CHAPTER 7

The Kalman filter: applications to forecasting and rational-expectations models

Robert F. Engle and Mark W. Watson

1 Introduction

Economics is not engineering; yet, perhaps, we can track the economy using the same tools used to track a spacecraft, an oil tanker, or a chemical reaction. In the 25 years since the publication of the original Kalman (1960) and Kalman and Bucy (1961) papers that introduced digital filters for nonstationary problems, economists have been studying these possibilities, and the presence of the August 1985 session of the World Congress of the Econometric Society suggests that it is still a question of great interest.

The initial attempts to apply these methods to economic problems immediately faced a major difficulty. Engineers usually had quantitative theories that described the equations of motion of physical systems and were primarily interested in estimates of the "state" of the system obtained from noisy measurements. The extraction of estimates of such signals from noise was called the estimation, or "state estimation," problem. Economists, however, knew far less about the fundamental laws of motion of economic systems and were therefore particularly interested in discovering such laws of motion from the noisy data rather than in merely estimating the state of the economy. Since the Kalman filter takes the parameters of the process as given in estimating the state, it appeared that there would be little possibility to apply such methods in economics.

However in the mid-1970s a number of economists were able to change the focus of the estimation problem and, applying Schweppe's (1965) result, used the Kalman filter as a computational tool to evaluate the likelihood function in complex cases. Therefore, the parameters of the process

This chapter was prepared for the World Congress of the Econometric Society, Cambridge, Massachusetts, August 1985. The authors thank the National Science Foundation for financial support.

could be estimated using maximum-likelihood methods, as in more standard econometrics. As a by-product, estimates of the state conditional on these fixed parameters were obtained, and sometimes these too were of interest.

The new interpretation of the state space framework allowed the unobserved states to have a wide variety of interpretations, and the model became the natural generalization of latent variable models to a full dynamic framework. In the state space model, it was possible to specify far more complex dynamic error structures. In many cases, these would be suggested by the economics of the situation where a disturbance would have an impact on different dependent variables in different time periods. The structure unified estimation of regressions with ARMA errors, time-varying parameters (TVPs), seasonal adjustment, dynamic factor analysis, and dynamic multiple-indicator, multiple-cause models.

A wide variety of potential applications became possible. A short taxonomy as in Engle and Watson (1980) will prove useful in discussing the successes and failures of these models:

1. Univariate: one dependent variable.
   - (a) Time-varying parameter models: One or more regression coefficients from a linear model are allowed to evolve stochastically, possibly with causal variables.
   - (b) Unobserved component models: The same as (a), except that the coefficients that vary are coefficients of the intercept. By specifying different structures, more than one component can be extracted. Both seasonality and trends can be analyzed in a regression context.
   - (c) UCARIMA models: The same as (b), except that the components are each ARIMA models. This is useful for seasonal adjustment and stochastic detrending.
2. Multivariate: more than one dependent variable.
   - (a) Varying coefficient regressions: One or more of the coefficients of a set of regression equations are allowed to vary.
   - (b) Unobserved component models: The varying coefficients are again on the intercept, but the important features are that the same components may appear in several equations.
   - (c) Dynamic factor analysis: The covariance matrix of a set of dynamic regressions has a factor structure that may have the latent variables entering different equations with different lags – a special case of (b).

(d) DYMIMIC: The latent variable in a set of dynamic regressions influences several dependent variables and is caused by several independent variables – a special case of (b).

This outline provides a framework for listing many of the successful empirical applications of the Kalman filter in economics. Most of the references on this list are of the unobserved components type. UCARIMA models have been successfully used for signal extraction by Pagan (1975), for forecasting by Harvey and Todd (1983) and Harvey (1984), and for seasonal adjustment by Nerlove, Grether, and Carvallo (1979), Engle (1979), and Hausman and Watson (1985). Howrey (1978) and Conrad and Corrado (1979) use the unobserved components model to estimate data revisions and to combine forecasts, whereas Harvey and Phillips (1979) point out that it can be used to estimate regression models with ARMA errors. Dynamic factor analysis was utilized by Engle and Watson (1981). Watson and Kraft (1984), Hamilton (1985), and Engle, Lilien, and Watson (1985) used the DYMIMIC model.

Although time-varying parameters have been used in many studies, these typically are interpreted as tests for the stability of a regression equation [see, e.g., Garbade (1977); Laumas and Mehra (1976); Rauser and Laumas (1976)]. Much less common are TVP models, either univariate or multivariate, where the variation is interpreted as economically important. One case is in finance, where the beta of the market equation is allowed to vary [see Ohlson and Rosenberg (1982); Alexander, Benson, and Eger (1982); Bos and Newbold (1984)]. Other recent exceptions are Doan, Litterman, and Sims (1984) in the multivariate context and EPRI (1983), which will be discussed in more detail later in this chapter.

This chapter will present two serious applications of the Kalman filter. These applications were selected to illustrate the power of the state space formulation in solving a wide variety of problems. The first application uses the TVP model to forecast and "weather normalize" electricity sales, whereas the second uses the multivariate unobserved components model to build a rational-expectations model of the relation between dividends and equity prices.

2 The time-varying parameter model

The varying coefficient regression model was introduced into economics by Cooley and Prescott (1973a, b, 1976) and Rosenberg (1972, 1973). Recent surveys by Pagan (1980), Nichols and Pagan (1983), Chow (1983a), Swamy and Tinsley (1980), and Beck (1983) have extended the range of models and results available. Harvey (1981a, b) provides a textbook version for econometricians. Los and Kell (1985) present some Monte Carlo evidence on the behavior of the methods. In each of these models, one or more of the parameters is allowed to evolve over time, and estimation techniques are explored for both the fixed parameters and the varying parameters. In these papers, a variety of large-sample estimation results, identification procedures, and diagnostic tests are presented and discussed. However, as will be mentioned below, the most important cases are yet to be given a satisfactory treatment.

To consider the simplest possible case, let y_t and x_t be scalar observable data series and let β_t and ϵ_t be scalar unobservable series with properties to be defined. Suppose z_t to be a vector of observed variables with γ the regression coefficients. Let

$$y_t = x_t \beta_t + z_t \gamma + \epsilon_t \tag{2.1}$$

define the relation between the series, and suppose that

$$y_t \mid \{x_t, z_t, \beta_t, \text{ the past of } y, x, z\} \sim N(x_t \beta_t + z_t \gamma, \sigma^2)\} \tag{2.2}$$

with x_t and z_t as weakly exogenous for the parameters of interest. The model is clearly unidentified as β_t is not observable. In this setup, β_t is the state to be estimated, and the measurements that are to be used to estimate this state are the y_t. Equation (2.1), or its more statistical counterpart (2.2), is therefore defined as the measurement equation since it defines how the measurements are derived from the unobserved states.

To complete the model, a generating equation for the state β_t is needed. This is called the transition equation since it explains how the state evolves. A parameter is inherently constant, at least if it is a parameter of taste or technology and therefore is invariant to structural change, as in Engle, Hendry, and Richard (1983). Any model of varying coefficients requires explanation. If the coefficient varies, then it must also follow some model, and one might hope that the parameters of this model would be constant.

A simple example of a transition equation is

$$\beta_t = \phi \beta_{t-1} + z_t \theta + \eta_t \quad \eta_t \sim \text{IN}(0, \lambda^{-1}\sigma^2) \tag{2.3}$$

with higher-order processes as direct generalizations. Our empirical experience and the theoretical arguments in this and the following section suggest that, in most cases, the model for β ought to have a unit root. For many data sets, the simple random walk with $\phi = 1$ and $\theta = 0$ performs well:

$$\beta_t = \beta_{t-1} + \eta_t \quad \eta_t \sim \text{IN}(0, \lambda^{-1}\sigma^2) \tag{2.4}$$

To develop theoretical models for the transition equations, three types of economic models will be suggested: behavioral, unobserved causes, and misspecification. A famous ***behavioral*** motivation for parameter variation is the Lucas (1976) critique of economic policy analysis. If the policy regime is changed, then agents will adjust their behavior and the coefficients that held in the previous regime will change to new values. This observation could be built into a structural model of changing coefficients. Of course, such a theory must include a formulation of how agents learn about the new model, and the transition equation could represent such a learning process. Plausibly, one would change the estimate of the state only when new information becomes available, thus suggesting the unit root. This model might be sensible for a variety of processes whereby policy regimes are shifted, although suboptimal for known types of changes.

Many other simpler theories might suggest changes in parameters following particular processes. The rational-expectations example presented later in this chapter is a prime example. In that case, the states include expectations of future prices and dividends, and the rational-expectations hypothesis gives specific content to the relationships. A similar strategy is followed by Burmeister and Wall (1982) and Hamilton (1985). Another example of interest is in Engle, Lilien, and Watson (1985), where the unobserved state is the rate at which housing rents are capitalized into asset prices. Here the capitalization rate can be calculated in steady state where inflation rates, interest rates, and tax rates are constant, but it is not clear how the market evaluates new information and changes in these variables. Thus, the steady state provides constraints on the form of the transition equation in the long run whereas the data are allowed to determine the short-run behavior.

The *unobserved causes* model recognizes that if we know why a parameter is changing, we can usually rewrite the model with constant parameters. For example, suppose $\beta_t = w_t\theta$; then (2.1) becomes

$$y_t = x_t w_t \theta + z_t \gamma + \epsilon_t$$

and least squares can be used directly. However, if w is unobserved, then it can be replaced only by its time series representation. Thus, if $\phi(B)w_t = \eta_t$, then the transition equation becomes

$$\phi(B)\beta_t = \eta_t \theta \tag{2.5}$$

and the parameters of the polynomial $\phi(B)$ can be interpreted in terms of the process of the unobserved w. For most economic variables, therefore, this process should be nonstationary and slowly evolving. Often, we know that it should be very smooth and have a unit root. Sometimes, we might know a series that causes w_t, and this too can be used to tighten the specification. The electricity forecasting example in this chapter is based upon the unobserved causes model where appliance saturations (particularly measured in efficiency units) are not observed.

The third model that appears to produce time-varying parameters is the omnipresent *misspecification*. The misspecification of a regression relation will generally lead to nonwhite residuals that may be partly "explained" by allowing some of the parameters of the model to be time varying. It is in this sense that the model has been suggested as a test procedure, particularly through the analysis of *recursive residuals* (see, e.g., Brown, Durbin, and Evans 1975; Harvey and Collier 1977). The peril of the time-varying parameter model is in the interpretation of the result, if indeed misspecification is the source of the variation. It remains an open question whether allowing for time variation in parameters of a misspecified model would improve forecasts of the model. One might suspect that in the majority of cases this would be true, but to our knowledge, this has not been investigated.

The interest in stochastic detrending of economic series in Harvey and Todd (1983), Watson (1985b), and Engle, Brown, and Stern (1985) can be viewed as correcting for misspecified trend components in a regression. The assumption of a unit root in the transition equation effectively distinguishes the trend from the cyclical and noise components.

In each of these cases, there is a suggestion that the transition equation ought to have a unit root. Further justification will arise from the discussion of the spline model in the next section. A consequence of this assumption, however, is that the asymptotic theory is not yet fully developed. It is not clear that maximum likelihood will have its usual properties in the presence of unit roots, even for the remaining parameters. In fact, in a recent paper on splines, Rice (1986) shows that when part of a model is parametric and the rest is nonparametric and fit by smoothing splines, the rate of convergence of the parametric part may be slower than for a pure parametric problem. This is an important area for further research.

3 The three statistical models: MLE, Bayes, spline

In this section, three statistical models will be developed that support estimation of the TVP model: the classical model, the Bayesian model, and the smoothing spline model. The key differences hinge on the way to control the degree of parameter variability allowed and the model used for the parameter process. Short of these differences, the estimates obtained for each method will be identical. The connections between these methods have been known for many years (see, e.g., Kimeldorf and Wahba 1970; Wahba and Wold 1975) and more recently exploited by Wecker and

Ansley (1983) and Ansley and Wecker (1983) [with discussion by Gersch (1983) and Dempster and Jonas (1983)] and by Wahba (1983). Most of this literature adapts classical approaches to the smoothing spline setup. Thus far, there seems little return fertilization. The possibilities for such a development will be discussed later. These depend somewhat on the particular setup.

The first statistical model, which is the one most often associated with TVP, is classical time series analysis leading to a maximum-likelihood estimate of λ. The log-likelihood function can be written as the sum of the log likelihoods of the conditional densities of y, which in turn are simply the densities of the innovations ν_t. Since these are linear in y, each contribution to the log likelihood is simply a Gaussian density:

$$L = -\sum (\log h_t + \nu_t^2/h_t) \tag{3.1}$$

For this to properly be interpreted as the log likelihood, the entire model, including the distributional assumptions, must be specified correctly. In particular, the transition equation is viewed as a true data generation equation. [It is worth pointing out, however, that some optimal properties of the Kalman filter can be developed even without the normality assumption. See, e.g., Anderson and Moore (1979).] This likelihood function can be maximized over the three parameters: γ, σ^2, and λ. The maximum-likelihood estimate of λ then regulates the trade-off between fitting y and having a constant β.

The second statistical model is the Bayesian model, where the transition equation for β_t is treated as a prior. That is, the prior is assumed to have the same distribution as a random variable following such a process. For the random-walk model in (2.4), the prior can be written as

$$\beta = R^{-1}\eta \tag{3.2}$$

with

$$\beta = (\beta_1, \beta_2, \ldots, \beta_T)'$$

$$\eta = (\eta_1, \eta_2, \ldots, \eta_T)'$$

$$R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

so that

$$\beta|\lambda, \sigma^2 \sim N(0, \lambda^{-1}\sigma^2 R^{-1}R'^{-1}) \tag{3.3}$$

Writing y and z as the $T\times 1$ and $T\times K$ vectors of dependent and independent variables and X and the $T\times T$ diagonal matrix with tth diagonal element x_t, the log of the density for y, conditional on β, can be written as

$$\log F(y\mid\beta) = -T/2\log\sigma^2 - (y-X\beta-z\gamma)'(y-X\beta-z\gamma)/2\sigma^2 \tag{3.4}$$

so that the log joint density of y and β is

$$\begin{aligned} L^* = {} & -T\log\sigma^2 + T/2\log\lambda \\ & -[(y-X\beta-z\gamma)'(y-X\beta-z\gamma)+\lambda\beta'R'R\beta]/2\sigma^2 \end{aligned} \tag{3.5}$$

and the log of the marginal density of y is

$$L' = -T/2\log\sigma^2 - 1/2\log|\Gamma| - (y-z\gamma)'\Gamma^{-1}(y-z\gamma)/2 \tag{3.6}$$

with

$$\Gamma = (I+\lambda^{-1}XR^{-1}R'^{-1}X')$$

From this observation, it is clear that if (3.2) is taken to be the true generating equation for β, then (3.6) is the log-likelihood function, so that $L=L'$.

Since the joint distribution is normal, the mode of the joint will give the Bayesian estimate of β for a wide variety of loss functions. In the case where $\gamma=0$, this is simply

$$\hat\beta = (X'X+\lambda R'R)^{-1}X'y = E(\beta\mid y) \tag{3.7}$$

the familiar Bayesian formula for the conjugate normal problem. It can be expressed as the expected value of β given the full data set and the prior and is therefore comparable with the smoothed estimate of β from the maximum-likelihood estimate (MLE). In fact, this is an alternative expression for the smoothed estimates.

In the Bayesian case, the selection of λ is typically viewed as a prior value for the parameter process and is not estimated from the data. As the sample grows, the size of $X'X$ as well as $R'R$ will grow. Multiplying both sides of (3.7) by $(X'X+\lambda R'R)$ and recognizing that X is diagonal and R has at most one off-diagonal element,

$$\begin{bmatrix} 2\lambda+x_1^2 & -\lambda & 0 & \cdots & 0 & 0 \\ -\lambda & 2\lambda+x_2^2 & -\lambda & & 0 & 0 \\ 0 & & \vdots & & \vdots & \vdots \\ \vdots & & -\lambda & & 2\lambda+x_{T-1}^2 & -\lambda \\ 0 & \cdots & 0 & \cdots & -\lambda & \lambda+x_T^2 \end{bmatrix}\hat\beta = X'y$$

an expression for $\hat\beta$ can be found:

$$-\lambda\hat\beta_{t-1} + (x_t^2+2\lambda)\hat\beta_t - \lambda\hat\beta_{t+1} = x_t y_t \tag{3.8}$$

This shows that the importance of λ and R does not vanish as the sample size grows. A small value of λ produces $\hat{\beta}_t = y_t/x_t$ and therefore fits each point exactly. A large value of λ constrains successive triples of coefficients to lie along a straight line with the boundary condition $\hat{\beta}_{T-1} = \hat{\beta}_T$. Thus, the parameter is constant.

When $\gamma \neq 0$, a similar expression can be written. The parameters β and γ are estimated by

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'X + \lambda R'R & X'z \\ z'X & z'z \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ z'y \end{bmatrix}$$

or

$$\hat{\beta} = (\hat{X}'\hat{X} + \lambda R'R)^{-1}\hat{X}'y \quad \hat{X} = [I - z(z'z)^{-1}z']X \tag{3.9}$$

The Bayesian who does not wish to impose a value for λ can instead formulate a prior on λ and let the data select the particular value that maximizes the posterior. If this prior is uniform, it would coincide with the MLE as usual.

The third model is the discrete-time version of smoothing splines as developed in the statistical literature for curve fitting. See, for example, Wahba and Wold (1975), Craven and Wahba (1979), and Rice and Rosenblatt (1983), and, in economics, Shiller (1984), Wecker and Ansley (1983), Ansley and Wecker (1983), and Engle et al. (1985). In this case, there is assumed to be some fixed, but unknown, set of β_t that are taken to be smooth in some sense to be specified. The estimation problem is to achieve a compromise between fidelity to the data (a good fit) and fidelity to the smoothness criterion (a smooth path for the β_t). Letting $e = y - X\beta - z\gamma$, fidelity to the data is defined as minimizing $e'e$. Smoothness criteria are often taken to be the integral of the squared mth derivative from zero to T. This can be approximated by the squared mth differences in discrete time with equally spaced intervals. For the time-varying parameter problem, the appropriate notion of smoothness is an unchanging parameter rather than a linearly changing one. Thus, again, we are led to a natural model with $m = 1$. Much of the smoothing spline literature and applications take $m = 2$, but for time-varying parameter models, a linear function is a more appropriate smoothness objective.

Estimation is accomplished by minimizing

$$L^{**} = e'e + \lambda\beta'R'R\beta \tag{3.10}$$

over β and γ, taking λ as given. It is clear that this will have the same maximand as L^* and consequently L once σ^2 is concentrated out. Once again, (3.7) and (3.9) provide expressions for the estimated $\hat{\beta}$. The statistical model underlying this is rather different, as now the β's are fixed values that are being estimated using a biased estimation procedure designed to substantially reduce variance leading to reduced mean squared error.

The smoothness prior is not assumed to be a correct representation of the true data generation process in the standard case. However, as pointed out above for the Bayesian estimation, the effects of the choice of λ and R do not vanish asymptotically. Many of the asymptotic results developed for nonparametric regression models of this kind assume that additional data are obtained using more frequent sampling over the same interval, and consequently consistency is not difficult to achieve (see, e.g., Wahba 1983).

Selection of the appropriate value for λ is a central feature of any operational nonparametric method. This is equivalent to the choice of bandwidth for kernel methods of estimation. Wahba and Wold (1975) proposed using cross-validation to pick λ, and in Craven and Wahba (1979), generalized cross-validation (GCV) was introduced and applied. This turns out to be a close relative of Akaike's (1969, 1973) finite prediction error (FPE) criterion and Akaike information criterion (AIC) and several others suggested by Shibata (1981) and surveyed by Atkinson (1981). Essentially, all attempt to use within-sample information to obtain estimates of mean square prediction error and then minimize this with respect to λ. These criteria are therefore designed to be optimal for a mean square error measure of risk. GCV can be defined as

$$\mathrm{GCV} = e'e/[1-\mathrm{tr}(A)/T]^2 = y'(I-A)^2y/[\mathrm{tr}(I-A)/T]^2 \qquad (3.11)$$

where $\hat{y} = A(\lambda)y$ and $A(\lambda)$ is the projection matrix

$$A = \hat{X}(\hat{X}'\hat{X} + \lambda R'R)^{-1}\hat{X}' + z(z'z)^{-1}z' \qquad (3.12)$$

and $e = y - X\hat{\beta} - z\hat{\gamma}$. If λ is very small, then the sum of squared residuals will be very small but $\mathrm{tr}(A)$ will be approximately T, thus leading to large values of GCV. Large values of λ decrease the trace of A but impose tighter constraints on the data, restricting the fit and increasing the sum of squared residuals. In Engle et al. (1985), the results using GCV to select λ are very satisfactory.

The classical investigator will therefore choose λ by maximizing L, whereas the nonparametric statistician will minimize some criterion such as GCV. In a recent comparison of these criteria, Wahba (1983) found that in some cases, maximum likelihood will undersmooth, that is, choose too small a value of λ compared to an optimal selection from the point of view of mean square forecast error loss function. Since many of the suggested applications of the TVP model are forecasting applications, this suggests that it may be interesting to consider other criteria for selecting λ. Furthermore, since GCV is designed to select λ optimally, even when the data generation process of β is not $R\beta = \eta$, it may be more robust to misspecification.

4 Forecasting electricity sales with the TVP model

Recently these methods have been applied quite successfully in modeling the demand for electricity over time. Early results are described in EPRI (1983) and others in Engle, Brown, and Stern (1986) and Granger and Engle (1985). Particular features of these studies will be reported here.

The usage of electricity by residential customers is strongly influenced by the weather since much of the fluctuation in demand is due to heating and cooling requirements. The number of homes with central air conditioners has increased since the early sixties as the costs of such units have decreased and as this technology has been incorporated into new buildings. As the price of electricity increased in the middle to late seventies, many customers reduced their thermostats to reduce their bills and introduced new technologies such as additional insulation and thermal engineering. The net effect has been continual gradual shifts in the relationship between temperature and electricity usage. This reveals itself in the aggregate data as a shift between the winter and summer peaks; in fact, many utilities are now summer peaking that were previously winter peaking.

The discussion suggests that the coefficient of hot weather might be an excellent candidate for a varying coefficient. A second potential candidate for the TVP formulation is the trend. The growth rate of electricity sales has slowed from a constant 7 percent per year in the sixties and early seventies to nearly a zero growth rate. Presumably, this is due to the combined effect of macroeconomic recession and dramatically increasing energy prices; however, such variables do not adequately represent the process, and one is often forced to rely on measures of "patriotism" or "conservation ethic" to explain the slowdown.

Two time-varying parameter models were proposed for this data set: TVP-A and TVP-B. The first allows the coefficients on weather to vary, and the second estimates a stochastic trend. In each case, monthly data on residential sales per customer by state (Y) were regressed on monthly seasonal dummies (MONTHDUM), the real price of electricity for a customer using 750 kWh/month (RELP750), real personal income per customer (RPINC/C), and weather variables measured as heating degree days (HDDs) and cooling degree days (CDDs) each measured from the daily midpoint of high and low temperature with 65° as the base.[1] Because the bills for a month are for electricity consumed partly in the previous month

[1] Specifically, if f_t is the average of the high and low temperatures on day t, then

$$\text{HDD} = \sum \max(65 - f_t, 0) \qquad \text{CDD} = \sum \max(f_t - 65, 0)$$

where the sums are over the days of the month.

and sometimes for the month before that, potentially two-month lags of the weather may be important (see Train et al. 1984).

The TVP-A model can be written as

$$\begin{aligned} Y_t &= \beta 1_t \mathrm{CDDMA}_t + \beta 2_t \mathrm{HDDMA}_t + z_t \gamma + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \\ \beta 1_t &= \beta 1_{t-1} + \eta 1_t \\ \beta 2_t &= \beta 2_{t-1} + \eta 2_t \end{aligned} \tag{4.1}$$

with $z = \{\mathrm{RPINC/C, RELP750, MONTHDUM, CONSTANT}\}$ and $\mathrm{CDDMA} = \mathrm{CDD}_t/3 + \mathrm{CDD}_{t-1}/2 + \mathrm{CDD}_{t-2}/6$ with HDDMA similarly defined. The moving averages of heating and cooling degree days are merely to condense the critical weather variables into a single parameter so that the effect of changing appliance stocks is not allowed to change the distribution of the load over the billing cycle. When this model is used for forecasting, the b's will take on their last estimated value throughout the forecast horizon. First-order serial correlation is assumed to be sufficient for this model on the grounds that twelfth-order (as found in many other specifications) is due simply to the failure to allow the weather coefficient to vary.

The model for stochastic trend is TVP-B and can be defined as

$$\begin{aligned} Y_t &= \beta_t + z_t^* \gamma + u_t \\ \beta_t &= \beta_{t-1} + 0.9(\beta_{t-12} - \beta_{t-13}) + \eta_t \\ u_t &= \rho u_{t-1} + 0.9(u_{t-12} - \rho u_{t-13}) + \epsilon_t \end{aligned} \tag{4.2}$$

where $z^* = \{\mathrm{RPINC/C, RELP750, CONSTANT, CDD}, \mathrm{CDD}_{t-1}, \mathrm{CDD}_{t-2}, \mathrm{HDD}, \mathrm{HDD}_{t-2}\}$. The regressors differ from TVP-A by exclusion of the seasonal dummies but allowance for a more flexible lag structure on the weather variables. Of most importance, however, is the constant coefficient on the weather and the allowance for a varying intercept that is constrained to have a unit root. The error structure now allows for seasonal serial correlation.

Notice that the equations for β_t and u_t are of exactly the same form with a common seasonal factor. The only difference is the imposition of the unit root in β_t: this distinguishes the trend from the disturbance. The sum of $\beta_t + u_t$ is an ARIMA(1, 1, 1) times the seasonal factor $(1 - 0.9B^{12})$. The seasonal factor is set with a known parameter 0.9 only for convenience in estimation so that the entire equation could be seasonally quasi-differenced before the TVP routine was called.

When forecasting with the model, u_t will gradually damp to its unconditional mean, which is zero, whereas β_t will damp to its last estimated value. The trend for the current month will be 90 percent of whatever it was estimated to be the previous year. Similarly, the current-month disturbance will be forecast as ρ times its value the previous month plus 90 percent of $u_t - \rho u_{t-1}$ the previous year. If instead of 0.9 a coefficient of 1.0 had been used, then the trend would not damp out but would remain at exactly the rate from the preceding year throughout the forecast horizon. All forms of second differencing used in this forecasting experiment proved quite inferior predictors, particularly for multistep forecasts where they soon proceeded to go way off the mark.

If the stochastic trend were unnecessary for some data set, then the variance of η should be zero. This implies that the regressors z^* adequately model the trend, which is interpretable by saying that Y and z^* are cointegrated, as in Engle and Granger (1987).

These models were estimated by maximum likelihood using the EM algorithms as presented in Watson and Engle (1983) [which is derived from the general procedure of Dempster, Laird, and Rubin (1977)]. Each step of this procedure is relatively fast, although it often takes many steps to achieve convergence. The alternative of using scoring as employed in Engle and Watson (1981) requires more computations and is optimal only in the neighborhood of the maximum where the likelihood is nearly quadratic. In more recent experience, it appears that a series of EM steps followed by several steps of scoring provides a quicker and more sure route to the maximum.

Two benchmark models were also estimated so that the forecast performance of the TVP models could be carefully evaluated. The more sophisticated model, labeled AUTO-A, is defined as

$$
\begin{aligned}
Y_t &= z_t^{**}\gamma + u_t \\
u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-12} + \rho_3 u_{t-13} + \epsilon_t
\end{aligned}
\qquad (4.3)
$$

where $z^{**} = \{$RPINC/C, RELP750, CONSTANT, MONTHDUM, CDD, CDD_{t-1}, CDD_{t-2}, HDD, HDD_{t-1}, $\text{HDD}_{t-2}\}$, which includes all the regressors of both z and z^*. It differs from TVP-B by not having the imposed unit root trend term but does not constrain the autoregressive error process. It differs from TVP-A by allowing more flexible error terms but does not include time-varying weather sensitivity. A still simpler model, labeled NAIVE, assumes that u_t in the AUTO-A model is white noise.

These four models are estimated for ten states from January 1964 through December 1978 and then are used to forecast the 36 months through December 1981. This produces 36 rolling forecasts one step ahead but only 12 forecasts 24 months ahead. Both conditional forecasts and unconditional forecasts were constructed. In the latter case, the weather

Table 7.1. *Geometric means across states*

Method	1 month	12 months	24 months	Annual average
Conditional rms forecast errors				
Naive	69.7	72.2	77.5	39.9
Auto-A	26.3	35.2	46.2	17.4
TVP-A	31.9	34.8	43.2	24.1
TVP-B	25.0[b]	30.7[b]	37.1[b]	14.6[b]
Unconditional rms forecast errors				
Naive	72.9	75.4	86.8	40.9
Auto-A	29.2	43.1	50.3	19.1
TVP-A	31.0	40.8[b]	43.5[b]	23.1
TVP-B	28.8[b]	41.3	47.7	18.3[b]

[a] Revised May 1984.
[b] The best at that horizon.

and economic variables must themselves be forecast. For this purpose, the weather was assumed to maintain its long-run average patterns, whereas the economic variables were forecast by low-order Box–Jenkins models.

The geometric means across states of the root-mean-square (RMS) forecast errors are given in Table 7.1. These results point out the abilities of time-varying parameter models in forecasting, particularly for multistep forecasts. The TVP-B model proves to be the most successful of all these models for both short- and long-run forecasting conditional on the future economic and weather variables, and the TVP-A model is quite successful for forecasts of a year or more. It apparently suffers in the one-month-ahead forecast because there remains important twelfth-order serial correlation that is picked up by the AUTO-A and TVP-B models. This hurts in the yearly average forecast as well as in the one-month forecast. Both models perform dramatically better than the NAIVE model and slightly better than the AUTO-A model, which is clearly a very sophisticated competitor.

5 Applications to rational-expectations models

The flexibility of the state space model makes it an appropriate tool for estimating a wide class of econometric models. One important class of models that fits naturally into the state space framework is the class of dynamic linear models containing unobservables. In these models, the state space measurement equations are used to describe the relationship between the observed and the unobserved variables. The state transition equations are then used to model the evolution of the unobserved variables through time. The Kalman filter can be used to form the one-step-ahead prediction errors and prediction error variances of the observed data. These form the basis of a Gaussian likelihood or other objective function, which can be maximized with respect to any unknown parameters. This procedure is discussed in detail in Engle and Watson (1981) and Watson and Engle (1983).

In modern macroeconomic models, expectation variables are an important class of unobserved variables. This section will investigate the special structure of state space models that include these variables. This question was first investigated by Wall (1980), who showed special state space forms for models that included adaptive or rational expectations. Burmeister and Wall (1982) used a state space model to estimate a money-demand/money-supply model of the German hyperinflation. The unobservable in their model was a "stochastic bubble" that drove the hyperinflation. Hamilton (1985) has used a state space model to investigate the relationship between the rate of price inflation, nominal interest rates, expectations of inflation, and real interest rates. The latter two variables are unobservable in his model.

In any completely specified economic model in which expectations play a role, the expectation formulation mechanism is completely specified. This makes it possible (in principle at least) to solve out for the unobserved expectations, leaving a model in which only observed variables are present. The standard textbook treatment of adaptive expectations leading to a Koyck-lag in observed variables is a case in point (see, e.g., Johnston 1984, p. 348). When expectations are formed rationally rather than adaptively, the same thing is generally true,[2] but the solution procedure is typically more complicated. [Solution procedures are discussed in Blanchard and Kahn (1980), Chow (1983b), and Whiteman (1983).]

Since it is possible to explicitly solve out for the unobserved expectations, one might question the utility of using a model for unobservables. This question was addressed in Watson (1985a) in the context of a dynamic linear rational-expectations model. He argued that the Kalman filter was a useful device for recursively solving the model and that the flexibility of the state space model would be particularly important for applications in which issues such as errors in variables, temporal aggregation, missing data, or temporal instability were important. In the remainder of this chapter, we will investigate the usefulness of the state space representation and the Kalman filter for estimating this type of model.

[2] This may not be possible in models incorporating bubbles.

The specific model that we consider relates stock prices, expected future stock prices, and dividends. The basic structure of the model follows from the standard arbitrage condition assuming a constant real interest rate. The data are a subset of those used by Shiller (1981) in his investigation of the same relationship using a variety of variance bounds.

The remainder of this section is organized as follows. We begin by briefly reviewing a simplified version of the model and its state space representation. This allows us to highlight the basic features of the recursive solution technique and to discuss the rationale behind some constraints that will be placed on the parameters of the state space model. The expository simplification in this discussion involves postulating a stationary AR(1) generating process for dividends. In the next section, we drop this assumption and assume that dividends have an ARIMA representation and that they may depend on lagged values of stock prices and on other variables that we do not observe. This leads to our empirical specification. The results from this initial specification are not entirely satisfactory. The innovations appear to be heteroscedastic, and a simple vector autoregressive model with a co-integration constraint fits the data much better. The final three sections investigate possible causes for the poor performance of the model. First, we estimate a model incorporating time-varying discount rates; in the next section, we estimate a model that incorporates dynamic errors in variables; and in the final section, we present a model in which some of the disturbance processes are characterized by autoregressive conditional heteroscedasticity.

5.1 *Recursive solutions*

If we let p_t denote the real price of a share of stock and d_t the real value of dividends per share, then the familiar efficient-markets hypothesis can be written as

$$p^e_{t+1} = \mu p_t - d_t \qquad p^e_{t+1} = E[p_{t+1} \mid \Omega_t] \tag{5.1}$$

Where it is assumed that dividends are paid at the end of the period, μ is the time-invariant gross rate of return, and the information set Ω_t contains present and past variables of p, d, and any other relevant variables. Equation (5.1) is the relationship that we will be investigating throughout the remainder of this chapter. Before presenting the state space form of the model and discussing the recursive solution procedure, we need to specify a model for the generation of dividends. To simplify the presentation in this section, we will assume that dividends are generated by the AR(1) process:

$$d_t = \phi d_{t-1} + e_t^d \tag{5.2}$$

where e_t^d is white noise with $E(e_t^d \mid \Omega_{t-1}) = 0$, $\text{var}(e_t^d \mid \Omega_{t-1}) = \sigma_d^2$, $|\phi| < 1$, and d_0 given. (These assumptions will be modified for the empirical analysis.) The state space representation of the model is derived from the following "reduced form":

$$p_t = p_t^e + e_t^p \tag{5.3}$$

$$d_t = d_t^e + e_t^d \tag{5.4}$$

$$d_{t+1}^e = \phi d_t^e + \phi e_t^d \tag{5.5}$$

$$p_{t+1}^e = \mu p_t^e - d_t^e + \mu e_t^p - e_t^d \tag{5.6}$$

with

$$E[e_t^p \mid \Omega_{t-1}] = E[e_t^d \mid \Omega_{t-1}] = 0$$

This reduced form describes the evolution of variables at time t as functions of variables at time $t-1$ and innovations. Equations (5.3) and (5.4) follow from the rational expectations for p and d. Equations (5.5) and (5.6) describe the generation of the expectations. Equation (5.5) follows directly from the AR(1) generating equation for d_t. Equation (5.6) is merely a rewrite of the efficient-markets equation (5.1) replacing p_t and d_t with their decompositions given in (5.3) and (5.4). Equations (5.3)–(5.6) incorporate all of the information concerning the evolution of the data given in equations (5.1) and (5.2).

To characterize the generation of the data completely, we need to introduce three parameters not present in relationships (5.1) and (5.2). First, we need to parameterize the covariance structure of the innovations e^d and e^p. We project e^p onto e^d, which yields

$$e_t^p = \pi e_t^d + u_t$$

where π and σ_u^2 are thus far unrestricted. In addition, we need initial conditions for the expectations p_1^e and d_1^e. We have $d_1^e = \phi d_0$, but the value of p_1^e is left unrestricted by the model. The complete characterization of the data generation process requires specification of the parameters π, σ_u^2, and p_1^e as well as the parameters σ_d^2, d_0, ϕ, and μ from above.

To write the model in state space form, we use equations (5.3)–(5.6) as a set of transition equations. The measurement equations merely select p_t and d_t from the state vector. The model is

$$Y_t = HX_t$$

$$X_t = \Phi X_{t-1} + G\epsilon_t$$

with

$$Y_t' = (p_t d_t) \qquad X_t' = (p_t d_t d_{t+1}^e p_{t+1}^e) \qquad \epsilon_t' = (u_t e_t^d)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \phi & 0 \\ 0 & 0 & -1 & \mu \end{bmatrix} \qquad G = \begin{bmatrix} 1 & \pi \\ 0 & 1 \\ 0 & \phi \\ \mu & \mu\pi - 1 \end{bmatrix}$$

Initializing the Kalman filter with a state vector that has mean $X_{0/0} = (p_0 d_0 d_1^e p_1^e)'$ and variance $P_{0/0} = 0$, we can use the Kalman filter to generate the one-step-ahead forecast errors and forecast error variances of the observed data.[3] These can be used to form the Gaussian likelihood of the date, which can be maximized with respect to the unknown parameters (ϕ, μ, π, σ_u^2, σ_d^2, and p_1^e).

In the empirical example that follows, we impose a constraint on the parameters of the model. This constraint rules out explosive "backward-looking" behavior and deterministic or stochastic rational "bubbles." To motivate the constraint, consider the constructed variable

$$w_t = p_t - d_t(\mu - \phi)^{-1}$$

It is easy to show that w_t^e evolves as

$$w_{t+1}^e = \mu w_t^e + \mu[c e_t^d + u_t]$$

with $c = \pi + (\mu - \phi)^{-1}$. This implies that

$$E[w_{t+k} \mid \Omega_t] = \mu^{k-1} w_{t+1}^e \quad \text{for } k > 1$$

Since $\mu > 1$, this implies that w_t is expected to explode if $w_{t+1}^e \neq 0$ for any t. Since w is a linear combination of d and p, and d is stationary, p is expected to explode at the same rate as w. The constraint that we impose rules out this expected explosive behavior, by setting $w_t^e = 0$ for all t. It is easy to verify that this constraint will be satisfied if and only if

$$p_1^e = (\mu - \phi)^{-1} d_1^e \qquad \pi = -(\mu - \phi)^{-1}$$

and $\sigma_u^2 = 0$. These constraints are equivalent to forming the "forward-looking" solution of p_t; that is,

$$p_t = \mu^{-1} \sum_{i=1}^{\infty} \mu^{-i} E(d_{t+i} \mid \Omega_t)$$

[3] There are two reasonable interpretations for the initial value, p_0. The first assumes that equation (5.1) holds at time 0, so that $p_0 = \mu^{-1}(p_1^e + d_0)$. The second assumes that p_0 is an unrestricted nuisance parameter. In either case, given p_1^e and d_1^e, the likelihood of the data for $t = 1, 2, \ldots, T$ does not depend on the value of p_0.

This example has shown the basic features underlying the state space representation of the model. However, the particular form of the model that was described has no empirical relevance. Notice, for example, that the imposition of the forward-looking solution yields a model for the bivariate process of p and d that is generated by a single disturbance. It implies that $p_t = (\mu - \phi)^{-1} d_t$ for all t, so that the regression of p_t onto d_t has an R^2 of 1. We will now discuss a modification of the model that incorporates an additional source of uncertainty.

5.2 *The empirical model*

Our empirical specification is somewhat richer than the specification used in the last section. In particular, we assume that $(1-B)d_t$ is covariance stationary and is generated by

$$(1-B)d_t = x'_{t-1}\beta + \xi^d_t \tag{5.7}$$

where x_{t-1} is covariance stationary with $x_{t-1} \in \Omega_{t-1}$. The innovation ξ^d is white noise, with $E(\xi^d_t \mid \Omega_{t-1}) = 0$. In this application, we will be using data on prices and dividends only, so that we will treat some of the x variables as unobserved. Since we are concerned only with the relationship between prices and dividends, we project x'_{t-1} onto lagged values of prices and dividends. This yields a relationship of the form

$$\phi(B)(1-B)d_t = \gamma(B)(1-B)p_{t-1} + \theta(B)e^d_t + \omega(B)u_{t-1} \tag{5.8}$$

where all polynomials in B are one sided and finite order, and we normalize $\phi_0 = \theta_0 = 1$. The innovation e^d is now viewed as the innovation from the restricted information set, that is, the information set consisting of lagged values of prices and dividends only.[4] This relationship led to our initial empirical specification for the generation of expectations of dividends:

$$d^e_{t+1} = d^e_t + \phi_1(d^e_t - d_{t-1}) + \phi_2(d_{t-1} - d_{t-2}) + \gamma_0(p^e_t - p_{t-1}) + \lambda_1 u_t + \lambda_2 e^d_t$$

This follows directly from (5.8) by assuming $\phi(B)$, $\gamma(B)$, $\theta(B)$, and $\omega(B)$ are polynomials of order 2, 0, 1, and 1, respectively. The coefficients $\lambda_1 = \gamma_0 - \omega_0$ and $\lambda_2 = 1 + \phi_1 - \theta_1$. The state space representation for the model has the same basic structure as the model that was presented in the last section. Extra lags of prices and dividends must be added to allow for the more complicated dynamics. The state vector is now

$$X_t = (p_t d_t d_{t-1} d^e_{t+1} p^e_{t+1})'$$

[4] When prices and dividends are co-integrated, it is possible that terms such as $d_{t-i} - \tau p_{t-i}$ appear on the right-hand side of this equation. This possibility will be discussed in more detail in the text.

and the matrices appearing in the state transition equation are

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -\gamma_1 & \phi_2-\phi_1 & -\phi_2 & 1+\phi_1 & \gamma_1 \\ 0 & 0 & 0 & -1 & \mu \end{bmatrix} \qquad G = \begin{bmatrix} 1 & \pi \\ 0 & 1 \\ 0 & 0 \\ \lambda_1 & \lambda_2 \\ \mu & \mu\pi - 1 \end{bmatrix}$$

In the last section, we discussed a constraint placed on the parameters of the model that ruled out explosive behavior in price expectations. We want to impose the same constraint on this system. The form of the constraint is somewhat different because of the change in the dividend process. To derive the constraint for this model, let ρ be an eigenvector of Φ' corresponding to an eigenvalue greater than unity. (For most reasonable parameter values in the model, the matrix Φ will have exactly one root with a modulus greater than 1.) Then

$$\rho' X_t = \alpha(\rho' X_{t-1}) + (\rho' g_1)u_t + (\rho' g_2)e_t^d$$

where α is the eigenvalue corresponding to ρ, and g_1 and g_2 are the first and second columns of G. Since $|\alpha| > 1$, explosive behavior can be ruled out only if

$$\rho' X_0 = 0 \tag{5.9}$$

$$\rho' g_1 = 0 \tag{5.10}$$

$$\rho' g_2 = 0 \tag{5.11}$$

Equation (5.9) places one constraint on the initial value of the state vector, equation (5.10) imposes a constraint on the relationship between μ and λ_1, and equation (5.11) imposes a constraint on the relationship between π, λ_2, and μ. These constraints will be imposed on the model.

The data that we will use to estimate the model is the Standard and Poor's (S & P) data set described and analyzed in Shiller (1981). Stock prices represent annual observations of the S & P composite stock price index deflated by the producer price index (PPI). The dividend series represent dividends per share adjusted to the S & P index. These are four-quarter totals and are deflated by the annual average of the PPI. The sample period is 1871–1979. A more complete description of the data can be found in Shiller (1981).

In the first column of Table 7.2, we present the results from our initial specification.[5] [In all of the models estimated, the coefficient on $(1-B)p_{t-1}$

[5] The initial values for the elements of the state vector were chosen in the following way. Data from the first two time periods were dropped from the sample and used to initialize

Table 7.2. *Structural models relating stock prices and dividends*

	Base model	Measured error	ARCH
ϕ_1	1.112	1.236	1.097
	(0.071)	(0.138)	(0.080)
ϕ_2	−0.148	−0.252	−0.130
	(0.073)	(0.138)	(0.082)
μ	1.040	1.043	1.037
	(0.006)	(0.008)	(0.008)
π	23.658	11.534	18.750
	(4.080)	(9.273)	(2.422)
σ_u	0.055	0.052	–
	(0.003)	(0.004)	
σ_{ed}	0.002	0.001	0.002
	(0.004)	(0.0002)	(0.0001)
$d_{1/0}$	0.006	0.006	0.006
	(0.008)	(0.005)	(0.004)
$p_{1/0}$	0.096	0.203	0.103
λ_1	0.003	0.002	0.002
λ_2	1.132	1.218	1.105
ρ_m	–	0.784	–
		(0.160)	
σ_m	–	0.085	–
		(0.151)	
α_0	–	–	4.0×10^{-4}
			(3.0×10^{-4})
ρ	–	–	0.998
			(0.021)
α_1	–	–	0.076
			(0.005)
L	−583	−577	−555

appearing in the dividend equation was small and statistically insignificant. For the results reported in the table, the coefficient was constrained to equal zero.] The point estimates look sensible. The standard errors for the one-step-ahead forecasts of prices and dividends are 0.068 and 0.0017,

p_0, d_0, and d_{-1}. The initial value, d_1^e, was estimated as an unknown nuisance parameter, and p_1^e was determined from the "boundedness" condition (3.3). An alternative is to use the initial values of p and d to determine p_1^e from $p_1^e = \mu p_0 - d_0$. The value of d_1^e could then be determined from the boundedness condition. These approaches lead to different respectively. These can be compared with root-mean-squares for $(1-B)p$ and $(1-B)d$ of 0.077 and 0.0017. The estimate of μ implies an annual real interest rate of 4 percent. From the autoregressive polynomial for $(1-B)d$, it may appear as if a unit root is present, but this may be an artifact of the parameterization chosen. The presence of the coefficients λ_1 and λ_2 allow moving-average terms and feedback from prices to dividends. This may lead to near cancellation of the large autoregressive root in the dividend process.

Some caution must be exercised in conducting inference concerning μ because it is estimated as a simple transformation of an estimated co-integrating coefficient. [See Granger (1981) and Engle and Granger (1987) for a discussion of co-integrating processes.] To see why this is so, recall that equations (5.9)–(5.11) imposed the constraint $\rho' X_t = 0$. This imposes a constraint between the levels of the price and dividend processes. In the model under consideration, the largest eigenvalue of Φ is μ, and a little algebra shows that the constraint that is imposed on the system can be written as

$$\lim_{k \to \infty} E(p_{t+k} \mid \Omega_t) = (1-\mu)^{-1} E(d_{t+k} \mid \Omega_t)$$

Since dividends follow an integrated process, this is just a co-integration constraint, and the variable $\{p_t - (1-\mu)^{-1} d_t\}$ is covariance stationary. Stock (1984) discusses inference in models with co-integrating constraints.

Although the estimated model looks sensible, it does not stand up to careful scrutiny. Misspecified dynamics will lead to serially correlated innovations, and the estimated innovations can be used to test for this. The estimated innovations from the model showed no gross serial correlation. The only significant auto- or cross-correlation was the second autocorrelation in price. It had a value of -0.28. A more stringent test of the model can be constructed, however. The model is a constrained co-integrated vector ARIMA(2, 1, 1) model. We can compare the fit of this model to an unconstrained model. To avoid the computational complexity of estimating moving-average coefficients we have approximated the unconstrained model with a third-order vector autoregression relation $(1-B)p$ and $(1-B)d$. To incorporate the co-integrating constraint, we included error correction terms of the form $(p_{t-1} - \tau d_{t-1})$ in the vector autoregressive representation (VAR). The results for this model are shown in Table 7.3.[6] The co-integrating coefficient, τ, is estimated to be 24.5. This

Footnote 5 *(cont.)*
estimates of the initial expectations, but the asymptotic distribution of the other estimated parameters is unaffected.

[6] The model was estimated in a two-step procedure. The coefficient τ was estimated by regressing the level of price on the level of dividends. This estimated value of τ was then

Table 7.3. *Co-integrated VAR model*[a]

	Dividends	Price
$(1-B)d_{t-1}$	0.220	6.486
	(0.129)	(4.923)
$(1-B)d_{t-2}$	−0.274	−1.464
	(0.123)	(4.701)
$(1-B)d_{t-3}$	0.009	3.643
	(0.126)	(4.808)
$(1-B)p_{t-1}$	0.002	0.077
	(0.003)	(0.129)
$(1-B)p_{t-2}$	0.003	−0.273
	(0.003)	(0.119)
$(1-B)p_{t-3}$	0.005	0.115
	(0.003)	(0.126)
ec_{t-1}	−0.000	−0.120
	(0.001)	(0.069)
σ	0.0017	0.0643

[a] $ec_t = \text{price}_t - 24.51\,\text{div}_t$.

implies an annual real interest rate of 4 percent, essentially identical to the estimate found in the structural model. The results suggest little feedback from prices to dividends, and a significant effect of the error correction term on prices but not on dividends.

The log likelihood associated with this model is −567, compared to −583 for the structural model. In the structural model, we have estimated 5 parameters describing the evolution of $(1-B)d$ and $(1-B)p$, 1 parameter – the co-integrated coefficient – describing the long-run relationship between the levels of p and d, and one initial condition relating the level of p and d. In the VAR, we have estimated 17 parameters describing the evolution of $(1-B)p$ and $(1-B)d$, and 1 co-integrating coefficient. Neglecting the initial value (which is not estimated consistently), this suggests that the X^2_{12} distribution is a valid large-sample approximation to

imposed during the estimation of the VAR. The Durbin-Watson (DW) statistic associated with the levels regression was 0.311. Monte Carlo results presented in Engle and Granger (1987) suggest that if the series were not co-integrated, then there is only a 10 percent chance of finding a DW statistic this large. Indeed, after a correction for heteroscedasticity, the DW statistic increases to 0.40.

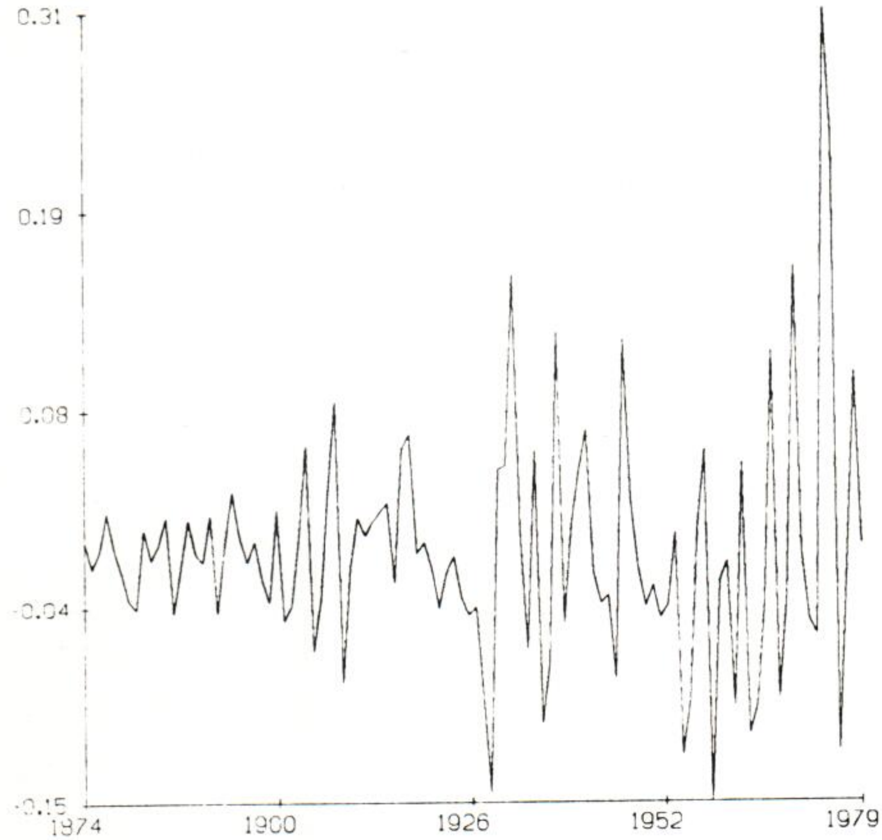

Figure 7.1 Price innovations, model 1.

the distribution of the likelihood ratio statistic.[7] The likelihood ratio statistic has a value of 32, which suggests that the model can be rejected at any reasonable confidence level.

A plot of the innovations from the model suggests another serious problem. These innovations, plotted in Figure 7.1, show clear heteroscedasticity. When the squared price innovations are regressed on a constant and the squares of lagged price, we find a positive coefficient and an R^2 of 0.19. This yields an LM heteroscedasticity test statistic of 19. The same exercise for dividends also yields a positive coefficient, but the R^2 is much lower; it is only 0.005.

At least three candidate sources of model misspecification come to mind that are consistent with the diagnostic test results. First, real interest rates,

[7] The distribution is complicated by the presence of the estimated co-integrating coefficient. Stock (1984) has shown that, for the VAR model, the large-sample distribution of the other parameters in the model is unaffected by using the estimated value of the co-integrating coefficient rather than the true value. The X^2 approximation used in the text is valid if this result carries over to our structural model.

and therefore μ, may not have been constant over the 109-year sample period. Second, the price, dividend, or the price deflator series may be contaminated with measurement error. Third, there may be persistent heteroscedasticity in the innovations arising from an unspecified source. This suggests that autoregressive conditional heteroscedasticity should be incorporated in the model. These candidates are investigated fully in the next three sections.

5.3 *Time-varying discount rates*

The incorporation of time-varying interest rates into our structural model is difficult. One approach would involve replacing our parameter μ in the state transition matrix with a set of time-varying μ_t's. Although this may appear to be a simple solution to the problem, two issues must be faced. First, how do we estimate the real interest rate – that is, what values do we use for μ_t? Second, how do we impose the nonexplosiveness or forward-looking solution on the model? Since μ is time varying, the procedure used in the last section is not appropriate. Rather than address these two difficult issues, we will temporarily abandon the full-information framework. Instead, we consider time variation in the efficient-markets relationship:

$$p_{t+1}^e = \mu p_t - d_t$$

If we let e_{t+1} denote the forecast error $p_{t+1} - p_{t+1}^e$, then we can rearrange this equation to form

$$p_{t+1} + d_t = \mu p_t + e_{t+1} \tag{5.12}$$

where e_{t+1} is uncorrelated with p_t. If the parameter μ is constant in equation (5.12) and the disturbance is homoscedastic, then the equation can be estimated efficiently by ordinary least squares (OLS). Estimating the equation by OLS yields

$$\mu = 1.029(0.016) \qquad \sigma_e = 0.0068 \qquad \mathrm{DW} = 1.86$$

The residuals from the OLS regression exhibit the same pattern of heteroscedasticity that was present in the model of the last section. We can carry out a test for time-varying parameters using the test proposed in Watson and Engle (1985). The alternative under consideration is

$$\mu_t - \mu = \phi(\mu_{t-1} - \mu) + \eta_t \tag{5.13}$$

with $\eta_t \sim \mathrm{NIID}(0, \sigma_\eta^2)$ and $|\phi| < 1$. Constant coefficients corresponds to the parameter restriction $\sigma_\eta^2 = 0$, and the test statistic is constructed as an LM test for this hypothesis. Since the transition parameter ϕ is unidentified under the null, the LM test statistic is calculated for a variety of values of ϕ, and the largest of these statistics is used to test the hypothesis. The square root of the test statistic takes on a value of 7.8 corresponding to $\phi = 0.05$. Under the null hypothesis, this statistic is the maximum of a set of correlated standard normal random variables, so that the value of the statistic appears to be very extreme. Using the approximation to the distribution of the test statistic suggested by Watson and Engle (which is not entirely appropriate because of the nonstationarity of the data), the test statistic has a probability value less than 0.3×10^{-14}. This is strong evidence against the null hypothesis of constant coefficients.

These test results suggest that time variation in μ may be important. A time-varying parameter model produces the following estimates:

$$p_{t+1} + d_t = \mu_t p_t \quad (\sigma = 0.0009)$$

$$\mu_t = \underset{(0.018)}{1.068} + \underset{(0.096)}{0.000}(\mu_{t-1} - 1.068) \quad (\sigma = 0.177)$$

The model appears to fit the data quite well, but the parameter estimates are difficult to interpret. The estimates imply that μ varies randomly around 1.068 with a standard deviation of 0.17. This standard deviation is quite large and implies that real annual interest rates as low as -29 percent and as high as 42 percent are not particularly surprising. In Figure 7.2, we plot the smoothed values of the discount rate implied by the model. These are the minimum mean square error estimates of μ_t conditional on the estimated parameters and all of the data.[8] The model suggests that real interest rates are very volatile, with annual real rates as high as 50 percent in 1936 and as low as -35 percent in 1932 and -33 percent in 1947 and 1975. These nonsensical results led us to consider the second candidate for the cause of the misspecification: errors in variables.

5.4 *Measurement error*

The description of the data given in Section 5.2 and more fully in Shiller (1981) suggests at least two important sources of measurement error. First, the portfolio of stocks used to construct the price and dividend indices is time varying. This implies that the time t expected future dividend series implicitly constructed by the model need not correspond to the time t portfolio of stocks. Second, the PPI used to deflate both prices

[8] These smoothed values appear to be quite accurate estimates of the μ_t's underlying the model. If we denote the smoothed estimates by $\mu_{t/T}$, then the RMS of $\mu_{t/T} - \mu_t$ varies between 0.009 and 0.001 over the sample period. These values are conditioned on the estimated parameters of the model.

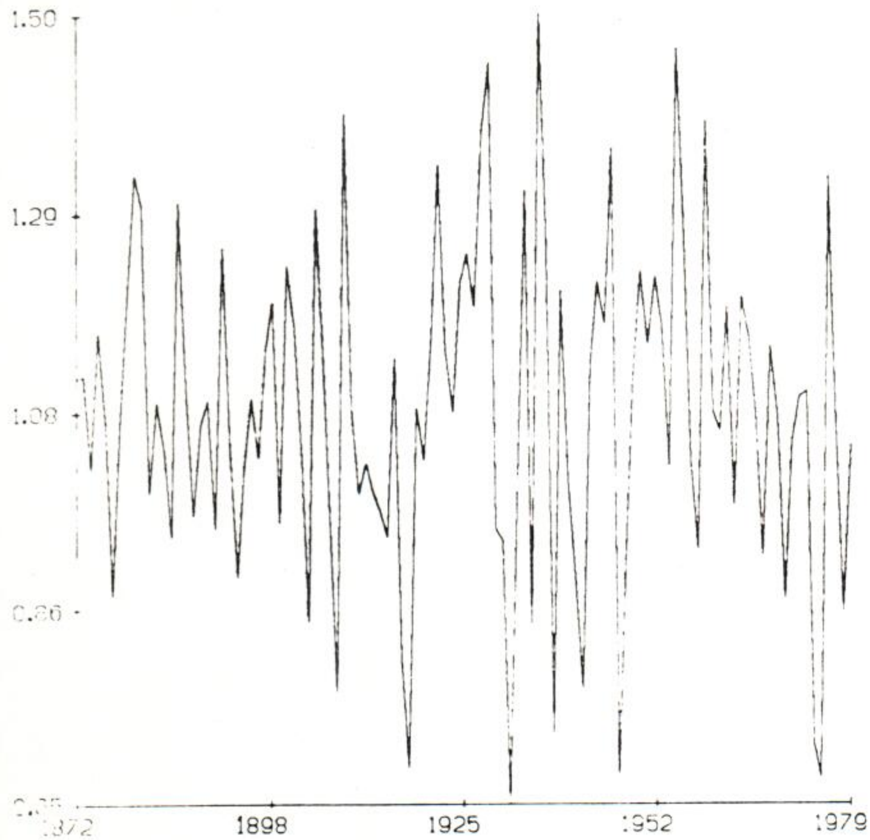

Figure 7.2 Time-varying gross rates of return implied by TVP model.

and dividends is not the ideal deflator. The ideal deflator is a perfectly measured index of prices for consumption goods. Trend movements in the PPI and this ideal measure are probably very similar, but there is no reason to believe that shorter-run, year-to-year movements are perfectly correlated. Since the structural model imposes very tight constraints on the long- and short-run movements in the data, these transitory deviations of the PPI from the ideal deflator may be very important.

To get a rough idea of the importance of possible measurement error in the model, we considered a modification of the basic model in which an additional AR(1) disturbance was appended to both p_t and d_t. These two measurement error processes were assumed to be independent of the structural disturbances u_t and e_t^d. This produced a model in which

$$p_t = p_t^* + m_t^p \qquad d_t = d_t^* + m_t^d$$

where p^* and d^* represent the true underlying values of price and dividends, and m^p and m^d are the measurement errors. The asterisks identify variables assumed to follow the structural model given in Section 5.2, and the measurement errors are assumed to follow univariate AR(1) models.

We allowed the innovations in the measurement errors to be correlated. The state space representation for this model is a slight modification of the model given in Section 5.3. The variables labeled p and d in the state vector of that model are now interpreted as the components with asterisks of p_t and d_t. The two measurement errors are also added to the state vector, which is now of dimension 7.

The estimates from this model were intriguing. The estimated autoregressive coefficients for the measurement error processes were very close to one another (0.93 and 0.90), and the correlation between the innovations in the measurement errors was estimated to be 1 (the boundary condition imposed during estimation). This suggests that there is one common factor missing from the original specification of the model. Can this additional common factor be interpreted as measurement error?

Suppose that the entire source of the measurement error is misspecification of the deflator. Let P and D denote the nominal values of stock prices and dividends, let PPI denote the producer price index, and let PC denote the ideal unobserved consumption deflator. We have $p=(P/\text{PPI})$, $d=(D/\text{PPI})$, and assume $p^*=(P/\text{PC})$, and $d^*=(D/\text{PC})$. A Taylor series approximation yields

$$p_t = p_t^* + m_t p_t^* \tag{5.14}$$

$$d_t = d_t^* + m_t d_t^* \tag{5.15}$$

where the measurement error $m_t=(\text{PPI}_t-\text{PC}_t)/\text{PC}_t$. Since the measurement error is common to both equations, this explains the common factor found above. In addition, the measurement error component is multiplied by the time-varying terms p_t^* and d_t^*. This explains the heteroscedasticity found in all of the models.

Before estimating the model using the Kalman filter, a slight modification is necessary. As they now stand, equations (5.14) and (5.15) are nonlinear in the unobservables m_t, p_t^*, and d_t^*. The likelihood function cannot be formed using the Kalman filter. However, an approximation to the model can be estimated. If we replace the terms $m_t p_t^*$ and $m_t d_t^*$ with $m_t p_{t/t-1}^*$ and $m_t d_{t/t-1}^*$, then the model is linear in the unobservables since $p_{t/t-1}^*$ and $d_{t/t-1}^*$ are predetermined functions of the observed data. The state space representation of this model is a slight modification of the basic model used in Section 5.3. The state vector is now

$$X_t = (p_t^* \quad d_t^* \quad d_{t-1}^* \quad d_{t+1}^{*e} \quad p_{t+1}^{*e} \quad m_t)'$$

The first five elements evolve according to the structural model in Section 5.2, and the measurement error evolves independently of the variables with asterisks as an AR(1) process. The measurement error innovation standard deviation will be denoted σ_m, and its autoregressive coefficient will be denoted ρ_m. Finally, the measurement equation is

$$Y_t = H_t X_t$$

with

$$Y_t' = (p_t d_t)$$

and

$$H_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & p^*_{t/t-1} \\ 0 & 1 & 0 & 0 & 0 & d^*_{t/t-1} \end{bmatrix}$$

The estimated parameters for this model are shown in the second column of Table 7.2. The results look sensible and are similar in many respects to the base model. The estimated gross rate of return has changed from 1.040 to 1.043. The estimated value of σ_m of 8.5 percent together with the estimated autoregressive coefficient of 0.78, imply that the standard error of the measurement error process is approximately 13 percent. The inclusion of measurement error improves the fit of the model significantly. The log likelihood has increased from -583 to -577. In Figures 7.3 and 7.4, we have plotted the actual values of p_t and d_t and the corresponding filtered estimates of the values with asterisks, which track the actual data quite closely during most of the sample period. The exception is the period 1974–9. From 1973 to 1975, the stock price index falls from 0.95 to 0.42. The model attributes some share of this 56 percent decline in share prices to measurement error. During this period the values with asterisks fall from 0.98 to 0.56, and the measurement error component must explain the remaining 0.14 drop in the index. This corresponds to a value of m_t of approximately 20 percent during this time period. One must suspect that the measurement error component is also capturing specification error during these years.

Diagnostic checks of the model provide a mixed picture. In this model, the innovations are heteroscedastic, but the predicted time-varying variances calculated by the model can be used to construct a set of normalized innovations. These normalized innovations should be independent white-noise processes with zero mean and unit variance. The normalized price innovation from the model has a mean of -0.08 and a standard deviation of 0.97, very close to the population values of 0 and 1. The corresponding values for dividends was -0.12 and 1.01. There are no significant cross-correlations between the innovations for leads and lags from 0 to 20 and no significant autocorrelation in the price innovations. There was some evidence of misspecification from the dividend innovations. These showed an estimated autocorrelation of -0.29 at lag 2 and -0.21 at lag 12

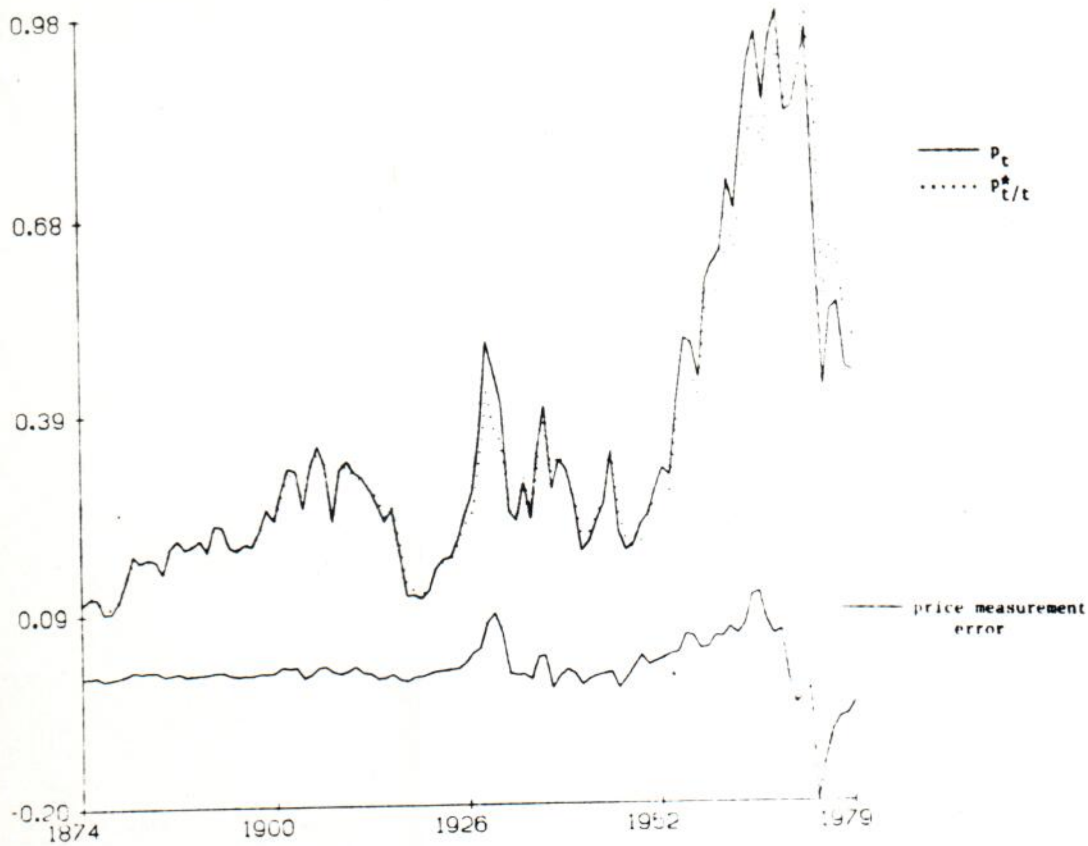

Figure 7.3 Decomposition of price, model 2.

but were otherwise close to zero. Two large correlations out of the 82 calculated does not suggest any serious model misspecification.

While the residuals appear to have the predicted variances and covariances, we also were interested in the performance of the model in capturing all of the heteroscedasticity in the data. If the model specification is correct, the innovations at time t should have a unit variance conditional on any of the data available at time $t-1$. Here the model does not seem to perform well. When we regress the squared, normalized innovation in price at time t on a constant and the squares of $p^*_{t/t-1}$, we find a positive coefficient and an R^2 of 0.13. This suggests that we have not completely captured all the heteroscedasticity in the data.

In the next section, we take an agnostic view on the source of the heteroscedasticity in the model. Rather than attempt to explain it in terms of time-varying interest rates or measurement error, as we have done in the last two sections, we merely attempt to incorporate its persistence in the model. To be specific, we model the innovation in price as a process with autoregressive conditional heteroscedasticity.

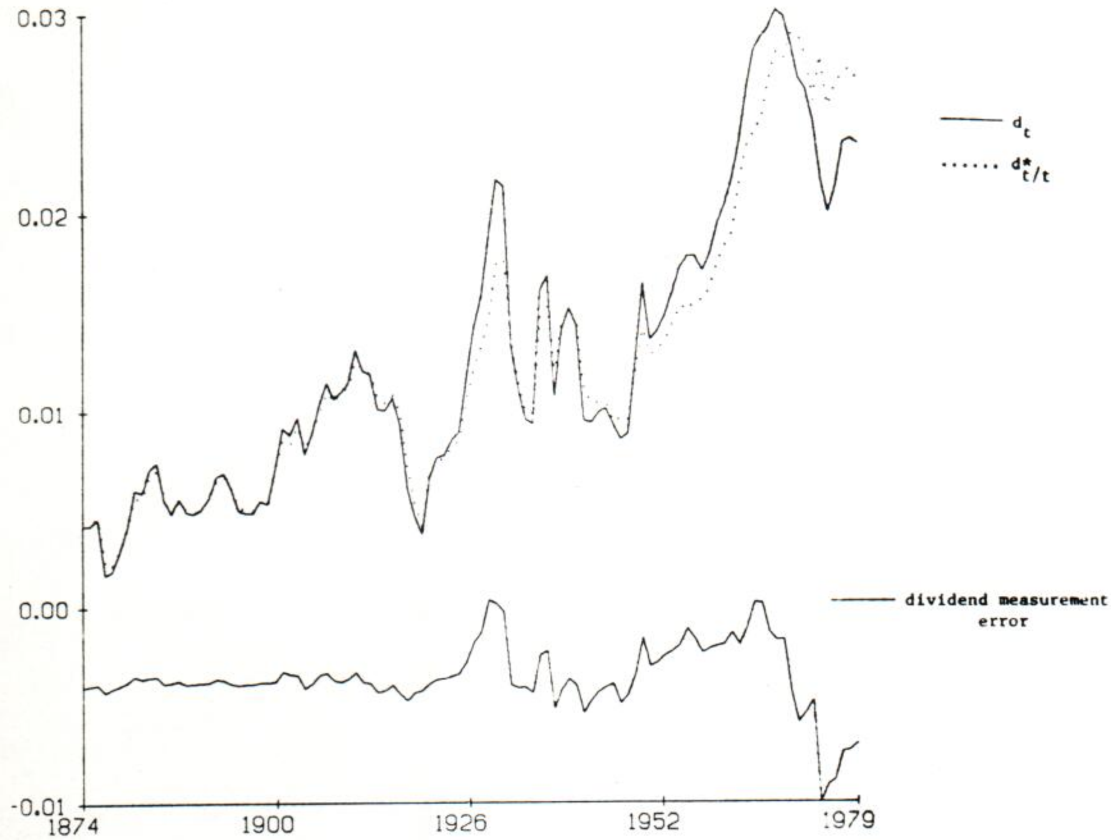

Figure 7.4 Decomposition of dividends, model 2.

5.5 *ARCH*

The heteroscedasticity in price innovations from model 1 appears to be quite persistent. A glance at Figure 7.1, which plots these innovations, should convince the reader of this persistence. The first autocorrelation of the squares of these innovations is 0.37, which suggests significant serial correlation in the variance of the price innovation. There does not seem to be any significant serial correlation in the dividend innovation. The first autocorrelation of the squared dividend innovations is only 0.03. In this section, we will estimate a model that incorporates ARCH in the price but not the dividend innovation.

Because the price and dividend processes are tightly connected by the structural model, some care must be taken in the specification of the ARCH process. Recall that we had normalized the model so that the price innovation was composed of two orthogonal components. The first was πe_t^d (where e^d is the dividend innovation), and the second was u_t. Heteroscedasticity in u_t will lead to heteroscedasticity in the price innovation but not in the dividend innovation. If we let $h_t = \text{var}(u_t \mid \Omega_{t-1})$, then the parameterization that we have chosen for the ARCH process can be written as

$$h_t = \alpha_0 + \rho h_{t-1} + \alpha_1 w_{t-1} \tag{5.16}$$

where

$$w_{t-1} = u_{t-1}^2 - h_{t-1}$$

Since $E(w_t \mid \Omega_{t-1}) = 0$, equation (5.16) can be viewed as an AR(1) process for the variance. The innovation in h_t is w_t, which is the deviation of u_t^2 from its expected value. Solving equation (5.16) yields

$$h_t = \alpha_0(1-e)^{-1} + \alpha_1 \sum_{i=0} \theta^i u_{t-1-i}^2$$

where

$$\theta = \rho - \alpha_1$$

so that h_t can be viewed as the exponential weighted average of the lagged squared errors. The homoscedastic model corresponds to $\alpha_1 = 0$, and the usual ARCH(1) model corresponds to $\rho = \alpha_1$.

The state space form of this model is nearly the same as model 1. The only modification is the incorporation of the time-varying variance for u_t. Since the likelihood value is formed sequentially using the Kalman filter, the incorporation of a time-varying variance is straightforward.

The estimated parameter values for this model are shown in Table 7.2 in the column labeled ARCH. Several results stand out. First, the fit of the model is very good. The value of the log likelihood has increased from -583 for model 1 to -555. The heteroscedasticity seems to be very persistent. The value of ρ is 0.998. Over the sample period, the variance of u_t appears to be a random walk with a small drift.[9] In Figure 7.5, we plot the implied standard deviation of the one-step-ahead forecast error in price. The standard deviation has increased markedly over the sample period. From 1871 to 1974, the standard deviation increased from 0.03 to 0.08. It increased dramatically (to 0.12) following the large unexpected decline in prices over the 1974–6 period.

The diagnostic checks suggested that the model was satisfactory. First, the ARCH process appeared to explain all of the heteroscedasticity. There was no significant autocorrelation in the (normalized) squared innovations. Both the (normalized) price and dividend innovations had sample

[9] One might question the large estimated value of ρ, in light of the reasonably low (0.37) first autocorrelation in the squared price innovations. If we assume that our model for h_t is correct, and that $u_t \mid \Omega_{t-1} \sim N(0, h_t)$, then a few lines of algebra shows that the population first autocorrelation is 0.38. (With these parameter values, however, the fourth moment of the random variable is infinite, so that the sample autocorrelations may be very imprecise estimates.)

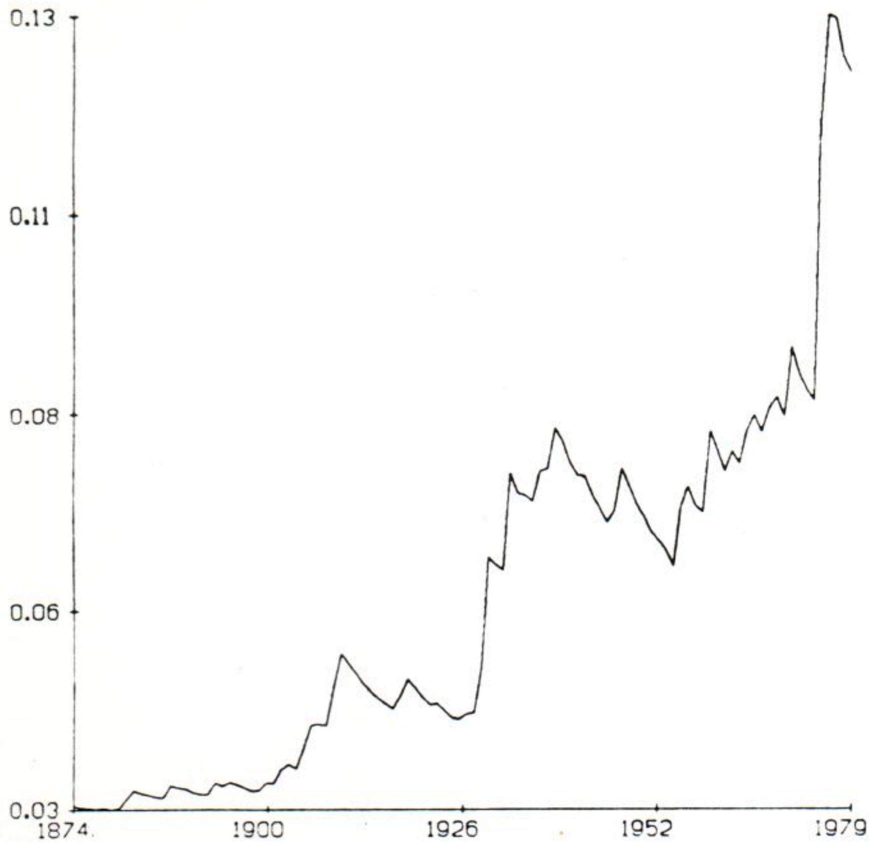

Figure 7.5 Standard deviation of price innovation, ARCH model.

standard deviations close to 1 (1.00 and 1.05). There was one large autocorrelation in the price innovations; the second autocorrelation had a value of 0.2, and there were two large-sample autocorrelations in the dividend innovations; the second and twelfth took on the values -0.28 and -0.2, respectively. Even these would probably lose importance when divided by estimated standard deviations.

What does this exercise suggest about the relationship between prices and dividends? First, the results from model 1 confirm results found elsewhere, that the relationship predicted by the model is not satisfied by the raw data. Second, our time-varying coefficient regression results suggest that the model cannot be salvaged by allowing time-varying interest rates. Our TVP model produced satisfactory statistical results but implied annual real rates of interest varying from -33 to -50 percent. Third, some of the friction between the data and the theory can be explained by the presence of measurement error in the price deflator. Our estimated model provided an improvement in the fit of the model and explained some of the heteroscedasticity in the errors. Our final model, which incorporated ARCH in the price innovations, was very successful in describing the data.

It implied a substantial increase in the variance of price innovations over the sample period.

6 Concluding remarks

At this point, it might be useful to reiterate the comments made by Andrew Harvey at the beginning of the August 1985 session of the World Congress of the Econometric Society. The Kalman filter, in and of itself, is of little interest to econometricians. The state space model that underlies the Kalman filter should be of great interest to econometricians. As our taxonomy in the introduction to this chapter showed, the state space model serves as a unifying model for all of the dynamic linear (and some nonlinear) models used in econometrics. Clearly, the strength of the state space is its flexibility. As the applications presented in the session suggest, this flexibility can profitably be explored by econometricians. Additional applications, utilizing the flexibility of the model, are important areas for future research.

However, one need not search for new applications of the model to find fertile areas for research. As our first application indicated, even the oldest application of the Kalman filter in econometrics – the TVP model – has important avenues open for future research. Of foremost importance is the need for distributional results for estimated parameters in TVP models with unit roots. Additionally, it would be useful to formally investigate the ability of linear TVP models to approximate unknown regression functions. Clearly, the quality of the approximation will depend on the form of the regression function and the time series structure of weakly exogenous variables. One can imagine cases in which the adaptive ability of the TVP model would provide a very good approximation.

The final application presented in this chapter – the use of the state space model and Kalman filter to solve a dynamic linear rational-expectations model – opens up a new direction for the use of the model in economics. As the simple example presented in this chapter indicates, the linear rational-expectations model places very tight constraints on the dynamic interaction of the variables in the model. When these constraints are tested using real data, they are usually rejected. The state space model may serve as a useful tool for loosening these constraints along some dimensions while preserving the basic structure of the rational-expectations model. An example of this is our empirical model, which included measurement error. There, the observed data were modeled as a linear combination of underlying latent variables (the variables with asterisks) that satisfied the rational-expectations model plus some additional noise. This additional noise helped describe the dynamics of the observed data but was of no economic interest. The linear rational-expectations model might still be viewed as an empirical success if it describes "most" of the features of the data and leaves very little to be captured by the additional noise.

Bibliography

Akaike, H. (1969), "Fitting Autoregressive Models for Prediction," *Annals of the Institute of Statistical Mathematics,* **21**, 243–7.

(1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory,* B. N. Petrov and F. Cxaki, eds. Budapest: Akademiai Kiado, pp. 267–81.

Alexander, G. J., P. G. Benson, and C. E. Eger (1982), "Timing Decisions and the Behavior of Mutual Fund Systematic Risk," *Journal of Financial and Quantitative Analysis,* **42**, 579–602.

Anderson, B. D. O., and J. B. Moore (1979), *Optimal Filtering,* Prentice-Hall, Englewood Cliffs, NJ.

Ansley, C. F., and W. E. Wecker (1983), "Extensions and Examples of the Signal Extraction Approach to Regression," in *Applied Time Series Analysis of Economic Data,* A. Zellner, ed., U.S. Bureau of the Census, U.S. Government Printing Office, Washington, DC, pp. 181–92.

Atkinson, A. C. (1981), "Likelihood Ratios, Posterior Odds, and Information Criteria," *Journal of Econometrics,* **16**, 15–20.

Beck, N. (1983), "Time Varying Parameter Regression Models," *American Journal of Political Science,* **27**(3), 557–600.

Blanchard, O. J., and C. M. Kahn (1980), "The Solution of Linear Difference Models under Rational Expectations," *Econometrica,* **48**, 1305–11.

Bos, T., and P. Newbold (1984), "An Empirical Investigation of the Possibility of Stochastic Systematic Risk in the Market Model," *Journal of Business,* **57**, 35–41.

Brown, R. L., J. Durbin, and J. M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships over Time (with Discussion)," *Journal of the Royal Statistical Society, Series B,* **37**, 149–92.

Burmeister, E., and K. D. Wall (1982), "Kalman Filtering Estimation of Unobserved Rational Expectations with an Application to the German Hyperinflation," *Journal of Econometrics,* **20**, 255–84.

Chow, G. C. (1983a), "Random and Changing Coefficient Models," in *Handbook of Econometrics,* Z. Griliches and M. D. Intrilligator, eds., North-Holland, Amsterdam, Chapter 21.

(1983b), *Econometrics,* McGraw-Hill, New York.

Conrad, W., and C. Corrado (1979), "Application of the Kalman Filter to Revisions in Monthly Retail Sales Estimates," *Journal of Economic Dynamics and Control,* **1**, 177–98.

Cooley, T. F., and E. C. Prescott (1973a), "An Adaptive Regression Model," *International Economic Review,* **14**, 364–71.

(1973b), "Tests of an Adaptive Regression Model," *Review of Economics and Statistics,* **55**, 248–56.

(1976), "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica,* **44**, 167–84.

Craven, P., and G. Wahba (1979), "Smoothing Noisy Data with Spline Functions," *Numerische Mathematik*, **31**, 377–403.

Dempster, A., N. M. Laird, and D. B. Rubin (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Dempster, A. P., and A. B. Jonas (1983), "Comments on 'Extensions and Examples of the Signal Extraction Approach to Regression' by Ansley and Wecker," in *Applied Time Series Analysis of Economic Data*, A. Zellner, ed., U.S. Bureau of the Census, U.S. Government Printing Office, Washington, DC, pp. 193–4.

Doan, T., R. Litterman, and C. Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, **3**, 1–100.

Engle, R. F. (1979), "Estimating Structural Models of Seasonality," in *Seasonal Analysis of Economic Time Series*, A. Zellner, ed., Research Report ER-1, U.S. Bureau of the Census, U.S. Government Printing Office, Washington, DC.

(1982a), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, **50**, 987–1008.

(1982b), "A General Approach to Lagrange Multiplier Model Diagnostics," *Journal of Econometrics*, **20**, 83–104.

(1983), "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," *The Handbook of Econometrics*, Vol. 2, Z. Griliches and M. Intrilligator, eds., North-Holland, Amsterdam.

Engle, R. F., S. Brown, and G. Stern (1986), "A Comparison of Adaptive Structural Forecasting Methods for Electricity Sales," Discussion Paper 86-12, University of California, San Diego.

Engle, R. F., and C. W. J. Granger (1987), "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, **55**, 251–76.

Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss (1986), "Nonparametric Regression Estimates of the Relation Between Electricity Sales and Weather," *Journal of the American Statistical Association*, **81**, 310–20.

Engle, R. F., D. Hendry, and J. F. Richard (1983), "Exogeneity," *Econometrica*, **51**, 277–304.

Engle, R. F., D. L. Lilien, and M. Watson (1985), "A DYMIMIC Model of Housing Price Determination," *Journal of Econometrics*, **28**, 307–26.

Engle, R. F., and M. Watson (1980), "A Time Domain Approach to Dynamic Factor Analysis and Mimic Models," *Les Cahiers du Seminaire d'Econometrie*, **22**, 109–22.

(1981), "A One Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association*, **76**, 774–81.

EPRI (1983), *Weather Normalization of Electricity Sales*, EA-3143, Final Report of Project 1922-1, prepared by Cambridge Systematics, Inc., and QUERI, Electric Power Research Institute, Palo Alto, CA.

Garbade, K (1977), "Two Methods of Examining the Stability of Regression Coefficients," *Journal of the American Statistical Association*, **72**, 54–63.

Gersch, W. (1983), "Comments on 'Extensions and Examples of the Signal Extraction Approach to Regression' by Ansley and Wecker," in *Applied Time Series Analysis of Economic Data*, A. Zellner, ed., U.S. Bureau of the Census, U.S. Government Printing Office, Washington, DC, pp. 195–8.

Granger, C. W. J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, **16**, 121–30.

Granger, C. W. J., and R. F. Engle (in press), "Econometric Forecasting – A Brief Survey of Current and Future Techniques," in *Proceedings of Interdisciplinary Forecasting Conference, Boulder, CO;* also UCSD Discussion Paper 84-6.

Hamilton, J. D. (1985), "Uncovering Financial Market Expectations of Inflation," *Journal of Political Economy,* **93**, 1224–41.

Harvey, A. C. (1981a), *The Econometric Analysis of Time Series,* Wiley, New York.

(1981b), *Time Series Models,* Phillip Allan, Oxford.

(1984), "A Unified View of Statistical Forecasting Procedures," *Journal of Forecasting,* **3**, 245–75.

Harvey, A. C., and P. Collier (1977), "Testing for Functional Misspecification in Regression Analysis," *Journal of Econometrics,* **6**, 103–20.

Harvey, A. C., and G. D. A. Phillips (1979), "The Estimation of Regression Models with Autoregressive-Moving Average Disturbances," *Biometrika,* **66**, 49–58.

Harvey, A. C., and P. H. J. Todd (1983), "Forecasting Economic Time Series with Structural and Box–Jenkins Models," *Journal of Business and Economic Statistics,* **1**, 299–315.

Hausman, J. A., and M. W. Watson (1985), "Errors-in-Variables and Seasonal Adjustment," *Journal of the American Statistical Association,* **80**, 531–40.

Howrey, E. P. (1978), "The Use of Preliminary Data in Econometric Forecasting," *Review of Economics and Statistics,* **60**, 193–201.

Johnston, J. (1984), *Econometric Methods,* 3rd ed. McGraw-Hill, New York.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Transactions ASME, Journal of Basic Engineering,* **82**(1), 35–45.

Kalman, R. E., and R. S. Bucy (1961), "New Results in Linear Filtering and Prediction Theory," *Transactions ASME, Journal of Basic Engineering,* **83**(3), 95–103.

Kimeldorf, G. S., and G. Wahba (1970), "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines," *Annals of Mathematical Statistics,* **41**, 495–502.

Laumas, G. S., and Y. P. Mehra (1976), "The Stability of the Demand for Money Function: The Evidence from Quarterly Data," *Review of Economics and Statistics,* **58**, 464–8.

Los, C. A., and C. M. Kell (1985), "Monte Carlo Studies of Econometric Estimators of Evolutionary Parameter Structures," Research Paper No. 8502, Federal Reserve Bank of New York.

Lucas, R. E., Jr. (1976), "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets,* K. Brunner and A. Meltzer, eds., Carnegie-Rochester Conference Series on Public Policy, 1, North-Holland, Amsterdam, pp. 19–46.

Nerlove, M., D. M. Grether, and J. L. Carvalho (1979), *Analysis of Economic Time Series: A Synthesis,* Academic, New York.

Nichols, D. F., and A. R. Pagan (1983), "Varying Coefficient Regression," in *Handbook of Statistics,* Vol. 5, E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, eds., North-Holland, Amsterdam, pp. 413–49.

Ohlson, J., and B. Rosenberg (1982), "Systematic Risk of the CRSP Equal-Weighted Common Stock Index: A History Estimated by Stochastic-Parameter Regression," *Journal of Business,* **55**, 121–45.

Pagan, A. R. (1975), "A Note on the Extraction of Components from Time Series," *Econometrica,* **43**, 163–8.

(1980), "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients," *Journal of Econometrics,* **13**, 341–64.

Rausser, G. C., and P. S. Laumas (1976), "The Stability of the Demand for Money in Canada," *Journal of Monetary Economics,* **2**, 367–80.

Rice, J. A. (1986), "Convergence Rates for Partially Splined Models," *Probability and Statistics Letters,* **4**, 203–8.

Rice, J., and M. Rosenblatt (1983), "Smoothing Splines: Regression, Derivatives and Deconvolution," *Annals of Statistics,* **11**, 141–56.

Rosenberg, B. (1972), "The Estimation of Stationary Stochastic Regression Paramaters Re-examined," *Journal of the American Statistical Association,* **67**, 650–4.

(1973), "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement,* **2**, 461–4.

Schweppe, F. C. (1965), "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Transactions on Information Theory,* **11**, 61–70.

Shibata, R. (1981), "An Optimal Selection of Regression Variables," *Biometrica,* **68**, 45–54.

Shiller, R. J. (1981), "Do Stock Prices Move too Much to Be Justified by Subsequent Changes in Dividends," *American Economic Review,* **71**(3), 421–36.

(1984), "Smoothness Priors and Nonlinear Regression," *Journal of the American Statistical Association,* **79**, 609–15.

Stock, J. (in press), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica.*

Swamy, P. A. V. B., and P. A. Tinsley (1980), "Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients," *Journal of Econometrics,* **12**, 103–42.

Train, K., P. Ignelzi, R. Engle, C. Granger, and R. Ramanathan (1984), "The Billing Cycle and Weather Variables in Models of Electricity Sales," *Energy,* **9**, 1041–7.

Wahba, G. (1983), "A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem," Technical Report No. 712, Department of Statistics, University of Wisconsin.

Wahba, G., and S. Wold (1975), "A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation," *Communications in Statistics,* **4**, 1–18.

Wall, K. (1980), "Generalized Expectations Modeling in Macroeconomics," *Journal of Economic Dynamics and Control,* **2**, 161–84.

Watson, M. W. (1985), "Recursive Solution Methods for Dynamic Linear Rational Expectations Models," Harvard University, manuscript.

(1986), "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics,* **18**, 49–75.

Watson, M., and R. Engle (1983), "The EM Algorithm for Dynamic Factor, MIMIC and Varying Coefficient Models," *Journal of Econometrics,* **23**, 385–400.

(1985), "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative," *Review of Economics and Statistics,* **67(2)**, 341–6.

Watson, M. W., and D. Kraft (1984), "Testing the Interpretation Indices in a Macroeconomic Index Model," *Journal of Monetary Economics,* **13**, 165–81.

Wecker, W., and C. Ansley (1983), "The Signal Extraction Approach to Linear and Nonlinear Regression Problems," *Journal of the American Statistical Association,* **78**, 81–9.

Whiteman, C. H. (1983), *Linear Rational Expectations Models,* University of Minnesota Press, Minneapolis.