

RECURSIVE SOLUTION METHODS FOR DYNAMIC LINEAR RATIONAL EXPECTATIONS MODELS*

Mark W. WATSON*

*Northwestern University, Evanston, IL 60208, USA
and
National Bureau of Economic Research*

This paper develops recursive solution methods for linear rational expectations models. The underlying structural model is transformed into a state-space representation, which can then be used to solve the model and to form the Gaussian likelihood function. The recursive solution method has several advantages over other approaches. First, the set of solutions to the model are summarized by a set of parameters that appear in the state-space representation but are unspecified by the structural model. Next, the likelihood function is formed as a byproduct of the solution to the model. Finally, modifications in the likelihood function necessary to incorporate complications arising from temporal aggregation, dynamic errors-in-variables, etc. are straightforward.

1. Introduction

Models in which rational expectations of future values of variables are simultaneously determined with current values of these variables play an important role in economics and finance. Linear versions of these models have been used to model a wide variety of phenomena, but have posed some unique problems. The presence of endogenous future expectations together with current values makes these models more difficult to solve than standard linear models; moreover, the solutions are often nonunique.¹ These models have also posed interesting econometric problems, which stem from the presence of unobserved expectations in the structural form of the model. Instrumental

*This paper has benefited from helpful comments and suggestions from seminar participants at Princeton, Rochester, Iowa, and Virginia. I owe particular thanks to Olivier Blanchard, Gary Chamberlain, Rob Engle, Charles Whiteman, and an anonymous referee. The financial support of the National Science Foundation is gratefully acknowledged.

¹A large number of solution procedures exist. See, for example, Blanchard and Kahn (1979), Whiteman (1983), Chow (1983), Gourieroux, Laffont, and Monfort (1982), and Broze, Gourieroux, and Szafarz (1985).

variable estimators provide one solution to the problem.² Full information methods have also been developed and applied. [Examples are Sargent (1979), Hansen and Sargent (1980), and Blanchard (1983).]

Full information methods are generally difficult to implement. The usual practice is to find an analytic solution to the model which produces a reduced form relating current and lagged values of the endogenous variables to current and lagged values of the exogenous variables, some disturbances, and the parameters of the model. This reduced form usually implies that the data can be represented as a vector ARMA process with complicated constraints connecting the ARMA coefficients. This constrained ARMA model can then be estimated by nonlinear maximum likelihood methods.

In this paper I propose an alternative procedure for solving and estimating dynamic linear rational expectations models. The method relies on the state-space representation of the model. The state-space approach to the formulation and estimation of dynamic linear rational expectations models has several advantages over other approaches. First, the constraints that the model places on the data are transparent. The complete set of solutions to the model is conveniently summarized by a set of parameters that is not determined in the structural model, nor by the process generating the exogenous variables. Next, the model is easily solved recursively by the Kalman filter; an algebraic solution to the model is unnecessary. As byproducts of the recursive solution procedure the Kalman filter produces the innovations and innovation variances of the observed data. These are the basic building blocks of the Gaussian likelihood function. Finally, and of particular importance in applied work, modifications in the empirical model that are necessary because of data limitations are easily incorporated. Modifications in the likelihood function necessary to incorporate complications arising from missing observations, temporal aggregation, interpolation, and dynamic errors-in-variables are straightforward. [The necessary modifications are analogous to those discussed in Harvey, McKenzie, Blake, and Desai (1981), Harvey and Pierse (1984), and Hausman and Watson (1985).]

This paper begins, in section 2, with a very simple model with one endogenous variable, one future expectation, and one exogenous variable. Most of the key features of the state-space representation of dynamic linear rational expectations models can be presented using this simple example. In section 3 we present the generalization of this model with more exogenous variables, longer future expectations, and lags of the endogenous variable, etc. In section 4 we discuss properties of the estimated parameters obtained by maximizing a Gaussian likelihood function, and show the importance of imposing stationar-

²The original IV estimator was suggested by McCallum (1976). A systematic discussion of IV estimators in dynamic linear rational expectations models can be found in Hansen and Sargent (1982).

ity on the solution of the model. An empirical study investigating the relationship between stock prices and dividends is presented in section 5, and some concluding remarks are offered in section 6.

2. A simple example

Many of the important characteristics of the recursive solution of linear dynamic rational expectations models can be illustrated with a simple example. In this section we analyze such a model in detail, and in the next section we discuss a generalization. The model that we consider is

$$y_{t+1/t} = \mu y_t + x_t, \quad t = 1, 2, \dots, \quad (2.1)$$

$$x_t = \phi x_{t-1} + e_t^x, \quad t = 1, 2, \dots, \quad (2.2)$$

$$x_0 = x0, \quad (2.3)$$

where $|\phi| < 1$ and e_t^x is iid with mean zero and variance σ_x^2 . Eq. (2.1) is a structural equation describing the relation between the expected value of y_{t+1} made at time t , denoted by $y_{t+1/t}$, the actual value of y at time t , denoted by y_t , and an exogenous variable x_t . Eq. (2.2) is the equation of motion for x_t , and (2.3) specifies the initial condition for x_t . We assume that the expectation $y_{t+1/t}$ is formed rationally:

$$y_{t+1/t} = E(y_{t+1} | \Omega_t), \quad (2.4)$$

where Ω_t is the set of information available at time t . Throughout this paper we use $z_{t/t-i}$ to denote $E(z_t | \Omega_{t-i})$, where the information set Ω_{t-i} will be clear from the context. We will assume that $(x_t, y_t) \in \Omega_t$ and $\Omega_{t-1} \subset \Omega_t$ for $t = 1, 2, \dots$.

Many models fit into this framework. Eq. (2.1) can be viewed as a (re-arranged) money demand equation in which y and x are the logarithms of the price level and the quantity of money, respectively. In this interpretation eq. (2.1) describes the demand for real balances as a function of expected price inflation. Alternatively, y_t can be viewed as the price of an asset, x_t as the value of services realized for the ownership of the asset, and (2.1) is a standard arbitrage equation.

The recursive solution for the model relies on a reduced form that expresses the variables at time t in terms of predetermined variables and innovations. This reduced form leads directly to the state-space representation of the model. It describes the evolution of all variables of interest, including the

expectational variables, and is written as

$$y_t = y_{t/t-1} + e_t^y, \quad E(e_t^y | \Omega_{t-1}) = 0, \quad (2.5)$$

$$x_t = x_{t/t-1} + e_t^x, \quad E(e_t^x | \Omega_{t-1}) = 0, \quad (2.6)$$

$$x_{t+1/t} = \phi x_{t/t-1} + \phi e_t^x, \quad (2.7)$$

$$y_{t+1/t} = \mu y_{t/t-1} + x_{t/t-1} + \mu e_t^y + e_t^x. \quad (2.8)$$

Eqs. (2.5) and (2.6) define rational expectations for y_t and x_t . Eqs. (2.7) and (2.8) describe the evolution of these expectations. Eq. (2.7) follows directly from (2.2) and (2.6); eq. (2.8) follows from (2.1) and (2.5)–(2.6). Other treatments of this example [Blanchard (1979), Gourieroux, Laffont, and Monfort (1982), Chow (1983), and Whiteman (1983)] skip over this reduced form and present what is essentially a final form of the model. Their methods eliminate the expectational terms $y_{t/t-1}$ and $x_{t/t-1}$ using (2.7) and (2.8) so that the observed data are expressed in terms of the innovations e_t^y and e_t^x . While this additional step is necessary for a closed form solution of the model, it is not necessary for a recursive solution, nor to form the likelihood of the observed data. As we show below, the likelihood can be obtained directly from (2.5)–(2.8) using standard recursive formulae.

Before writing the model in state–space form the properties of the process $\{e_t^x, e_t^y\}$ and initial values $x_{1/0}$ and $y_{1/0}$ must be specified. Our concern will be limited to the second-moment properties of the data, thus we need only specify the covariance properties of $\{e_t^x, e_t^y\}$.

The properties of $\{e_t^x\}$ and the initial value $x_{1/0}$ follow directly from the structural model. The disturbance e_t^x is serially uncorrelated with variance σ_x^2 . The initial value of $x_{1/0}$ is ϕx_0 . The model provides less information about the covariance properties of $\{e_t^y\}$ and the value of $y_{1/0}$. From eq. (2.4) we know that $\{e_t^y, \Omega_t\}$ forms a martingale difference sequence. Endowing e_t^y with second moments leads to the implication that e_t^y is uncorrelated with its own lagged values and the lagged values of e_t^x . The variance of e_t^y is left unrestricted by the model as is the covariance between e_t^y and e_t^x . Equivalently, projecting e_t^y onto e_t^x , we have $e_t^y = \pi e_t^x + u_t$ where π and σ_u^2 are left unrestricted by the model.³ The model also leaves $y_{1/0}$ unrestricted. Different values of the parameters π , σ_u^2 , and $y_{1/0}$ lead to different first and second moments for the $\{y_t, x_t\}$ process and correspond to different solutions of the model.

³Nothing in the model imposes the constraint that π and σ_u are time-invariant. Indeed, as pointed out by Pesaran (1981) and Broze, Gourieroux, and Szafarz (1985), the only restriction that is placed on the e_t^y sequence is the martingale difference restriction shown in eq. (2.5).

These three parameters completely characterize the set of solutions to (2.1)–(2.3) which have a time-invariant parameterization and describe the first two moments of $\{y_t, x_t\}$. It is instructive to interpret these parameters in terms of other solutions that have appeared in the literature. The disturbance u_t represents a shock to the expectations of future y_t 's that is uncorrelated with the fundamental driving variable x_t . It represents a 'stochastic bubble', discussed in Blanchard and Watson (1982), Grossman and Diba (1983), West (1986), and elsewhere. The parameter σ_u^2 indexes the relative importance of this stochastic bubble. The coefficient π transmits new information about the current values of x to the expectations of future values of y . It determines whether the process is 'forward looking' or 'backward looking' or some combination of both [see Blanchard (1979)]. Finally, the coefficient $y_{1/0}$ represents the initial condition for the expectations process. It reflects the initial information about the influence of x on future y and the influence of any 'deterministic bubbles' on future y . [See Flood and Garber (1980).]

Using this set of parameters, the state-space representation for the model is

$$Y_t = SX_t + Ae_t, \quad (2.9)$$

$$X_{t+1} = FX_t + Ge_t, \quad (2.10)$$

where

$$Y_t = (x_t \quad y_t)',$$

$$X_t = (x_{t/t-1} \quad y_{t/t-1})',$$

$$e_t = (e_t^x \quad u_t)',$$

$$S = I_2,$$

$$A = \begin{bmatrix} 1 & 0 \\ \pi & 1 \end{bmatrix}, \quad F = \begin{bmatrix} \phi & 0 \\ 1 & \mu \end{bmatrix}, \quad G = \begin{bmatrix} \phi & 0 \\ 1 + \mu\pi & \mu \end{bmatrix}.$$

Eq. (2.9) is the measurement equation; it relates the observed data, Y_t , to the state vector, X_t , and a vector of innovations. Eq. (2.10) describes the evolution of the state. The state vector contains all information from the past useful for predicting the future. Together with the initial condition for the state, eqs. (2.9)–(2.10) serve as a complete characterization of the first- and second-moment properties of the data.

In any empirical application, the parameters μ , ϕ , π , σ_x^2 , and σ_u^2 are unknown and must be estimated using data on Y_t . This can be done by maximizing the Gaussian likelihood function. The Gaussian likelihood can be formed from the data in a straightforward manner using the Kalman filter [see

Schweppe (1965) or Harvey (1982)]. The properties of estimators formed by maximizing this function are the subject of section 4.

In many specific applications of the model, the parameter μ is restricted *a priori* to be larger than one. For instance, in the stock price/dividend interpretation of the model, the parameter $\mu = (1 + r)$ where r is the interest rate. Since μ is an eigenvalue of the state transition matrix, F , this implies that a linear combination of the expectations $x_{t/t-1}$, $y_{t/t-1}$ is exploding. But with $|\phi| < 1$, x_t is stationary, so that any explosive behavior must be occurring in y_t . For a variety of reasons it may be reasonable to rule out this type behavior *a priori*. For instance, one might insist on a solution to the model in which stationary inputs (x) lead to stationary outputs (y). This can usually be accomplished by choosing a parameterization which annihilates the linear combination of the elements of the state vector corresponding to the explosive root. In this example the explosive linear combination is given by $w_{t+1/t} = y_{t+1/t} + x_{t+1/t}(\mu - \phi)^{-1}$. Substituting from (2.7) and (2.8) we have

$$w_{t+1/t} = \mu w_{t/t-1} + \mu [ce_t^x + u_t], \quad (2.11)$$

where $c = \pi + (\mu - \phi)^{-1}$. When $\mu > 1$, the expectations of future y will explode (almost surely) as t increases unless $w_{t+1/t} = 0$ for all t . This explosive behavior can be ruled out if and only if

$$w_{1/0} = 0 \quad \Rightarrow \quad y_{1/0} = -x_{1/0}(\mu - \phi)^{-1}, \quad (2.12)$$

$$c = 0 \quad \Rightarrow \quad \pi = (\phi - \mu)^{-1}, \quad (2.13)$$

$$u_t = 0, \quad t = 1, 2, \dots \quad \Rightarrow \quad \sigma_u^2 = 0. \quad (2.14)$$

This nonexplosiveness condition imposes three constraints on the model when $\mu > 1$. These three constraints, given in (2.12)–(2.13), exactly determine the three parameters $y_{1/0}$, π , and σ_u that characterize the set of solutions to the model (2.1)–(2.3). The uniqueness of the solution in this case has been pointed out in many places. It is implied by Proposition 1 of Blanchard and Kahn (1980), Lemma 1 of Whiteman (1983), and Theorem 5.5 of Broze, Gourieroux, and Szafarz (1985).

In this simple model, condition (2.14) implies that the bivariate (x_t, y_t) process is singular, and thus is a condition that would be rejected in any empirical setting. However, this singularity is an artifact of the very simple structure underlying the example used in this section. In the next section a more general structure is introduced and this singularity disappears. Stationarity in these more general structures does require knife-edge parametric constraints analogous to (2.12) and (2.13).

3. A generalization

In this section we present a generalization of last section's model. While the notation in this new model is more complicated than in the model of section 2, its state-space representation is essentially the same. The model that we will consider is

$$y_{t+r/t} = \sum_{i=1}^{r-1} \mu^i y_{t+r-i/t} + \sum_{j=0}^p \mu^{j+r} y_{t-j} + x_t' \beta + u_t^s, \tag{3.1}$$

where x_t is a k -vector of observed variables and u_t^s is an unobserved structural disturbance. Letting $z_t = (x_t' u_t^s)'$, the 'forcing variables' x_t and u_t^s are generated by the vector ARMA process

$$z_t = \sum_{i=1}^q \Phi_i z_{t-i} + \sum_{i=0}^{q-1} \Theta_i e_{t-i}^z, \tag{3.2}$$

where $\Theta_0 = I$, and we have assumed (without loss of generality) that the order of the AR is q and the order of the MA is $q - 1$. The vector e_t^z is white noise with mean zero and covariance matrix Σ_z , which we partition as

$$\text{var}(e_t^z) = \text{var} \begin{bmatrix} e_t^x \\ e_t^s \end{bmatrix} = \begin{bmatrix} \Sigma_x & 0 \\ 0 & \sigma_s^2 \end{bmatrix}.$$

We will find it useful to write the z_t process in first-order form as

$$Z_t = \Phi Z_{t-1} + H e_t^z, \tag{3.3}$$

where $Z_t' = [Z_t^1, Z_t^2, \dots, Z_t^{q'}]$. Each component, Z_t^i , is a $(k + 1)$ vector, and

$$\Phi = \begin{bmatrix} \Phi_1 & & & \\ \Phi_2 & & & \\ \vdots & & & \\ \Phi_q & & & 0 \end{bmatrix},$$

and $H = [I_{k+1} \Theta_1' \Theta_2' \dots \Theta_{q-1}']'$. The first component of Z_t , Z_t^1 , is z_t . The other $q - 1$ components are implicitly defined by eq. (3.3). It will also be convenient

to stack the endogenous variables into a vector

$$y_t^e = [y_{t+r-1/t}, y_{t+r-2/t}, \dots, y_{t+1/t}, y_t, y_{t-1}, \dots, y_{t-p}]'$$

Eq. (3.1) can then be written as

$$y_{t+1/t}^e = \mu y_t^e + \delta Z_t, \quad (3.4)$$

where

$$\mu = \left[\begin{array}{ccc|c} \mu_1 & \mu_2 & \dots & \mu^{p+r} \\ \hline & & & 0 \\ & I_{r+p-1} & & \vdots \\ & & & 0 \end{array} \right],$$

$$\delta = \left[\begin{array}{cc} \bar{\beta} & 0_{1 \times (q-1)(k+1)} \\ 0_{(r+p-1) \times (k+1)} & 0_{(r+p-1) \times (q-1)(k+1)} \end{array} \right],$$

where $\bar{\beta}$ is the $1 \times (k+1)$ vector $\bar{\beta} = [\beta' \ 1]$ and $0_{k \times j}$ is a $k \times j$ matrix of zeros. We assume that the initial condition for Z_t is

$$Z_0 = Z0. \quad (3.5)$$

Eqs. (3.3)–(3.5) are the multivariate analogues of (2.1)–(2.3). The reduced form of (3.3)–(3.5) is

$$y_t^e = y_{t/t-1}^e + \varepsilon_t^y, \quad E[\varepsilon_t^y | \Omega_{t-1}] = 0, \quad (3.6)$$

$$Z_t = Z_{t/t-1} + \varepsilon_t^z, \quad E[\varepsilon_t^z | \Omega_{t-1}] = 0, \quad (3.7)$$

$$Z_{t+1/t} = \Phi Z_{t/t-1} + \Phi \varepsilon_t^z, \quad (3.8)$$

$$y_{t+1/t}^e = \mu y_{t/t-1}^e + \delta Z_{t/t-1} + \mu \varepsilon_t^y + \delta \varepsilon_t^z, \quad (3.9)$$

where $\varepsilon_t^z = H \varepsilon_t^z$ from (3.3). Since the vector y_t^e contains lagged values of y_t , which are known at time $t-1$, many of the elements of ε_t^y are zero. We will write ε_t^y as

$$\varepsilon_t^{y'} = [\varepsilon_{r-1,t}^y, \varepsilon_{r-2,t}^y, \dots, \varepsilon_{0,t}^y, 0, 0, \dots, 0],$$

where

$$\epsilon_{j,t}^y = y_{t+j/t} - y_{t+j/t-1} \quad \text{for } j = 0, 1, \dots, r - 1.$$

This reduced form can be used to discuss the set of solutions to the model (3.3)–(3.5). We restrict ourselves to solutions with time-invariant parameterizations which describe the first and second moments of the $\{x_t, y_t\}$ process. Eqs. (3.7) and (3.8) together with Σ_z completely characterize the $\{x_t\}$ process. Eqs. (3.6) and (3.9) describe the generation of y_t . The y_t data are driven by the exogenous variables Z_t plus the innovations $\epsilon_{j,t}^y$. These $\epsilon_{j,t}^y$ innovations are martingale difference sequences with respect to Ω_{t-1} . Together with the set of initial conditions given by $y_{1/0}^e$ these sequences determine the set of solutions to the model [see Broze, Gourieroux, and Szafarz (1985)]. Since our concern is with the first and second moments of (x_t, y_t) we need (i) the initial conditions $y_{1/0}^e$ and (ii) the covariance properties of ϵ_t^y . These are left unspecified by the (3.3)–(3.5) so that the solution to the model characterized by (3.6)–(3.9) is not unique. The set of solutions can be parameterized by the $r + p$ initial conditions $y_{1/0}^e$, the $r(r + 1)/2$ parameters in the covariance matrix of ϵ_t^y , and the $r(k + 1)$ covariances between ϵ_t^z and ϵ_t^y . Alternatively, if we let $e_t^y = [\epsilon_{r-1,t}^y, \epsilon_{r-2,t}^y, \dots, \epsilon_{0,t}^y]'$ denote the nonzero elements of ϵ_t^y and project e_t^y onto e_t^z , we have

$$e_t^y = \tilde{\Pi} e_t^z + u_t, \tag{3.10}$$

where u_t and e_t^z are uncorrelated and $\text{var}(u_t) = \Sigma_u$. (The reason for the $\tilde{\cdot}$ over Π will soon be clear.) The set of solutions is characterized by the $(r + p) + r(k + 1) + r(r + 1)/2$ parameters in $y_{1/0}^e, \tilde{\Pi}$, and Σ_u .

The model (3.6)–(3.9) can now be written in state–space form. One state–space representation is

$$Y_t = SX_t + \tilde{A}\epsilon_t, \tag{3.11}$$

$$X_{t+1} = FX_t + \tilde{G}\epsilon_t, \tag{3.12}$$

where Y_t is the $k + 1$ vector $[x_t', y_t']$, X_t is the $n \times 1$ vector $[Z_{t/t-1}', y_{t/t-1}^e]'$ [where $n = (k + 1)q + r + p$], S is a $(k + 1) \times n$ selection matrix which selects $x_{t/t-1}$ and $y_{t/t-1}$ from the state vector. The innovation vector $\epsilon_t' = [e_t^{z'} u_t]$, and $\tilde{A}' = [\tilde{A}_1' \tilde{A}_2']$, where \tilde{A}_1 is $k \times (k + 1 + r)$ and \tilde{A}_2 is $1 \times (k + 1 + r)$. These blocks are $\tilde{A}_1 = [I_k \ 0_{k \times (r+1)}]$ and $\tilde{A}_2 = [\tilde{\pi}_r \ l_r]$, where $\tilde{\pi}_r$ is the r th row of $\tilde{\Pi}$ and l_r is an $r \times 1$ row vector with 0's in every entry except the last, which is equal to 1. Thus the innovation in x_t is $\tilde{A}_1 \epsilon_t = \epsilon_t^x$ and the innovation in y_t is $\epsilon_{0,t}^y$, which is written as $\epsilon_{0,t}^y = \tilde{\pi}_r \epsilon_t^z + l_r u_t$ from (3.10).

The matrices F and G are given by

$$F = \begin{bmatrix} \Phi & 0_{q(k+1) \times (r+p)} \\ \delta & \mu \end{bmatrix},$$

$$G = \begin{bmatrix} \Phi H & 0_{q(k+1) \times r} \\ \mu \tilde{\Pi}^a + \delta H & \mu L \end{bmatrix},$$

with

$$\Pi^a = \begin{bmatrix} \tilde{\Pi} \\ 0_{p \times (k+1)} \end{bmatrix}, \quad L = \begin{bmatrix} I_r \\ 0_{p \times r} \end{bmatrix}.$$

Together with the initial conditions the state-space model (3.11)–(3.12) completely characterizes the first and second moments of the observed data and can be used to form the Gaussian likelihood of the data. The unknown parameters could then be estimated by maximum likelihood methods, as described, for example, in Watson and Engle (1983).

For purposes of estimation an alternative state-space representation of the data is more convenient, since many of the parameters in the representation (3.11)–(3.12) have only a transient impact on the data. These parameters have no effect on the steady-state solution and cannot be consistently estimated from the data. In this sense the state-space representation given by (3.11) and (3.12) is overparameterized. Notice, for example, that the representation (3.11)–(3.12) characterizes the second moments of the $k+1$ vector Y_t using $r+k+1$ noises in ε_t , which in turn are characterized by $(r+k+1)(r+k+2)/2$ covariances. An alternative representation of the data would characterize the second moments of Y_t in terms of $k+1$ fundamental noises, characterized by $(k+1)(k+2)/2$ covariances, plus some initial conditions. This alternative characterization is the ‘innovations representation’ of the state-space model. To derive this representation, suppose that the data are passed through a Kalman filter. We would then obtain

$$Y_t = SX_{t/t-1} + v_t, \quad (3.13)$$

$$X_{t+1/t} = FX_{t/t-1} + K_t v_t, \quad (3.14)$$

where $v_t = Y_t - E(Y_t | \Omega_{t-1})$ and the information set Ω_t is composed of $(y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1)$ and the initial values of $Z_{1/0}$ ($= \Phi Z_0$) and $y_{1/0}^e$. The matrix K_t is the Kalman gain, and we'll let $h_t = \text{var}(v_t | \Omega_{t-1})$. Eqs. (3.13) and (3.14) have the same basic structure as (3.11) and (3.12). The main difference is that the $(k+1+r)$ shocks in ε_t which drive (3.11)–(3.12) have been replaced with the $(k+1)$ shocks in v_t . Unfortunately, the alternative

representation is not time-invariant since both h_t and K_t will, in general, depend on t . However, for our particular model it is possible to show that both h_t and K_t converge to steady state values and therefore eqs. (3.13) and (3.14) can serve as a convenient parsimonious steady-state characterization of the second-moment properties of the data. This result is stated formally in:

Theorem 1. For the model given by (3.13)–(3.14), with given initial conditions we have

- (i) $\lim_{t \rightarrow \infty} h_t = h,$
- (ii) $\lim_{t \rightarrow \infty} K_t = K,$
- (iii) h can be written as $AQA',$
- (iv) K can be written as $GA^{-1},$

where the parameterizations of h and K are chosen to mimic the structure of (3.11) and (3.12). In particular,

$$G = \begin{bmatrix} \Phi H \\ \mu \Pi^a + \delta H \end{bmatrix} \quad \text{with} \quad \Pi^a = \begin{bmatrix} \Pi \\ 0_{p \times (k+1)} \end{bmatrix}.$$

Π is an $r \times (k + 1)$ matrix, $A' = [A'_1 A'_2]$, where A_1 is $k \times (k + 1)$ and A_2 is $1 \times (k + 1)$. These blocks are $A_1 = [I_k \ 0]$ and $A_2 = \pi_r$, where π_r is the r th row of the matrix Π appearing in G . Finally, Q is a $(k + 1) \times (k + 1)$ matrix:

$$Q = \begin{bmatrix} \Sigma_x & 0 \\ 0 & \sigma^2 \end{bmatrix}.$$

Proof. See appendix.

The intuition underlying the theorem is straightforward. The representation of the model given in (3.11) and (3.12) describes the $k + 1$ elements in Y_t in terms of the $k + 1 + r$ elements of the noise vector ε_t . The multivariate Wold representation theorem suggests that we should be able to characterize the second-moment properties of the $k + 1$ elements in Y_t using a noise vector containing just $k + 1$ elements, and this is just what the innovations representation of the model does. The specific initial conditions assumed in (3.11) and (3.12) lead to some transient heteroskedasticity in the innovations.⁴

⁴Initial conditions in time series models often leads to transient heteroskedasticity in the innovation series. ARMA models provide one example.

When the data are stationary, assumptions concerning initial conditions have an $O_p(1)$ effect on the likelihood function, and therefore have no effect on the asymptotic distribution of the estimated parameters. This suggests that, for purposes of estimating the parameters of the model, the initial conditions can be ignored. The likelihood function can be formed using the steady-state solution to the model. A parameterization of the steady-state model, which parallels the state-space form in (3.11) and (3.12) follows from (iii) and (iv) of Theorem 1:

$$Y_t = SX_{t/t-1} + Aa_t, \quad (3.15)$$

$$X_{t+1/t} = FX_{t/t-1} + Ga_t, \quad (3.16)$$

with $\text{var}(a_t) = Q$, and initial conditions given by $X_{1/0} = [Z'_{1/0}, y'_{1/0}]$.

The advantage of using the steady-state parameterization of the model is that it abstracts from the set of parameters that have a transient effect on the second moments of Y_t through their effect on the initial conditions. Throughout the remainder of the paper we will deal with the representation of the model given in (3.15) and (3.16).

4. Estimation

When the data are stationary and stationarity is imposed on the solution, it is a straightforward exercise to derive the limiting distribution of the estimated parameters. When the data are not stationary, or are stationary and stationarity is not imposed on the solution, the properties of the estimated parameters are less clear. Because of the importance of imposing stationarity on the data, we begin this section with a discussion of the mechanics for doing so. We then present a set of conditions which guarantee that the estimated parameters will be consistent with the usual asymptotic normal distribution. We end this section with a discussion of the consequences of estimating the parameters without imposing the stationary solution.

When the eigenvalues of F are inside the unit circle the data are covariance-stationary. The data may also be covariance stationary when some of the eigenvalues of F are outside the unit circle. This occurs when the parameter values of the model are such that the linear combinations of the elements of the state corresponding to the explosive eigenvalues are identically equal to zero. Imposing this condition is straightforward.

Let d_i be an eigenvector of F' corresponding to a root λ_i , with $|\lambda_i| > 1$. We have $d_i'X_{t+1/t} = \lambda_i d_i'X_{t/t-1} + d_i'Ga_t$. Since the model is completely detectable (see the proof of Theorem 1), this explosive linear combination of the X 's will lead to explosive behavior in the y 's. The explosive behavior in the y 's can

only be ruled out if $d'_i X_{t+1/t} = 0$ for $t = 1, 2, \dots$. But this will occur only if $d'_i X_{1/0} = 0$ and $d'_i G = 0$. These two conditions must be imposed on the model for every root greater than one in modulus. Each root of F larger than one imposes one linear constraint on the set of initial values and one linear constraint on each of the $k + 1$ columns of G . Recall from Theorem 1 that the lower block of G contains the $r \times (k + 1)$ parameter matrix Π , so that the restriction places $k + 1$ restrictions on Π . Each root larger than one will impose a new set of restrictions. Let the set of eigenvalues with modulus greater than one be denoted $\lambda_1, \lambda_2, \dots, \lambda_j$ and let $R = [c_1, c_2, \dots, c_j]$ be the matrix composed of the corresponding eigenvectors. Assume that R has rank \bar{j} . The stationarity assumption will then place \bar{j} restrictions on the initial values of $X_{1/0}$ and $\bar{j}(k + 1)$ restrictions on the matrix of coefficients Π . When $\bar{j} = r$ there are $r(k + 1)$ restrictions placed on the otherwise arbitrary $r \times (k + 1)$ matrix Π and r restrictions placed on the otherwise arbitrary initial conditions $y_0^e, y_1^e, \dots, y_{r-1}^e$. Thus the solution is unique. [This is a restatement of Proposition 1 in Blanchard and Kahn (1980).]

We will now consider properties of estimates of the unknown parameters formed by maximizing a Gaussian likelihood function under the assumption of stationarity. Some additional notation will facilitate the discussion. Let θ denote the $m \times 1$ vector of unknown parameters and let $v_t = Aa_t$ denote the innovation in Y_t . The variance of v_t is $A(\theta)Q(\theta)A(\theta)'$, which we denote by $h(\theta)$. The moving average representation for Y_t can be written as

$$Y_t = C_0 \bar{a}_t + C_1 \bar{a}_{t-1} + C_2 \bar{a}_{t-2} + \dots,$$

where $\bar{a}_t = h(\theta)^{-1/2} v_t$, $C_0 = h(\theta)^{1/2}$, $C_i = SF^{i-1}GA^{-1}h^{1/2}$ for $i = 1, 2, \dots$. We will make the following assumptions:

- A.1. The data Y_t are stationary and ergodic with a state-space representation given by (3.15)–(3.16).
- A.2. θ_0 is in the interior of W^θ , a compact subset of R^m .
- A.3. If $\theta' \neq \theta_0$, then $C_i(\theta') \neq C_i(\theta_0)$ for some i .
- A.4. The functions $F(\theta)$, $G(\theta)$, $S(\theta)$, $A(\theta)$, and $Q(\theta)$ have continuous first and second derivatives.
- A.5. The eigenvalues of $F(\theta_0)$ are inside the unit circle.
- A.6. The eigenvalues of $[F(\theta_0) - G(\theta_0)A(\theta_0)^{-1}S(\theta_0)]$ are inside the unit circle.

A.7. Let B_t denote the Borel subfield of all events generated by v_j , $j \leq t$, then:

- (i) $E(v_t | B_{t-1}) = 0$ (a.s.),
- (ii) $E(v_t v_t' | B_{t-1}) = h(\theta_0)$ (a.s.),
- (iii) v has finite fourth moments.

Assumptions A.1 and A.2 are standard. Assumption A.3 is an identifiability assumption. Assumption A.4 allows us to use Taylor series expansions of the system matrices. Assumption A.5 implies that the X process is stationary, and assumption A.6 implies that the vector ARMA representation of the Y_t process is invertible. Assumption A.7 allows the use of a martingale central limit theorem.

Using these assumptions we can discuss the properties of $\hat{\theta}_T$, where $\hat{\theta}_T$ solves

$$\max_{\theta} \sum_{t=1}^T \frac{1}{2} \left[\log |h(\theta)| - v_t(\theta)' h(\theta)^{-1} v_t(\theta) \right].$$

These properties are given in:

Theorem 2. Given assumptions A.1–A.7,

- (i) $\hat{\theta}_T \xrightarrow{\text{as}} \theta_0$,
- (ii) $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_\theta)$,

where an expression for Σ_θ is given in the appendix.

Proof. See appendix.

The theorem follows fairly directly from the properties of estimators in vector ARMA models. The only complication is that there are constraints across parameters characterizing the conditional mean and variance of Y_t , which complicates the expression for Σ_θ .

When stationarity is not imposed on the solution, the properties of the estimated parameters are less obvious. While we can say little about these estimators in general, we will offer a suggestive example. The equivalence of the state–space representation with an ARMA representation suggests that, in some situations, the estimated parameters may have distributions similar to those in explosive AR models [e.g., Stigum (1974)]. However, this analogy with explosive AR models may be misleading because of the nonlinear restrictions across the parameters in the ARMA representation. The following example demonstrates the complications that can arise because of these nonlinear

restrictions. In this example the data are stationary, but the model is estimated allowing for the possibility of a ‘bubble’ component. This results in an inconsistency among the estimated parameters.

Consider

$$y_{t+1/t} = \mu y_t + 1 + e_t, \quad (4.1)$$

where $\mu > 1$ and the stationarity condition is satisfied by the data. Imposing the stationarity assumption yields the solution

$$y_t = (1 - \mu)^{-1} + a_t, \quad (4.2)$$

where $a_t = -\mu^{-1}e_t$. The ‘constrained’ estimate of μ is $\hat{\mu}_c = 1 - (\bar{y})^{-1}$, where \bar{y} is the sample mean of the y_t ’s. Under very general assumptions about the disturbance terms, this estimate of μ will be consistent and asymptotically normal.

The solution given in (4.2) imposes the constraint that $y_{1/0} = (1 - \mu)^{-1}$. Suppose this constraint was not imposed; we then have

$$y_t = (1 - \mu)^{-1} + \gamma \mu^t + a_t, \quad (4.3)$$

where $\gamma = y_{1/0} - (1 - \mu)^{-1}$, so that γ_0 , the true value of γ , is zero. Let $\alpha = (1 - \mu)^{-1}$ and consider the least squares estimates of γ and α . The first-order conditions are

$$\sum \hat{a}_t \hat{\mu}^t = 0, \quad (4.4)$$

$$\sum \hat{a}_t [1 + \hat{\gamma} t \hat{\mu}^{t-1} \hat{\alpha}^{-2}] = 0, \quad (4.5)$$

where $\hat{\alpha} = (1 - \hat{\mu})^{-1}$, $\hat{a}_t = y_t - \hat{\alpha} - \hat{\gamma} \hat{\mu}^t$, and the summations run from $t = 1$ to T . From (4.4),

$$\hat{\gamma} = \left[\sum (y_t - \hat{\alpha}) \hat{\mu}^t \right] \left[\sum \hat{\mu}^{2t} \right]^{-1},$$

so that $\hat{\mu}^T \hat{\gamma} \sim O_p(1)$. Rearranging (4.5) and using $y_t = \alpha + a_t$ yields

$$\begin{aligned} \hat{\alpha} - \alpha &= T^{-1} \sum a_t - (\hat{\gamma} \hat{\mu}^T) (T^{-1} \sum \hat{\mu}^{-T+t}) \\ &\quad - \hat{\mu}^{-1} \hat{\alpha}^{-2} (\hat{\gamma} \hat{\mu}^T) \left[T^{-1} \sum \hat{a}_t t \hat{\mu}^{-T+t} \right] \\ &= d1_T + d2_T + d3_T. \end{aligned}$$

The first term, $d1_T$, represents the difference between $\hat{\alpha}$ and α when the stationarity constraint is imposed. The second term will vanish in probability when $\hat{\mu} > 1$. The final term, $d3_T$, will not vanish in probability; each of the terms $\hat{\mu}^{-1}\hat{\alpha}^{-2}$, $(\hat{\gamma}\hat{\mu}^T)$, $[T^{-1}\sum\hat{\alpha}_t\hat{\mu}^{-T+t}]$ is $O_p(1)$ but none are $o_p(1)$. Thus the discrepancy $\hat{\alpha} - \alpha$ will be bounded in probability but will not converge to zero. The cause of the inconsistency in $\hat{\alpha}$ can be seen by rewriting the first-order condition (4.5) as

$$\sum \hat{\alpha}_t [1 + c_t \hat{\mu}^{t-T}] = 0, \quad (4.6)$$

where $c_t = \hat{\alpha}^{-2}\hat{\gamma}\hat{\mu}^{T-1} \sim O_p(1)$. Eq. (4.6) shows that $\hat{\alpha}$ is found by setting a weighted sum of the residuals equal to zero. Since the weights are $O_p(t\hat{\mu}^{t-T})$, nearly all of the weight is placed on the last few observations; in essence, $\hat{\alpha}$ is determined by a small sample, even asymptotically. Note, however, that even though α and thus μ are not estimated consistently, $\hat{\gamma}$ can be expected to be very close to zero when $\hat{\mu} > 1$, since $\hat{\mu}^T\hat{\gamma}$ is bounded in probability.

5. Stock prices and dividends

One interpretation of the model presented in section 2 is an arbitrage relationship between the price of a portfolio of stocks and the dividends paid on the portfolio. Letting p_t denote the portfolio price, d_t denote the value of dividend accruing to portfolio, and μ denote the (constant) real gross rate of interest, this arbitrage relationship can be written as

$$E(p_{t+1}|\Omega_t) = \mu(p_t - d_t). \quad (5.1)$$

This follows from the assumption of risk neutrality on the part of investors who share a common set of information Ω_t . One solution to eq. (5.1) is the 'forward looking' or 'fundamental' solution given by

$$p_t = \sum_{i=0}^{\infty} \mu^{-i} E(d_{t+i}|\Omega_t). \quad (5.2)$$

This expresses p_t as the discounted present value of expected future dividends.⁵ In this section we carry out a test of (5.1) and the particular solution given in (5.2) using the methods developed in the previous sections.

⁵The relationship given in (5.2) has been the subject of a large number of empirical tests. The recent literature includes Shiller (1981), Blanchard and Watson (1982), Engle and Watson (1985), Campbell and Shiller (1986), and West (1986).

To begin, an equation of motion for d_t is necessary. We assume that d_t is generated by

$$d_{t+1} = x_t' \beta + \xi_{t+1}, \tag{5.3}$$

where x_t is a set of variables known to the investor at time t and ξ_{t+1} is an innovation which is uncorrelated with all information in Ω_t . Our empirical analysis will use data on price and dividends only, so that the implications of (5.1)–(5.3) for the joint process describing p_t and d_t must be derived. We will assume that dividends follow an integrated process and that Δd_t is covariance-stationary. As Engle and Watson (1985) and Campbell and Shiller (1986) point out, the relationship given in (5.2) implies that p_t is also integrated and that p_t and d_t are cointegrated. The series $p_t - \zeta d_t$, with $\zeta = \mu(\mu - 1)^{-1}$, is covariance-stationary.

To derive the bivariate process describing p_t and d_t , project x_t onto current and lagged values of d_t and p_t which yields

$$\begin{aligned} \phi(B) \Delta d_{t+1} &= \gamma(B) \Delta p_t + \alpha_1 (p_t - \zeta d_t) \\ &+ \omega(B) \varepsilon_{t+1}^d + \rho(B) u_t, \end{aligned} \tag{5.4}$$

where $\phi(B)$, $\gamma(B)$, $\omega(B)$, and $\rho(B)$ are one-sided polynomials in the back-shift operator B . The error ε_{t+1}^d is the innovation in d_{t+1} that arises from an information set consisting of past dividends and prices. The scale of ε_{t+1}^d is identified using the normalization $\omega_0 = 1$. Following the notation introduced in the previous sections we let ε_t^p denote the innovation in p_t formed from this restricted information set. Projecting ε_t^p onto ε_t^d yields

$$\varepsilon_t^p = \pi \varepsilon_t^d + u_t,$$

where u_t and ε_t^d are uncorrelated white noise processes. The error correction term $(p_t - \zeta d_t)$ may appear in (5.4) because of the cointegration implied by (5.2) [see Engle and Granger (1987)].

In our empirical model we assume that $\phi(B)$ is a polynomial of order two, $\omega(B)$ is a polynomial of order one, $\gamma(B) = \gamma$, and $\rho(B) = \rho$. This implies

$$\begin{aligned} d_{t+1/t} &= (1 + \phi_1 - \alpha_1 \zeta) d_{t/t-1} + (\phi_2 - \phi_1) d_{t-1} - \phi_2 d_{t-2} \\ &+ (\gamma + \alpha_1) p_{t/t-1} - \gamma p_{t-1} + \lambda_1 \varepsilon_t^d + \lambda_2 u_t, \end{aligned} \tag{5.5}$$

where

$$\lambda_1 = -\omega_1 + (1 + \phi_1 - \alpha_1 \zeta) + \pi(\gamma + \alpha_1),$$

$$\lambda_2 = \rho + \gamma + \alpha_1.$$

The model can then be written in state-space form given in (3.15) and (3.16) with

$$Y_t = \begin{bmatrix} d_t \\ p_t \end{bmatrix}, \quad a_t = \begin{bmatrix} \varepsilon_t^d \\ u_t \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 \\ \pi & 1 \end{bmatrix},$$

$$X_{t+1/t} = \begin{bmatrix} d_{t+1/t} \\ d_t \\ d_{t-1} \\ p_{t+1/t} \\ p_t \end{bmatrix}, \quad G = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 0 \\ 0 & 0 \\ \mu(\pi - 1) & \mu \\ \pi & 1 \end{bmatrix},$$

$$F = \begin{bmatrix} 1 + \phi_1 - \alpha_1 \zeta_1 & \phi_2 - \phi_1 & -\phi_2 & \gamma + \alpha_1 & -\gamma \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -\mu & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The parameter μ is the gross rate of interest so that $\mu > 1$, and for reasonable values of the parameters ϕ_1 , ϕ_2 , α_1 , ζ , and γ , the matrix F will have one eigenvalue with modulus greater than 1. The forward looking solution given in (5.2) is chosen by annihilating the linear combination of the elements of the state vector associated with this explosive root. This imposes one constraint on the vector of initial conditions $X_{1/0}$ and one constraint on each column of G .

The data that are used in the analysis are the Dow Jones annual data used in Shiller (1981). The sample period is 1928–1978. We let 1930 correspond to $t = 1$ and use the data from 1928 and 1929 to serve as initial conditions d_0 , d_{-1} , and p_{-1} . The initial values of $d_{1/0}$ and $p_{1/0}$ were estimated imposing the constraint mentioned in the last paragraph.

Results are given in table 1. Choosing the forward looking stable solution makes the parameters λ_1 and λ_2 implicit functions of π and μ because of the constraints imposed on the columns of G . In addition, the initial condition $p_{1/0}$ is an implicit function of $d_{1/0}$, d_0 , d_{-1} , and p_0 , because of the constraint imposed on $X_{1/0}$. The two coefficients γ and α_1 were very small and insignificant in every model estimated. The results shown in table 1 constrain these coefficients to equal zero. In the bottom panel of the table we show the autocorrelations of the fitted innovations from the model. These should be vector white noise, and the results suggest that this is not too far from the truth. There are two slightly troublesome values: $\text{cor}(p_t p_{t-2}) = -0.29$ and $\text{cor}(p_t d_{t-1}) = 0.27$, but the sample size is only 49. Unlike the results for a

Table 1
Results for the restricted model.

Parameter	Estimate	Standard error		
ϕ_1	1.082	0.141		
ϕ_2	-0.299	0.113		
μ	1.042	0.002		
π	15.120	2.105		
σ_{ed}	0.012	0.002		
σ_u	0.613	0.053		
$d_{1/0}$	0.109	0.002		
$p_{1/0}$	2.330	—		
λ_1	1.190	—		
λ_2	0.010	—		
Log likelihood value = 189.78				
Autocorrelations of innovations				
	j			
Series	1	2	3	4
$\text{cor}(p_t, p_{t-j})$	-0.11	-0.29	0.12	0.25
$\text{cor}(d_t, d_{t-j})$	0.16	-0.11	0.10	0.19
$\text{cor}(p_t, d_{t-j})$	0.27	-0.16	-0.07	0.02
$\text{cor}(d_t, p_{t-j})$	-0.13	0.10	0.12	0.21

similar model estimated in Engle and Watson (1985), the innovations in this data set show no evidence of heteroskedasticity.

A variety of other tests of the model suggest themselves. First, as pointed out above, eq. (5.2) implies that p_t and d_t are cointegrated, so that the autoregressive representation of $p_t - \zeta d_t$ should not contain a unit root. Letting $\hat{\zeta} = \hat{\mu}(\hat{\mu} - 1)^{-1}$, the first autocorrelation coefficient of $p_t - \hat{\zeta}d_t$ is 0.53, suggesting that no unit root is present.

We have also estimated an unrestricted version of the model. This version replaces eq. (5.1) with

$$\begin{aligned}
 p_{t+1/t} = & p_{t/t-1} + v(d_{t/t-1} - d_{t-1}) + \alpha_2(p_{t/t-1} - \zeta d_{t/t}) \\
 & + \lambda_3 \varepsilon_t^d + \lambda_4 u_t.
 \end{aligned} \tag{5.6}$$

The state-space representation of this unrestricted model is a minor modification of the representation for the restricted model. The state equation describing the evolution of $p_{t+1/t}$ in the restricted model is replaced with eq. (5.6). In this representation of the model, none of the roots of F are outside the unit

Table 2
Results for the unrestricted model.

Parameter	Estimate	Standard error		
ϕ_1	0.543	0.413		
ϕ_2	-0.480	0.131		
ν	-16.880	2.582		
α_2	-0.046	0.251		
π	19.040	2.414		
σ_{ed}	0.010	0.002		
σ_u	0.574	0.064		
$d_{1/0}$	0.115	0.011		
$p_{1/0}$	2.417	0.432		
λ_1	1.187	0.365		
λ_2	0.013	0.003		
λ_3	13.612	10.140		
λ_4	0.953	0.230		
Log likelihood value = 201.57				
Autocorrelations of innovations				
	j			
Series	1	2	3	4
$\text{cor}(p_t, p_{t-j})$	-0.05	-0.05	0.18	0.19
$\text{cor}(d_t, d_{t-j})$	0.02	-0.05	0.00	0.03
$\text{cor}(p_t, d_{t-j})$	0.02	-0.09	0.15	0.09
$\text{cor}(d_t, p_{t-j})$	-0.02	0.08	-0.14	0.11

circle so that no constraints need to be imposed on the initial value of the state nor the columns of G .

The results from this model are shown in table 2. We report results with the parameters γ and α_1 constrained to equal zero. When the model was estimated relaxing this constraint, the estimated values of these parameters were very small in magnitude and statistically insignificant. We have also imposed the constraint that the cointegrating parameter ξ equal its implied model from table 1 [i.e., $\hat{\mu}(\hat{\mu} - 1)^{-1}$]. The results of Stock (1987) imply that this estimator is $O_p(T^{-1})$, and hence is a different order of magnitude than the other estimated parameters.

Comparing the likelihood values of the two models we see that the unrestricted model fits the data much better. It includes seven additional parameters, and the likelihood ratio statistic is 23.58 which can be compared to the 1% critical value of the $\chi^2(7)$ distribution of 18.5. There is reason to doubt the quality of the $\chi^2(7)$ approximation for the distribution of the likelihood ratio statistic. In particular, one of the restrictions imposed under the null is a restriction on the initial values $p_{1/0}$ and $d_{1/0}$ which, as initial conditions for

an integrated process, are not consistently estimated under the null or the alternative. Hence the usual argument used to derive the large sample distribution of the likelihood ratio statistic is not appropriate. Since these initial conditions have an $o_p(1)$ effect on the average log likelihood one might conjecture that the likelihood ratio statistic is better approximated by the $\chi^2(6)$ distribution than the $\chi^2(7)$. This would imply a more significant departure from the null.

The reason for the large likelihood ratio statistic can be seen by substituting estimated coefficients into eq. (5.6). One has

$$p_{t+1/t-1} = 1.05(p_{t/t-1} - d_{t/t-1}) - 14.7d_{t/t-1} + 15.9d_{t-1} \\ + 13.6\epsilon_{t-1}^d + 0.95u_{t-1}. \quad (5.7)$$

On the other hand, eq. (5.1) implies

$$p_{t+1/t-1} = \mu(p_{t/t-1} - d_{t/t-1}). \quad (5.8)$$

Since $\hat{\mu} = 1.04$, the first term on the left-hand sides of (5.7) and (5.8) are essentially the same. However, the unrestricted model includes additional terms. Interestingly, the cause of the rejection of (5.1) and (5.2) does not appear to stem from (5.2), but rather from (5.1). That is, it is not the particular 'forward looking' solution given in (5.2) that is rejected by the data, but rather the arbitrage relation in (5.1).

6. Concluding remarks

In this paper we have presented an alternative solution procedure for dynamic linear rational expectations models. This alternative solution technique is based on the state-space representation of the model. It has several advantages over other methods. First, it is a capital rather than a labor-intensive solution procedure. The algebra underlying the analytic solution to the model doesn't need to be done. The model is solved recursively while the likelihood function is being evaluated, rather than algebraically solving the model and then recursively evaluating the likelihood function. Second, and probably most important, once the model is cast in state-space form, the flexibility of that representation can be exploited to handle problems arising from missing data, temporal aggregation, dynamic errors in variables, or other problem-specific data limitations. Finally, the literature on identification in state-space models [e.g., Glover and Willems (1974)] serves as a useful addition to the literature on identification in linear rational expectations models [e.g., Pesaran (1981) and Blanchard (1982)].

Appendix

Proof of Theorem 1

The theorem follows directly from Anderson and Moore (1979, pp. 77–80) given the following lemma.

Lemma. *The pair (F, S') is completely detectable.*

Proof. Let W be an eigenvector of F corresponding to an eigenvalue λ , with $|\lambda| > 1$. The pair (F, S') is completely detectable if $SW \neq 0$. Since F is lower triangular with diagonal blocks Φ and μ and all of the eigenvalues of Φ are inside the unit circle, W is of the form $W = [0'_{q(k+1)} \ \omega']$, where ω is a eigenvector of μ . Since S is a selection matrix which selects the first k elements and the $(q(k+1) + r)$ th element from an arbitrary $q(k+1) + r + p$ vector, $SW = \omega_r$, where ω_r is the r th element of ω . The form of μ implies that the elements of ω satisfy

$$\sum_{i=1}^{p+r} \mu_i \omega_i = \lambda \omega_1,$$

$$\omega_1 = \lambda \omega_2,$$

$$\omega_2 = \lambda \omega_3,$$

$$\vdots$$

$$\omega_{p+r-1} = \lambda \omega_p.$$

This implies that $SW = \omega_r = 0$ if and only if $\omega_i = 0$ for all i . But this cannot be true since ω is an eigenvector. Thus proving the lemma.

The remainder of the proof parallels Anderson and Moore (1979, pp. 77–80) and is omitted.

Proof of Theorem 2

The assumptions of the theorem imply that the model is a stationary and invertible finite-order vector ARMA model [see Akaike (1974)]. The strong consistency of the estimators follows directly from Dunsmuir and Hannan (1976). The model does not satisfy the conditions of their theorem for asymptotic normality because both the conditional mean of Y_t and its variance depend on the same set of parameters. This merely complicates the

expression for the asymptotic covariance matrix. A sketch of the proof of asymptotic normality follows.

The objective function is

$$L(\theta) = \sum_{t=1}^T -\frac{1}{2} [\log |h(\theta)| - v_t' h(\theta)^{-1} v_t],$$

where $v_t = Y_t - S(\theta)X_{t/t-1}$. The i th element of the score vector is

$$\partial L(\theta_0)/\partial \theta_i = \sum_{t=1}^T (L1_{i,t} + L2_{i,t}),$$

where

$$L1_{i,t} = -\frac{1}{2} TR \left\{ \left[h(\theta_0)^{-1} \partial h(\theta_0)/\partial \theta_i \right] \left[I - h(\theta_0)^{-1} v_t v_t' \right] \right\}$$

and

$$L2_{i,t} = -(\partial v_t/\partial \theta_i)' h(\theta_0)^{-1} v_t.$$

Note that $L1_{i,t}$ is a stationary and ergodic martingale difference with finite variance; it follows from Billingsley (1961) that

$$T^{-1/2} \sum_{t=1}^T L1_t \xrightarrow{\mathcal{L}} N(0, D),$$

with $d_{ij} = \text{cov}(L1_{i,t}, L1_{j,t})$.

The $L2_{i,t}$ term is only slightly more trouble. It is straightforward to show that $\partial v_t/\partial \theta_i$ can be written as

$$\partial v_t/\partial \theta_i = \sum_{k=0}^{t-2} M_{i,k} v_{t-1-k} + N_{i,t} X_1,$$

where

$$M_{i,k} = -S_i F^k V - S A_1^k V_i - S \sum_{j=0}^{k-1} A_1^j A_2 F^{k-j-1} V$$

and

$$N_{i,t} = -S_i F^{t-1} - S \sum_{k=0}^{t-2} A_1^k A_2 F^{t-2-k},$$

with

$$S_i = \partial S(\theta_0)/\partial \theta_i, \quad V = G(\theta_0)A(\theta_0)^{-1}, \quad V_i = \partial V(\theta_0)/\partial \theta_i,$$

$$A_1 = F - VS, \quad A_2 = F_i - VS_i \quad \text{with} \quad F_i = \partial F(\theta_0)/\partial \theta_i,$$

and the final term in $M_{i,k}$ is zero for $k=0$. Since both F and A_1 have eigenvalues which all are less than one in modulus, $N_{i,t}$ and $M_{i,k}$ converge to zero at an exponential rate as t and $k \rightarrow \infty$. This implies that

$$T^{-1/2} \sum L2_t - T^{-1/2} \sum \tilde{L}2_t \xrightarrow{p} 0,$$

where

$$\tilde{L}2_t = \left(\sum_{k=0}^{\infty} M_{i,k} v_{t-1-k} \right)' h(\theta_0) v_t$$

is a stationary and ergodic martingale difference sequence with finite variance. This implies

$$T^{-1/2} \sum L2_t \xrightarrow{\mathcal{L}} N(0, W),$$

with $w_{ij} = \text{cov}(\tilde{L}2_{i,t}, \tilde{L}2_{j,t})$. If we let v_{ij} denote the covariance between $L1_{i,t}$ and $\tilde{L}2_{j,t}$, then $T^{-1/2} \partial L(\theta_0) / \partial \theta$ converges to a normal random vector with mean zero and covariance matrix U , with $u_{ij} = d_{ij} + w_{ij} + 2v_{ij}$.

The usual Taylor series argument expanding $T^{-1/2} \partial L(\hat{\theta}_T) / \partial \theta$ about $T^{-1/2} \partial L(\theta_0) / \partial \theta$ completes the proof, yielding

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_\theta),$$

with $\Sigma_\theta = J U J'$, with $J^{-1} = \text{plim}(T^{-1} \partial^2 L(\theta_0) / \partial \theta \partial \theta')$.

References

- Akaike, H., 1974, Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, *Annals of the Institute of Statistical Mathematics* 20, 363–388.
- Anderson, B.D.O. and J.B. Moore, 1979, *Optimal filtering* (Prentice-Hall, Englewood Cliffs, NJ).
- Billingsley, P., 1961, The Lindeberg–Levy theorem for martingales, *Journal of the American Mathematical Society* 12, 788–792.
- Blanchard, O.J., 1979, Backward and forward solutions for economies with rational expectations, *American Economic Review* 69, 114–118.
- Blanchard, O.J., 1982, Identification in dynamic linear models with rational expectations, HIER discussion paper no. 888 (Harvard University, Cambridge, MA).
- Blanchard, O.J., 1983, The production and inventory behavior of the American automobile industry, *Journal of Political Economy* 91, 365–400.
- Blanchard, O.J. and C.M. Kahn, 1980, Solution of linear difference models under rational expectations, *Econometrica* 38, 1305–1311.
- Blanchard, O.J. and M.W. Watson, 1982, Bubbles, rational expectations, and financial markets, in: Paul Wachtel, ed., *Crisis in the economic and financial structure: Bubbles, bursts and shocks* (Lexington Books, Lexington, MA).

- Broze, L., C. Gourieroux, and A. Szafarz, 1985, Solutions of linear rational expectations models, *Econometric Theory* 1, 341–368.
- Campbell, J. and R.J. Shiller, 1986, Cointegration and tests of present value models, NBER working paper no. 1885.
- Chow, G.C., 1983, *Econometrics* (McGraw-Hill, New York, NY).
- Diba, B. and H. Grossman, 1983, Rational asset price bubbles, NBER working paper no. 1059.
- Dunsmuir, W. and E.J. Hannan, 1976, Vector linear time series models, *Advances in Applied Probability* 8, 339–364.
- Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation, and testing, *Econometrica* 55, 251–276.
- Engle, R.F. and M.W. Watson, 1985, The Kalman filter: Applications to forecasting and rational expectations models, in: T. Bewley, ed., *Advances in econometrics, Fifth world congress* (Cambridge University Press, Cambridge).
- Flood, R.P. and P.M. Garber, 1980, Market fundamentals versus price-level bubbles: The first tests, *Journal of Political Economy* 88, 745–780.
- Glover, K. and J.C. Willmoms, 1974, Paramaterizations of linear dynamical systems: Canonical forms and identifiability, *IEEE Transactions on Automatic Control* AC-19, 640–645.
- Gourieroux, C., J.J. Laffont, and A. Monfort, 1982, Rational expectations in linear models: Analysis of solutions, *Econometrica* 50, 409–425.
- Hansen, L.P. and T.J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control* 2, 7–46.
- Hansen, L.P. and T.J. Sargent, 1982, Instrumental variables procedures for estimating linear rational expectations models, *Journal of Monetary Economics* 9, 263–296.
- Harvey, A.C., 1981, *Time series models* (Halstead Press, New York, NY).
- Harvey, A.C. and R.G. Pierse, 1984, Estimating missing observations in economic time series, *Journal of the American Statistical Association* 79, 125–131.
- Harvey, A.C., C.R. McKenzie, D.P.C. Blake, and M.J. Desai, 1981, Irregular data revisions, in: Arnold Zellner, ed., *Applied time series analysis of economic data*, Economic research report no. ER-5 (U.S. Department of Commerce, Washington, DC).
- Hausman, J.A. and M.W. Watson, 1985, Errors-in-variables and seasonal adjustment procedures, *Journal of the American Statistical Association* 80, 531–540.
- McCallum, B.T., 1976, Rational expectations and the natural rate hypothesis: Some consistent estimates, *Econometrica* 44, 43–52.
- Sargent, T.J., 1979, A note on the maximum likelihood estimation of the rational expectations model of the term structure, *Journal of Monetary Economics* 5, 133–138.
- Pesaran, M.H., 1981, Identification of rational expectations models, *Journal of Econometrics* 16, 375–398.
- Schweppe, F., 1965, Evaluation of likelihood functions for Gaussian signals, *IEEE Transactions on Information Theory* 11, 61–70.
- Shiller, R.J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends, *American Economic Review* 71, 421–436.
- Stigum, B.P., 1974, Asymptotic properties of dynamic stochastic parameter estimates III, *Journal of Multivariate Analysis* 4, 47–88.
- Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* 55, 1035–1056.
- West, K.D., 1986, *Speculative bubbles and stock price volatility*, Mimeo. (Princeton University, Princeton, NJ).
- Whiteman, C.H., 1983, *Linear rational expectations models* (University of Minnesota Press, Minneapolis, MN).