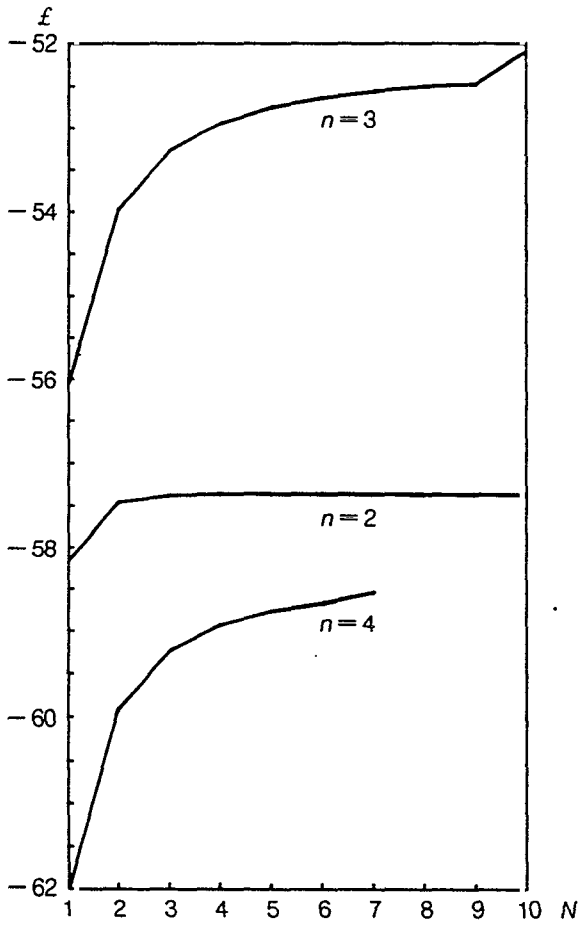

Fig. 3. Price response to price innovations. Curves labeled "unrestricted" are the ψ_{22} column in table 1; the others are $\tilde{\psi}_{22}$ as defined in (40).

Fig. 3a: $n = 2$ Fig. 3b: $n = 3$ Fig. 3c: $n = 4$

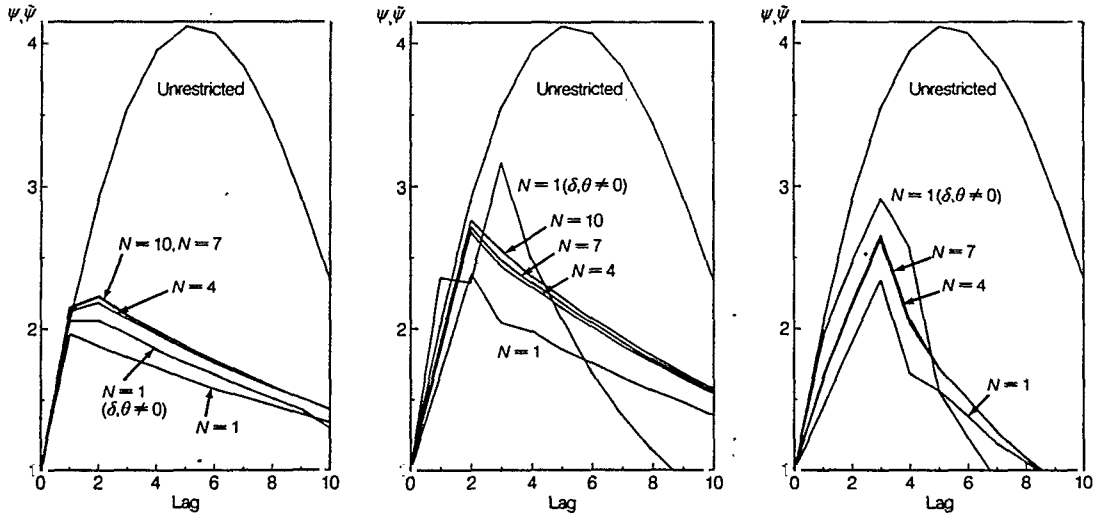

Fig. 4. Price response to output innovations. Curves labeled "unrestricted" are the ψ_{21} column in table 1; the others are $\tilde{\psi}_{21}$ as defined in (40).

Fig. 4a: $n = 2$

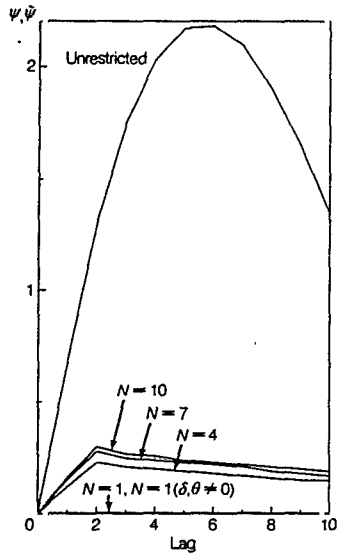

Fig. 4b: $n = 3$

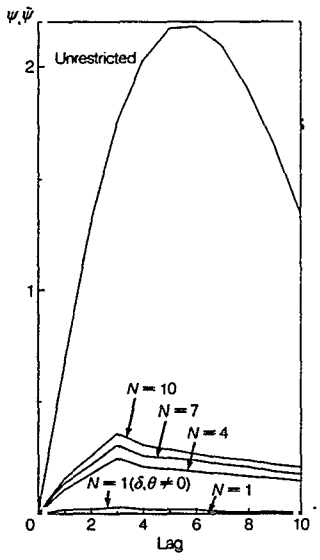

Fig. 4c: $n = 4$

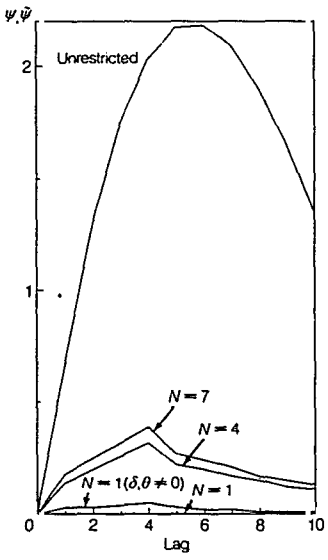

Fig. 5. Output response to output innovations. Curves labeled "unrestricted" are the ψ_{11} column in table 1; the others are $\tilde{\psi}_{11}$ as defined in (40).

Fig. 5a: $n = 2$ Fig. 5b: $n = 3$ Fig. 5c: $n = 4$

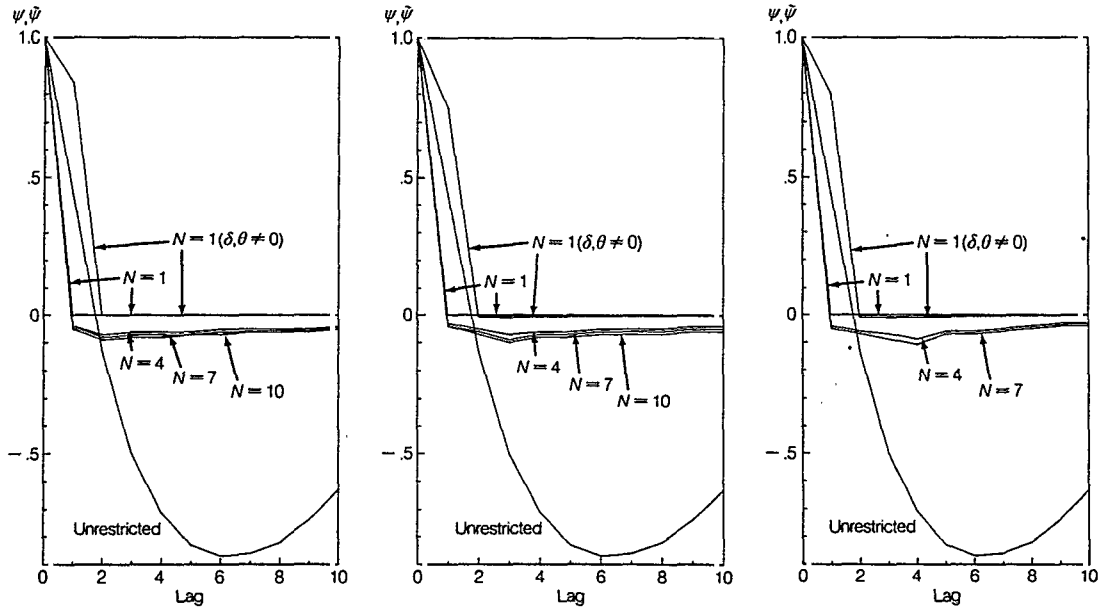

Fig. 6. Output response to price innovations. Curves labeled "unrestricted" are the ψ_{12} column in table 1; the others are $\tilde{\psi}_{12}$ as defined in (40).

Fig. 6a: $n = 2$

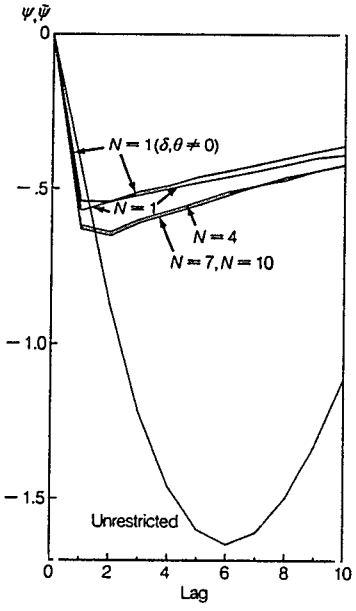

Fig. 6b: $n = 3$

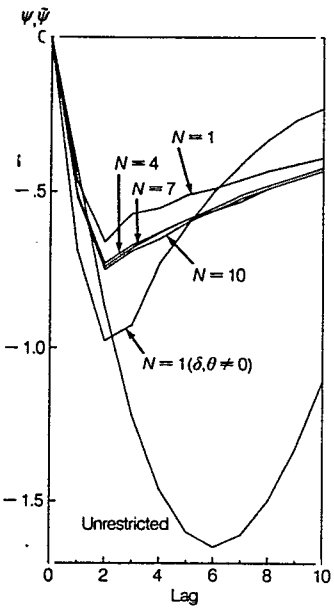

Fig. 6c: $n = 4$

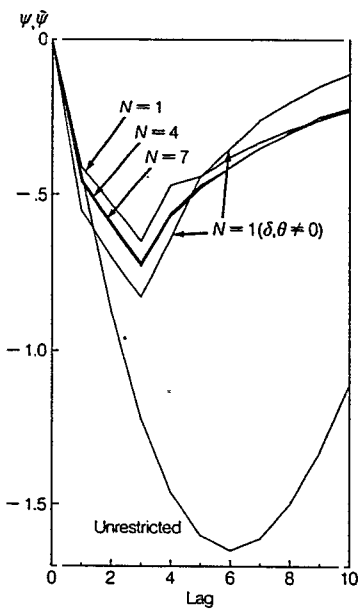

Table 1

MA representation of unrestricted VAR. [a/]

Lag	Output effect of shock to: Output ψ_{11}	Output effect of shock to: Price level ψ_{12}	Price effect of shock to: Output ψ_{21}	Price effect of shock to: Price level ψ_{22}
0	1.000	0.000	0.000	1.000
1	.442	-.431	.673	2.040
2	-.134	-.872	1.298	2.911
3	-.503	-1.221	1.748	3.552
4	-.714	-1.462	2.029	3.952
5	-.826	-1.601	2.166	4.120
6	-.870	-1.646	2.180	4.073
7	-.864	-1.607	2.087	3.837
8	-.817	-1.496	1.904	3.445
9	-.736	-1.326	1.651	2.932
10	-.629	-1.111	1.346	2.335

[a/] Write the model in eq. (38) as $z_t = A_1 z_{t-1} + A_2 z_{t-2} + u_t$, where $z_t = (y_t, p_t)^T$ and $u_t = (u_{1t}, u_{2t})^T$. Then $\psi_0 + \psi_1 L + \psi_2 L^2 + \ldots = (I - A_1 L - A_2 L^2)^{-1}$.

Table 2

Summary of MA representations.

Notation	Description	Where defined
Γ	MA representation at fine level of temporal aggregation, constrained by theory	Eqs. (39a)-(39d)
$\tilde{\psi}$	MA representation of sampled, averaged data, generated by Γ	Eq. (40)
ψ	Unconstrained MA representation of sampled data	Note to table 1

Table 3

Benchmark version of Taylor's model
($n = 2$; $N = 1$; $\delta, \theta = 0$). [a/]

	Output effect of shock to		Price effect of shock to	
Lag	Output $\tilde{\psi}_{11}$	Price level $\tilde{\psi}_{12}$	Output $\tilde{\psi}_{21}$	Price level $\tilde{\psi}_{22}$
0	1.0	0.0	0.0	1.0
1	0.0	$-\beta(1+a_1)$	0.0	$1 + a_1$
2	0.0	$-\beta a_1(1+a_1)$	0.0	$a_1(1+a_1)$
3	0.0	$-\beta a_1^2(1+a_1)$	0.0	$a_1^2(1+a_1)$
4	0.0	$-\beta a_1^3(1+a_1)$	0.0	$a_1^3(1+a_1)$
5	0.0	$-\beta a_1^4(1+a_1)$	0.0	$a_1^4(1+a_1)$

[a/] The $\tilde{\psi}$'s are computed from (40), taking into account that when $N = 1$, $\tilde{\psi}(\cdot) = \Gamma(\cdot)$.

Table 4

Taylor's modified model ($n = 2$; $N = 1$; $\delta, \theta \neq 0$). [a/]

Lag	Output effect of shock to: Output $\tilde{\psi}_{11}$	Output effect of shock to: Price level $\tilde{\psi}_{12}$	Price effect of shock to: Output $\tilde{\psi}_{21}$	Price effect of shock to: Price level $\tilde{\psi}_{22}$
0	1.0	0.0	0.0	1.0
1	$\theta - \beta k_1$	$-\beta(1+a_1+k_2)$	k_1	$1 + a_1 + k_2$
2	$-\beta(1+a_1)k_1$	$-\beta(a_1+k_2)(1+a_1)$	$(1+a_1)k_1$	$(a_1+k_2)(1+a_1)$
3	$-\beta a_1(1+a_1)k_1$	$-\beta a_1(a_1+k_2)(1+a_1)$	$a_1(1+a_1)k_1$	$a_1(a_1+k_2)(1+a_1)$
4	$-\beta a_1^2(1+a_1)k_1$	$-\beta a_1^2(a_1+k_2)(1+a_1)$	$a_1^2(1+a_1)k_1$	$a_1^2(a_1+k_2)(1+a_1)$
5	$-\beta a_1^3(1+a_1)k_1$	$-\beta a_1^3(a_1+k_2)(1+a_1)$	$a_1^3(1+a_1)k_1$	$a_1^3(a_1+k_2)(1+a_1)$

[a/] The $\tilde{\psi}$'s are computed from (40), taking into account that when $N = 1$, $\tilde{\psi}(\cdot) = \Gamma(\cdot)$.

Table 5

Maximized values of the likelihood function $\mathfrak{L}$ ($\delta = \theta = 0$). [a/]

Timing interval index N	Contract length n (years)		
	2	3	4
1	-58.17980	-56.08670	-62.11900
2	-57.44960	-53.95340	-59.89430
3	-57.37090	-53.26730	-59.21900
4	-57.36090	-52.95050	-58.92500
5	-57.36130	-52.77220	-58.76620
6	-57.36340	-52.65910	-58.66850
7	-57.36550	-52.58050	-58.53960
8	-57.36730	-52.52440	-- [b/]
9	-57.36860	-52.49170	--
10	-57.36980	-52.09910	--

[a/] These are the maximized values of $\mathfrak{L}$ in (28), conditional on the indicated values of n and N.

[b/] Computational difficulties prevented assembling results for cases when $nN > 30$.

Table 6

Restricted maximum likelihood estimates ($\delta = \theta = 0$). [a/]

Timing interval index N	Parameters: Demand β	Wage γ	Innovation variances V_{11}	V_{22}	V_{12}	Correlation of variances ρ
			Two-year contracts (n = 2)			
1	.2897 (.0690) [b/]	.0031 (.0080)	5.212 (1.537)	12.067 (3.611)	.449 (1.663)	.057
4	.2911 (.0442)	.0036 (.0089)	21.983 (6.541)	28.690 (8.726)	8.678 (5.820)	.346
7	.2917 (.0748)	.0037 (.0117)	39.156 (11.743)	45.691 (14.887)	15.883 (10.033)	.376
10	.2916 (.0528)	.0038 (.0092)	56.304 (16.875)	62.621 (18.994)	22.877 (14.190)	.385
			Three-year contracts (n = 3)			
1	.2771 (.0671)	.0119 (.0279)	5.113 (1.508)	21.610 (6.413)	1.501 (2.235)	.143
4	.2736 (.0724)	.0129 (.0358)	21.166 (6.149)	59.307 (18.078)	14.853 (7.711)	.419
7	.2731 (.0146)	.0142 (.0133)	37.527 (10.871)	96.199 (27.439)	27.366 (13.132)	.455
10	.2721 (.0093)	.8632 (.0219)	54.177 (15.722)	111.812 (33.367)	37.532 (17.202)	.482
			Four-year contracts (n = 4)			
1	.2792 (.0765)	.1018 (.0765)	4.925 (1.459)	65.328 (19.023)	2.918 (4.091)	.163
4	.2751 (.0191)	.1223 (.0196)	20.960 (6.124)	216.887 (60.740)	30.304 (14.940)	.449
7	.2749 (.0119)	.2394 (.0550)	37.418 (10.896)	357.690 (100.537)	55.847 (25.584)	.483

[a/] The parameters are computed by maximizing the log-likelihood function (28) with respect to β, γ, and V holding $\delta = \theta = 0$. The correlation of the variances $\rho = V_{12}(V_{11}V_{22})^{-1/2}$.

[b/] The numbers in parentheses are asymptotic standard errors, computed from (29).

Table 7

Unrestricted maximum likelihood estimates (N = 1; δ, $\theta \neq 0$). [a/]

Contract length n (years)	Parameters: Wage shock δ	Demand shock θ	Demand β	Wage γ	Innovation variances: V_{11}	V_{22}	V_{12}	Log-Likelihood function $\mathfrak{L}$
2	.0573 (.2786)[b/]	.8410 (.0815)	.2617 (.0832)	.0052 (.0138)	2.7489 (.8616)	11.3858 (3.3708)	-.8150 (1.3218)	-51.0571
3	.6012 (.8305)	.7493 (.1524)	.2937 (.0788)	.1313 (.2290)	2.8493 (.8594)	13.9760 (4.1496)	.1123 (1.7550)	-45.3254
4	.5367 (.3307)	.8083 (.1259)	.2846 (.0902)	.4380 (.3331)	2.6110 (.8206)	42.4373 (12.3862)	.6461 (2.5174)	-51.4181

[a/] The parameters are computed by maximizing the log-likelihood function (28) with respect to δ, θ, β, γ, V, and $\mathfrak{L}$ with the timing interval index N = 1.

[b/] The numbers in parentheses are asymptotic standard errors, computed from (29).

Appendix A

Proofs of Propositions

This appendix supplies proofs of Propositions 1-3 in section 4.2 of the paper. The proof of Proposition 1 requires two lemmas.

Lemma 1. Suppose

(i) $R(\tau) = \sum_{k=1}^{q} W_k \lambda_k^{\tau}$ for $\tau = 0, 1, 2, \ldots$ are given

(ii) the λ_k's are distinct, but not zero and

(iii) the W_k's are not zero.

Then

(iv) the following m x m matrix is nonsingular:

$$\begin{bmatrix} R(m-1) & R(m-2) & \cdots & R(0) \\ R(m) & R(m-1) & \cdots & R(1) \\ \vdots & \vdots & & \vdots \\ R(2m-2) & R(2m-3) & \cdots & R(m-1) \end{bmatrix}.$$

Proof. Suppose the lemma is false and the matrix is singular. Then there exists a set of numbers--$\alpha_0, \ldots, \alpha_{m-1}$--not all zero, such that

$$\sum_{i=0}^{m-1} \alpha_i R(\tau-i) = 0 \qquad \tau = m-1, m, \ldots, 2m-2.$$

Alternatively,

$$\sum_{k=1}^{m} W_k \lambda_k^{\tau} \alpha(\lambda_k^{-1}) = 0 \qquad \tau = m-1, m, \ldots, 2m-2 \tag{A.1}$$

where $\alpha(\lambda_k^{-1}) = \alpha_0 + \alpha_1 \lambda_k^{-1} + \ldots + \alpha_{m-1} \lambda_k^{-(m-1)}$.

Choose the indexes so that $\alpha(\lambda_k^{-1}) \neq 0$ for $k = 1, \ldots, n$, and if $n < m$, then $\alpha(\lambda_k^{-1}) = 0$ for $n < k \leq m$. Note that n cannot be zero since $\alpha_0 + \alpha_1 z + \ldots + \alpha_{m-1} z^{m-1}$ has at most $m - 1$ zeros. Use this to rewrite (A.1):

$$\sum_{k=1}^{n} W_k \lambda_k^{\tau} \alpha(\lambda_k^{-1}) = 0 \qquad \tau = m - 1, m, \ldots, 2m - 2. \tag{A.2}$$

The following $n - 1$ steps successively eliminate $W_1\lambda_1^{\tau}$, $W_2\lambda_2^{\tau}$, $\ldots$, $W_{n-1}\lambda_{n-1}^{\tau}$ from the first n equations in (A.2).

Step 1

a. Divide (A.2) by λ_1^{τ} to get

$$W_1 \alpha(\lambda_1^{-1}) + \sum_{k=2}^{n} W_k \left(\frac{\lambda_k}{\lambda_1}\right)^{\tau} \alpha(\lambda_k^{-1}) = 0 \qquad \tau = m - 1, m, \ldots, 2m - 2.$$

b. Subtract the $\tau = m$ equation from the $\tau = m - 1$ equation, the $\tau = m + 1$ equation from the $\tau = m$ equation, and so on, to get

$$\sum_{k=2}^{n} W_k \left(\frac{\lambda_k}{\lambda_1}\right)^{\tau} \left(1 - \frac{\lambda_k}{\lambda_1}\right) \alpha(\lambda_k^{-1}) = 0 \qquad \tau = m - 1, m, \ldots, 2m - 3. \tag{A.3}$$

Step 2

a. Multiply (A.3) by $(\lambda_1/\lambda_2)^{\tau}$ to get

$$W_2 \left(1 - \frac{\lambda_2}{\lambda_1}\right) \alpha(\lambda_2^{-1}) + \sum_{k=3}^{n} W_k \left(\frac{\lambda_k}{\lambda_2}\right)^{\tau} \left(1 - \frac{\lambda_k}{\lambda_1}\right) \alpha(\lambda_k^{-1}) = 0$$

$\tau = m - 1, m, \ldots, 2m - 3.$

b. Subtract the $\tau = m$ equation from the $\tau = m - 1$ equation, the $\tau = m + 1$ equation from the $\tau = m$ equation, and so on, to get

$$\sum_{k=3}^{n} W_k \left(\frac{\lambda_k}{\lambda_2}\right)^{\tau} \left(1 - \frac{\lambda_k}{\lambda_1}\right)\left(1 - \frac{\lambda_k}{\lambda_2}\right) \alpha(\lambda_k^{-1}) = 0 \qquad \tau = m - 1, m, \ldots, 2m - 4.$$

Proceeding in this way eventually yields the following:

Step n - 1

$$W_n\left(\frac{\lambda_n}{\lambda_{n-1}}\right)^{\tau}\left(1-\frac{\lambda_n}{\lambda_1}\right)\ \ldots\ \left(1-\frac{\lambda_n}{\lambda_{n-1}}\right)\alpha(\lambda_n^{-1}) = 0 \qquad (A.4)$$

$\tau = m - 1, m, \ldots, 2m - n - 1.$

By hypothesis, $\alpha(\lambda_n^{-1}) \neq 0$. Therefore, (A.4) implies one or more of these conditions: $W_n = 0$, $\lambda_n = 0$, $\lambda_n = \lambda_i$ for some $i \in \{1,\ldots,n-1\}$. But each of these contradicts the assumptions of Lemma 1. Q.E.D.

Lemma 2. If

(i) $W_k + W_k^T \neq 0$ for $k = 1, \ldots, m$,

then

(ii) there exists a scalar α such that

$$(\alpha,1-\alpha)W_k\begin{pmatrix}\alpha\\1-\alpha\end{pmatrix} \neq 0 \qquad k = 1, \ldots, m.$$

Proof. Define the function $f(\alpha,\ell)$: $R^1 \times \{1,2,\ldots,m\} \to R^1$ as follows:

$$f(\alpha,\ell) = (\alpha,1-\alpha)W_\ell\begin{pmatrix}\alpha\\1-\alpha\end{pmatrix} = \frac{1}{2}(\alpha,1-\alpha)[W_\ell+W_\ell^T]\begin{pmatrix}\alpha\\1-\alpha\end{pmatrix}$$

$\ell = 1, 2, \ldots, m, \alpha \in R^1$.

Under condition (i), no more than two values of α exist such that $f(\alpha,\ell) = 0$. Denote this (possibly empty) set Ω_ℓ for $\ell = 1, \ldots, m$. Let Ω be the union of $\Omega_1, \ldots, \Omega_m$. The value of α referred to by (ii) may be chosen from the set of real numbers, excluding the finite set Ω. Q.E.D.

Proof of Proposition 1. First get ρ_k^N for $k = 1, \ldots, m$. Note that

$$\prod_{i=1}^{m}(1-\rho_i^N L)\sum_{k=1}^{m}\tilde{W}_k\rho_k^{N\tau} = \sum_{k=1}^{m}\tilde{W}_k\left[1-\frac{\rho_k^N}{\rho_k^N}\right]\prod_{\substack{i=1\\i\neq k}}^{m}\left[1-\frac{\rho_i^N}{\rho_k^N}\right]\rho_k^{N\tau} = 0$$

where $L\rho_k^{N\tau} \equiv \rho_k^{N(\tau-1)}$. Let $\tilde{A}(z) \equiv 1 - \tilde{a}_1 z - \tilde{a}_2 z - \ldots - \tilde{a}_m z^m = \Pi_{k=1}^{m}(1-\rho_k^N z)$. Then

$$R_{\underline{y}}(\tau,\phi) = \tilde{a}_1 R_{\underline{y}}(\tau-1,\phi) + \ldots + \tilde{a}_m R_{\underline{y}}(\tau-m,\phi) \qquad (A.5)$$

$$\tau = m + 2, m + 3, \ldots, 2m + 1.$$

[The $R_{\underline{y}}$'s are defined in the paper's eqs. (26a)-(26d).]

Since $\tilde{W}_k$ is proportional to W_k [eq. (26d)], condition (i) implies that (i) of Lemma 2 holds. Thus, a 2 x 1 vector γ can be chosen so that $\gamma^T \tilde{W}_k \gamma \neq 0$ for $k = 1, \ldots, m$. Write $R(\tau) \equiv \gamma^T R_{\underline{y}}(\tau,\phi)\gamma$ for $\tau = m + 2, m + 3, \ldots, 2m + 1$. Then pre- and postmultiplying (A.5) by γ^T and γ, respectively, yields $R(\tau) = \tilde{a}_1 R(\tau-1) + \ldots + \tilde{a}_m R(\tau-m)$ for $\tau = m + 2, m + 3, \ldots, 2m + 1$. In matrix notation:

$$\begin{pmatrix} R(m+2) \\ R(m+3) \\ \vdots \\ R(2m+1) \end{pmatrix} = \begin{bmatrix} R(m+1) & \cdots & R(2) \\ R(m+2) & \cdots & R(3) \\ \vdots & & \vdots \\ R(2m) & \cdots & R(m+1) \end{bmatrix} \begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_m \end{pmatrix}.$$

Condition (ii) and the fact that $\gamma^T \tilde{W}_k \gamma \neq 0$ for $k = 1, \ldots, m$ guarantee Lemma 1, so the matrix to the right of the equality possesses a unique inverse. Consequently, the above equation can be solved uniquely for $\tilde{a}_1, \ldots, \tilde{a}_m$. The ρ_k^N's are calculated to satisfy

$$\tilde{A}(z) = \prod_{k=1}^{m} (1-\rho_k^N z).$$

The $\tilde{W}_k$'s are obtained from

$$\begin{pmatrix} R_{\underline{y}}(2,\phi) \\ \vdots \\ R_{\underline{y}}(m+1,\phi) \end{pmatrix} = \begin{bmatrix} \rho_1^{2N_I} & \cdots & \rho_m^{2N_I} \\ \vdots & & \vdots \\ \rho_1^{(m+1)N_I} & \cdots & \rho_m^{(m+1)N_I} \end{bmatrix} \begin{pmatrix} \tilde{W}_1 \\ \vdots \\ \tilde{W}_m \end{pmatrix}.$$

The matrix to the right of this equality also has a unique inverse, if the ρ_k's are distinct, as required by (ii). Q.E.D.

Proof of Proposition 2. Suppose a set ρ_1, ρ_2, ρ_3 is available which satisfies the paper's eq. (11); that is,

$$-\frac{4+\beta\gamma}{4-3\beta\gamma} = \lambda[1 + (\rho_1+\rho_2+\rho_3)^2 + (\rho_1\rho_2+\rho_2\rho_3+\rho_1\rho_2)^2 + (\rho_1\rho_2\rho_3)^2] \tag{A.6}$$

$$\frac{1}{4} = \lambda[-(\rho_1+\rho_2+\rho_3)(1+\rho_1\rho_2+\rho_2\rho_3+\rho_1\rho_3) - (\rho_1\rho_2+\rho_2\rho_3+\rho_1\rho_3)(\rho_1\rho_2\rho_3)] \tag{A.7}$$

$$\frac{1}{6} = \lambda[\rho_1\rho_2 + \rho_2\rho_3 + \rho_1\rho_3 + (\rho_1+\rho_2+\rho_3)(\rho_1\rho_2\rho_3)] \tag{A.8}$$

$$\frac{1}{12} = -\lambda\rho_1\rho_2\rho_3 \tag{A.9}$$

where $A(B) = (1-\rho_1B)(1-\rho_2B)(1-\rho_3B) = 1 - a_1B - a_2B^2 - a_3B^3$ and $b(B) = -[(4+\beta\gamma)/(4-3\gamma\beta)] + (1/4)(B+B^{-1}) + (1/6)(B^2+B^{-2}) + (1/12)(B^3+B^{-3})$. Here

$$D(\phi) = \{\rho_1,\rho_2,\rho_3;-\rho_1,\rho_2,\rho_3;\rho_1,-\rho_2,\rho_3;\rho_1,\rho_2,-\rho_3;$$

$$-\rho_1,-\rho_2,\rho_3;\rho_1,-\rho_2,-\rho_3;-\rho_1,\rho_2,-\rho_3;-\rho_1,-\rho_2,-\rho_3\}.$$

I show that every element of D except ρ_1, ρ_2, ρ_3 violates one of (A.6)-(A.9). Consider three cases:

Case 1. Replace (ρ_1,ρ_2,ρ_3) by $(-\rho_1,\rho_2,\rho_3)$. Divide (A.8) by (A.9) to get

$$2 = \frac{\rho_1\rho_2 + \rho_2\rho_3 + \rho_1\rho_3}{\rho_1\rho_2\rho_3} - (\rho_1+\rho_2+\rho_3). \tag{A.10}$$

Suppose $-\rho_1$, ρ_2, ρ_3 satisfy (A.6)-(A.9). Then

$$2 = \frac{-\rho_1\rho_2 + \rho_2\rho_3 - \rho_1\rho_3}{\rho_1\rho_2\rho_3} - (-\rho_1+\rho_2+\rho_3). \qquad (A.11)$$

Subtract (A.11) from (A.10) to get $0 = (-2/\rho_1) - 2\rho_1$, or $\rho_1 = i$, the imaginary number. This is a contradiction, since ρ_1, ρ_2, and ρ_3 lie inside the unit circle.

Case 2. Replace (ρ_1,ρ_2,ρ_3) by $(-\rho_1,-\rho_2,\rho_3)$. Divide (A.9) into (A.7) to get

$$3 = \frac{(\rho_1+\rho_2+\rho_3)(1+\rho_1\rho_2+\rho_2\rho_3+\rho_1\rho_3)}{\rho_1\rho_2\rho_3} + (\rho_1\rho_2+\rho_2\rho_3+\rho_1\rho_3). \qquad (A.12)$$

If (A.6)-(A.9) hold for $(-\rho_1,-\rho_2,\rho_3)$, then

$$3 = \frac{(-\rho_1-\rho_2+\rho_3)(1+\rho_1\rho_2-\rho_2\rho_3-\rho_1\rho_3)}{\rho_1\rho_2\rho_3} + (\rho_1\rho_2-\rho_2\rho_3-\rho_1\rho_3). \qquad (A.13)$$

Subtract (A.13) from (A.12) to get

$$0 = 2(\rho_2\rho_3+\rho_1\rho_3) + \frac{2(\rho_1+\rho_2)(1+\rho_1\rho_2) + 2\rho_3(\rho_2\rho_3+\rho_1\rho_3)}{\rho_1\rho_2\rho_3}$$

or $0 = 2(1+\rho_3^2)(1+\rho_1\rho_2)$, which implies $\rho_3 = i$, a contradiction. This rules out the case in which the sign on any two of ρ_1, ρ_2, ρ_3 can be changed.

Case 3. Replace (ρ_1,ρ_2,ρ_3) by $(-\rho_1,-\rho_2,-\rho_3)$. If $(-\rho_1,-\rho_2,-\rho_3)$ satisfy (A.6)-(A.9), then they satisfy (A.10), so that

$$2 = -\frac{\rho_1\rho_2 + \rho_2\rho_3 + \rho_1\rho_3}{\rho_1\rho_2\rho_3} + (\rho_1+\rho_2+\rho_3). \qquad (A.14)$$

Adding (A.10) to (A.14) gives $4 = 0$, a contradiction.

Conclude that the only element in $D(\phi)$ which satisfies the restrictions of Taylor's model, (11), is (ρ_1,ρ_2,ρ_3). Q.E.D.

Proof of Proposition 3. Suppose the contrary. That is, there exist ϕ, $\Phi' \varepsilon \Phi$ for $\phi \neq \phi'$ such that $\{K(\phi'),\tilde{K}(\phi'),W_k(\phi'),\rho_k(\phi');k=1,\ldots,m\} = \{K(\phi),\tilde{K}(\phi), W_k(\phi),\rho_k(\phi);k=1,\ldots,m\}$. Then there must be no $\tau \varepsilon \{0,\pm 1/N,\pm 2/N,\ldots\}$ such that $R_y(\tau,\phi') \neq R_y(\tau,\phi)$. But this contradicts condition (i). Q.E.D.

Appendix B

U.S. Data Used in the Paper

Year	Real gross national product		Consumer price index	
	Log (y_t^*)	Detrended log (y_t)	Log (p_t^*)	Detrended log (p_t)
1954	6.420	-0.309	-0.217	9.529
1955	6.484	2.688	-0.221	6.032
1956	6.505	1.318	-0.206	4.392
1957	6.523	-0.375	-0.171	4.769
1958	6.521	-4.066	-0.144	4.336
1959	6.580	-1.707	-0.136	2.017
1960	6.602	-2.942	-0.120	0.484
1961	6.627	-3.947	-0.110	-1.631
1962	6.683	-1.796	-0.099	-3.645
1963	6.722	-1.403	-0.087	-5.563
1964	6.774	0.238	-0.074	-7.387
1965	6.831	2.475	-0.057	-8.804
1966	6.889	4.770	-0.028	-9.111
1967	6.915	3.970	0.000	-9.395
1968	6.958	4.767	0.041	-8.405
1969	6.984	3.816	0.093	-6.295
1970	6.980	0.006	0.151	-3.668
1971	7.010	-0.529	0.193	-2.583
1972	7.066	1.569	0.226	-2.463
1973	7.119	3.396	0.286	0.452
1974	7.105	-1.492	0.390	7.736
1975	7.092	-6.259	0.477	13.358
1976	7.148	-4.188	0.534	15.843

Lucas, Robert E., Jr., 1973, Some international evidence on output-inflation tradeoffs, American Economic Review 63, 326-334.

Mundlak, Yair, 1961, Aggregation over time in distributed lag models, International Economic Review 2, 154-163.

Pham-Dinh, Tuan, 1978, Estimation of parameters in the ARMA model when the characteristic polynomial of the MA operator has a unit zero, Annals of Statistics 6, 1369-1389.

Phillips, P. C. B., 1973, The problem of identification in finite parameter continuous time models, Journal of Econometrics 1, 351-362.

Rozanov, Yu A., 1967, Stationary random processes (Holden-Day, San Francisco).

Sargan, J. D. and Alok Bhargava, 1983, Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle, Econometrica 51, 799-820.

Sargent, Thomas J. and Neil Wallace, 1975, "Rational" expectations, the optimal monetary instrument, and the optimal money supply rule, Journal of Political Economy 83, 241-254.

Sims, Christopher A., 1971, Discrete approximations to continuous time distributed lags in econometrics, Econometrica 39, 545-563.

Taylor, John B., 1980a, Output and price stability: An international comparison, Journal of Economic Dynamics and Control 2, 109-132.

Taylor, John B., 1980b, Aggregate dynamics and staggered contracts, Journal of Political Economy 88, 1-23.

Zellner, Arnold, 1968, Note on the effect of temporal aggregation on estimation of a stock adjustment model, Manuscript (Graduate School of Business, University of Chicago).

Zellner, Arnold and Claude Montmarquette, 1971, A study of some aspects of temporal aggregation problems in econometric analyses, Review of Economics and Statistics 53, 335-342.