Unobserved Components Models in Economics and Finance

THE ROLE OF THE KALMAN FILTER IN TIME SERIES ECONOMETRICS

ANDREW HARVEY
and SIEM JAN KOOPMAN

Economic time series display features such as trend, seasonal, and cycle that we do not observe directly from the data. The cycle is of particular interest to economists as it is a measure of the fluctuations in economic activity. An *unobserved components model* attempts to capture the features of a time series by assuming that they follow stochastic processes that, when put together, yield the observations. The aim of this article is thus to illustrate the use of unobserved components models in economics and finance and to show how they can be used for forecasting and policy making.

Setting up models in terms of components of interest helps in model building; see the discussions in [1] and [2] for a comparison with alternative approaches. A detailed treatment of unobserved components models is given in [3]. The statistical treatment of unobserved components models is based on the state-space form. The unobserved components, which depend on the state vector, are related to the observations by a measurement equation.

The Kalman filter is the basic recursion for estimating the state, and hence the unobserved components, in a linear state-space model (see "Kalman Filter"). The estimates, which are based on current and past observations, can be used to make predictions. Backward recursions yield smoothed estimates of components at each point in time based on past, current, and future observations.

A set of one-step-ahead prediction errors, called innovations, is produced by the Kalman filter. In a Gaussian model, the innovations can be used to construct a likelihood function that can be maximized numerically with respect to unknown parameters in the system; see [4]. Once the parameters are estimated, the innovations can be used to construct test statistics that are designed to assess how well the model fits. The STAMP package [5] embodies a model-building procedure in which test statistics are produced as part of the output.

Digital Object Identifier 10.1109/MCS.2009.934465

Kalman Filter

Consider a linear Gaussian state-space model of which the local level model (1), (2) is a simple example. The Kalman filter is a recursive algorithm that computes the minimum mean square estimate (MMSE) of the state vector based on past observations. In the local level model (1), (2), the Kalman filter evaluates the MMSE of the level, $\tilde{\mu}_{t|t-1} = E(\mu_t|y_1, \ldots, y_{t-1})$, together with its mean square error, for $t = 2, \ldots, T$. The likelihood function of the model can be obtained from the prediction errors $v_t = y_t - \tilde{\mu}_{t|t-1}$ for $t = 2, \ldots, T$. The algorithm can further deal with unknown initial conditions, fixed regression effects, missing values, and forecasting. Related smoothing algorithms have been developed for evaluating MMSEs of μ_t based on all observations $y_1, \ldots, y_T$. When the Gaussian assumption is not valid, the methods produce minimum mean square linear estimators. Reviews of the Kalman filter and related methods are presented in [1] and [3].

We begin by describing the application of state-space methods to two economic time series, the rate of inflation and the national income in the United States. The goal of this article is to show how simple models are able to capture long-term and short-term movements that have an economic interpretation. We then illustrate the modeling of a component that allows seasonal patterns to evolve over time. Multivariate models are considered next. The joint modeling of two or more time series offers several advantages and insights. In particular, common factors in trends and cycles can be identified and estimated. For illustration, we consider a model of inflation and national income.

One of the attractions of the state-space form is that it allows data irregularities to be handled. Data irregularities often occur in economic time series, where observations may be missing, obtained at mixed frequencies, revised, or collected from a survey. We discuss a multivariate example involving the estimation of the underlying change in the level of unemployment.

In the 1990s, rapid developments in computing power led to significant advances in the statistical treatment of nonlinear and non-Gaussian models. Classical and Bayesian approaches moved closer together because both draw on computer-intensive techniques, such as Markov chain Monte Carlo methods and importance sampling (see "Importance Sampling"). In the present article we apply nonlinear models to financial data to demonstrate what they can achieve. For example, the prices of many financial assets depend on the volatility, that is, the variance, of stock returns. Volatility changes over time but is not directly observable from daily returns. We thus use a nonlinear unobserved components model to estimate volatility from the data.

UNIVARIATE LINEAR MODELS

The Rate of Inflation

Informed economic policy draws on knowledge of the growth rates of time series of key indicators. When the time series observations Y_t for $t = 1, \ldots, T$ are expressed in logarithms, the underlying growth rate of the series is the change in the level. The first differences of the time series

Importance Sampling

Importance sampling is a Monte Carlo simulation technique for numerically evaluating integrals. In the context of nonlinear and non-Gaussian state-space analysis, we need to evaluate integrals that appear in formulas for likelihood functions and in conditional expectations. Consider the non-Gaussian density $p(y|\theta) = \prod_{t=1}^{T} p(y_t|\theta_t)$, where $p(y_t|\theta_t)$ is given by (29) and $y = (y'_1, \ldots, y'_T)'$ and $\theta = (\theta'_1, \ldots, \theta'_T)'$. The likelihood function is

$$p(y) = \int p(y, \theta)\, d\theta = \int p(y|\theta) p(\theta)\, d\theta.$$

The Monte Carlo estimator $\hat{p}(y) = M^{-1}\sum_{i=1}^{M} p(y|\theta^i)$, with $\theta^i \sim p(\theta)$ for $i = 1, \ldots, M$ is not efficient since most draws θ^i do not support y. Therefore, we introduce an importance density $g(\theta|y)$ to obtain

$$p(y) = \int \frac{p(y, \theta)}{g(\theta|y)} g(\theta \mid y) d\theta = g(y) \int \frac{p(y, \theta)}{g(y, \theta)} g(\theta|y)\, d\theta.$$

In practice we let $g(\theta|y)$ be Gaussian. The densities $g(y, \theta)$ and $g(y)$ are therefore also Gaussian. For the purpose of numerically evaluating $p(y)$, a Gaussian approximating model relating y to θ is constructed. The evaluation of $g(y)$ as the likelihood function of the approximating model is carried out by the Kalman filter. The importance sampling estimator of the likelihood function is then given by

$$\hat{p}(y) = g(y) M^{-1} \sum_{i=1}^{M} \frac{p(y, \theta^i)}{g(y, \theta^i)},$$

where $\theta^i \sim g(\theta|y)$ is computed using a simulation smoothing algorithm. To obtain an effective importance density, we require an accurate approximation of the nonlinear non-Gaussian model. An approximating model can be obtained by choosing $g(\theta|y)$ such that its first moment equals the mode of $p(\theta|y)$. The details are discussed in [1] and [16].

are defined as the current value Y_t minus the previous value Y_{t-1}, that is, $y_t = Y_t - Y_{t-1}$ for $t = 2, \ldots, T$. The differenced series is a direct measure of the growth rate. Unfortunately, this series may be noisy, as in the case of the first difference of the logarithm of the quarterly price level, which, when multiplied by four, is the annual rate of inflation. The seasonal difference of Y_t, which is defined as the current value minus the value of a year ago, that is, $z_t = Y_t - Y_{t-4}$ for a quarterly time series, yields a more stable measure of the growth rate and is often quoted by government agencies as an indicator of the rate of inflation. However, the seasonal difference is just the sum of four consecutive first differences, that is, $z_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}$, and weighting the observations in this way may be regarded as somewhat arbitrary.

By accounting for the dynamic properties of the series, an unobserved components model provides a method for estimating the underlying rate of inflation. The Gaussian *local level* model for a series y_t consists of a random walk component μ_t plus an irregular term, that is,

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \qquad t = 1, \ldots, T, \tag{1}$$

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \tag{2}$$

where the disturbances ε_t and η_t are mutually independent and $\text{NID}(0, \sigma^2)$ denotes a normally distributed white noise process with mean zero and variance σ^2. When σ_η^2 is zero, the random walk component μ_t becomes constant. The signal-to-noise ratio $q = \sigma_\eta^2/\sigma_\varepsilon^2$ plays a key role in weighting observations for prediction and signal extraction. Specifically, in the steady state, the predictions from the Kalman filter are implicitly formed as an exponentially weighted moving average of past observations, that is,

$$\tilde{y}_{T+1|T} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j y_{T-j}. \tag{3}$$

The weights of past observations are determined by λ, and this parameter is directly related to the signal-to-noise ratio by the formula $\lambda = (-q + \sqrt{q^2 + 4q})/2$ [3]. Note that a higher value of q implies a more rapid discounting of past observations in (3).

Figure 1 shows the rate of inflation in the United States, that is, $y_t = Y_t - Y_{t-1}$, where Y_t is the logarithm of the consumer price index multiplied by 400. The filtered estimates $\hat{\mu}_t$ of μ_t are obtained from a local-level model with q estimated, by maximum likelihood, to be 0.22 for $t = 2, \ldots, T$. The estimates $\hat{\mu}_t$ are also displayed in Figure 1. Let τ be the time index for the first quarter of 1983. The estimate $\hat{\mu}_\tau = 4.4\%$, with a root mean square error (RMSE) of one fifth of $\hat{\mu}_\tau$. On the other hand, the figure for inflation in the first quarter of 1983 is $y_\tau = 3.1\%$. This figure is well below $\hat{\mu}_\tau$, but three quarters later its value is $y_{\tau+3} = 4.6\%$, which is close to $\hat{\mu}_\tau$. These results illustrate the volatile nature of y_t and thus the need to base policy on filtered estimates rather than raw data.

Trend-Cycle Decompositions

The *local linear trend* model generalizes the local level by introducing into (2) a stochastic slope β_t, which itself follows a random walk. Thus,

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \tag{4}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \qquad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2), \tag{5}$$

where the disturbances ε_t, η_t, and ζ_t are mutually independent. If the variances σ_η^2 and σ_ζ^2 are both zero, then the trend is deterministic. When only σ_ζ^2 is zero, the slope is fixed, and the trend reduces to a random walk with drift. Allowing σ_ζ^2 to be positive, but setting σ_η^2 to zero, gives an *integrated random walk* trend, which, when estimated by signal extraction, tends to evolve slowly over time.

The stochastic trend can be combined with a stochastic cycle ψ_t to provide the basis for the trend-cycle decomposition model, that is,

$$y_t = \mu_t + \psi_t + \varepsilon_t, \qquad t = 1, \ldots, T. \tag{6}$$

The stochastic cycle is

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1, \ldots, T, \tag{7}$$

where λ_c is a parameter in the range $0 \le \lambda_c \le \pi$, ρ is a damping factor, and κ_t and κ_t^* are mutually independent Gaussian white noise disturbances with zero mean and common variance σ_κ^2. Assuming that the initial vector $(\psi_0, \psi_0^*)'$ has zero mean and covariance matrix $\sigma_\psi^2 \mathbf{I}_2$ and that $0 \le \rho < 1$, it follows that ψ_t is stationary. The spectrum of the stochastic cycle (7) displays a peak that becomes sharper as ρ moves closer to one; see [3, p. 60]. In the special case when $\sigma_\kappa^2 = 0$ and $\rho = 1$, the cycle is deterministic.

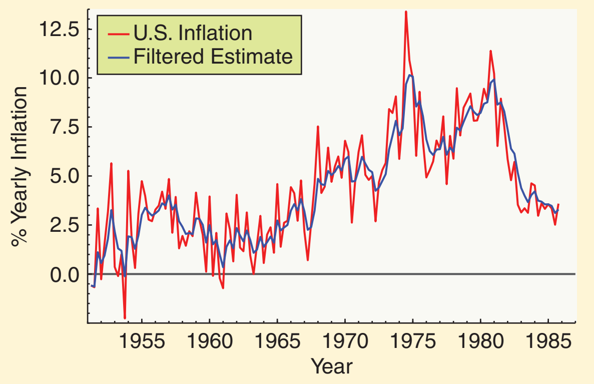

FIGURE 1 Quarterly rate of inflation in the United States from 1951.1 to 1985.4. The filtered estimate of the level μ_t is obtained from the Kalman filter applied to the local level model (1) and (2) with the maximum likelihood estimate $q = \sigma_\eta^2/\sigma_\varepsilon^2$ given as 0.22.

An interesting application of trend-cycle decompositions in economics is to the national income, specifically, the gross domestic product (GDP) of a country. The cycle indicates the extent to which income is above or below the equilibrium, that is, the long-run level, of income. When the economy is below equilibrium, there are unused resources, while, if the economy is above equilibrium, inflationary pressure is likely to occur. The extracted cycle can be regarded as a measure of the *output gap*, which is closely watched by a central bank when it is deciding how to set interest rates.

Fitting a trend-cycle decomposition model (6) provides a basis for identifying turning points in the state of the economy and assessing their significance. Various definitions of turning points can be considered, for example, a change in the sign of the cycle, a change in the sign of the slope, or both together.

Figure 2 shows the trend-cycle decomposition for the logarithm of the quarterly U.S. GDP time series y_t. The unknown parameters, including ρ and λ_c, are estimated by maximum likelihood. Figure 3 shows the annualized underlying growth rate, which is defined as the estimate of the slope times 400. The time series of seasonal differences $z_t = Y_t - Y_{t-4}$ is also shown. This series, which includes the effect of temporary growth emanating from the cycle, is fairly noisy, although it is much smoother than the first difference series $y_t = Y_t - Y_{t-1}$. The growth rate from the model, on the other hand, indicates how the prolonged upswings of the 1960s and 1990s are assigned to the trend rather than to the cycle.

Seasonality

Economic time series recorded at quarterly or monthly intervals typically exhibit seasonal behavior. Over a long period of time, seasonal patterns are likely to change. Unobserved components models can deal with such changes by means of a stochastic seasonal component.

A seasonal component γ_t can be added to a model consisting of a trend and irregular component to give

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \ldots, T. \tag{8}$$

The fixed seasonal pattern can be modeled as

$$\gamma_t = \sum_{j=1}^{s} \gamma_j z_{jt}, \tag{9}$$

where s is the number of seasons, and the dummy variable z_{jt} is one in season j and zero otherwise. The coefficients γ_j, for $j = 1, \ldots, s$, are constrained to sum to zero, so that the seasonal effects are eliminated from annualized observations.

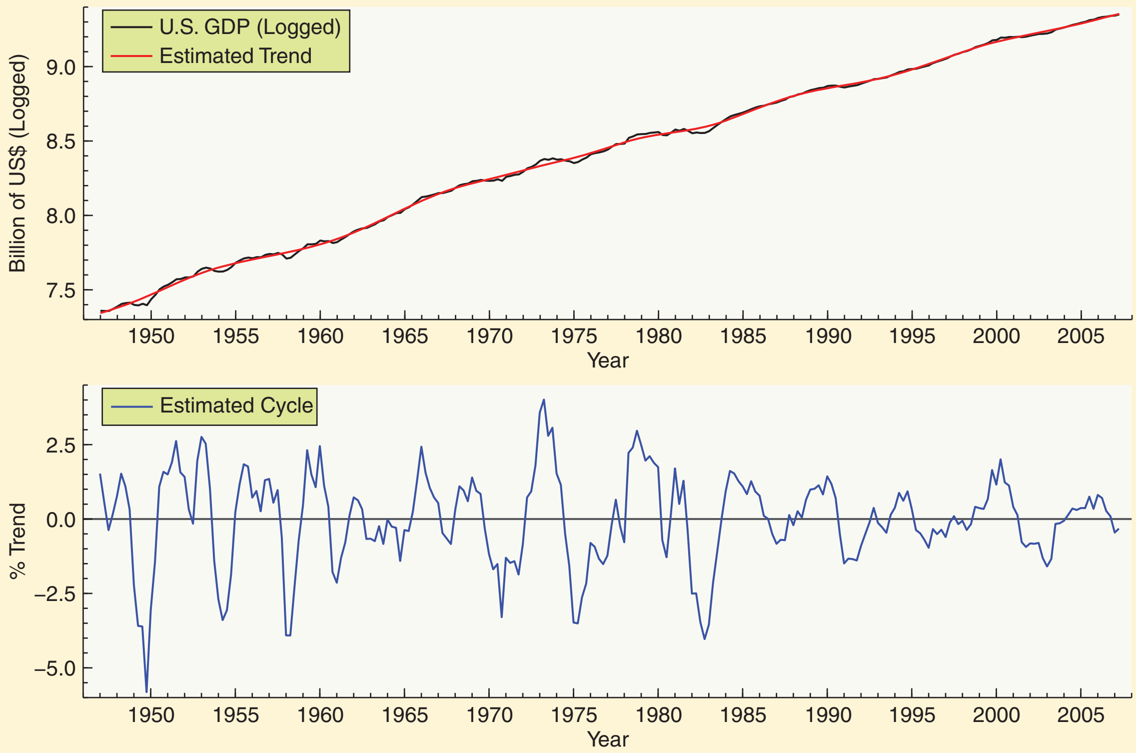

FIGURE 2 The logarithm of the quarterly gross domestic product in the United States from 1947.1 to 2007.2. The smoothed estimates of the trend μ_t and cycle ψ_t are obtained from the Kalman filter and a backward smoothing recursion applied to the trend-cycle model (6) with variances estimated by maximum likelihood.

The goal of this article is to show how simple models are able to capture long-term and short-term movements that have an economic interpretation.

The seasonal pattern can be allowed to change over time as follows. Let γ_{jt} denote the effect of season j at time t and allow the vector $\boldsymbol{\gamma}_t = (\gamma_{1t}, \ldots, \gamma_{st})'$ to evolve as the multivariate random walk

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t, \quad t = 1, \ldots, T, \tag{10}$$

where $\boldsymbol{\omega}_t = (\omega_{1t}, \ldots, \omega_{st})'$ is a zero-mean disturbance vector with

$$\mathrm{Var}(\boldsymbol{\omega}_t) = \sigma_\omega^2(\mathbf{I} - s^{-1}\mathbf{1}\mathbf{1}'), \tag{11}$$

where σ_ω^2 is the variance of each element in $\boldsymbol{\omega}_t$ (scaled by $(s-1)/s$) and $\mathbf{1}$ denotes the $s \times 1$ vector of ones. Although all s seasonal components are continually changing, only one component affects the observation at any particular point in time, that is, $\gamma_t = \gamma_{jt}$ when season j is prevailing at time t. The requirement that the seasonal components always sum to zero is enforced by the restriction that the disturbances sum to zero at each point in time. This restriction is implemented by the correlation structure in (11), where $\mathrm{Var}(\mathbf{1}'\boldsymbol{\omega}_t) = 0$, and coupled with initial conditions requiring that the seasonal components sum to zero at $t = 0$. Specifically, it is assumed that, at $t = 0$, $\boldsymbol{\gamma}_t$ has mean zero and a covariance matrix proportional to (11). Hence $E(\mathbf{1}'\boldsymbol{\gamma}_0) = \mathrm{Var}(\mathbf{1}'\boldsymbol{\gamma}_0) = 0$, and thereafter the fact that $\mathbf{1}'\boldsymbol{\omega}_t = 0$ ensures that $\mathbf{1}'\boldsymbol{\gamma}_t = 0$.

In the *basic structural model*, the time series is decomposed as in (8), where μ_t is the stochastic trend of (4) and (5), γ_t is the stochastic seasonal component of (10) and (11), and the irregular component ε_t is assumed to be random. The disturbances in all three components are mutually independent. The signal-to-noise ratio $q_\omega = \sigma_\omega^2/\sigma_\varepsilon^2$ associated with the seasonal component determines how rapidly the seasonal component changes relative to the irregular component. Figure 4 shows the trend and seasonal components extracted from a quarterly series on the consumption of gas in the United Kingdom by "other final users." The analysis is carried out using the STAMP 8 package [5]. The model is estimated in logarithms, and the smoothed components are then exponentiated. The plot of individual seasonal components shows a sharp change around 1970. This change is due to the availability of inexpensive natural gas from the North Sea, leading to greater use of gas for heating than for cooking. Hence, there is a shift to higher consumption in the winter.

Sometimes the seasonal component is estimated by signal extraction and removed to produce a *seasonally adjusted* series. Seasonally adjusted series are often provided by statistical agencies since the adjusted series are easier for users to interpret. Unobserved components models provide one way of carrying out this task. Figure 4(c) shows the seasonally adjusted gas series.

The time-series analysis of weekly, daily, and hourly observations often requires the treatment of calendar effects arising from weekends, holidays, and moving festivals, such as Easter. Calendar effects can be addressed by using an unobserved components approach. For example, a weekly seasonal pattern is treated in [6] by constructing a parsimonious but flexible model for the U.K. money supply based on time-varying splines and related mechanisms.

MULTIVARIATE LINEAR MODELS

Assuming that observations on N time series are available, the local linear trend model can be generalized to become

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathrm{NID}(0, \Sigma_\varepsilon), \quad t = 1, \ldots, T, \tag{12}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$ and $\boldsymbol{\epsilon}_t$ is the $N \times 1$ irregular vector. The $N \times 1$ trend vector $\boldsymbol{\mu}_t$ depends on the $N \times 1$ disturbance vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{\zeta}_t$. The dynamic specification for the trend component $\boldsymbol{\mu}_t$ is given by

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathrm{NID}(\mathbf{0}, \Sigma_\eta), \tag{13}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim \mathrm{NID}(\mathbf{0}, \Sigma_\zeta), \tag{14}$$

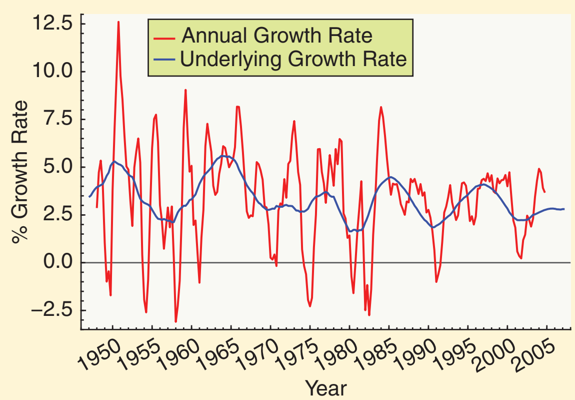

FIGURE 3 Annual differences (×100) of the logarithm of the quarterly gross domestic product in the United States from 1947.1 to 2007.2. The smoothed estimates of the slope β_t in (5) are obtained as described in Figure 2 and multiplied by 400.

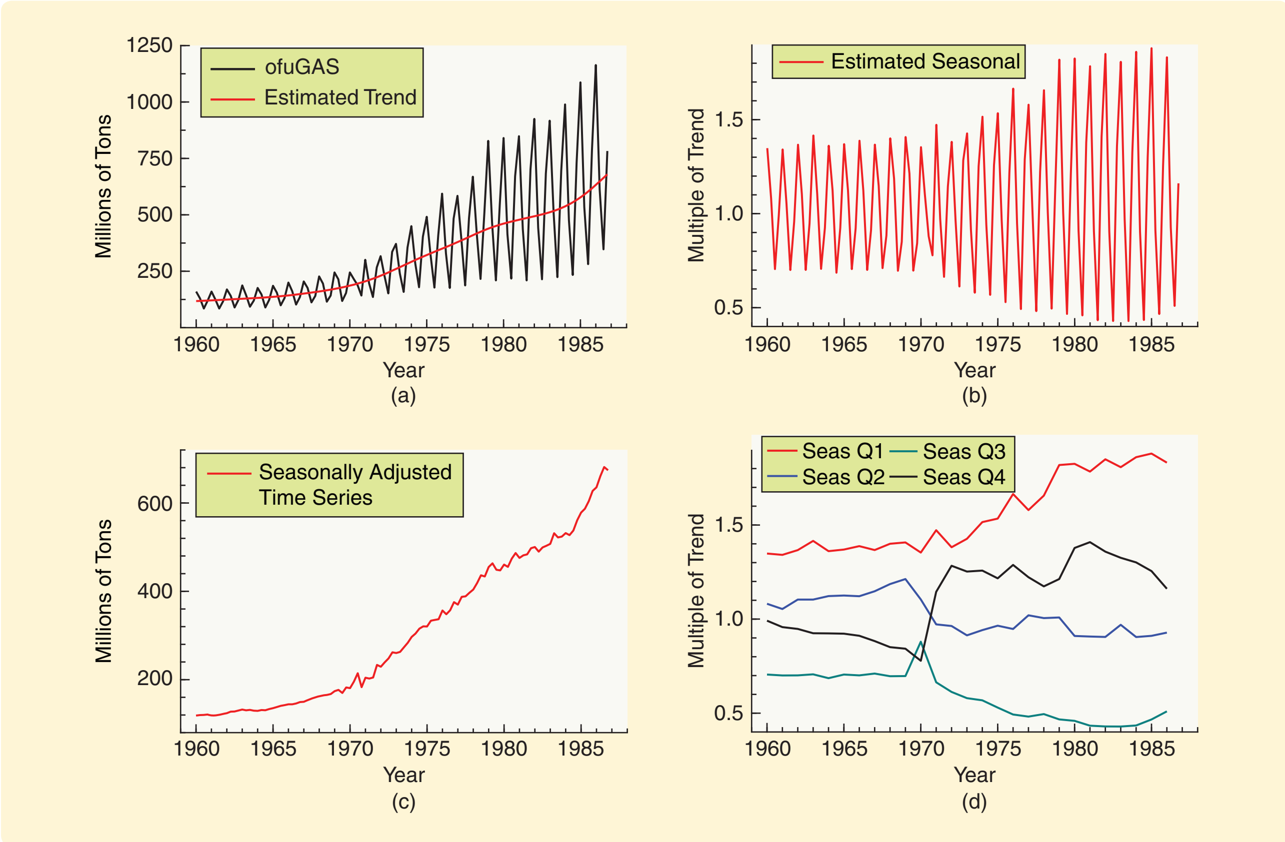

FIGURE 4 Quarterly gas consumption in millions of tons of "other final users" in the United Kingdom from 1960.1 to 1986.4. The smoothed estimates of trend and seasonal are obtained from state-space methods applied to the basic Gaussian structural model (8) for logarithms with parameters estimated by maximum likelihood. The seasonally adjusted time series is defined as the observations with the estimated seasonal component subtracted. The individual estimated seasonal effects for quarters 1, 2, 3, and 4 are displayed in (d).

where $\boldsymbol{\beta}_t$ is the $N \times 1$ vector of slopes. The correlations implied by the covariance matrices Σ_η and Σ_ζ capture the relationships between the permanent, that is, underlying, parts of the series, while the correlations in Σ_ε reflect transitory, that is, short term, associations.

The model (12) can be extended by including a stationary vector autoregressive process to capture short-term interactions between the series. Another possibility is to add a multivariate cycle $\boldsymbol{\psi}_t$ specified as

$$\begin{bmatrix} \boldsymbol{\psi}_t \\ \boldsymbol{\psi}_t^* \end{bmatrix} = \left[\rho\begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \otimes \mathbf{I}_N\right]\begin{bmatrix} \boldsymbol{\psi}_{t-1} \\ \boldsymbol{\psi}_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad (15)$$

where κ_t and κ_t^* are $N \times 1$ vectors of Gaussian disturbances such that $E(\kappa_t \kappa_t') = E(\kappa_t^* \kappa_t^{*'}) = \Sigma_\kappa$, $E(\kappa_t \kappa_t^{*'}) = \mathbf{0}$, and Σ_κ is an $N \times N$ covariance matrix [9], [10]. The operator $\otimes$ in (15) denotes the Kronecker matrix product. Since the parameters ρ and λ_c are the same in all N time series, the cycles have the same spectrum, and thus are referred to as *similar* cycles.

Common Trends and Co-Integration

Important special cases of the trend model in (13), (14) arise when variance matrices Σ_η or Σ_ζ are of less than full rank. To simplify matters suppose that the slopes $\boldsymbol{\beta}_t$ are constant over time ($\Sigma_\zeta = 0$ leading to $\boldsymbol{\beta}_t = \boldsymbol{\beta}_1$ for all t) and that the rank of Σ_η is $K < N$. A model with $\boldsymbol{\mu}_t$ defined as a $K \times 1$ vector of *common trends* can be written as

$$\mathbf{y}_{1t} = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_{1t}, \quad (16)$$
$$\mathbf{y}_{2t} = \boldsymbol{\Pi}\boldsymbol{\mu}_t + \boldsymbol{\gamma} + \boldsymbol{\epsilon}_{2t}, \quad (17)$$

where the $N \times 1$ vector $\mathbf{y}_t$ is partitioned into a $K \times 1$ vector $\mathbf{y}_{1t}$ and an $R \times 1$ vector $\mathbf{y}_{2t}$, and the disturbance vector $\boldsymbol{\epsilon}_t$ is similarly partitioned. Furthermore, $\boldsymbol{\Pi}$ is an $R \times K$ matrix of coefficients, $\boldsymbol{\gamma}$ is an $R \times 1$ vector of constants, and $\boldsymbol{\mu}_t$ follows a multivariate random walk with drift, that is,

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathrm{NID}(\mathbf{0}, \Sigma_\eta), \quad (18)$$

where $\boldsymbol{\eta}_t$ and $\boldsymbol{\beta}$ are $K \times 1$ vectors and Σ_η is a $K \times K$ positive-definite matrix.

The presence of common trends implies co-integration [7], [8]. Two time series are *co-integrated* if each time series must be differenced d times to become stationary, where $d \geq 1$, but there is a linear combination of the time series that is stationary after differencing $d - b$ times, where

State-space methods permit a flexible treatment of unobserved components models.

$d \geq b \geq 1$. The *co-integrating vectors* in (16) are the rows of an $R \times N$ matrix $\mathbf{A}$ with the property that the $R \times 1$ vector $\mathbf{A}\mathbf{y}_t$ is stationary. By partitioning $\mathbf{A}$ as $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2]$, where $\mathbf{A}_1 + \mathbf{A}_2\mathbf{\Pi} = \mathbf{0}$, the common trend system can be transformed to an equivalent co-integrating system by premultiplying $\mathbf{y}_t$ by the $N \times N$ matrix

$$\begin{bmatrix} \mathbf{I}_K & \mathbf{0} \\ \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \tag{19}$$

to give

$$\mathbf{y}_{1t} = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_{1t}, \tag{20}$$

$$\mathbf{A}_1\mathbf{y}_{1t} + \mathbf{A}_2\mathbf{y}_{2t} = \mathbf{A}_2\boldsymbol{\gamma} + \boldsymbol{\epsilon}_t, \tag{21}$$

where $\boldsymbol{\epsilon}_{\mathbf{t}} = \mathbf{A}_1\boldsymbol{\epsilon}_{1t} + \mathbf{A}_2\boldsymbol{\epsilon}_{2t}$. The equations in (21) are the co-integrating relationships, while (20) contains the common trends and generates nonstationary series for $\mathbf{y}_{1t}$. The simplest choice for the matrix $\mathbf{A}$ is $[-\mathbf{\Pi}\ \mathbf{I}_R]$, but alternative matrices may have a more natural economic interpretation.

In the bivariate model

$$y_{1t} = \mu_t + \varepsilon_{1t}, \tag{22}$$

$$y_{2t} = \pi\mu_t + \gamma + \varepsilon_{2t}, \tag{23}$$

where

$$\mu_t = \mu_{t-1} + \beta + \eta_t, \tag{24}$$

for $t = 1, \ldots, T$, the common trend in the two time series y_{1t} and y_{2t} is μ_t. Premultiplying the observation vector by $(-\pi, 1)$ yields the co-integrating relationship

$$y_{2t} = \pi y_{1t} + \gamma + \varepsilon_t, \tag{25}$$

where $\varepsilon_t = \varepsilon_{2t} - \pi\varepsilon_{1t}$. When $\pi = 1$, the trend in y_{2t} is always at a constant distance γ from the trend in y_{1t}. This case is known as *balanced growth*. When each series is modeled in logarithms, the level of the first trend must be multiplied by $\exp(\gamma)$ to give the second trend. Equation (25) tells us that the difference between the two series is stationary.

The original inspiration for the development of co-integration is the observation that national income and consumption tend to move together in the long run, with consumption a fixed proportion of income [7]. The comovement may not be sustained over a long period of time. However, if income and consumption are assumed to have time-varying slopes, as illustrated for income in Figure 3, then an unobserved components model can be adapted so that co-integration can be plausibly made to appear in the slopes but not in the levels. Thus, in (13), Σ_η can be of full rank, whereas Σ_ζ is not. If the slopes in all of the series are the same, then $\Sigma_\zeta = \sigma_\zeta^2\mathbf{1}\mathbf{1}'$ or, alternatively, $\boldsymbol{\beta}_t = \mathbf{1}\beta_t$, where β_t is a scalar random walk with noise variance σ_ζ^2. When this restriction is imposed, the long-run forecasts of the logarithms of the series are parallel straight lines; in other words, the predicted growth rates of consumption and income are the same.

Inflation and the Output Gap

The simple univariate model fitted to inflation in Figure 1 can be extended by including a serially correlated component to better capture short-term movements. One possibility is a stochastic cycle. But what explains the cycle? In the late 1950s the engineer-turned-economist A.W. Phillips noted that wages tended to rise faster in times of low unemployment. The reason is that when labor is scarce, competing employers bid up wages to attract workers. This relationship is known as the Phillips curve [11]. A similar relationship applies to prices and the output gap, leading to the regression model

$$\pi_t = \mu + \beta x_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \quad t = 1, \ldots, T, \tag{26}$$

where π_t is the rate of inflation, x_t is a measure of the output gap, such as detrended GDP, and μ is a constant. Equation (26) started to break down in the mid-1960s as inflation rose and inflationary expectations began to be built into wage negotiations. Economists tried to rescue the Phillips curve in several ways, first by including lagged dependent variables and then by including terms to capture the expectations formed by economic agents who are assumed to process all available information in a rational manner. A popular model, which assumes that inflation depends on the expectation of the next period's inflation together with inflation from the previous time period, is the hybrid new Keynesian Phillips curve [11].

Unobserved components models offer a straightforward way to capture a changing underlying level in inflation by replacing the constant term in (26) by a random walk. Thus,

$$\pi_t = \mu_t + \beta x_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \quad t = 1, \ldots, T, \tag{27}$$

where $\mu_t = \mu_{t-1} + \eta_t$ and $\eta_t \sim \mathrm{NID}(0, \sigma_\eta^2)$. The estimation of (27) can be done in two steps, the first being to extract the output gap from GDP by a trend-cycle decomposition based on (6).

A multivariate solution is to model inflation and GDP jointly as

$$\begin{pmatrix} \pi_t \\ y_t \end{pmatrix} = \begin{pmatrix} \mu_t^\pi \\ \mu_t^y \end{pmatrix} + \begin{pmatrix} \psi_t^\pi \\ \psi_t^y \end{pmatrix} + \begin{pmatrix} \varepsilon_t^\pi \\ \varepsilon_t^y \end{pmatrix}, \quad t = 1, \ldots, T, \quad (28)$$

where μ_t^π is a random walk and μ_t^y is an integrated random walk. These stochastic trends are assumed to be independent of each other. The irregular disturbances ε_t^π and ε_t^y can be correlated, as can the stochastic cycles $\boldsymbol{\psi}_t = (\psi_t^\pi, \psi_t^y)'$, which are "similar cycles" defined in (15). If the cycles are perfectly correlated, then ψ_t^π is proportional to ψ_t^y and, substituting in the inflation equation, which is the first equation in (28), yields an equation that corresponds to (27) when ψ_t^y is set to x_t.

Figure 5 shows smoothed components estimated using the STAMP 8 package [5] with U.S. data from the first quarter of 1986. Since there is almost perfect correlation between the stochastic cycles, the implied equation for π_t is close to (27). Figure 5(a) and (c) shows the effect of the output gap on inflation. The estimated correlation matrix of the cycle yields an estimate of β equal to 0.52. The interpretation is that an output gap of 2% above trend is associated with an annual inflation rate that is 1% above the underlying trend in inflation.

A multivariate unobserved components model linking inflation to unemployment is adopted in [12]. The motivation is similar to the motivation in (28) in that high unemployment implies underused resources and hence a gap between actual and potential output. The specification of the model differs from the specification in (28) in that both trends are random walks, while $\boldsymbol{\psi}_t$ is a vector autoregression.

Data Irregularities

Some of the most striking benefits of unobserved component models become apparent only when more complex problems are considered. Parsimonious multivariate models can be formulated that provide interpretation of the components and insight into the value of, for example, using auxiliary series to improve the efficiency of forecasting a target series. Furthermore, the Kalman filter offers the flexibility needed for dealing with data irregularities, such as missing observations and observations at mixed frequencies. The study reported in [13] on the measurement of British unemployment provides an illustration. The challenge is to obtain timely estimates of the

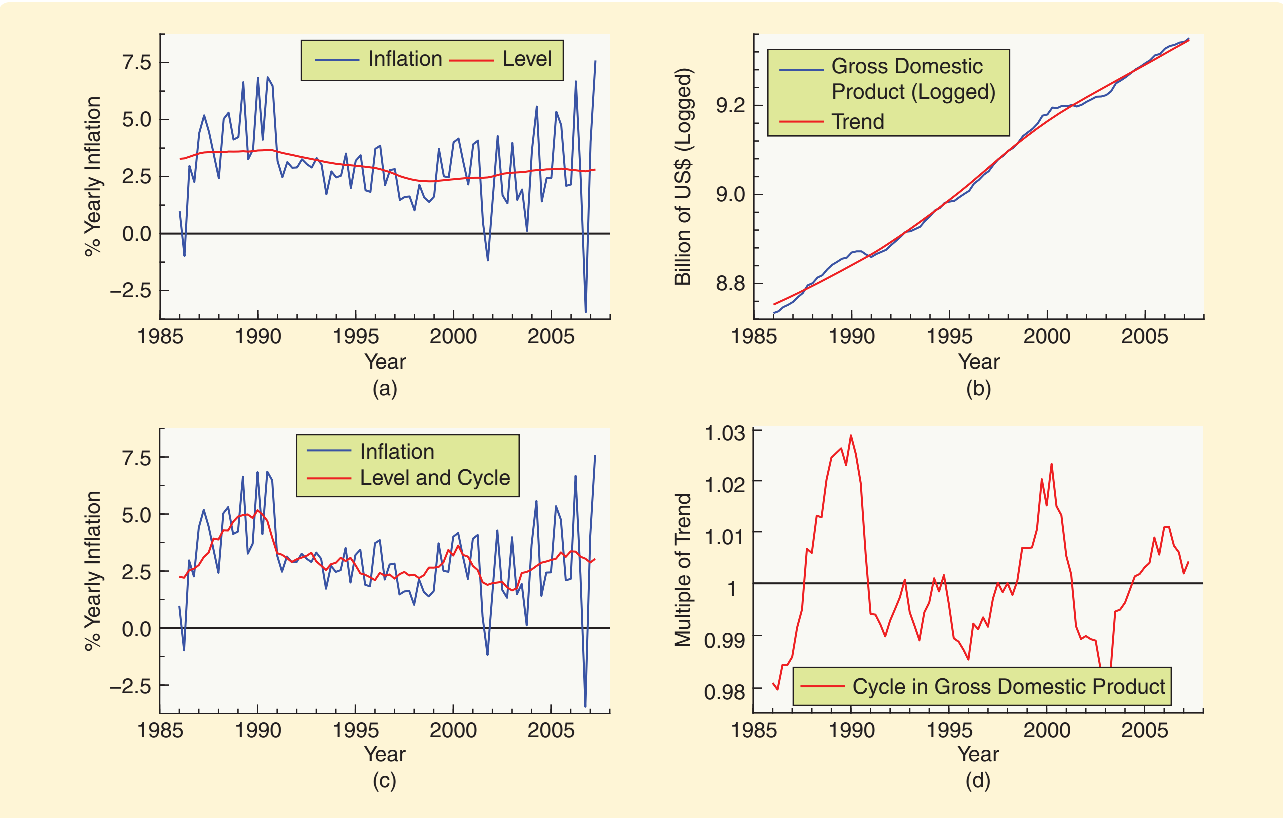

FIGURE 5 Empirical results from a bivariate model fitted to quarterly inflation and the logarithms of gross domestic product in the U.S. from 1985.1 to 2007.2. The rate of inflation is an annual percentage. The smoothed estimates of the trend and cycle components are obtained from state-space methods applied to the bivariate trend-cycle model (28), where the parameters are estimated by maximum likelihood.

underlying change in unemployment. Estimates of the numbers of unemployed according to the International Labor Organization (ILO) definition have been published on a quarterly basis since the spring of 1992. From 1984 to 1991 figures were produced for the spring quarter only. The estimates are obtained from the Labour Force Survey (LFS), which consists of a rotating sample of approximately 60,000 households. Another measure of unemployment, based on administrative sources, is the number of people claiming unemployment benefits. This measure, known as the claimant count, is highly accurate and available monthly. Although the claimant count does not correspond to the ILO definition, Figure 6 shows that it moves roughly in the same way as the LFS figure.

There are thus two issues to be addressed. The first is how to extract the best estimate of the underlying monthly change in a series that is subject to sampling error and that may not have been recorded every month. The second is how to use a related series to improve this estimate. The models constructed in [13] use the state-space form to handle the complicated error structure coming from the rotating sample. Using the claimant count as an auxiliary series in the multivariate model of [13] decreases the root mean square error of the estimator of the underlying change in unemployment by 53.6%.

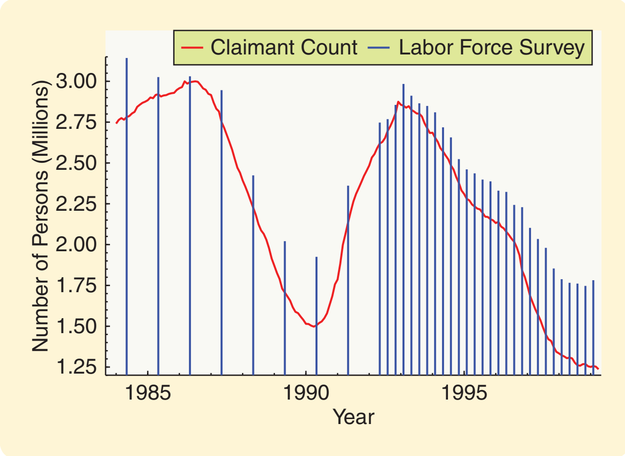

FIGURE 6 British unemployment measures. The vertical lines represent annual observations from 1984 to 1991 and quarterly observations from 1992 to 1999 of the numbers of unemployed obtained from the British labor force survey. The solid line represents the monthly claimant counts (in millions) between 1984 and 1999.

Real-Time Estimation

Tracking the output gap in real time is of interest to economists [14]. Real-time estimation of unobserved components is an exercise in filtering. However, as new observations become available, the estimates can be improved by smoothing. A Bayesian approach [10] has the advantage of giving statistics, such as the probability that the output gap is increasing.

A further complication can arise in real-time estimation, namely, that data are often subject to revisions. The optimal use of different vintages of observations in constructing the best estimate of a series, or the underlying level, can be achieved through the Kalman filter; see [3, pp. 337–341].

NONLINEAR AND NON-GAUSSIAN MODELS

Nonlinearities can be introduced into unobserved components models in various ways. First, the time variation in the measurement and transition equations may depend on past observations. This extension opens up a wide range of possibilities for modeling parameters as function of past observations. The Kalman filter yields optimal estimates for a model that is Gaussian conditional on past observations [15]. Second, nonlinearity arises in an obvious way when the measurement or transition equations have a nonlinear functional form. Finally, a model can be non-Gaussian. For example, the measurement equation may have disturbances generated by a *t*-distribution. More fundamentally, nonnormality may be intrinsic to the observations, as with count data, where the number of events occurring in each time period is recorded. When these numbers are small, a normal approximation is unreasonable, and a model must explicitly account for the fact that the observations are nonnegative integers.

Count Data and Categorical Observations

Nonstationary time series models are designed to deal with series in which the changing mean cannot be captured by observable variables and thus must be modeled by a stochastic mechanism. The unobserved components approach explicitly takes into account the notion that two sources of randomness may be present, one affecting the underlying mean and the other coming from the distribution of the observations around that mean. We therefore aim to develop a model for count data in which the distribution of an observation conditional on the mean is Poisson or negative binomial, while the mean itself evolves as a stochastic process that is always positive. The same ideas can be applied to categorical variables.

The exponential family of distributions contains many of the distributions used for modeling count variables. A general model specification for the $N \times 1$ vector of time series $\mathbf{y}_t$ is implied by the conditional density function

$$p(\mathbf{y}_t \,|\, \boldsymbol{\theta}_t) = \exp\{\mathbf{y}_t' \boldsymbol{\theta}_t - b_t(\boldsymbol{\theta}_t) + c(\mathbf{y}_t)\}, \quad t = 1, \ldots, T, \quad (29)$$

where $\boldsymbol{\theta}_t$ is an $N \times 1$ vector of signals, $b_t(\boldsymbol{\theta}_t)$ is a twice differentiable function of $\boldsymbol{\theta}_t$, and $c(\mathbf{y}_t)$ is a function of $\mathbf{y}_t$ only. The vector $\boldsymbol{\theta}_t$ is related to the mean of the distribution by means of the function $b_t(\boldsymbol{\theta}_t)$, while $\boldsymbol{\theta}_t$ may depend on a state vector that changes over time. Dependence of $\boldsymbol{\theta}_t$ on past observations and explanatory variables can be introduced as well. The statistical treatment of the model is by simulation methods. Markov chain Monte Carlo methods are adopted in [16], while importance sampling and antithetic variables are advocated in [1]. Both techniques can also be applied in a Bayesian framework.

An illustration of an analysis based on (29) is provided in [17], where the quarterly number of defaults in a panel of U.S. firms for various rating classes and industries are modeled over the period 1981–2005. Defaults are regarded as realizations from a binomial distribution, conditional on a systematic, dynamic risk factor that represents the economic and business climate. We have

$$y_{it} \sim \text{Binomial}(p_{it}, k_{it}), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{30}$$

where y_{it} is the number of defaults for U.S. firms in class i and quarter t, k_{it} is the number of firms in class i at the beginning of quarter t, and p_{it} is the associated probability of default. A class can be defined by firms with the same initial rating and in the same industry, but alternative characteristics can also be used. An appropriate link function is given by the signal $\theta_{it} = \log\{p_{it}/(1 - p_{it})\}$, which is modeled by

$$\theta_{it} = \lambda_i + \beta_i f_t, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{31}$$
$$f_t = \phi_1 f_{t-1} + \phi_2 f_{t-2} + \eta_t, \quad t = 1, \ldots, T, \tag{32}$$

where f_t is the systematic risk factor. The constant default intensity is represented by λ_i, and the dependence on the risk factor is measured by β_i. A range of models is studied in [17] for analyzing $N = 364$ different classes of U.S. firms for $T = 101$ quarters. The estimation of λ_i and β_i is carried out by Monte Carlo maximum-likelihood methods. The resulting estimates yield some interesting insights. For example, the extracted systematic factor f_t presented in Figure 7 appears to be subject to cyclical behavior. Furthermore, this study demonstrates that the timing and duration of high-default periods can be characterized by a common systematic factor f_t together with a set of industry-specific factors.

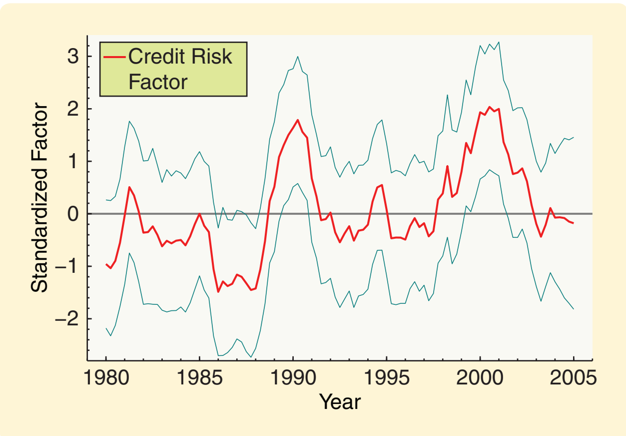

FIGURE 7 The smoothed estimate of the systematic factor f_t. This estimate, which is obtained from the Kalman filter and importance sampling methods applied to the model (30), (31), is based on a panel data set of quarterly defaults for 364 classes of U.S. firms between 1980 and 2005. The estimates are presented with their 95% confidence intervals.

Heavy-Tailed Distributions and Robustness

Simulation techniques are easy to use when the measurement and transition equations are linear but the disturbances are non-Gaussian. Allowing the disturbances to have heavy-tailed distributions provides a robust method for dealing with outliers and structural breaks. While outliers and breaks can be handled by inserting dummy variables after the event, a robust model offers a possible solution for coping with outliers and breaks in future observations.

An example of a time series with breaks and outliers is provided by the quarterly gas consumption in the United Kingdom. The estimated seasonal component from the Gaussian basic structural model, which is presented in Figure 4(b), reveals a rather unappealing wobble at the time North Sea gas was introduced in 1970. When the irregular component in the basic structural model is allowed to follow a t-distribution, the number of degrees of freedom is estimated to be 13; see [1, pp. 233–235]. The robust treatment of the atypical observations in 1970 produces a more satisfactory estimate of the seasonal component.

Stochastic Volatility

Although many financial variables, such as stock returns, are serially uncorrelated over time, their squares are not. Serial correlation in squared observations indicates serial correlation in variance, while time-varying variance has implications for the pricing of options and other financial derivatives. The most common way to model the serial correlation in variance is by means of the generalized autoregressive conditional heteroskedasticity (GARCH) class in which it is assumed that the conditional variance of the observations is an exact function of the squares of past observations and previous variances. An alternative approach is to model the variance as an unobserved component. This approach leads to *stochastic volatility* (SV) models. A review of GARCH and SV models is given in [18]; see also [19].

The SV model has two attractions. The first is that the SV model is the discrete-time analogue of the continuous-time model used in work on option pricing. The second is that the statistical properties of SV models can be determined straightforwardly, and extensions, such as the introduction of seasonal components, are readily handled. The disadvantage with respect to GARCH is that, while GARCH can be estimated relatively easily by maximum likelihood, the SV model requires the use of more computer-intensive methods [1].

The Gaussian discrete-time stochastic volatility model for a de-meaned series of returns y_t can be written as

$$y_t = \sigma_t \varepsilon_t = \sigma \exp(h_t/2)\, \varepsilon_t, \;\; \varepsilon_t \sim \text{NID}(0, 1), \quad t = 1, \ldots, T, \tag{33}$$

where σ is a scale parameter, and the movements in variance are captured by modeling h_t as a stationary first-order autoregressive process, that is,

$$h_{t+1} = \phi h_t + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2). \tag{34}$$

If ε_t and η_t are correlated, the model can pick up leverage effects, the phenomenon whereby an increase in volatility is more likely to be associated with negative returns than positive ones.

Squaring the observations in (33) and taking logarithms yields

$$\log y_t^2 = \omega + h_t + \xi_t, \quad (35)$$

where $\xi_t = \log \varepsilon_t^2 - E(\log \varepsilon_t^2)$ and $\omega = \log \sigma^2 + E(\log \varepsilon_t^2)$, so that ξ_t has zero mean by construction. Taken together, (35) and (34) make up a linear state-space model, but the distribution of the measurement equation noise ξ_t is far from normal. The quasi-maximum-likelihood estimators of the parameters ϕ and σ_η^2 are obtained by treating both ξ_t and η_t as normal random variables. More advanced computational techniques are required if estimation is to be based on the original stochastic volatility specification in (33) and (34) [20].

CONCLUSIONS

State-space methods permit a flexible treatment of unobserved components models. Furthermore, data irregularities such as missing observations are easily handled. For example, irregularly spaced observations can be dealt with since, as discussed in [3, Chap. 3], unobserved components models can be set up in continuous time, and the implied discrete-time state-space form derived.

Current theoretical and empirical research in time series econometrics focuses on non-Gaussian and nonlinear models that follow functional forms suggested by economic and finance theory. For example, many central banks are developing dynamic stochastic general equilibrium models using state-space methods. These models are based on unobserved components, and estimation is by maximum likelihood or Bayesian methods [21].

AUTHOR INFORMATION

Andrew Harvey (andrew.harvey@econ.cam.ac.uk) obtained the B.A. degree in economics and statistics from the University of York in 1968 and the M.Sc. in statistics from the London School of Economics in 1969. He worked for two years in the Central Bureau of Statistics in Kenya before taking up an academic post at the University of Kent at Canterbury. After spending a year at the University of British Columbia, he moved to the Statistics Department at the London School of Economics, where he was a professor of econometrics from 1984 to 1996. In 1996 he became a professor of econometrics in the Faculty of Economics at the University of Cambridge and a Fellow of Corpus Christi College. He has published over 100 academic papers and is the author of two textbooks, *Time Series Models* and *The Econometric Analysis of Time Series*, and of the monograph *Forecasting, Structural Time Series Models and Kalman Filter*. He is a Fellow of the Econometric Society and of the British Academy. He can be contacted at the Faculty of Economics, Cambridge University, Sidgwick Ave., Cambridge CB3 9DD U.K.

Siem Jan Koopman obtained the B.Sc. degree in economics from the University of Amsterdam in 1988 and the Ph.D. in statistics from the London School of Economics (LSE) in 1992. He was employed at LSE as a research officer and lecturer from 1992 to 1997. In 1998 and 1999 he was a research fellow at Tilburg University. In 1999 he became a professor in econometrics at the Vrije Universiteit in Amsterdam. In the same year he became an econometric research coordinator of the Tinbergen Institute. In 2001 he was the ASA/NSF Census research fellow at the U.S. Bureau of the Census in Washington, DC. He has produced more than 70 publications and is a coauthor of *Time Series Analysis by State Space Methods* and *An Introduction to State Space Time Series Analysis*.

REFERENCES

[1] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. Oxford: Oxford Univ. Press, 2001.

[2] A. C. Harvey, "Forecasting with unobserved components time series models," in *The Handbook of Economic Forecasting*, G. Elliot, C. W. J. Granger, and A. Timmermann, Eds. Amsterdam: North-Holland, 2006, pp. 327–412.

[3] A. C. Harvey, *Forecasting, Structural Time Series Models and Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[4] F. Schweppe, "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. Inform. Theory*, vol. 11, pp. 61–70, 1965.

[5] S. J. Koopman, A. C. Harvey, J. A. Doornik, and N. Shephard, *STAMP 8.0 Structural Time Series Analyser, Modeller, and Predictor.* London: Timberlake Consultants Ltd., 2007.

[6] A. C. Harvey, S. J. Koopman, and M. Riani, "The modeling and seasonal adjustment of weekly observations," *J. Bus. Econ. Statist.*, vol. 15, pp. 354–368, 1997.

[7] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, vol. 55, pp. 251–276, 1987.

[8] J. H. Stock and M. W. Watson, "Testing for common trends," *J. Amer. Statist. Soc.*, vol. 83, pp. 1097–1107, 1988.

[9] J. Valle e Azevedo, S. J. Koopman, and A. Rua, "Tracking the business cycle of the Euro area: A multivariate model-based bandpass filter," *J. Bus. Econ. Statist.*, vol. 24, pp. 278–290, 2006.

[10] A. C. Harvey, T. Trimbur, and H. van Dijk, "Trends and cycles in economic time series: A Bayesian approach," *J. Econometrics*, vol. 140, pp. 618–649, 2007.

[11] J. Gali and M. Gertler, "Inflation dynamics: A structural econometric analysis," *J. Monetary Econ.*, vol. 44, pp. 195–222, 1999.

[12] J. Lee and C. R. Nelson, "Expectation horizon and the Phillips curve: The solution to an empirical puzzle," *J. Appl. Econometrics*, vol. 22, pp. 161–178, 2007.

[13] A. C. Harvey and C.-H. Chung, "Estimating the underlying change in unemployment in the UK (with discussion)," *J. R. Statist. Soc. A*, vol. 163, pp. 303–339, 2000.

[14] A. Orphanides and S. van Norden, "The unreliability of output-gap estimates in real time," *Rev. Econ. Statist.*, vol. 84, pp. 569–583, 2002.

[15] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes II: Applications (Transl.: A. B. Aries)*. New York: Springer-Verlag, 1978.

[16] N. Shephard and M. K. Pitt, "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, vol. 84, pp. 653–667, 1997.

[17] S. J. Koopman and A. Lucas, "A non-Gaussian panel time series model for estimating and decomposing default risk," *J. Bus. Econ. Statist.*, vol. 26, pp. 510–525, 2008.

[18] T. G. Andersen, T. Bollersev, T. Christoffersen, and F. X. Diebold, "Volatility and correlation forecasting," in *The Handbook of Economic Forecasting*, G. Elliot, C. W. J. Granger, and A. Timmermann, Eds. Amsterdam: North-Holland, 2006, pp. 777–878.

[19] N. Shephard, *Stochastic Volatility*. Oxford: Oxford Univ. Press, 2005.

[20] S. Kim, N. Shephard, and S. Chib, "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Rev. Econ. Studies*, vol. 65, pp. 361–393, 1998.

[21] F. Canova, *Methods for Applied Macroeconomic Research*. Princeton, NJ: Princeton Univ. Press, 2007.