A dynamic linear demand schedule for labor is estimated and tested. The hypothesis of rational expectations and assumptions about the orders of the Markov processes governing technology impose overidentifying restrictions on a vector autoregression for straight-time employment, overtime employment, and the real wage. The model is estimated by the full-information maximum-likelihood method. The model is used as a vehicle for reexamining some of the paradoxical cyclical behavior of real wages described in the famous Dunlop–Tarshis–Keynes exchange.

Both Keynes (1939) and various classical writers asserted that real wages would move countercyclically as employers moved along downward-sloping demand schedules relating the employment-capital ratio to the real wage. Dunlop (1938) and Tarshis (1939) described evidence which they interpreted as failing to confirm a countercyclical pattern of real-wage movements. That and much subsequent evidence on the question, which is reviewed and extended by Bodkin (1969), consisted mostly of simple contemporaneous regressions between real wages and some measure of the stage of the business cycle. By and large that evidence was regarded as rejecting the view that the data can be described as observations falling along an aggregate demand schedule for employment. This view of the evidence in large measure stimulated attempts to describe aggregate employment and real wages by “disequilibrium models,” the work of Barro and Grossman (1971) and Solow and Stiglitz (1968) being two prominent examples.

This paper aims to provide a framework for reexamining some of this...
evidence within the context of a stochastic and dynamic aggregate demand schedule for labor. The old evidence is simply not decisive because the view that the aggregate data lie along the type of demand schedule considered in this paper places no restrictions on the simple contemporaneous regressions in the studies summarized by Bodkin (1969); however, under certain conditions, that view does place restrictions on aggregate real wages and employment as a vector stochastic process. The plan of this paper is to extract and test these implications.

This paper starts from the findings of the recent paper by Salih Neftci (1978), which computed long two-sided distributed lags between aggregate employment and real wages for post–World War II data for the United States. Neftci found that there were complicated and economically significant dynamic interactions between real wages and employment and that there was much stronger evidence for Granger (1969) causality flowing from real wages to employment than for Granger causality in the reverse direction. Further, the influence of real wages on employment was predominantly negative.

To represent Neftci’s findings in a slightly different form than he did, table 1 reports estimates of a fourth-order bivariate autoregression for quarterly aggregate measures of real wages $w$ and employment $n_1$, both seasonally unadjusted. The theory of vector autoregressions and moving averages is reviewed briefly in the Appendix. The data are a straight-
TABLE 2
VECTOR-MOVING AVERAGE REPRESENTATION OF REAL WAGE AND AGGREGATE
EMPLOYMENT  
(1948I–1972IV)*

| Lag | 1  | 2   | 3   | 4   |
|-----|----|-----|-----|-----|
| 0   | .3697 | 0   | 0   | .0150 |
| 1   | .5897 | -.00126 | -.0231 | .0146 |
| 2   | .5946 | -.00177 | -.0287 | .0140 |
| 3   | .5464 | -.00253 | -.0840 | .0149 |
| 4   | .5025 | -.00329 | -.1553 | .0139 |
| 5   | .4444 | -.00359 | -.2103 | .0130 |
| 6   | .3741 | -.00370 | -.2580 | .0123 |
| 7   | .3080 | -.00371 | -.2983 | .0115 |
| 8   | .2520 | -.00359 | -.3262 | .0108 |
| 9   | .2048 | -.00339 | -.3426 | .0102 |
| 10  | .1661 | -.00316 | -.3493 | .0096 |
| 11  | .1357 | -.00291 | -.3480 | .0091 |
| 12  | .1125 | -.00266 | -.3406 | .0086 |
| 13  | .0950 | -.00243 | -.3285 | .0081 |
| 14  | .0820 | -.00221 | -.3135 | .0076 |
| 15  | .0724 | -.00202 | -.2966 | .0071 |
| 16  | .0652 | -.00184 | -.2789 | .0067 |
| 17  | .0597 | -.00169 | -.2611 | .0063 |
| 18  | .0554 | -.00155 | -.2436 | .0059 |
| 19  | .0519 | -.00143 | -.2269 | .0055 |
| 20  | .0488 | -.00132 | -.2110 | .0051 |
| 21  | .0460 | -.00123 | -.1962 | .0048 |
| 22  | .0434 | -.00114 | -.1824 | .0045 |
| 23  | .0410 | -.00106 | -.1696 | .0042 |

Note.—Col. 1: Response of employment to 1 SD innovation in employment; col. 2: Response of real wage to 1 SD innovation in employment; col. 3: Response of employment to 1 SD innovation in real wage; col. 4: Response of real wage to 1 SD innovation in real wage. Correlation of innovations in employment and real wage is .2442.

* Observation period for left-hand-side variables. For method of construction of vector-moving average, see Appendix.

time wage index in manufacturing divided by the consumer price index, measured in 1967 dollars, and number of employees on nonagricultural payrolls, measured in millions of men. The data are described more in Section 3 below. The $F$-statistic pertinent for testing the null hypothesis that lagged real wages have zero coefficients in the vector autoregression for employment has a marginal significance level of .091. The $F$-statistic pertinent for testing the hypothesis that lagged levels of employment have zero coefficients in the vector autoregression for the real wage has a marginal significance level of .869.1 This pattern is consistent with Neftci's finding much stronger evidence of Granger causality extending from real wages to employment than in the other direction.

Table 2 reports estimates of the moving average representation implied by the autoregressions in table 1. Table 2 depicts the matrix of responses

1 For data on the left-side variable extending from 1951I–1972IV, which more closely matches Neftci's period than mine, the marginal significance level for testing the null hypothesis that real wages do not Granger-cause employment is .0745, and for the null hypothesis that employment does not Granger-cause the real wage the marginal significance level is .5012. These autoregressions included constant, trend, and three seasonal dummies.
to one-standard-deviation innovations in the real wage and employment respectively. A one-standard-deviation innovation in employment leads to a strong, sustained increase in employment and a small (relative to the response to its own innovation) sustained decrease in the real wage. A one-standard-deviation innovation in the real wage leads to a sustained and sizable decrease in employment and a sustained and sizable increase in the real wage. The response of employment to the real-wage innovation is of the same order of magnitude as it is to its own innovation, in contrast to the response of the real wage to the employment innovation. The magnitude of the estimated response of employment to real-wage innovations seems of substantial economic significance.

Tables 3 and 4 report two alternative decompositions of the variances of the $k$-step-ahead forecast errors of the $(n_1, w)$ process into parts attributable to variance in the "orthogonalized innovations" in employment and the real wage. As indicated in the Appendix, these decompositions are not unique, which accounts for the two tables. However, since the innovations in employment and the real wage in table 1 have only a moderate correlation of .2442, the differences between the decompositions in tables 2 and 3 are bound to be modest, as they are. The tables reveal that a substantial percentage (40 or 48) of the 35-quarter-ahead forecast-error variance in employment (which approximates the steady-state variance in the indeterministic part of employment) is accounted for by innovations in the real wage. Only a small percentage (1 or 6) of the 35-quarter-ahead forecast-error variance in the real wage is accounted for by the innovation in employment.

Two characteristics of these results are particularly important for purposes of this study. First, there do appear to be some complicated dynamic interactions between aggregate employment and these real-wage data that might be susceptible to analysis with a dynamic model of the demand for employment. Second, these data seem to be consistent with the assumption that the real wage is not Granger-caused by employment. This assumption, which will be imposed below, substantially simplifies the modeling task.

The plan of this paper is to estimate a dynamic aggregative demand schedule for employment for postwar U.S. data. While the demand model makes employment depend inversely on the appropriate real wage, as does the static theory, a potentially rich dynamic structure is introduced into that dependence because firms are assumed to face costs of rapidly adjusting their labor force and so find it optimal to take into account future expected values of the real wage in determining their current employment. The model imposes overidentifying restrictions on a vector of stochastic processes composed of employment, a measure of overtime employment, and the real wage. The aim is to test the adequacy of these overidentifying restrictions.
### Table 3

**Decompositions of Variance of Forecast Errors**

| | Variance of k-step-ahead Forecast Error Explained by Orthogonalized Innovation in: |
|---|---|---|
| | Employment | Real Wage |
| Employment: | | |
| k = 1 | .1367 | 94.03 | 5.96 |
| k = 2 | .4783 | 95.24 | 4.75 |
| k = 3 | .8244 | 95.59 | 4.40 |
| k = 4 | 1.1076 | 96.50 | 3.49 |
| k = 5 | 1.3462 | 97.04 | 2.95 |
| k = 6 | 1.5423 | 96.74 | 3.25 |
| k = 7 | 1.7017 | 95.41 | 4.58 |
| k = 8 | 1.8407 | 93.06 | 6.93 |
| k = 9 | 1.9705 | 89.96 | 10.03 |
| k = 10 | 2.0956 | 86.47 | 13.52 |
| k = 11 | 2.2169 | 82.91 | 17.08 |
| k = 12 | 2.3334 | 79.51 | 20.48 |
| k = 20 | 2.9620 | 64.14 | 35.85 |
| k = 35 | 3.2381 | 59.18 | 40.81 |
| Real Wage: | | |
| k = 1 | .00022 | 0 | 100.00 |
| k = 2 | .00043 | .34 | 99.65 |
| k = 3 | .00062 | .71 | 99.28 |
| k = 4 | .00083 | 1.25 | 98.74 |
| k = 5 | .00101 | 2.02 | 97.97 |
| k = 6 | .00117 | 2.78 | 97.21 |
| k = 7 | .00132 | 3.45 | 96.54 |
| k = 8 | .00145 | 4.04 | 95.95 |
| k = 9 | .00156 | 4.53 | 95.46 |
| k = 10 | .00166 | 4.91 | 95.08 |
| k = 11 | .00175 | 5.20 | 94.79 |
| k = 12 | .00183 | 5.41 | 94.58 |
| k = 20 | .00220 | 5.87 | 94.12 |
| k = 35 | .00238 | 5.91 | 94.08 |

* The orthogonalized innovation in employment here equals the innovation in employment, while the orthogonalized innovation in the real wage equals that part of the innovation in the real wage that is orthogonal to the innovation in employment.

† SE of orthogonalized innovation in employment = .01505.

‡ SE of orthogonalized innovation in real wage = .3586.

The model is formed by blending the costly adjustment model of Lucas (1967), Treadway (1969), and Gould (1968) with Lucas’s static model of overtime work and capacity (1970). The model is formulated so that it delivers linear decision rules relating the demand for straight-time employment and overtime employment each to the real-wage process. The model imposes the rational-expectations hypothesis, since firms are supposed to use the true moments of the real-wage process in forming forecasts. The rational-expectations hypothesis is a main source of the overidentifying restrictions imposed by the model.
### TABLE 4

**Decompositions of Variance of Forecast Errors**

| Variance of k-step-ahead Forecast Error Explained by Orthogonalized Innovation in: | Employment | Real Wage |
|---------------------------------|-----------|----------|
| Employment:                     |           |          |
| $k = 1$                          | .136      | 100.0    |
| $k = 2$                          | .478      | 99.8     |
| $k = 3$                          | .824      | 99.8     |
| $k = 4$                          | 1.107     | 99.2     |
| $k = 5$                          | 1.346     | 97.7     |
| $k = 6$                          | 1.542     | 95.3     |
| $k = 7$                          | 1.701     | 92.0     |
| $k = 8$                          | 1.840     | 88.1     |
| $k = 9$                          | 1.970     | 83.8     |
| $k = 10$                         | 2.095     | 79.5     |
| $k = 11$                         | 2.216     | 75.4     |
| $k = 12$                         | 2.333     | 71.8     |
| $k = 20$                         | 2.962     | 56.6     |
| $k = 35$                         | 3.238     | 51.7     |
| Real wage:                      |           |          |
| $k = 1$                          | .00022    | 5.96     |
| $k = 2$                          | .00043    | 4.34     |
| $k = 3$                          | .00062    | 3.47     |
| $k = 4$                          | .00083    | 2.74     |
| $k = 5$                          | .00101    | 2.24     |
| $k = 6$                          | .00117    | 1.95     |
| $k = 7$                          | .00132    | 1.78     |
| $k = 8$                          | .00145    | 1.67     |
| $k = 9$                          | .00156    | 1.61     |
| $k = 10$                         | .00166    | 1.56     |
| $k = 11$                         | .00175    | 1.52     |
| $k = 12$                         | .00183    | 1.48     |
| $k = 20$                         | .00220    | 1.26     |
| $k = 35$                         | .00238    | 1.17     |

*The orthogonalized innovation in the real wage here just equals the innovation in the real wage, while the orthogonalized innovation in employment equals that part of the employment innovation that is orthogonal to the innovation in the real wage.

† SE of orthogonalized innovation in employment = .0146.

‡ SE of orthogonalized innovation in real wage = .3697.

---

In addition to providing some new evidence in the Dunlop-Tarshis tradition, this paper illustrates a technology for maximum-likelihood estimation of decision rules under the hypothesis that expectations are rational. That technology potentially has a variety of applications.²

²Applications of related methods are contained in Sargent (1977, 1978a). John Taylor (1978) uses a minimum-distance estimator to estimate a macroeconomic model subject to rational-expectations restrictions.
I. The Demand for Employment

The model is formed by taking Lucas's model of overtime work and capacity (1970) and amending it to permit potentially different adjustment costs to be associated with rapidly changing straight-time and overtime labor.\(^3\) It is widely asserted that it is much cheaper to adjust the overtime labor force quickly than it is to adjust the straight-time labor force; consequently, it is alleged that overtime labor responds rapidly to the market signals that the firm receives, while the straight-time labor force responds more sluggishly. The model is designed to represent this phenomenon and to provide a framework for estimating its dimensions and testing it.

I shall work with a representative firm, although as I shall remark below, the model can handle certain kinds of diversity across firms. Following Lucas, suppose that the representative firm faces the instantaneous production function:

\[
y(t + \tau) = f[n(t + \tau), k(t + \tau)]; f_n, f_k, f_{nk} > 0; f_{nn}, f_{kk} < 0
\]

\[
t = 0, 1, 2, 3, \ldots
\]

\[
\tau \in [0, 1).
\]

Here \(y(t + \tau)\) is the rate of output per unit time at instant \(t + \tau\), \(n(t + \tau)\) is the number of employees at instant \(t + \tau\), and \(k(t + \tau)\) is the stock of capital at \(t + \tau\). The length of the "day" is one, so that \(t\) indexes days and \(\tau\) indexes moments within the day. The firm is assumed to have a constant capital stock over the day so that \(k(t + \tau) = k(t) \equiv k_i\) for \(\tau \in [0, 1)\). The firm is assumed to be able to hire workers for a straight-time shift of fixed length \(h_1 < 1\) at the real wage \(w_i\) during day \(t\). During the overtime shift of length \(h_2 = 1 - h_1\), the firm can hire all the labor it wants during day \(t\) at the real wage \(pw_i\), where \(p \approx 1.5\) is an overtime premium. Thus, for the first \(h_1\) moments of day \(t\) the firm must pay workers \(w_i\), while for the remaining \(h_2\) moments it must pay \(pw_i\). Confronted with these market opportunities, it is optimal for the firm to choose to set \(n(t + \tau) = n_{1t}\) for \(\tau \in [0, h_1]\) and \(n(t + \tau) = n_{2t}\) for \(\tau \in (h_1, 1)\). That is, it is optimal for the firm to choose a single level of straight-time employment \(n_{1t}\) during \(t\) and a single level of overtime employment of \(n_{2t}\) during the day \(t\).

\(^3\) Restrictions on the production function required to permit Lucas's static model to account for the cyclical behavior of labor productivity and real average hourly earnings were discussed by Sargent and Wallace (1974). Adding differential costs for adjusting straight-time and overtime labor would widen the class of production functions that could lead to procyclical movements of average hourly earnings and labor productivity.
The firm's output over the "day" is then
\[ y_t = \int_0^1 y(t + \tau) \, d\tau \]
\[ = h_1 f(n_{1t}, k_t) + h_2 f(n_{2t}, k_t). \]

I take several steps to specialize this setup further. First, to simplify things, I assume that capital is constant over time so that \( k_t \) can be dropped as an argument from \( f(\cdot, \cdot) \). (In the econometric work below, steps are taken to detrend the data prior to estimation partly in order to minimize the damage caused by this approximation.) Second, I assume a quadratic production function and write instantaneous output on the first and second shifts as
\[ f(n_{1t}, k) = (f_0 + a_1) n_{1t} - \frac{(f_1/2)}{2} n_{1t}^2 \]
\[ f(n_{2t}, k) = (f_0 + a_2) n_{2t} - \frac{(f_1/2)}{2} n_{2t}^2 \]
where \( f_0, f_1 > 0 \), and where \( a_1 \) and \( a_2 \) are exogenous stochastic processes affecting productivity of straight-time and overtime employment. I assume that \( E a_{1t} = E a_{2t} = 0 \). The stochastic processes \( a_{1t} \) and \( a_{2t} \) will be required to satisfy certain regularity conditions to be specified below.

The firm is assumed to bear daily costs of adjusting its straight-time labor force of \( (d/2)(n_{1t} - n_{1t-1})^2 \) and to bear daily costs of adjusting its overtime labor force of \( (e/2)(n_{2t} - n_{2t-1})^2 \). It is widely believed that it is substantially more expensive to adjust the straight-time labor force so that \( d \gg e \). The firm faces an exogenous stochastic process for the real wage \((w_t)\). The firm's straight-time and overtime wage bills are, respectively, \( w_t h_1 n_{1t} \) and \( pw_t h_2 n_{2t} \).

The firm chooses contingency plans for \( n_{1t} \) and \( n_{2t} \) to maximize its expected real present value:
\[ v_t = E_t \sum_{j=0}^{\infty} b^j \left[ (f_0 + a_{1t+j} - w_{t+j}) h_1 n_{1t+j} - \frac{(f_1/2)}{2} h_1 n_{1t+j}^2 \right. \]
\[ - \frac{d}{2} (n_{1t+j} - n_{1t+j-1})^2 + (f_0 + a_{2t+j} - pw_{t+j}) h_2 n_{2t+j} \]
\[ - \frac{(f_1/2)}{2} h_2 n_{2t+j}^2 - \frac{e}{2} (n_{2t+j} - n_{2t+j-1})^2 \left. \right] \]
\[ f_{0t}, f_1, d, e > 0, p > 1, 0 < b < 1, \]

Optimization problems of this form are discussed by Holt, Modigliani, Muth, and Simon (1960), Graves and Telser (1972), and Kwakernaak and Sivan (1972). The treatment here closely follows that of Sargent (1978b). It would be straightforward to carry along \( n \) firms, each facing the same wage process and operating under the same functional form for its objective function (1), yet each having different values for the parameters \( f_0, f_1, d, \) and \( e \). It would then be straightforward to aggregate the Euler equations and their solutions (7). Thus, assuming a representative firm is only a convenience, as the model admits a tidy theory of aggregation.
where $n_{1t-1}$ and $n_{2t-1}$, as well as the stochastic processes for $w$, $a_1$, and $a_2$, are given to the firm. Here $b$ is a real discount factor that lies between zero and one. The operator $E_t$ is defined by $E_t x = E x | \Omega_t$, where $x$ is a random variable, $E$ is the mathematical expectation operator, and $\Omega_t$ is an information set available to the firm at time $t$. I assume that $\Omega_t$ includes at least $\{n_{1t-1}, n_{2t-1}, a_{1t}, a_{1t-1}, \ldots, a_{2n}, a_{2t-1}, \ldots, w_t, w_{t-1}, \ldots\}$. The firm is assumed to maximize (1) by choosing stochastic processes for $n_1$ and $n_2$ from the set of stochastic processes that are (nonanticipative) functions of the information set $\Omega_t$. (Below, I will further restrict the class of stochastic processes over which the optimization is carried out.) I assume that the stochastic processes $w_t, a_{1t}$, and $a_{2t}$ are of exponential order less than $1/b$, which means that for some $K > 0$ and some $x$ such that $1 < x < 1/b$, $|E_t w_{t+j}| < K(x)^{-j+t}$, $|E_t a_{1t+j}| < K(x)^{-j+t}$, and $|E_t a_{2t+j}| < K(x)^{-j+t}$, for all $t$ and all $j \geq 0$. I further assume that all random variables have finite first- and second-order moments.

First-order necessary conditions for the maximization of (1) consist of a set of “Euler equations” and a pair of transversality conditions. The Euler equations for $\{n_{1t}\}$ and $\{n_{2t}\}$ are

$$bE_{t+j} n_{1t+j+1} + \phi_1 n_{1t+j} + n_{1t+j-1} = (h_1/d)(w_{t+j} - a_{1t+j} - f_0)$$

$$j = 0, 1, 2, \ldots, \quad (2)$$

$$bE_{t+j} n_{2t+j+1} + \phi_2 n_{2t+j} + n_{2t+j-1} = (h_2/e)(p w_{t+j} - a_{2t+j} - f_0)$$

$$j = 0, 1, 2, \ldots, \quad (3)$$

where

$$\phi_1 = -[(f_1 h_1/d) + (1 + b)]$$

$$\phi_2 = -[(f_1 h_2/e) + (1 + b)]. \quad (4)$$

The transversality conditions are

$$\lim_{T \to \infty} b^T E_{t+T} n_{1t+T} = \lim_{T \to \infty} b^T E_{t+T} n_{2t+T} = 0. \quad (5)$$

To solve the Euler equations for the optimum contingency plans, first obtain the factorizations

$$\left(1 + \frac{\phi_1}{b} z + \frac{1}{b} z^2\right) = (1 - \delta_1 z)(1 - \delta_2 z), \quad (6)$$

$$\left(1 + \frac{\phi_2}{b} z + \frac{1}{b} z^2\right) = (1 - \mu_1 z)(1 - \mu_2 z). \quad (7)$$

Given the assumptions about the signs and magnitudes of the parameters composing $b$, $\phi_1$, and $\phi_2$, it follows that factorizations exist with $0 < \delta_1$

---

5 See Sargent (1978b), chaps. 9 and 14.
< 1 < (1/b) < δ_2 and 0 < μ_1 < 1 < (1/b) < μ_2. It then follows that solutions of the Euler equations that satisfy the transversality conditions and the initial conditions are given by

\[ n_{1t} = \delta_1 n_{1t-1} - (\delta_1 h_1 / d) \sum_{i=0}^{\infty} \left( \frac{1}{\delta_2} \right)^i E_t(w_{t+i} - a_{1t+i} - f_0), \]  

(7a)

\[ n_{2t} = \mu_1 n_{2t-1} - (\mu_1 h_2 / e) \sum_{i=0}^{\infty} \left( \frac{1}{\mu_2} \right)^i E_t(pw_{t+i} - a_{2t+i} - f_0). \]  

(7b)

It can be verified directly that these solutions satisfy the Euler equations and the transversality conditions. The polynomial equation (5) implicitly defines δ_1 and δ_2 as functions of (f_1 h_1 / d). By studying this polynomial, it is possible to show that δ_1 is a decreasing function of (f_1 h_1 / d) and that (1/δ_2) = b δ_1. It follows that δ_1 and (1/δ_2) both increase with increases in the adjustment-cost parameter d. Reference to equation (7a) then shows that increases in the adjustment-cost parameter d, by increasing δ_1 and (1/δ_2), decrease the speed with which the firm responds to the real-wage and productivity signals that it receives. Similarly, μ_1 and (1/μ_2) are decreasing functions of (f_1 h_2 / e) and (1/μ_2) = b μ_1.

Equations (7a) and (7b) are decision rules for setting n_{1t} and n_{2t} as linear functions of n_{1t-1}, n_{2t-1}, and the conditional expectations E_t m_{t+i}, E_t a_{1t+i}, and E_t a_{2t+i}, i = 0, 1, 2, . . . . However, in general, these conditional expectations are nonlinear functions of the information in Ω_t. Given particular stochastic processes for w_t, a_{1t}, and a_{2t}, equations (7a) and (7b) can be solved for decision rules expressing n_{1t} and n_{2t} as in general, nonlinear functions of Ω_t.

For the purposes of empirical work, it is convenient to restrict ourselves to the class of decision rules that are linear functions of Ω_t. The optimal linear decision rules can be obtained by replacing the conditional mathematical expectations in (7a) and (7b) with the corresponding linear least-squares projections on the information set Ω_t. Accordingly, henceforth, in all forecasting formulas, I will replace the mathematical expectation operator E by the linear least-squares projection operator ̂E.  

To derive from (7a) and (7b) explicit decision rules for n_{1t} and n_{2t} as functions of Ω_t, it is necessary further to restrict the stochastic processes w_t, a_{1t}, and a_{2t}. I assume that a_{1t} and a_{2t} are each first-order Markov processes for which

\[ ̂E_t a_{1t+i} = \rho_1^i a_{1t}, \quad i \geq 0 \]

\[ ̂E_t a_{2t+i} = \rho_2^i a_{2t}, \quad i \geq 0, \]  

(8)

6 See Sargent (1978b).

7 See Sargent (1978b). The solution (7) clearly exhibits the certainty-equivalence or separation property. That is, the same solution for n_{1t} and n_{2t} would emerge if we maximized the criterion formed by replacing (a_{1t+j}, a_{2t+j}, w_{t+j}) by (E_t a_{1t+j}, E_t a_{2t+j}, E_t w_{t+j}) and dropping the operator ̂E, from outside the sum in (1).

8 In the statistical literature the linear least-squares projection operator ̂E is often referred to as the "wide sense expectation" operator.
where \( |\rho_1| < 1/b, |\rho_2| < 1/b \). That is, I assume that \( a_{1t} \) and \( a_{2t} \) are generated by the stochastic processes

\[
\begin{align*}
a_{1t} &= \rho_1 a_{1t-1} + \xi_{1t}, \\
a_{2t} &= \rho_2 a_{2t-1} + \xi_{2t},
\end{align*}
\]

where \( \xi_{1t} \) and \( \xi_{2t} \) are least-squares residuals with finite variances and \( \mathbb{E}[\xi_{1t} | \Omega_{t-1}] = \mathbb{E}[\xi_{2t} | \Omega_{t-1}] = 0 \). Although (9) permits \( \xi_{1t} \) and \( \xi_{2t} \) to be arbitrarily correlated contemporaneously, it does in effect rule out correlation between them at any nonzero lags. I assume that \( w_t \) is an \( n \)-th

order Markov process

\[
w_t = v_0 + v_1 w_{t-1} + v_2 w_{t-2} + \cdots + v_n w_{t-n} + \xi_{3t},
\]

where \( \xi_{3t} \) is a least-squares disturbance that satisfies \( \mathbb{E}[\xi_{3t} | \Omega_{t-1}] = 0 \). The condition that \( \mathbb{E}[\xi_{3t} | \Omega_{t-1}] = 0 \) means that \( \xi_{3t} \) is serially uncorrelated and that \( w_t \) is not caused in Granger’s (1969) sense, by \( n_1 \) or \( n_2 \). That the lack of Granger causality from \( n_1 \) or \( n_2 \) to \( w \) is a workable approximation for the data to be studied here is supported by the empirical results of Neftci (1978), which are summarized above. It is convenient to represent the \( n \)-th-order process (10) as the \( (n+1) \)-vector first-order Markov process, \( x_t = A x_{t-1} + \epsilon_t \), where

\[
x_t = \begin{bmatrix} w_t \\
w_{t-1} \\
w_{t-2} \\
\vdots \\
w_{t-n} \\
1 \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \xi_{3t} \\
0 \\
0 \\
\vdots \\
0 \\
0 \end{bmatrix}
\]

\[
A = \begin{bmatrix}
v_1 & v_2 & \cdots & v_n & v_0 \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 0 \end{bmatrix}.
\]

We can write,

\[
\begin{align*}
x_{t+1} &= A x_t + \epsilon_{t+1}, \\
x_{t+2} &= A^2 x_t + \epsilon_{t+2} + A \epsilon_{t+1} \\
\vdots \\
x_{t+j} &= A^j x_t + \epsilon_{t+j} + A \epsilon_{t+j-1} + \cdots + A^{j-1} \epsilon_{t+1}.
\end{align*}
\]

Since \( \mathbb{E}[\epsilon_{t+k}] = 0 \) for \( k \geq 1 \), we have \( \mathbb{E}_t x_{t+j} = A^j x_t \). Assume that the eigenvalues of \( A \) are distinct so that \( A \) can be written as \( A = \Lambda P^{-1} \), where the columns of \( P \) are the eigenvectors of \( A \) and \( \Lambda \) is the diagonal
matrix whose elements are the eigenvalues of $A$. Then we have $\hat{E}_t x_{t+j} = P A^j P^{-1} x_t$. Finally, let $c$ be the $1 \times (n + 1)$ row vector $(1, 0, 0, \ldots, 0)$ so that $w_t = c x_t$. We thus have that

$$E_t w_{t+j} = c P A^j P^{-1} x_t.$$ (11)

Substituting from (8) and (11) into (7a) gives

$$n_{1t} = \delta_1 n_{1t-1} - \frac{\delta_1 h_1}{d} c P \sum_{i=0}^{\infty} \left( \frac{1}{\delta_2^i} \right)^i P^{-1} x_t$$

$$+ \frac{\delta_1 h_1}{d} \left[ f_0 / \left( 1 - \frac{1}{\delta_2} \right) \right] + \frac{\delta_1 h_1}{d} \left[ 1 / \left( 1 - \frac{\rho_1}{\delta_2} \right) \right] a_{1t}.$$ (12)

Let $\lambda_i$ be the $i$th element of $\Lambda$. Since $\delta_2 = (1/\delta_1 b)$, we have that $|\lambda_i / \delta_2| = |\lambda_i b| < 1$ by virtue of the assumption that $w_t$ is of exponential order less than $1/b$, that is, that $|\lambda_i b| < 1$. Then the infinite sum above converges and we can write

$$n_{1t} = \delta_1 n_{1t-1} - \frac{\delta_1 h_1}{d} c P \left[ \frac{1}{1 - \lambda_i b} \right] P^{-1} x_t$$

$$+ \frac{\delta_1 h_1}{d} \left[ f_0 / \left( 1 - \frac{1}{\delta_2} \right) \right] + \frac{\delta_1 h_1}{d} \left[ 1 / \left( 1 - \frac{\rho_1}{\delta_2} \right) \right] a_{1t},$$ (12)

where $[1/(1 - \lambda_i b)]_{ii}$ is a diagonal matrix with $[1 - (\lambda_i b)]^{-1}$ as the $i$th diagonal element.

Let us write (12) as

$$n_{1t} = \delta_1 n_{1t-1} + \alpha_1 w_t + \alpha_2 w_{t-1} + \cdots + \alpha_n w_{t-n+1} + \alpha_0$$

$$+ \frac{\delta_1 h_1}{d} \left( \frac{f_0}{1 - \delta_1 b} \right) + a'_{1t},$$ (13)

where

$$(\alpha_1, \alpha_2, \cdots, \alpha_n, \alpha_0) = - \frac{\delta_1 h_1}{d} c P \left[ \frac{1}{1 - \lambda_i b} \right] P^{-1}$$

$$a'_{1t} = \frac{\delta_1 h_1}{d} \left( \frac{1}{1 - \rho_1 \delta_1 b} \right) a_{1t}.$$ (14)

The assumption that $w_t$ is of exponential order less than $(1/b)$ implies that the max $|\lambda_i| < (1/b)$, where $\lambda_i$ is the $i$th element of $\Lambda$.

Here I am using that

$$\left( \sum_{i=0}^{\infty} \left( \frac{1}{\delta_2} \right)^i \right) \rho_1 a_{1t} = \left[ 1 / \left( 1 - \frac{\rho_1}{\delta_2} \right) \right] a_{1t},$$

since $|\rho_1| < 1/b$ and $|\delta_2| > 1/b$, so that the infinite sum converges.

Engineers directly obtain solutions of the form (13) by solving matrix Ricatti equations, e.g., see Kwakernaak and Sivan (1972). In their jargon, our system is "not controllable" but is "stabilizable" and "detectable" so that convergence of iterations on the Ricatti equation is assured. The stabilizability of our system depends on $\{a_{1t}\}, \{a_{2t}\}$, and $\{w_t\}$ being of exponential order less than $(1/b)$. 

The assumption that $w_t$ is of exponential order less than $(1/b)$ implies that the max $|\lambda_i| < (1/b)$, where $\lambda_i$ is the $i$th element of $\Lambda$.
Proceeding in the same way, we can write the decision rule for $n_{2t}$ as

$$n_{2t} = \mu_1 n_{2t-1} + \beta_1 w_t + \beta_2 w_{t-1} + \cdots + \beta_n w_{t-n+1} + \beta_0$$

$$+ \frac{\mu_1 h_2}{e} \left( \frac{f_0}{1 - \mu_1 b} \right) + a'_{2t},$$

where

$$\left( \beta_1, \beta_2, \cdots, \beta_n, \beta_0 \right) = -p \frac{\mu_1 h_2}{e} cP \left[ \frac{1}{1 - \lambda_1 \mu_1 b} \right] P^{-1},$$

$$a'_{2t} = \frac{\mu_1 h_2}{e} \left( \frac{1}{1 - \rho_2 \mu_1 b} \right) a_{2t}.$$

Equations (14) and (16) succinctly summarize how the distributed lag coefficients, the $\alpha$'s and $\beta$'s, reflect the combination of forecasting (through the parameters of $P$ and $\Lambda$) and optimization (through the parameters $d$, $\delta$, and $\mu$) elements. Clearly, the decision rules (13) and (15) are not invariant with respect to changes in the stochastic process for real wages (10), a general characteristic of optimum decision rules whose far-reaching implications for econometric policy evaluation have been stressed by Robert E. Lucas, Jr. (1976).

Since I will fit the model to data that are deviations from means and trends, I shall henceforth drop the constants from (13), (15), and (10). Substitute (10) for $w_t$ and subtract $\rho_1 a_{1t-1}'$ from both sides of (13) to get

$$n_{1t} = (\delta_1 + \rho_1) n_{1t-1} - \rho_1 \delta_1 n_{1t-2} + (\alpha_2 + \alpha_1 v_1 - \alpha_1 \rho_1) w_{t-1}$$

$$+ (\alpha_3 + \alpha_1 v_2 - \alpha_2 \rho_1) w_{t-2} + \cdots + (\alpha_n + \alpha_1 v_{n-1} - \alpha_{n-1} \rho_1) w_{t-n+1}$$

$$+ (\alpha_1 v_n - \alpha_n \rho_1) w_{t-n} + [\alpha_1 \xi_{3t} + (a'_{1t} - \rho_1 a'_{1t-1})].$$

From our earlier assumptions, $\bar{E}_{t-1} [\alpha_1 \xi_{3t} + (a'_{1t} - \rho_1 a'_{1t-1})] = 0$, so that (17) is the (vector) autoregression for $n_{1t}$. In particular, we have

$$\bar{E}_{t-1} n_{1t} = (\delta_1 + \rho_1) n_{1t-1} - \rho_1 \delta_1 n_{1t-2} + (\alpha_2 + \alpha_1 v_1 - \rho_1 \alpha_1) w_{t-1}$$

$$+ (\alpha_3 + \alpha_1 v_2 - \alpha_2 \rho_1) w_{t-2}$$

$$+ \cdots + (\alpha_n + \alpha_1 v_{n-1} - \alpha_{n-1} \rho_1) w_{t-n+1}$$

$$+ (\alpha_1 v_n - \alpha_n \rho_1) w_{t-n}.$$ (18)

Similarly, we have for $n_{2t}$

$$n_{2t} = (\mu_1 + \rho_2) n_{2t-1} - \rho_2 \mu_1 n_{2t-2} + (\beta_2 + \beta_1 v_1 - \beta_1 \rho_2) w_{t-1}$$

$$+ (\beta_3 + \beta_1 v_2 - \beta_2 \rho_2) w_{t-2}$$

$$+ \cdots + (\beta_n + \beta_1 v_{n-1} - \beta_{n-1} \rho_2) w_{t-n+1}$$

$$+ (\beta_1 v_n - \beta_n \rho_2) w_{t-n} + [\beta_1 \xi_{3t} + (a'_{2t} - \rho_2 a'_{2t-1})].$$ (19)
We can now write the complete three-variate vector autoregression for \( n_t, n_{2t}, w_t \) as

\[
\begin{align*}
    n_{1t} & = (\delta_1 + \rho_1) n_{1t-1} - \rho_1 \delta_1 n_{1t-2} + (\alpha_2 + \alpha_1 v_1 - \alpha_1 \rho_1) w_{t-1} \\
          & + (\alpha_3 + \alpha_1 v_2 - \alpha_2 \rho_1) w_{t-2} \\
          & + \cdots + (\alpha_n + \alpha_1 v_{n-1} - \alpha_{n-1} \rho_1) w_{t-n+1} \\
          & + (\alpha_1 v_n - \alpha_n \rho_1) w_{t-n} + u_{1t}, \\
    n_{2t} & = (\mu_1 + \rho_2) n_{2t-1} - \rho_2 \mu_1 n_{2t-2} + (\beta_2 + \beta_1 v_1 - \beta_1 \rho_2) w_{t-1} \\
          & + (\beta_3 + \beta_1 v_2 - \beta_2 \rho_2) w_{t-2} \\
          & + \cdots + (\beta_n + \beta_1 v_{n-1} - \beta_{n-1} \rho_2) w_{t-n+1} \\
          & + (\beta_1 v_n - \beta_n \rho_2) w_{t-n} + u_{2t}, \\
    w_t & = v_1 w_{t-1} + v_2 w_{t-2} + \cdots + v_n w_{t-n} + u_{3t},
\end{align*}
\]

(20a)

(20b)

(20c)

where

\[
\begin{bmatrix}
    u_{1t} \\
    u_{2t} \\
    u_{3t}
\end{bmatrix} = \begin{bmatrix}
    \alpha_1 \xi_{3t} + (\alpha_1' \xi_{1t-1}) \\
    \beta_1 \xi_{3t} + (\alpha_2' \xi_{2t-1}) \\
    \mu_1 \xi_{3t} + (\alpha_n \xi_{n-1})
\end{bmatrix} = \begin{bmatrix}
    n_{1t} - \hat{E}_{t-1} n_{1t} \\
    n_{2t} - \hat{E}_{t-1} n_{2t} \\
    w_t - \hat{E}_{t-1} w_t
\end{bmatrix}.
\]

Here \( u_t \) is the vector of innovations, that is, errors in predicting \( n_{1t}, n_{2t}, w_t \) from past information. There are \( (3n + 4) \) regressors in (20), that is, \( w_{t-1}, \ldots, w_{t-n} \), each of which appear three times, and \( n_{1t-1}, n_{1t-2}, n_{2t-1}, \) and \( n_{2t-2} \), each of which appears once. The free parameters of the model are \( f_1, d, e, \rho_1, \rho_2, \beta_1, \ldots, \beta_n \), so that there are \( (n + 5) \) parameters to be estimated. As it turns out, the model is overidentified for \( n > 1 \).

Collecting the equations that summarize the restrictions that the model imposes on the vector autoregression (20), we have

\[
\begin{align*}
    \phi_1 & = -\left( f_1 \frac{h_1}{d} + (1 + b) \right) \\
    \phi_2 & = -\left( f_1 \frac{h_2}{e} + (1 + b) \right) \\
    \left( 1 + \phi_1 \frac{b}{z} + \frac{1}{b} z^2 \right) & = (1 - \delta_1 z)(1 - \delta_2 z) \\
    \left( 1 + \phi_2 \frac{b}{z} + \frac{1}{b} z^2 \right) & = (1 - \mu_1 z)(1 - \mu_2 z) \\
    (\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_0) & = -\frac{\delta_1 h_1}{d} c \left( \frac{1}{1 - \lambda_1 b} \right)_{tt} P^{-1} \\
    (\beta_1, \beta_2, \ldots, \beta_n, \beta_0) & = -\frac{\mu_1 h_2}{e} c \left( \frac{1}{1 - \lambda_1 \mu_1 b} \right)_{tt} P^{-1} \\
    A & = PAP^{-1}.
\end{align*}
\]

(21)
Estimates of the free parameters \( \theta = (f_1, d, e, \rho_1, \rho_2, v_1, \ldots, v_n) \) are obtained by using the method of maximum likelihood to estimate the vector autoregression (20a), (20b), and (20c), subject to (21).\(^{12}\) Let \( \hat{u}_t = (\hat{u}_{1t}, \hat{u}_{2t}, \hat{u}_{3t})' \) be the sample residual vector associated with the parameter values \( \theta \). Under the assumption that \( u_t \) is a trivariate normal vector with \( E u u'_t = V \), the likelihood function of a sample of observations on the residuals extending over \( t = 1, \ldots, T \) is

\[
L(\theta) = (2\pi)^{-T/2} |V|^{-T/2} \exp \left( -\frac{1}{2} \sum_{t=1}^{T} \hat{u}_t V^{-1} \hat{u}_t \right).
\]

As shown by Wilson (1973) and Bard (1974), maximum-likelihood estimates of \( \theta \) with \( V \) unknown can be obtained by minimizing \( |\hat{V}| \) with respect to \( \theta \), where \( \hat{V} \) is the sample covariance matrix of \( u_t \),

\[
\hat{V} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t.
\]

The matrix \( \hat{V} \) is the maximum-likelihood estimator of \( V \) (see Wilson [1973] or Bard [1974, pp. 62–66]).\(^{13}\) The value of the likelihood function turns out to be \( \log L(\theta) = -(1/2) m T \log (2\pi) - (1/2) T \{ \log |\hat{V}| + m \} \), where \( m \) is the number of variates, equal to three in the present model.

Now consider the unconstrained version of the vector autoregression (20) in which each of the \( (3n + 4) \) regressors has its own free parameter. Let \( L_u \) be the value of the likelihood function at its unrestricted maximum, that is, the maximum obtained by permitting each of the \( (3n + 4) \) regressors to have its own free parameter. Let \( L_r \) be the value of the likelihood under the restrictions (21). Then \(-2 \log_e (L_r/L_u)\) is asymptotically distributed as \( \chi^2(q) \) where \( q = (3n + 4) - (n + 5) \) is the number of restrictions imposed by the theory. High values of the likelihood ratio lead to rejection of the restrictions that the theory imposes on the vector autoregression. Using the calculations of Wilson (1973, p. 80) or Bard (1974), it can be shown that the likelihood ratio is equal to

\[
T \{ \log_e |\hat{V}_r| - \log_e |\hat{V}_u| \},
\]

where \( \hat{V}_r \) and \( \hat{V}_u \) are the restricted and unrestricted estimates of \( V \), respectively.

I also used a likelihood-ratio statistic to test the constrained vector autoregression ([20a], [20b], and [20c]) against a second and even less constrained alternative, namely, an unconstrained trivariate vector autoregression with \( n \) lagged values of \( n_1, n_2, \) and \( w \) on the right-hand side of

\(^{12}\) The parameters \( f_0 \) and \( v_0 \) are dropped because the data are in the form of deviations from means and trend terms. The parameters \( b, p, h_1, \) and \( h_2 \) will be fixed a priori.

\(^{13}\) The likelihood function was maximized by using a derivative-free hill-climbing method with a Davidon-Fletcher-Powell algorithm for updating the Hessian. The complicated nature of the restrictions (21) led me to opt for a derivative-free method over an algorithm that required even analytical first derivatives. My attempts numerically to calculate asymptotic standard errors from the inverse of the information matrix were unsuccessful as one or two diagonal elements turned out to be negative.
each equation. Let \( \hat{P}_u \) be the estimated sample covariance matrix of the residuals in the unrestricted vector autoregression. Then the appropriate likelihood-ratio statistic is given by \( T\{\log_e|\hat{P}_r| - \log_e|\hat{P}_u|\} \). Since the unconstrained parameterization now has \( 9n \) free parameters, the likelihood ratio is asymptotically distributed as \( \chi^2 \) with \( \{9n - (n + 5)\} \) degrees of freedom.

II. Alternative Estimation Strategies

It should be stressed that the vector autoregression ([20a], [20b], and [20c]) which builds in the cross-equation restrictions implied by the model has been obtained under the assumption (8) that the productivity shocks \( a_{1t} \) and \( a_{2t} \) are first-order Markov processes. The forms of the vector autoregressions ([20a], [20b], and [20c]) would be altered had we assumed other forms for the \( a_{1t} \) and \( a_{2t} \) processes, as the reader can verify by calculations paralleling those above.

An alternative estimation strategy is available that avoids the necessity to make specific assumptions about the forms of the stochastic processes for the disturbances \( a_{1t} \) and \( a_{2t} \), only requiring that these processes be covariance stationary. The alternative estimator requires instead that the \( \omega_t \) process be strictly econometrically exogenous with respect to \( n_{1t} \) and \( n_{2t} \), in particular requiring that \( E_{\omega_t}a_{1s} = E_{\omega_t}a_{2s} = 0 \) for all \( t \) and \( s \). Under that assumption, the model (7a) and (7b) can readily be shown to place restrictions on the projections of \( n_{1t} \) and \( n_{2t} \), respectively, on the entire \( \{\omega_s\} \) process. The structure of those restrictions parallels those worked out by Sargent (1978a) for a consumption function example. An asymptotically efficient estimator such as “Hannan’s efficient estimator,” which allows for complicated serial-correlation patterns in the disturbances, could then be applied to estimating the projections with and without the restrictions imposed by the model.

This alternative estimation strategy gets along with much weaker assumptions about the serial-correlation properties of the disturbance processes \( \{a_{1t}\} \) and \( \{a_{2t}\} \) at the cost of making somewhat more stringent assumptions about the exogeneity of \( \omega_t \), that is, about the correlation between \( \omega_t \) and the \( a_{js} \)’s. The original estimator proposed that operates on (20a), (20b), and (20c) does assume that \( \{\omega_t\} \) is a process that is not caused in Granger’s (1969) sense by \( n_{1t} \) or \( n_{2t} \), that is, that \( \hat{E}(\omega_t|\omega_{t-1}, \omega_{t-2}, \ldots, n_{1t-1}, n_{1t-2}, \ldots, n_{2t-1}, n_{2t-2}, \ldots) = \hat{E}(\omega_t|\omega_{t-1}, \omega_{t-2}, \ldots) \). Now Sims’s (1972) theorems assure us that if \( \omega_t \) is not Granger-caused by \( n_{1t} \) or \( n_{2t} \), then there exists a statistical representation in which \( \omega_t \) is strictly econometrically exogenous with respect to \( n_{1t} \) or \( n_{2t} \). However, this statistical representation need not correspond with the appropriate economic behavioral relationship. It is possible for \( n_{1t} \) or \( n_{2t} \) to fail to cause \( \omega_t \), and yet for “instantaneous causality” to flow from \( n_{1t} \) or \( n_{2t} \) to \( \omega_t \) so
that \( w_t \) is not strictly exogenous in the appropriate model. See Sargent (1977a) for an example of this phenomenon within the context of Cagan's model of hyperinflation. The "autoregressive estimator" based on (20a), (20b), and (20c) permits arbitrary correlation between the innovations to \( n_1 \) or \( n_2 \) and \( w_t \) and makes no assumption about which pattern of instantaneous causality explains those correlations. On the other hand, the alternative "projection estimator" attributes all of those correlations to the workings of the demand schedules for \( n_1 \) and \( n_2 \) ([(7a) and (7b)].

For the present application, I prefer the estimator that makes the weaker assumption about the correlations between innovations to employment and the real wage.

The reader by now will have understood that optimizing, rational-expectations models do not entirely eliminate the need for side assumptions not grounded in economic theory. Some arbitrary assumptions about the nature of the serial-correlation structure of the disturbances and/or about strict econometric exogeneity are necessary in order to proceed with estimation.

Perhaps I should conclude this section by pointing to another source of arbitrariness, namely, the latitude at our disposal in specifying the firm's optimization problem. For example, adding terms like \(- (d_2/2)(n_{1t-1} - n_{1t-2})^2\) to the firm's daily profits would lead to Euler equations that are fourth-order stochastic difference equations and would lead to decision rules that depend on two lagged values of employment. Such specifications would seem plausible and would lead to materially different restrictions than those above on vector autoregressions (or projections of \( n \) on \( w_t \), as the case may be). There are clearly limits set by the requirements of econometric identification on our ability to estimate such complicated adjustment-cost parameterizations. Identification problems in such models have as yet received little attention at a general level.

The general theme of this section has been to issue a warning that rational-expectations, optimizing models will not be able to save us entirely from the ad hoc assumptions and interpretations made in applied work. However, this is not to deny that the rational-expectations hypothesis seems promising as a device for organizing restrictions on parameterizations of econometric models.

III. Parameter Estimates

The model was estimated using quarterly data on total civilian employment and a straight-time real-wage index, with the period of observation extending from 1947I through 1972IV, of which \( n \) observations at the beginning of the sample are lost when the order of the wage autoregression is set at \( n \). The variable \( n_{1t} \) was in the first instance measured by the seasonally adjusted BLS series "Employees on Nonagricultural Payrolls,
Private and Government.” To get a measure of \( n_{2t} \), the following procedure was used. I defined the variable \( \hat{h}_t \) to be average weekly hours, a series measured by the seasonally adjusted BLS series “Average Weekly Hours in Manufacturing.” I then estimated total man-hours by \( \hat{h}_t n_{1t} \). Finally, I measured \( n_{2t} \) by

\[
n_{2t} = \frac{\hat{h}_t n_{1t} - h_1 n_{1t}}{h_2},
\]

where \( h_1 \) and \( h_2 \) were set a priori at 37 and 17, respectively.\(^{14}\) The real wage \( w_t \) was measured by deflating the seasonally unadjusted BLS series “Average Hourly Earnings: Straight-Time Manufacturing Production Workers” by the seasonally unadjusted consumer price index \((1967 = 100)\).

I also created seasonally unadjusted measures of \( n_{1t} \) and \( n_{2t} \), by taking as a measure of \( n_{1t} \) the seasonally unadjusted BLS series “Employees on Private Nonagricultural Payrolls” and then using the preceding procedure to create estimates of \( n_{2t} \), by using the seasonally unadjusted average weekly hours series. The data are quarterly averages of monthly data. Notice that \( h_1 \) and \( h_2 \) are constants that are independent of time.

For reasons developed in Sargent (1976), I would argue that seasonally unadjusted data are the ones that ought to be used. Briefly, this view follows from the assumption that agents are themselves observing and responding to the seasonally unadjusted variates, so that the cross-equation restrictions delivered by the model pertain to the seasonally unadjusted data. Seasonal adjustment of the data could cause rejection of the cross-equation rational-expectations restrictions when they are in fact true. However, arguments have been made against this position in advocacy of seasonally adjusted data in exactly the present context (see Sims [1976]). For this reason, I report some results for both seasonally adjusted and unadjusted data.

I begin by describing the estimates obtained using the seasonally adjusted employment series together with the seasonally unadjusted real-wage series. (Later I will describe the results obtained with the seasonally unadjusted series for all variables.) Before estimating the model, the data on \( n_{1t} \) and \( n_{2t} \) were each detrended by regressing them on a constant, linear trend and trend squared, and then using the residuals from those regressions as the data for estimating the model.\(^{15}\) The data on \( w_t \) were formed as the residuals from a regression on a constant, linear trend, trend squared,

\(^{14}\) That these values for \( h_1 \) and \( h_2 \) do not add to unity, as in the theoretical presentation of the model, amounts only to a harmless renormalization. I guessed at these values for \( h_1 \) and \( h_2 \). The guess for \( h_1 \) measured in hours per week seemed reasonable after having inspected the time series for average weekly hours. For purposes of constructing the data on \( n_{2t} \), the choice of both \( h_1 \) and \( h_2 \) matters. For the purpose of estimating the demand functions, given the data on \( n_1 \) and \( n_2 \), only the ratio of \( h_1 \) to \( h_2 \) matters, as proportional changes in \( d \) and \( e \) can cancel the effects of proportionate increases in \( h_1 \) and \( h_2 \).

\(^{15}\) With the seasonally unadjusted employment data, I first regressed each of \( n_{1t} \), \( n_{2t} \), and \( w_t \) against a constant, trend, trend squared, and three seasonal dummies and used the residuals from those regressions as the data.
and three seasonal dummies. Two reasons can be given for detrending in this way prior to fitting the model. First, the model ignores the effects of capital on employment, except to the extent that these can be captured by the productivity processes \(a_1\) and \(a_2\). Second, the theory predicts that any deterministic components of the employment and real-wage processes will not be related by the same distributed lag model as are their indeterministic parts. Detrending prior to estimation is a device designed to isolate the indeterministic components. The real wage is measured in 1967 dollars, while employment is measured in millions of men.

Table 5 reports estimates of the model for \(n = 4\) for the seasonally adjusted data. The free parameters were \(f_1\), \(d\), \(e\), \(\rho_1\), \(\rho_2\), \(v_1\), \(v_2\), \(v_3\), and \(v_4\) with \(b\) being fixed at .95, \(h_1\) at 37, \(h_2\) at 17, and the premium \(p\) at 1.5. Since \(n = 4\), for the more constrained of our two alternative hypotheses, the likelihood-ratio statistic is asymptotically distributed as \(\chi^2\) with \(q = (3n + 4) - (n + 5) = 7\) degrees of freedom. The likelihood ratio is 9.53, which has a "marginal confidence level" of .783. The marginal confidence level is defined as follows. Let \(X\) be a \(\chi^2\) random variable with \(q\) degrees of freedom. Let \(x\) be the value of the likelihood-ratio statistic. Then the marginal confidence level is defined as \(\text{Prob}(X < x)\) under the null hypothesis. High values of the confidence level lead to rejecting the hypothesis. The likelihood-ratio statistic in this case indicates that the hypothesis cannot be rejected at marginal significance levels below .20. However, versus the less-constrained alternative hypothesis, the marginal confidence level is .9864, which indicates that the data do contain substantial evidence against the hypothesis. Notice the different lag shapes

| TABLE 5 |
|-----------------------------------------------|
| **FIRST SOLUTION OF LIKELIHOOD EQUATIONS,** |
| **SEASONALLY ADJUSTED DATA (\(n = 4\)**) |
| (19481–19721V)* |

| \(f_1\) | 19.80 |
| \(d\) | 2377.90 |
| \(e\) | 104.02 |
| \(\delta_1\) | .5886 |
| \(\alpha_1\) | -.0185 |
| \(\beta_1\) | -.0600 |
| \(\alpha_2\) | .0009 |
| \(\beta_2\) | -.0001 |
| \(\alpha_3\) | .0017 |
| \(\beta_3\) | -.0004 |
| \(\alpha_4\) | .0044 |
| \(\beta_4\) | .0026 |

\[
V = \begin{pmatrix}
.9220E - 01 & .2000E + 00 & .1298E - 02 \\
.7747E + 00 & .2077E - 02 \\
.1949E - 03 & .1949E - 03
\end{pmatrix}
\]

\[
B^{-1}VB^{-1'} = \begin{pmatrix}
.9225E - 01 & .2002E + 00 & .1301E - 02 \\
.7749E + 00 & .2089E - 02 \\
.1949E - 03 & .1949E - 03
\end{pmatrix}
\]

| \(|\hat{P}_r| = .5497E - 05\) | \(|\hat{P}_u| = .4998E - 05\) | \(|\hat{P}_u| = .3474E - 05\) |
|--------------------------|--------------------------|--------------------------|
| \(T(\log |\hat{P}_r| - \log |\hat{P}_u|) = 9.5271\) | \(T(\log |\hat{P}_r| - \log |\hat{P}_u|) = 45.881\) |

Marginal confidence level = .7830  
Marginal confidence level = .9864

*Period of observation on the dependent variables.
and the magnitudes of the distributed lag coefficients of straight-time employment and overtime employment in the real wage, the $\alpha$'s and $\beta$'s, respectively. Overtime employment is estimated to be more responsive to the real wage. Further, the straight-time adjustment cost parameter $d$ is estimated to be much larger than the overtime adjustment cost parameter $e$. That is why $n_{1t}$ depends more strongly on $n_{1t-1}$ than $n_{2t}$, does on $n_{2t-1}$, that is, why $\delta_1$ is estimated to exceed $\mu_1$.

Since the likelihood ratio test assumes that the $u$'s are serially uncorrelated, table 5 also reports three statistics, $KS(n_1)$, $KS(n_2)$, and $KS(w)$, which are Kolmogorov-Smirnov statistics from the cumulated periodograms for $u_1$, $u_2$, and $u_3$, that is, for the estimated innovations for $n_1$, $n_2$, and $w$, respectively, implied by the vector autoregression constrained by the model. The Kolmogorov-Smirnov statistic recorded is the maximum absolute deviation of the cumulated periodogram of the disturbance from its theoretical value under the assumption that the disturbances are serially uncorrelated. Durbin (1969) reports tables for the distribution of this statistic, though they are not applicable where lagged dependent variables are included as regressors, as in the present case. It is nevertheless of some comfort that the Kolmogorov-Smirnov statistics in table 5 and in subsequent tables do not signal dangerous levels of serial correlation. Notice that the Kolmogorov statistics are greater for the $n_1$ and $n_2$ innovations than for the $w$ innovation. This is symptomatic of the fact that the model fits an $n$th-order Markov process in $w$ but only permits two lagged own-values to enter the autoregressions for $n_1$ and $n_2$, thereby leaving it more likely that the model will neglect some higher-order serial correlation for $n_1$ and $n_2$. This pattern for the Kolmogorov-Smirnov statistics repeats itself in the subsequent tables.

Table 5 also reports the estimated covariance matrix of the innovations $V = E(u'u'_t)$. Recall that

$$
\begin{bmatrix}
  u'_{1t} \\
  u'_{2t} \\
  u'_{3t}
\end{bmatrix} =
\begin{bmatrix}
  1 & 0 & \alpha_1 \\
  0 & 1 & \beta_1 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  \xi_{1t} \\
  \xi_{2t} \\
  \xi_{3t}
\end{bmatrix} = B \xi_t
$$

where $\xi_{1t} = a'_{1t} - \rho_1 a'_{1t-1}, \xi_{2t} = a'_{2t} - \rho_2 a'_{2t-1}$, and where

$$
B =
\begin{bmatrix}
  1 & 0 & \alpha_1 \\
  0 & 1 & \beta_1 \\
  0 & 0 & 1
\end{bmatrix},
\xi_t =
\begin{bmatrix}
  \xi_{1t} \\
  \xi_{2t} \\
  \xi_{3t}
\end{bmatrix}.
$$

Then, since $\xi_t = B^{-1} u_t$, the covariance matrix of $\xi_t$ can be estimated from $E \xi_t \xi'_t = B^{-1} V B^{-1'}$, an estimate of which is also reported in table 5. The correlation between the innovations to $a'_{1t}$ and to $a'_{2t}$, that is, $\xi_{1t}$ and $\xi_{2t}$, is estimated to be .748. The correlation between the innovations to $a'_{1t}$ and $w_t$, that is, $\xi_{1t}$ and $\xi_{3t}$, is .3061, while that between $\xi_{2t}$ and $\xi_{3t}$ is .1700. I had expected $\xi_{1t}$ and $\xi_{2t}$ to be even more highly correlated than they are.
As it happens, the estimates reported in table 5 correspond to the higher of two local maxima of the likelihood function which I found. The parameter estimates associated with the lower of these two local maxima are reported in table 6. In view of the form of the vector autoregression ([20a], [20b], and [20c]), it is not at all surprising that the likelihood function should exhibit multiple maxima. In particular, notice that the coefficients in (20a), (20b), and (20c) on \( n_{1t-1}, n_{1t-2}, n_{2t-1}, n_{2t-2} \) are, respectively, \( (\delta_1 + \rho_1), -\delta_1 \rho_1, (\mu_1 + \rho_2), \) and \( -\mu_1 \rho_2). \) If it were not for the constraints across \( \mu_1 \) and the \( \beta \)'s and across \( \delta_1 \) and the \( \alpha \)'s and the appearance of \( \rho_1 \) and \( \rho_2 \) elsewhere on the right-hand side of (20a), (20b), and (20c), the parameters \( \delta_1, \rho_1, \mu_1, \) and \( \rho_2 \) would not be identified, since it would be impossible to distinguish the effects of \( \delta_1 \) from \( \rho_1 \) and the effects of \( \mu_1 \) from \( \rho_2 \). The presence of lagged \( w \)'s on the right-hand side of (20a), (20b), and (20c) and the aforementioned constraints resolve this identification problem but leave a vestige of it in the form of probable multiple peaks in the likelihood function with small samples. Comparing the parameter estimates in tables 5 and 6 shows that table 5 is a high \( (\rho_1, \rho_2) \)-low \( (\delta_1, \mu_1) \) solution, while table 6 reports the high \( (\delta_1, \mu_1) \)-low \( (\rho_1, \rho_2) \) solution. Notice that for the table 6 estimates, \( \rho_1 + \delta_1 = 1.528 \) and \( \rho_1 \delta_1 = .555 \), while for the table 5 estimates, \( \rho_1 + \delta_1 = 1.526 \) while \( \rho_1 \delta_1 = .552 \).

Figures 1 and 2 depict two views of the likelihood surface as a function of \( \delta_1 \) and \( \rho_1 \). The likelihood surface has a ridge and is characterized by two
VALUES OF DELTA
VALUES OF ROEI

Fig. 1.—Likelihood surface

VALUES OF DELTA
VALUES OF ROEI

Fig. 2.—Likelihood surface
local maxima. Figure 3 depicts iso-likelihood contours in the \((\delta_1, \rho_1)\) plane. These figures emphasize the weakness of the identification of \((\delta_1, \rho_1)\) and of \((\mu_1, \rho_2)\).

The presence of multiple maxima of the likelihood function means that caution is called for in interpreting the test statistics reported, since the asymptotic distribution on which the test is computed does not predict multiple maxima for the likelihood function and so does not provide a very good approximation for the sample size that we are studying. The presence of multiple maxima of the likelihood function also argues for starting the nonlinear estimation from several different initial parameter estimates. I followed this practice in each case reported below.

Table 7 reports the estimates for the seasonally unadjusted data with \(n = 4\). The estimates indicate \(d \gg e\) and are qualitatively similar to those described above. For testing the model versus the more constrained of the two alternative hypotheses, the marginal confidence level is .53. Versus the less constrained alternative, the marginal confidence level is .68. These results indicate that the sample does not contain strong evidence against the hypothesis.

Table 8 reports estimates of the model for the seasonally unadjusted data with \(n = 8\). The likelihood-ratio statistic for testing against the more constrained alternative hypothesis is now distributed asymptotically as \(\chi^2\) with 15 degrees of freedom under the null hypothesis that the model is correct. Once again, both likelihood ratios indicate that the sample does not contain much evidence against the model. For the seasonally unadjusted data with \(n = 8\), table 9 reports the maximum-likelihood
Fig. 3—Iso-likelihood contours.
TABLE 8
Seasonally Unadjusted Data (n = 8) (1949I-1972IV)*

| Parameter | Value |
|-----------|-------|
| $f_1$     | .3612 |
| $d$       | 3266.29 |
| $e$       | 75.6750 |
| $\rho_1$  | 4.094 |
| $\rho_2$  | .0571 |
| $\delta_1$| .9569 |
| $\mu_1$   | .7687 |
| $\alpha_1$| -.3790 |
| $\alpha_2$| -.0745 |
| $\alpha_3$| -.0045 |
| $\alpha_4$| -.1010 |
| $\alpha_5$| -.0989 |
| $\alpha_6$| -.0787 |
| $\alpha_7$| -.0448 |
| $\alpha_8$| -.0448 |
| $\beta_1$ | -.7970 |
| $\beta_2$ | -.0417 |
| $\beta_3$ | .0211 |
| $\beta_4$ | .1232 |
| $\beta_5$ | -.0335 |
| $\beta_6$ | -.0203 |
| $\beta_7$ | .0356 |
| $\beta_8$ | -.0057 |
| $\nu_1$   | .8719 |
| $\nu_2$   | .0982 |
| $\nu_3$   | .1183 |
| $\nu_4$   | -.2537 |
| $\nu_5$   | .0322 |
| $\nu_6$   | .0795 |
| $\nu_7$   | -.1688 |
| $\nu_8$   | .0397 |

$V = \begin{pmatrix} .1355E + 00 & .2721E + 00 & .9675E + 03 \\ .8420E + 00 & .1147E - 02 & .1791E - 03 \end{pmatrix}$

$B^{-1}VB^{-1'} = \begin{pmatrix} .1362E + 00 & .2734E + 00 & .1035E - 02 \\ .8439E + 00 & .1289E - 02 & .1791E - 03 \end{pmatrix}$

$|\tilde{\nu}_r| = .6802E - 05, |\tilde{\nu}_w| = .6163 - 05, |\tilde{\nu}_u| = .3399E - 05$

$T|\log|\tilde{\nu}_r| - \log|\tilde{\nu}_u|| = 9.4610, T|\log|\tilde{\nu}_r| - |\tilde{\nu}_u|| = 66.59937$

Marginal confidence level = .1478, Marginal confidence level = .7680

* Period of observation on the dependent variables.

Estimates of the vector autoregression (20a), (20b), and (20c), both unconstrained and constrained by the restrictions of the model (21). The constrained and unconstrained estimates are close except in one respect: The model-constrained vector autoregressions for $n_1$ and $n_2$ have coefficients on lagged w's that are generally much smaller in absolute value than their unconstrained counterparts. This pattern is also reflected in tables 10 through 14. Table 10 shows the vector moving average representation implied by the model-constrained estimates while table 11 shows a decomposition of variance of the 35-quarter-ahead forecast-error variances. Tables 12 and 13 show the corresponding moving average representation and decomposition of variance for the unconstrained estimates that are reported in table 9. Comparison of tables 10 and 12, on one hand, and tables 11 and 13, on the other, indicates that while the constrained model captures the same response of $n_1$ and $n_2$ to their own innovations that is depicted in the unconstrained estimates, the constrained model substantially underestimates the responses of $n_1$ and $n_2$ to innovations in $w$. The moving average representation implied by the model-constrained estimates have one-standard-deviation wage innovations giving rise to much smaller movements in $n_1$ and $n_2$ than are those associated with one-standard-deviation own innovations in $n_1$ and $n_2$. Contrast this with the relatively sizable responses of $n_1$ and $n_2$ to real-wage innovations in the unconstrained estimates. The decompositions of
TABLE 9
VECTOR AUTOREGRESSIONS (n = 8)
SEASONALLY UNADJUSTED DATA
(1949I-1972IV)

|                | Unconstrained | Constrained by (21) |
|----------------|---------------|---------------------|
| \( n_{1t-1} \) | 1.4040        | 1.3663              |
| \( n_{1t-2} \) | -.4305        | -.3918              |
| \( w_{t-1} \)  | -.7105        | -.2498              |
| \( w_{t-2} \)  | -2.8277       | -.0515              |
| \( w_{t-3} \)  | -2.3616       | -.0309              |
| \( w_{t-4} \)  | 1.0005        | -.0031              |
| \( w_{t-5} \)  | 6.8486        | -.0698              |
| \( w_{t-6} \)  | -.1849        | -.0683              |
| \( w_{t-7} \)  | -6.2083       | -.0543              |
| \( w_{t-8} \)  | .1310         | .0466               |

(20b):
| \( n_{2t-1} \) | .8361         | .8258               |
| \( n_{2t-2} \) | -.0710        | -.0439              |
| \( w_{t-1} \)  | -.5042        | -.6911              |
| \( w_{t-2} \)  | -10.2706      | -.0548              |
| \( w_{t-3} \)  | .2643         | .0277               |
| \( w_{t-4} \)  | -6.8648       | .1616               |
| \( w_{t-5} \)  | 17.6580       | -.0440              |
| \( w_{t-6} \)  | -.7399        | -.0266              |
| \( w_{t-7} \)  | 6.7369        | .0468               |
| \( w_{t-8} \)  | -12.0341      | -.0268              |

(20c):
| \( w_{t-1} \)  | .8557         | .8719               |
| \( w_{t-2} \)  | .1021         | .0982               |
| \( w_{t-3} \)  | .0699         | .1183               |
| \( w_{t-4} \)  | -.2183        | -.2536              |
| \( w_{t-5} \)  | .7217         | -.0322              |
| \( w_{t-6} \)  | .9416         | .0795               |
| \( w_{t-7} \)  | -.2382        | -.1688              |
| \( w_{t-8} \)  | .0574         | -.0997              |

The variance in tables 11 and 13 indicate the extent to which the constrained model attributes less of a role to real-wage innovations in driving \( n_1 \) and \( n_2 \).

Notice how both tables 10 and 12 show \( n_2 \) responding more quickly to an own innovation than does \( n_1 \).

The estimates in tables 10–13 came from the data that are residuals from regressions on constant, trend, trend squared, and three seasonal dummies. Table 14 is the counterpart of table 13 where trend squared has been omitted. The effect of dropping trend squared is to increase somewhat the percentage of the variance of the 35-quarter-ahead prediction error in \( n_1 \) or \( n_2 \) that is explained by innovations in the real wage. The results in table 14 are presented to form a bridge to the estimates of Neftci and those summarized in the introduction, which included trend but not trend-squared terms.

The vector autoregressions summarized in tables 9–14 all impose the
TABLE 10
MOVING AVERAGE REPRESENTATION IMPLIED
BY MODEL FOR SEASONALLY UNADJUSTED (TABLE 8) ESTIMATES

| Lag | $n_1$  | $w$  | $n_2$ |
|-----|--------|------|-------|
| 0   | 0.4324 | 0.0144 | 0 |
| 1   | 0.5909 | 0.0125 | 0 |
| 2   | 0.6379 | 0.0123 | 0 |
| 3   | 0.6401 | 0.0137 | 0 |
| 4   | 0.6247 | 0.0110 | 0 |
| 5   | 0.6028 | 0.0096 | 0 |
| 6   | 0.5788 | 0.0095 | 0 |
| 7   | 0.5547 | 0.0060 | 0 |
| 8   | 0.5312 | 0.0044 | 0 |
| 9   | 0.5085 | 0.0030 | 0 |
| 10  | 0.4366 | 0.0007 | 0 |
| 11  | 0.4657 | 0.0003 | 0 |
| 12  | 0.4456 | -0.0012 | -0.0099 |
| 13  | 0.4264 | -0.0012 | -0.0095 |
| 14  | 0.4081 | -0.0032 | -0.0060 |
| 15  | 0.3905 | -0.0028 | -0.0030 |
| 16  | 0.3737 | -0.0028 | -0.0043 |
| 17  | 0.3576 | -0.0028 | -0.0004 |
| 18  | 0.3422 | -0.0029 | -0.0016 |
| 19  | 0.3275 | -0.0029 | 0.0031 |
| 20  | 0.3134 | -0.0029 | 0.0041 |
| 21  | 0.3018 | -0.0029 | 0.0046 |
| 22  | 0.2869 | -0.0029 | 0.0047 |
| 23  | 0.2746 | -0.0029 | 0.0045 |
| 24  | 0.2628 | -0.0029 | 0.0040 |
| 25  | 0.2514 | -0.0029 | 0.0034 |
| 26  | 0.2406 | -0.0029 | 0.0027 |
| 27  | 0.2302 | -0.0029 | 0.0019 |
| 28  | 0.2203 | -0.0029 | 0.0018 |
| 29  | 0.2108 | -0.0029 | 0.0019 |
| 30  | 0.2018 | -0.0029 | 0.0020 |
| 31  | 0.1931 | -0.0029 | 0.0021 |

Response to a one-standard-deviation innovation in $n_1$:

Response to a one-standard-deviation innovation in $w$:
TABLE 10 (Continued)

|   |   |   |
|---|---|---|
| 26.   | -0.0208 | 0.0004 | 0.0012 |
| 27.   | -0.0198 | 0.0003 | 0.0006 |
| 28.   | -0.0190 | 0.0006 | 0.0009 |
| 29.   | -0.0184 | 0.0005 | -0.0003 |
| 30.   | -0.0179 | 0.0006 | -0.0007 |
| 31.   | -0.0176 | 0.0005 | -0.0009 |

Response to a one-standard-deviation innovation in \( n_2 \):

|   |   |   |
|---|---|---|
| 0   | 0 | 0 | 1.080 |
| 1   | 0 | 0 | 0.892 |
| 2   | 0 | 0 | 0.892 |
| 3   | 0 | 0 | 0.892 |
| 4   | 0 | 0 | 0.892 |
| 5   | 0 | 0 | 0.892 |
| 6   | 0 | 0 | 0.892 |
| 7   | 0 | 0 | 0.892 |
| 8   | 0 | 0 | 0.892 |
| 9   | 0 | 0 | 0.892 |
| 10  | 0 | 0 | 0.892 |
| 11  | 0 | 0 | 0.892 |
| 12  | 0 | 0 | 0.892 |
| 13  | 0 | 0 | 0.892 |
| 14  | 0 | 0 | 0.892 |
| 15  | 0 | 0 | 0.892 |
| 16  | 0 | 0 | 0.892 |
| 17  | 0 | 0 | 0.892 |
| 18  | 0 | 0 | 0.892 |
| 19  | 0 | 0 | 0.892 |
| 20  | 0 | 0 | 0.892 |
| 21  | 0 | 0 | 0.892 |
| 22  | 0 | 0 | 0.892 |
| 23  | 0 | 0 | 0.892 |
| 24  | 0 | 0 | 0.892 |
| 25  | 0 | 0 | 0.892 |
| 26  | 0 | 0 | 0.892 |
| 27  | 0 | 0 | 0.892 |
| 28  | 0 | 0 | 0.892 |
| 29  | 0 | 0 | 0.892 |
| 30  | 0 | 0 | 0.892 |
| 31  | 0 | 0 | 0.892 |

Correlation matrix of innovations:

\[
\begin{array}{ccc}
    n_1 & w & n_2 \\
    n_1 & 1.00 & .197 & .808 \\
    w & .197 & 1.000 & .135 \\
    n_2 & .808 & .135 & 1.000 \\
\end{array}
\]

TABLE 11

Variance Decompositions for Forecast Errors Implied by Model (Tables 8 and 9 Estimates): Seasonally Unadjusted Data

|   |   |   |
|---|---|---|
| 98.3 | 1.71 |   |
| 0 | 100. |   |
| 63.79 | 1.38 | 34.84 |

Note.—Percentage of 35-step-ahead forecast error variance in \( x \) accounted for by "orthogonalized innovations" in \( n_1, w, n_2 \). Orthogonalization order: \( w, n_1, n_2 \). Orthogonalization order refers to the procedure described in the Appendix of defining an orthogonal process from \( u_t = F_n \). If the orthogonalization order is \( n_1, w, n_2 \), then the "orthogonalized \( n_1 \) innovation" is simply the \( n_1 \) innovation; the "orthogonalized \( w \) innovation" is the part of \( w \) innovation orthogonal to the \( n_1 \) innovation; the "orthogonalized \( n_2 \) innovation" is the part of the \( n_2 \) innovation that is orthogonal to both the \( n_1 \) and \( w \) innovations.
| Lag | $n_1$ | $w$ | $n_2$ |
|-----|-------|-----|-------|
| 0   | 0.4005| 0   | 0     |
| 1   | 0.5623| 0   | 0     |
| 2   | 0.6170| 0   | 0     |
| 3   | 0.6241| 0   | 0     |
| 4   | 0.6106| 0   | 0     |
| 5   | 0.5886| 0   | 0     |
| 6   | 0.5635| 0   | 0     |
| 7   | 0.5377| 0   | 0     |
| 8   | 0.5123| 0   | 0     |
| 9   | 0.4877| 0   | 0     |
| 10  | 0.4642| 0   | 0     |
| 11  | 0.4417| 0   | 0     |
| 12  | 0.4203| 0   | 0     |
| 13  | 0.3999| 0   | 0     |
| 14  | 0.3805| 0   | 0     |
| 15  | 0.3621| 0   | 0     |
| 16  | 0.3445| 0   | 0     |
| 17  | 0.3278| 0   | 0     |
| 18  | 0.3119| 0   | 0     |
| 19  | 0.2967| 0   | 0     |
| 20  | 0.2823| 0   | 0     |
| 21  | 0.2686| 0   | 0     |
| 22  | 0.2556| 0   | 0     |
| 23  | 0.2432| 0   | 0     |
| 24  | 0.2314| 0   | 0     |
| 25  | 0.2202| 0   | 0     |
| 26  | 0.2095| 0   | 0     |
| 27  | 0.1993| 0   | 0     |
| 28  | 0.1896| 0   | 0     |
| 29  | 0.1804| 0   | 0     |
| 30  | 0.1717| 0   | 0     |
| 31  | 0.1633| 0   | 0     |

Response to a one-standard-deviation innovation in $n_1$:

Response to a one-standard-deviation innovation in $w$:
TABLE 12 (Continued)

|   | 0.1573 | 0.0000 | 0.0267 |
|---|--------|--------|--------|
| 24 | -0.1449 | 0.0002 | 0.0357 |
| 25 | -0.1333 | 0.0005 | 0.0400 |
| 26 | -0.1234 | 0.0007 | 0.0397 |
| 27 | -0.1158 | 0.0007 | 0.0386 |
| 28 | -0.1090 | 0.0008 | 0.0343 |
| 29 | -0.1036 | 0.0007 | 0.0284 |
| 30 | -0.0996 | 0.0006 | 0.0224 |
| 31 | -0.0960 | 0.0005 | 0.0156 |

Response to a one-standard-deviation innovation in \( n_2 \):

|   |   |   | 1.0595 |
|---|---|---|--------|
| 0 | 0 | 0 | 0.8859 |
| 1 | 0 | 0 | 0.6654 |
| 2 | 0 | 0 | 0.4934 |
| 3 | 0 | 0 | 0.3653 |
| 4 | 0 | 0 | 0.2704 |
| 5 | 0 | 0 | 0.2001 |
| 6 | 0 | 0 | 0.1481 |
| 7 | 0 | 0 | 0.1096 |
| 8 | 0 | 0 | 0.0811 |
| 9 | 0 | 0 | 0.0600 |
| 10 | 0 | 0 | 0.0444 |
| 11 | 0 | 0 | 0.0329 |
| 12 | 0 | 0 | 0.0243 |
| 13 | 0 | 0 | 0.0180 |
| 14 | 0 | 0 | 0.0133 |
| 15 | 0 | 0 | 0.0098 |
| 16 | 0 | 0 | 0.0073 |
| 17 | 0 | 0 | 0.0054 |
| 18 | 0 | 0 | 0.0040 |
| 19 | 0 | 0 | 0.0029 |
| 20 | 0 | 0 | 0.0021 |
| 21 | 0 | 0 | 0.0016 |
| 22 | 0 | 0 | 0.0012 |
| 23 | 0 | 0 | 0.0008 |
| 24 | 0 | 0 | 0.0006 |
| 25 | 0 | 0 | 0.0004 |
| 26 | 0 | 0 | 0.0003 |
| 27 | 0 | 0 | 0.0002 |
| 28 | 0 | 0 | 0.0001 |
| 29 | 0 | 0 | 0.0001 |
| 30 | 0 | 0 | 0.0001 |
| 31 | 0 | 0 | 0.0001 |

Correlation matrix of innovations:

|   | \( n_1 \) | \( w \) | \( n_2 \) |
|---|----------|--------|--------|
| \( n_1 \) | 1.00     | 0.2066 | 0.7971 |
| \( w \)   | 1.0000   | 1.350  | 1.0000 |
| \( n_2 \) | 1.0000   |        |        |

extensive zero restrictions incorporated in (20a), (20b), and (20c), for example, lagged \( n_2 \)'s do not appear in the autoregression for \( n_1 \). Tables 15 and 16 report summary statistics for fourth-order vector autoregressions with no such zero restrictions built in, that is, four lags of each variable appear in the autoregression for each of \( n_1 \), \( n_2 \), and \( w \). A constant, trend, and three seasonal dummies are also included in the regressions. Table 15 reports marginal significance levels appropriate for testing the null hypo-
### TABLE 13

**Decomposition of Variance of Forecast Error for Unconstrained Estimates: Seasonally Unadjusted Data (1948I–1972IV)**

| Orthogonalization order* = n₁, w, n₂: | n₁ | w | n₂ |
|--------------------------------------|----|---|----|
| x = n₁                                 | 74.71 | 25.29 | 0 |
| x = w                                 | 4.27 | 95.73 | 0 |
| x = n₂                                 | 49.14 | 20.12 | 30.73 |

Orthogonalization order = n₁, n₂, w:

| x = n₁                                 | 74.71 | 25.23 | .06 |
| x = w                                 | 4.27 | 95.49 | .24 |
| x = n₂                                 | 49.14 | 19.03 | 31.83 |

Orthogonalization order = w, n₁, n₂:

| x = n₁                                 | 87.76 | 12.24 | 0 |
| x = w                                 | 0 | 100 | 0 |
| x = n₂                                 | 52.22 | 17.05 | 30.73 |

**Note.**—Percentage of 35-step-ahead forecast error variance in x accounted for by "orthogonalized innovations" in n₁, w, n₂.

* Orthogonalization order refers to the procedure described in the Appendix of defining an orthogonal u process from \( u_t = F_n e_t \). If the orthogonalization order is n₁, w, n₂, then the "orthogonalized n₁ innovation" is simply the n₁ innovation; the "orthogonalized w innovation" is the part of the w innovation orthogonal to the n₁ innovation; the "orthogonalized n₂ innovation" is the part of the n₂ innovation that is orthogonal to both the n₁ and w innovations.

### TABLE 14

**Decomposition of Variance of Forecast Errors, Seasonally Unadjusted Data (1948I–1972IV)**

| Orthogonalization order* = n₁, w, n₂: | n₁ | w | n₂ |
|--------------------------------------|----|---|----|
| x = n₁                                 | 50.74 | 49.26 | 0 |
| x = w                                 | 3.58 | 96.42 | 0 |
| x = n₂                                 | 43.68 | 24.71 | 29.62 |

Orthogonalization order = n₁, n₂, w:

| x = n₁                                 | 50.74 | 49.23 | .03 |
| x = w                                 | 3.58 | 96.36 | .66 |
| x = n₂                                 | 43.68 | 24.31 | 30.01 |

Orthogonalization order = w, n₁, n₂:

| x = n₁                                 | 64.66 | 35.34 | 0 |
| x = w                                 | 0 | 100 | 0 |
| x = n₂                                 | 46.80 | 23.59 | 29.62 |

**Note.**—Percentage of 35-step-ahead forecast error variance in x accounted for by "orthogonalized innovations" in n₁, w, n₂. Data are residuals from regressions on constant, trend, and three seasonal dummies, with no trend-squared terms, in contradistinction to the table 13 results.

* Orthogonalization order refers to the procedure described in the Appendix of defining an orthogonal u process from \( u_t = F_n e_t \). If the orthogonalization order is n₁, w, n₂, then the "orthogonalized n₁ innovation" is simply the n₁ innovation; the "orthogonalized w innovation" is the part of the w innovation orthogonal to the n₁ innovation; the "orthogonalized n₂ innovation" is the part of the n₂ innovation that is orthogonal to both the n₁ and w innovations.
TABLE 15

SUMMARY STATISTICS FOR FOURTH-ORDER VECTOR
AUTOREGRSSIONS FOR \((n_1, n_2, w)\)
SEASONALLY UNADJUSTED DATA*
(1948I–1972IV)

|                | \(n_1\) | \(n_2\) | \(w\) |
|----------------|---------|---------|-------|
| \(x = n_1\)    | .0000   | .0000   | .2000 |
| \(x = n_2\)    | .0395   | .0000   | .0446 |
| \(x = w\)      | .6857   | .6128   | .0000 |

* Regressions included a constant, trend, and three seasonal dummies.
† Where \(f\) is the calculated value of the pertinent \(F\)-statistic, the marginal significance level is defined as prob\(\{F > f\}\) under the null hypothesis.

TABLE 16

DECOMPOSITION OF VARIANCE OF FORECAST ERROR
IMPLIED BY VECTOR AUTOREGRESSION FOR \((n_1, n_2, w)\)
SEASONALLY UNADJUSTED DATA
(1948I–1972IV)

|                | \(n_1\) | \(w\) | \(n_2\) |
|----------------|---------|-------|--------|
| Orthogonalization order* = \(n_1, w, n_2\): |         |       |        |
| \(x = n_1\)    | 21.82   | 48.74 | 29.44  |
| \(x = w\)      | .76     | 98.39 | .85    |
| \(x = n_2\)    | 23.64   | 16.25 | 60.25  |
| Orthogonalization order = \(w, n_1, n_2\): |         |       |        |
| \(x = n_1\)    | 26.90   | 43.66 | 29.44  |
| \(x = w\)      | 2.11    | 97.03 | .85    |
| \(x = n_2\)    | 20.82   | 18.93 | 60.24  |

Note.—Percentage of 35-quarter ahead forecast error variance in \(x\) accounted for by "orthogonalized innovation" in \(n_1, w, n_2\).

* Orthogonalization order refers to the procedure described in the Appendix of defining an orthogonal \(w\) process from \(u_t = F_{et}\). If the orthogonalization order is \(n_1, w, n_2\), then the "orthogonalized \(n_1\) innovation" is simply the \(n_1\) innovation; the "orthogonalized \(w\) innovation" is the part of the \(w\) innovation orthogonal to the \(n_1\) innovation; the "orthogonalized \(n_2\) innovation" is the part of the \(n_2\) innovation that is orthogonal to both the \(n_1\) and \(w\) innovations.

thesis that \(n_1\) or \(n_2\) or \(w\) fails to Granger-cause each of the other variables. These \(F\)-statistics are consistent with Neftci's results and indicate stronger evidence for Granger causality flowing from \(w\) to \(n_1\) and \(n_2\) than from \(n_1\) or \(n_2\) to \(w\). However, the statistics also indicate Granger causality from \(n_1\) to \(n_2\) and from \(n_2\) to \(n_1\), patterns which are ruled out by the model (20a), (20b), (20c), and (21). The data indicate dynamic interactions between \(n_1\) and \(n_2\) that the model in its present form cannot account for. The decompositions of variance of 35-quarter-ahead forecast errors in
table 16 once again reinforce Neftci's findings in confirming that substantial percentages of the variance in employment forecasting errors are attributable to innovations in the real wage.

In summary, while the model usually passes the likelihood-ratio tests I have calculated, it does seem to do violence to two aspects of the data. First, the model generates estimates that seem to understate the magnitude of the inverse influence exerted by the real wage on employment. Second, a priori the model neglects dynamic interactions between $n_1$ and $n_2$ that seem to be there. On the first point, the maximum-likelihood estimates of the parameters $d$ and $e$, which also influence the response to $w$ of $n_1$ and $n_2$, respectively, seem mainly to have been chosen to permit the model to capture the response of $n_1$ and $n_2$ to their own innovations. As a by-product, this involved understating the responses of $n_1$ and $n_2$ to $w$, which seems less costly in terms of the likelihood function than misstating the response to own innovations. Perhaps a richer specification of the Markov processes for $a_{1t}$ and $a_{2t}$, say permitting them to be second-order processes, would permit enough flexibility to remedy this feature. Permitting the Markov processes for $a_{1t}$ and $a_{2t}$ to depend on lagged cross terms $a_2$ and $a_1$, respectively, would provide one way to remedy the second deficiency of the model, for it would potentially permit dynamic interactions between $n_1$ and $n_2$ of the kind revealed by table 15. Another way to account for those dynamic interactions would be to let costs of adjustment for $n_1$ depend on the level of $n_2$, and vice versa. This could be done while remaining within the linear-quadratic framework of this paper. However, extensions in each of these directions, while feasible, are costly both in the sense that they reduce the degree of overidentification of the model and in the sense that they make maximum-likelihood estimation more expensive.

IV. Conclusions

The simple contemporaneous correlations that formed the evidence in the original Dunlop-Tarshis-Keynes exchange and also in much of the follow-up empirical work done to date are not sufficient to rule on the question of whether the time series are compatible with a model in which firms are always on their demand schedules for employment. This is true according to virtually any dynamic and stochastic theory of the demand for employment. In this paper, I have tried to indicate one way in which the time-series evidence can be brought to bear on the question in the context of a simple dynamic, stochastic model. The empirical results are moderately comforting to the view that the employment–real-wage observations lie along a demand schedule for employment. It is important to emphasize that this view has content (i.e., imposes overidentifying restrictions) because I have a priori imposed restrictions on the orders
of the adjustment-cost processes and on the Markov processes governing disturbances. At a general level without such restrictions, it is doubtful whether the equilibrium view has content.

Appendix

Vector Autoregressions and Moving Averages

Let $x_t$ be an $(nx1)$ vector jointly covariance stationary, linearly indeterministic stochastic process. The $m$th-order vector autoregression for this process is

$$x_t = \alpha + \sum_{j=1}^{m} A_j x_{t-j} + \epsilon_t^m,$$  \hspace{1cm} (A1)

where $\epsilon_t^m$ is an $(nx1)$ vector of least-squares disturbances. Here $\alpha$ is an $(nx1)$ vector and the $A_j$'s are $nxn$ matrices that under mild regularity conditions are uniquely determined by the population orthogonality conditions $E\epsilon_t^m = 0$ and $E\epsilon_t^m x_{t-j} = 0_{nxn}, j = 1, 2, \ldots, m$. The $\epsilon_t^m$ process is termed the process of innovations, the parts of $x_t$ that cannot be predicted linearly from $m$ lagged $x_t$'s; $\epsilon_t^m$ is the process of one-step-ahead prediction errors. If $m = \infty$, the orthogonality conditions imply that $E\epsilon_t^m \epsilon_{t-s} = 0$ for $s \neq 0$, having the practical implication that if $m$ is taken to be big enough, as we shall assume, the $\epsilon_t^m$ vector is serially uncorrelated. If we solve the vector difference equation (A1) for $x_t$ backward in terms of the $\epsilon$ process and ignore transient terms, we get the vector moving average representation

$$x_t = \alpha' + \sum_{j=0}^{\infty} C_j \epsilon_{t-j},$$  \hspace{1cm} (A2)

where $\alpha'$ is an $(nx1)$ vector of constants, where $C_j$ is an $(nxn)$ matrix and $C_0 \equiv I$. The matrix Fourier transforms of the $A_j$'s and $C_j$'s are related by

$$(I - A_1 e^{-iw} - \ldots - A_m e^{-iw^m})^{-1} = \sum_{j=0}^{n} C_j e^{-iw^j}.$$  

The $(nx1)$ vector process $\epsilon_t^m$ is composed of disturbances that are mutually orthogonal at all nonzero lags and leads (by the orthogonality conditions), but

$$E\epsilon_t^m \epsilon_{t}' = \sum$$

is not in general diagonal. To illustrate how to construct a moving average representation with a disturbance process that is orthogonal contemporaneously as well as at all lags, let $n = 2$ and consider the transformation

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = F u_t,$$

where $\rho = E\epsilon_{1t} \epsilon_{2t}/E\epsilon_{1t}^2$. Here we are choosing $u_{1t} = \epsilon_{1t}$ and are decomposing $\epsilon_{2t}$ by the least-squares projection equation $\epsilon_{2t} = \rho \epsilon_{1t} + u_{2t}$, where the least-squares orthogonality condition $E u_{2t} \epsilon_{1t} = 0$ implies that $\rho = E\epsilon_{1t} \epsilon_{2t}/E\epsilon_{1t}^2$. Here $u_{2t}$ is the part of $\epsilon_{2t}$ that is orthogonal to $\epsilon_{1t}$. By construction, $u_{1t}$ and $u_{2t}$ are orthogonal. Therefore, a new moving average representation in terms of mutually orthogonal disturbances at all lags is given by

$$x_t = \alpha' + \sum_{j=0}^{\infty} C_j Fu_{t-j}$$

$$= \alpha' + \sum_{j=0}^{\infty} D_j u_{t-j},$$
where \( D_j = C_j F \). Of course, there is more than one such choice of \( u_t \) processes that does the job. For example, in the \( n = 2 \) example, we could have selected \( u_{1t} = \varepsilon_{2t} \) and then chosen \( u_{2t} \) as the part of \( \varepsilon_{1t} \) that is orthogonal to \( \varepsilon_{2t} \). In the text, for the \( n = 2 \) case, I have calculated moving average representations for both of the ways of choosing \( u_t \) discussed above. More generally any choice of \( u_t = F^{-1} \varepsilon_t \) that makes \( E u_t u_t' = F^{-1} \sum F''^{-1} \) a diagonal matrix can be used to deliver a moving average representation in terms of a \( u \) process that is orthogonal contemporaneously as well as at all leads and lags.

In the \( n = 2 \) case, the first-mentioned way of defining \( u_t \) is equivalent with changing the form of the vector autoregression (A1) by adding current \( x_{1t} \) to the right-hand side of the autoregression for \( x_{2t} \) and then solving the vector-difference equation for a moving average representation in terms of the vector of residuals from this pair of autoregressions. The second-mentioned way of defining \( u_t \) amounts to changing the form of the vector autoregression (A1) by adding current \( x_{2t} \) to the right-hand side of the autoregression for \( x_{1t} \) (leaving current \( x_{1t} \) excluded from the autoregression for \( x_{2t} \)) and calculating the moving average in terms of the residuals from these equations.

The \( k \)-step-ahead error in forecasting \( x_t \) linearly from its own past is given by

\[
E_{t-k} x_t = C_0 e_t + \cdots + C_{k-1} e_{t-k+1} + D_0 u_t + \cdots + D_{k-1} u_{t-k+1},
\]

where \( E_{t-k} x_t \) is the linear least-squares forecast of \( x_t \) given \( x_{t-k}, x_{t-k+1}, \ldots \). From the extensive orthogonality least-squares estimates of the vector autoregression (A1) are known to be statistically consistent (Anderson and Taylor [1976] and Ljung [1976]). For a more extensive discussion of vector stochastic processes and some macroeconomic applications, see Sargent (1978b).

References

Anderson, T. W., and Taylor, John B. "Conditions for Strong Consistency of Least Squares Estimates in Linear Models." Technical Report no. 213, Stanford Univ. Inst. Math. Studies Soc. Sci., August 1976.

Bard, Yonathan. Nonlinear Parameter Estimation. New York: Academic Press, 1974.

Barro, Robert J., and Grossman, H. I. "A General Disequilibrium Model of Income and Employment." A.E.R. 61 (March 1971): 82–93.

Bodkin, Ronald G. "Real Wages and Cyclical Variations in Employment: A examination of the Evidence." Canadian J. Econ. 2 (August 1969): 353–74.

Dunlop, John T. "The Movement of Real and Money Wage Rates." Econ. J. 48 (September 1938): 413–34.

Durbin, J. "Tests for Serial Correlation in Residual Analysis Based on the Periodogram of Least-Squares Residuals." Biometrika 56 (March 1969): 1–15.

Gould, John P. "Adjustment Costs in the Theory of Investment of the Firm." Rev. Econ. Studies 35 (January 1968): 47–55.

Granger, C. W. J. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." Econometrica 37, no. 3 (July 1969): 424–38.
Graves, R., and Telser, L. Functional Analysis in Mathematical Economics: Optimization over Infinite Horizons. Chicago: Univ. Chicago Press, 1972.

Holt, Charles C.; Modigliani, F.; Muth, T. F.; and Simon, H. A. Planning Production, Inventories, and Work Force. Englewood Cliffs, N.J.: Prentice-Hall, 1960.

Keynes, John Maynard. "Relative Movements of Real Wages and Output." *Econ. J.* 49 (March 1939): 34–51.

Kwakernaak, Huibert, and Sivan, Raphael. Linear Optimal Control Systems. New York: Wiley Interscience, 1972.

Ljung, L. "Consistency of the Least Squares Identification Method." *IEEE Transactions on Automatic Control* (October 1976), pp. 779–81.

Lucas, Robert E. "Adjustment Costs and the Theory of Supply." *J.P.E.* 75, no. 4, pt. 1 (August 1967): 321–34.

———. "Capacity, Overtime, and Empirical Production Functions." *A.E.R. Papers and Proc.* 60, no. 2 (May 1970): 23–27.

———. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets*, edited by Karl Brunner and Allan H. Meltzer. The Carnegie-Rochester Conferences on Public Policy, a Suppl. Series to the *J. Monetary Econ.*, vol. 1. Amsterdam: North-Holland, 1976.

Neftci, Salih N. "A Time Series Analysis of the Real Wages-Employment Relationship." *J.P.E.* 86, no. 2, pt. 1 (April 1978): 281–91.

Phelps, Edmund S. "A Note on Short-Run Employment and Real Wage Rate under Competitive Commodity Markets." *Internat. Econ. Rev.* 10, no. 2 (June 1969): 220–32.

Sargent, Thomas J. "Comment on Seasonal Adjustment and Multiple Time Series Analysis by Kenneth Wallis." Paper presented at NBER/Bureau of Census Conference on Seasonal Adjustment of Economic Time Series, Washington, September 9–10, 1976.

———. "The Demand for Money during Hyperinflations under Rational Expectations: I." *Internat. Econ. Rev.* 18 (February 1977): 59–82.

———. "Rational Expectations, Econometric Exogeneity, and Consumption." *J.P.E.* 86, no. 4 (August 1978): 673–700. (a).

———. *Notes on Macroeconomic Theory.* New York: Academic Press, 1978, in press. (b)

Sargent, Thomas J., and Wallace, Neil. "The Elasticity of Substitution and Cyclical Behavior of Productivity, Wages, and Labor's Share." *A.E.R. Papers and Proc.* 64 (May 1974): 257–63.

Sims, Christopher A. "Money, Income, and Causality." *A.E.R.* 62, no. 4 (September 1972): 540–52.

———. "Response to Sargent's Comment on Wallis' Paper." Manuscript. Univ. Minnesota, November 1976.

Solow, Robert M., and Stiglitz, J. E. "Output, Employment, and Wages in the Short Run." *Q.J.E.* 82 (November 1968): 537–60.

Tarshis, Lorie. "Changes in Real and Money Wages." *Econ. J.* 49 (March 1939): 150–54.

Taylor, John B. "Estimation and Control of a Macroeconomic Model with Rational Expectations." Manuscript. Columbia Univ., February 1978.

Treadway, Arthur B. "On Rational Entrepreneurial Behavior and the Demand for Investment." *Rev. Econ. Studies* 36, no. 106 (April 1969): 227–39.

Wilson, G. T. "The Estimation of Parameters in Multivariate Time Series Models," *J. Royal Statist. Soc. ser.* B, no. 1 (1973), pp. 76–85.