Statistical Modelling 2010; **10**(**4**): 353–374

The vector innovations structural time series framework: a simple approach to multivariate forecasting

Ashton de Silva[1], Rob J Hyndman[2] and Ralph Snyder[2]
[1]School of Economics, Finance and Marketing, RMIT University, Australia
[2]Department of Econometrics and Business Statistics, Monash University, Australia

Abstract: The vector innovations structural time series framework is proposed as a way of modelling a set of related time series. As with all multivariate approaches, the aim is to exploit potential inter-series dependencies to improve the fit and forecasts. The model is based around an unobserved vector of components representing features such as the level and slope of each time series. Equations that describe the evolution of these components through time are used to represent the inter-temporal dependencies. The approach is illustrated on a bivariate dataset comprising Australian exchange rates of the UK pound and US dollar. The forecasting accuracy of the new modelling framework is compared to other common uni- and multivariate approaches in an experiment using time series from a large macroeconomic database.

Key words: exponential smoothing; forecast comparison; multivariate time series; state space model; vector autoregression; vector innovations structural time series model

Received July 2007; **revised** March 2008; **accepted** November 2008

1 Introduction

A new multivariate time series approach is proposed based on three special cases of the vector innovations state space model (Anderson and Moore, 1979), we call it the vector innovations structural time series (VISTS) framework. The distinctive features of this approach are that its dynamics are expressed in terms of the unobserved components, called the level and growth rate and it has only a single source of error for each series. A theory of estimation and prediction is presented. Moreover, a comparison with other common multivariate time series approaches is made using some common economic time series.

The paper is structured as follows. A brief overview of existing time series methods is given in Section 2. The VISTS framework is introduced in Section 3, where a maximum likelihood estimation method based on exponential smoothing is proposed, instead of the Kalman Filter (Harvey, 1986). In Section 4, the approach is compared

Address for correspondence: Ashton de Silva, School of Economics, Finance and Marketing, RMIT University, VIC 3000, Australia. E-mail: Ashton.desilva@rmit.edu.au

 10.1177/1471082X0901000401

to standard econometric time series techniques using exchange rate data. The results of a forecast accuracy study, which comprises 2000 bivariate time series selected from a large macroeconomic database (Watson, 2003), are presented in Section 5. Finally, we give some brief concluding comments in Section 6.

2 Overview

The autoregressive integrated moving average (ARIMA) class of models were popularized by Box and Jenkins (1970) and have become dominant in time series analysis. They are sufficiently flexible that they can model a wide range of time series characteristics, including stationarity, non-stationarity, seasonality, cyclic and other characteristics.

A characteristic often observed when analyzing time series data (especially macroeconomic variables) is non-stationarity; i.e., changes in the distribution of the process over time. The ARIMA approach to non-stationarity is to 'difference' the data by calculating the difference between successive observations. Sometimes, more than one round of differencing is required in order to obtain a stationary series (with an unchanging distribution over time). Various tests have been developed to determine whether or not a series is non-stationary; two of the most common are the Dickey–Fuller (Dickey and Fuller, 1979) and KPSS (Kwiatkowski *et al.*, 1992) tests. Determining the degree of differencing required is the first step in fitting an ARIMA process.

A univariate ARIMA process can be expressed as

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i} + e_t, \tag{2.1}$$

where y_t denotes the (possibly differenced) observation and e_t denotes an independent and identically distributed disturbance at time t. The parameters ϕ_i and θ_i are estimated (usually by maximum likelihood) subject to some constraints (Hamilton, 1994). The process can be expressed as an ARIMA(p, d, q) model, where d denotes the number of differences required and p and q specify the number of 'autoregressive' and 'moving average' terms in (2.1), respectively. There are automatic identification procedures for determining the values of p and q (e.g., see Brockwell and Davis, 1991, 301–05).

The multivariate version of the ARIMA process is

$$\nabla^d \boldsymbol{y}_t = \sum_{i=1}^{p} \Phi_i \boldsymbol{y}_{t-i} + \sum_{i=1}^{q} \Theta_i \boldsymbol{e}_{t-i} + \boldsymbol{e}_t, \tag{2.2}$$

where $\boldsymbol{y}_t$ denotes the observation at time t, ∇ is the differencing operator, the bolded characters are N-vectors and the coefficient matrices are of dimension $N \times N$, where N denotes the number of individual time series.

The multivariate specification in equation (2.2) is commonly referred to as a vector autoregressive integrated moving average (VARIMA) process. It was developed, in part, by Sims (1980). A fully specified VARIMA specification has been difficult to develop, due to identification issues (Hannan, 1969; Tiao and Tsay, 1989). Consequently, the special case of a vector autoregression (VAR) model (with $q = 0$) is most often applied. However, recent developments by Athanasopoulos and Vahid (2008) have made the full VARIMA model more tractable.

One of the main criticisms of ARIMA modelling is that it does not provide any insight into the data generating process. An alternative modelling framework that is designed to provide such insight is the decomposition approach, which is based on the premise that an observation is the aggregate of a set of latent components. For example

$$y_t = T_t + C_t + S_t + I_t,$$

where T_t, C_t, S_t and I_t denote the trend, cyclic, seasonal and irregular components at time t, respectively. The state space time series specification is one example of this approach.

We believe that the state space approach is more general, flexible and transparent than the ARIMA alternative. For example, it can be generalized to include more than one variable, or a collection of exogenous variables, in a relatively simple manner. In addition, the framework is able to handle data irregularities such as missing observations simply. Finally, the latent components of the series are estimated directly, providing the analyst with a clear breakdown of how each component contributes to the observed series. The advantages of the state space approach are discussed in more detail by Durbin and Koopman (2001: 51–53).

A linear state space model can be written as

$$y_t = \boldsymbol{H}\boldsymbol{x}_{t-1} + \epsilon_t, \tag{2.3}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t, \tag{2.4}$$

where $\boldsymbol{x}_t$ denotes a vector of unobserved components. Equation (2.3) is called the measurement or observation equation and equation (2.4) is called the transition or state equation. The matrices $\boldsymbol{H}$ and $\boldsymbol{F}$ are coefficient matrices that are either wholly or partially pre-determined. The disturbance terms (ϵ_t and $\boldsymbol{\eta}_t$) are assumed to be normally distributed and are usually assumed to be independent of each another.

One popular special case of state space models is the class of structural models. For these models, the vector $\boldsymbol{x}_t$ denotes a vector of states corresponding to the trend, cycle and seasonal components (see Harvey, 1989, Chapter 2).

The first appearance of a multivariate structural time series formulation was in Jones (1966). It was subsequently investigated by Enns *et al.* (1982); Harvey (1986), Harvey (1989) and Pfeffermann and Allon (1989), all of whom adopt the assumption that the disturbances are independent. We shall refer to this as the multiple source of error (MSOE) structural model.

In contrast, we assume that the same disturbance appears in both equations. This was first proposed by Anderson and Moore (1979) and is known as the innovations

form of state space models. However, it was not until the publication of the paper by Snyder (1985) that the link between exponential smoothing (Brown, 1959) and the innovations specification became apparent. Moreover, Snyder (1985) and Ord *et al.* (1997) argue that innovations structural models are particularly important because they provide a statistical foundation for the linear versions of exponential smoothing.

It is often thought that the MSOE models provide a better modelling framework than the innovations models, because the multiple sources of error appear to allow greater generality. In fact, in the univariate context, the additional disturbances associated with the unobserved components are essentially redundant (Anderson and Moore, 1979; Hannan and Deistler, 1988), and equivalent results can be obtained using a corresponding innovations structural model (Hyndman *et al.*, 2008).

We believe that our proposed approach is an important development, as it provides a useful alternative multivariate time series specification. One important advantage of this new framework is that it can be implemented using a standard maximum likelihood estimation procedure. This is unlike the MSOE specification, which requires the Kalman Filter. In addition, it is not characterized by the difficulties associated with the identification of an appropriate structure, as the VARIMA alternative is (Harvey and Koopman, 1997; Hyndman, 2001).

3 Vector innovations structural framework

In this section, the VISTS model is introduced and compared with other common multivariate approaches.

The key feature of the proposed framework, as of all structural time series models, is that it allows the unobserved components of a time series to change randomly over time. Specifically, it is assumed that the random N-vector of observations $\boldsymbol{y}_t$ is a linear function of a k-vector of unobserved components $\boldsymbol{x}_{t-1}$ plus error. This linear relationship, called the *measurement equation*, is

$$\boldsymbol{y}_t = \boldsymbol{H}\boldsymbol{x}_{t-1} + \boldsymbol{e}_t,$$

where $\boldsymbol{H}$, the so-called *structure matrix*, is fixed, with elements which are normally 1's and zeros. The term $\boldsymbol{H}\boldsymbol{x}_{t-1}$ encapsulates the effect of the history of the process; the innovations $\boldsymbol{e}_t$, on the other hand, represent the effects of new forces that are at work on the series. The reason for the lag in the index of the components vector is that the latter is deemed to represent the state of the process at the beginning of period t, i.e., at time $t-1$.

The innovations $\{\boldsymbol{e}_t\}$ are inter-temporally uncorrelated and are governed by a common $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ distribution. In this paper, it is assumed that the variance matrix $\boldsymbol{\Sigma}$ is diagonal, meaning that contemporaneous innovations are also uncorrelated. The diagonal elements of $\boldsymbol{\Sigma}$ are typically unknown.

Restricting $\boldsymbol{\Sigma}$ to be diagonal has intuitive appeal for the following reasons. First, it is more consistent with the spirit of state space modelling, in that the relationships are modelled by the state equations only. Second, when we initially applied these

models, we allowed $\boldsymbol{\Sigma}$ to be unrestricted; when we subsequently constrained it to be diagonal, we found that there was no deterioration in the forecasting performance.

The evolution of the unobserved components is governed by the first-order Markovian relationship

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{G}\boldsymbol{e}_t.$$

This is called the *transition equation*. The fixed $k \times k$ matrix $\boldsymbol{F}$ is referred to as the *transition matrix*; its elements are also typically zeros and 1's, but occasionally they may be unknown damping factors. The $k \times N$ matrix $\boldsymbol{G}$ is the *impact* matrix. Many of its elements are typically unknown; they determine the effects of the innovations on the process beyond the period in which they occur. When $\boldsymbol{G} = \boldsymbol{0}$, the innovations have no impact on the components of the time series—any change in the components is deterministic, and hence completely predictable. When $\boldsymbol{G}$ is diagonal with non-zero diagonal elements, each innovation has a persistent effect on its own series, but no effect on the other series. When $\boldsymbol{G}$ has non-zero off-diagonal elements, an innovation will then have an impact on other series as well as its own.

3.1 Special cases of the innovations structural framework

For each of the special cases considered, the general structure of the ith series can be written as:

$$y_{i,t} = \boldsymbol{h}_i'\boldsymbol{x}_{i,t-1} + e_{i,t},$$
$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{G}\boldsymbol{e}_t,$$

where $y_{i,t}$, $\boldsymbol{h}_i$, $\boldsymbol{x}_{i,t-1}$ and $e_{i,t}$ represent suitably commensurate sub-vectors and sub-matrices of the vectors and matrices in the general form. The special case where the off-diagonals of $\boldsymbol{G}$ (and, when applicable, $\boldsymbol{F}$ also) are zero corresponds to the situation where there are no inter-series dependencies, and where the framework reduces to N univariate innovations structural models. Thus, the particular form of inter-series dependence allowed for in this paper occurs when at least some of the $g_{ij} \neq 0$ for $i \neq j$.

We now discuss three particular cases of this model that are especially useful.

3.1.1 Vector local level model

The simplest univariate innovations structural model relies on a single unobserved component called the local level, which follows a random walk (RW) over time. Its vector analogue, which has a random N-vector $\boldsymbol{\ell}_t$ of levels for the N-series in period t, is

$$\boldsymbol{y}_t = \boldsymbol{\ell}_{t-1} + \boldsymbol{e}_t, \tag{3.1}$$
$$\boldsymbol{\ell}_t = \boldsymbol{\ell}_{t-1} + \boldsymbol{A}\boldsymbol{e}_t, \tag{3.2}$$

where $\boldsymbol{A}$ is a matrix of impact parameters designated by α_{ij}. Interdependencies between the series are reflected by non-zero off-diagonal elements in the matrix $\boldsymbol{A}$.

A more traditional perspective of the model is obtained from its reduced form. By first differencing equation (3.1) and using equation (3.2) to eliminate the levels, the model can be written as a VARIMA(0,1,1) model $\nabla \boldsymbol{y}_t = \boldsymbol{\Theta} \boldsymbol{e}_{t-1} + \boldsymbol{e}_t$, where $\boldsymbol{\Theta} = \boldsymbol{A} - \boldsymbol{I}$ and $\boldsymbol{I}$ is an identity matrix. A unique value of $\boldsymbol{\Theta}$ is associated with a given matrix $\boldsymbol{A}$, and vice versa. Therefore, the vector local level (VLL) model (equations (3.1) and (3.2)) is equivalent to a VARIMA(0,1,1) model. This suggests the possibility of a close relationship between the vector innovations structural models and VARIMA models in general.

3.1.2 Vector local trend model

The local levels in the vector innovations local level model can be augmented by a random N-vector of growth rates, $\boldsymbol{b}_t$, to give the vector innovations local trend model

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\ell}_{t-1} + \boldsymbol{b}_{t-1} + \boldsymbol{e}_t, \\ \boldsymbol{\ell}_t &= \boldsymbol{\ell}_{t-1} + \boldsymbol{b}_{t-1} + \boldsymbol{A}\boldsymbol{e}_t, \\ \boldsymbol{b}_t &= \boldsymbol{b}_{t-1} + \boldsymbol{B}\boldsymbol{e}_t, \end{aligned}$$

where the element β_{ij} in $\boldsymbol{B}$ is the effect of the jth innovation on the growth of series i.

The reduced form is found by double differencing the measurement equation, and then using the transition equations to eliminate the levels and growth rates, to give the VARIMA(0,2,2) model

$$\nabla^2 \boldsymbol{y}_t = \boldsymbol{\Theta}_2 \boldsymbol{e}_{t-2} + \boldsymbol{\Theta}_1 \boldsymbol{e}_{t-1} + \boldsymbol{e}_t,$$

where $\boldsymbol{\Theta}_1 = \boldsymbol{A} + \boldsymbol{B} - 2\boldsymbol{I}$ and $\boldsymbol{\Theta}_2 = \boldsymbol{I} - \boldsymbol{A}$. Again, given both $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are uniquely determined, and vice versa. The vector local trend and VARIMA(0,2,2) models are equivalent.

3.1.3 Vector damped local trend model

In practice, the growth rate may be more appropriately modelled as a stationary process rather than a RW (Gardner and McKenzie, 1985). The growth equation in the vector local trend (VLT) model can be modified to incorporate damping factors. It is referred to as the vector damped local trend (VDLT) model. The revised model takes the form

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\ell}_{t-1} + \boldsymbol{b}_{t-1} + \boldsymbol{e}_t, \\ \boldsymbol{\ell}_t &= \boldsymbol{\ell}_{t-1} + \boldsymbol{b}_{t-1} + \boldsymbol{A}\boldsymbol{e}_t, \\ \boldsymbol{b}_t &= \boldsymbol{\Phi}\boldsymbol{b}_{t-1} + \boldsymbol{B}\boldsymbol{e}_t, \end{aligned}$$

where $\boldsymbol{\Phi}$ is a diagonal matrix formed from the damping factors. Its reduced form is a VARIMA(1,1,2) model

$$\boldsymbol{z}_t = \boldsymbol{\Phi}\boldsymbol{z}_{t-1} + \boldsymbol{\Theta}_2 \boldsymbol{e}_{t-2} + \boldsymbol{\Theta}_1 \boldsymbol{e}_{t-1} + \boldsymbol{e}_t,$$

where $z_t = y_t - y_{t-1}$, $\Theta_2 = A + B - I - \Phi$ and $\Theta_1 = \Phi(I - A)$. Provided that Φ is non-singular, unique values of Θ_1 and Θ_2 can be determined for given values of A and B, and vice versa. The VDLT and VARIMA(1,1,2) models are equivalent.

3.2 Comparison with common alternatives

The conventional multivariate structural time series specification (Harvey, 1989) has multiple sources of randomness for each series. It takes the form:

$$y_t = \bar{H}x_t + u_t,$$
$$x_t = Fx_{t-1} + v_t,$$

where the N-vector u_t and the k-vector v_t are independent disturbances that act as $N + k$ primary sources of randomness. Unlike the innovations form, the unobserved components vector is not lagged in the measurement equation. Typically, the structure matrices of the two models are related by $H = \bar{H}F$. The disturbance vectors are contemporaneously and inter-temporally uncorrelated.

We now compare the different specifications on a case-by-case basis and show that the MSOE models are equivalent to restricted forms of the innovations models presented in the previous section. We begin with the local level case, which in MSOE form is

$$y_t = \ell_t + u_t \qquad u_t \sim N(0, R),$$
$$\ell_t = \ell_{t-1} + v_t \quad v_t \sim N(0, Q).$$

Although there are some close parallels with the innovations form, the links are not as direct. The levels can be eliminated, to give the reduced form $\nabla y_t = u_t - u_{t-1} + v_t$. The right-hand side of this reduced form is the sum of two moving average processes. According to the Granger–Newbold (Granger and Newbold, 1986) theorem, the sum of moving average processes is itself a moving average process. Moreover, the auto-covariance function of the sum is the sum of the auto-covariances of the component moving average processes. Thus, this reduced form is also a VARIMA(0,1,1) process.

A comparison of the auto-covariance structure, however, shows that the MSOE local level model is less general than the VARIMA (0,1,1) model. Specifically, the first-order auto-covariance of the MSOE formulation is

$$\mathrm{E}(\nabla y_t, \nabla y_{t-1}) = R,$$

the variance matrix of the transition equation, which is, by definition, a symmetric positive definite matrix. Therefore, the first-order auto-covariance must be positive. In contrast, the first-order auto-covariance of the VARIMA (0,1,1) model is $\Theta\Sigma$, where the coefficient matrix Θ is not constrained to be symmetric positive definite.

A similar conclusion is reached when comparing local trend models. The MSOE VLT model is

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\ell}_t + \boldsymbol{u}_t, \\ \boldsymbol{\ell}_t &= \boldsymbol{\ell}_{t-1} + \boldsymbol{b}_{t-1} + \boldsymbol{v}_t, \\ \boldsymbol{b}_t &= \boldsymbol{b}_{t-1} + \boldsymbol{w}_t, \end{aligned}$$

where $\boldsymbol{w}_t$ denotes an N-vector of disturbances. This can be reduced to an equivalent VARIMA(0,2,2) model $\nabla^2 \boldsymbol{y}_t = \boldsymbol{w}_t + (\boldsymbol{v}_t - \boldsymbol{v}_{t-1}) + (\boldsymbol{u}_t - 2\boldsymbol{u}_{t-1} + \boldsymbol{u}_{t-1})$. Again, using the Granger–Newbold addition theorem, it can readily be established that the first-order auto-covariance is always non-positive, and that the second-order auto-covariance is always positive. It is therefore equivalent to a restricted VARIMA(0,2,2) model. Given that the reduced form of the innovations vector local trend is not restricted, it can be concluded that the innovations local trend model is more general than its MSOE counterpart.

Once again, we arrive at the same conclusion when we consider the damped local trend model. The MSOE form of the VDLT model is:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\ell}_t + \boldsymbol{u}_t, \\ \boldsymbol{\ell}_t &= \boldsymbol{\ell}_{t-1} + \boldsymbol{b}_{t-1} + \boldsymbol{v}_t, \\ \boldsymbol{b}_t &= \boldsymbol{\Phi}\boldsymbol{b}_{t-1} + \boldsymbol{w}_t. \end{aligned}$$

It is straightforward to show that this model can be reduced to a *restricted* VARIMA(1,1,2) model. Given that the reduced form of the innovations VLT model is not restricted, it can be concluded that the innovations damped local trend model is more general than its MSOE counterpart.

In general, in can be concluded that for any MSOE model, there is an equivalent innovations form, but not vice versa (see Hyndman *et al*., 2008, Chapter 13). Therefore, the MSOE model is more restrictive. Specifically, the MSOE model is more restrictive in terms of parameter space and dimension (see Harvey, 1989: 432).

The other common alternative, the VARIMA model, has the general form

$$\boldsymbol{\Phi}(L)\boldsymbol{z}_t = \boldsymbol{\Theta}(L)\boldsymbol{e}_t, \tag{3.3}$$

where L is a lag operator, $\boldsymbol{z}_t = (1 - L)^d \boldsymbol{y}_t$ and $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ are matrix polynomial functions of the lag operator satisfying the usual stationarity and invertibility conditions. The relationship between this and the special cases of the VISTS framework has already been outlined. It is worthwhile noting that $\boldsymbol{e}_t$ is an innovations vector that corresponds to the innovations vector used in the innovations structural model. Despite the close links, the frameworks differ, in that equation (3.3) contains no unobserved components, giving the VISTS framework an interpretative advantage.

An interesting issue arises when the variables under consideration correspond to different degrees of non-stationarity. We do not explicitly deal with this issue in this paper. However, it can be argued that models such as the vector local trend (both SSOE and MSOE versions) imply two unit roots in each series, and if a series

genuinely has a unit root, its part of the model should automatically collapse back to something resembling a local level model (which implies one unit root). Therefore, there is some scope in our framework for the automatic handling of series with mixed levels of non-stationarity.

Interestingly, Hyndman (2001) shows that although the (univariate) linear innovations models considered in the paper are mathematically equivalent to selected ARIMA processes, the innovations framework produces better forecasts. Naturally, this would be expected to hold for the multivariate case also.

3.3 Estimation

The matrices $\boldsymbol{H}$, $\boldsymbol{F}$, $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$ in the vector innovations state space model potentially depend on a vector of unknown parameters designated by $\boldsymbol{\theta}$. We outline a maximum likelihood procedure for estimating $\boldsymbol{\theta}$. The development of this procedure is hampered by the existence of non-stationary states, which imply that the variances of some of the elements of $\boldsymbol{x}_0$ are infinite, so that the Gaussian density of the sample $\boldsymbol{y}_1, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ degenerates to zero through the entire sample space. A common strategy, when all the states are non-stationary, is to redefine the likelihood function in terms of the conditional density

$$p(\boldsymbol{y}_{k+1}, \boldsymbol{y}_{k+2}, \ldots, \boldsymbol{y}_T | \boldsymbol{\theta}, \boldsymbol{y}_1, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k). \tag{3.4}$$

A second possibility is to condition on a fixed but unknown value of $\boldsymbol{x}_0$, rather than on the initial series values; in other words, use the conditional Gaussian density

$$p(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T | \boldsymbol{\theta}, \boldsymbol{x}_0). \tag{3.5}$$

It transpires that the pursuit of the first conditional likelihood eventually leads to the need for an augmented Kalman Filter (Ansley and Kohn, 1985; de Jong, 1991) to evaluate the likelihood function. In the second conditional likelihood, the elements of $\boldsymbol{x}_0$ effectively become parameters. Then, the conditional likelihood, viewed as a function of $\boldsymbol{\theta}$ and $\boldsymbol{x}_0$, can be represented by $L(\boldsymbol{\theta}, \boldsymbol{x}_0 | \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T) = p(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T | \boldsymbol{\theta}, \boldsymbol{x}_0)$. Using conventional conditional probability theory, this version of the conditional likelihood function can be written as the product of the one-step-ahead prediction distributions as follows:

$$L(\boldsymbol{\theta}, \boldsymbol{x}_0 | \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T) = \prod_{t=1}^{T} p(\boldsymbol{y}_t | \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{x}_0).$$

The moments of the prediction distributions are

$$\mathrm{E}(\boldsymbol{y}_t | \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{x}_0) = \boldsymbol{H}\boldsymbol{x}_{t-1}$$

and

$$\mathrm{Var}(\boldsymbol{y}_t | \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \sigma, \boldsymbol{x}_0) = \boldsymbol{\Sigma}.$$

The state vectors are calculated using the general linear exponential smoothing recursions

$$\begin{aligned}\hat{\boldsymbol{y}}_t &= \boldsymbol{H}\boldsymbol{x}_{t-1},\\ \boldsymbol{e}_t &= \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t,\\ \boldsymbol{x}_t &= \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{G}\boldsymbol{e}_t.\end{aligned}$$

The log-likelihood function is

$$\log L(\boldsymbol{\theta}, \boldsymbol{x}_0) = -\frac{T}{2}\left[\log(2\pi) + \sum_{i=1}^{N}\log(\sigma_i^2)\right] - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N} e_{it}^2/\sigma_i^2,$$

where σ_i^2 is the ith diagonal element of $\boldsymbol{\Sigma}$. The maximum likelihood estimate of the variance is

$$\sigma_i^2 = \sum_{t=1}^{T} e_{it}^2/T.$$

The vector $\boldsymbol{\theta}$ is restricted to satisfy various invertibility and stationarity conditions that are specific to the particular model under consideration.

An optimizer requires start-up values for $\boldsymbol{x}_0$ and $\boldsymbol{\theta}$. These depend on the particular model being considered. The start-up values for $\boldsymbol{x}_0$ can be determined by heuristics as follows

VLL Model: Start-up values for the initial levels $\boldsymbol{\ell}_0$ equal the average of the first 10 observations for each series.

VLT Model: The first 10 observations of each series are regressed against time. The intercept and slope estimates provide approximations of the values of $\boldsymbol{\ell}_0$ and $\boldsymbol{b}_0$, respectively.

The start-up values for the elements of the parameter matrices are determined as follows:

$\boldsymbol{A}$	Diagonal elements set to 0.33,	off-diagonal elements set to 0;
$\boldsymbol{B}$	Diagonal elements set to 0.5,	off-diagonal elements set to 0;
$\boldsymbol{\Phi}$	Diagonal elements set to 0.9,	off-diagonal elements set to 0.

3.4 Prediction

Being uncertain, future series values are governed by probability distributions, referred to as prediction distributions. Ignoring the uncertainty in estimating the parameters, the model equations suggest that these distributions are Gaussian. Let $\boldsymbol{\mu}_{T+j|T}$ denote the mean of the jth-step-ahead prediction distribution, with the

forecast origin being at the end of period T, and let $\boldsymbol{V}_{T+j|T}$ be the variance matrix. Also, let $\boldsymbol{m}_{T+j|T}$ and $\boldsymbol{W}_{T+j|T}$ be the moments of distribution of the state vector in period $T + j$. Then the moments of these future distributions can be computed recursively using the following formulae:

$$\begin{aligned}
\boldsymbol{\mu}_{T+j|T} &= \boldsymbol{H}\boldsymbol{m}_{T+j-1|T}, \qquad j = 1, 2, \ldots, h,\\
\boldsymbol{V}_{T+j|T} &= \boldsymbol{H}\boldsymbol{W}_{T+j-1|T}\boldsymbol{H}' + \boldsymbol{\Sigma},\\
\boldsymbol{m}_{T+j|T} &= \boldsymbol{F}\boldsymbol{m}_{T+j-1|T},\\
\boldsymbol{W}_{T+j|T} &= \boldsymbol{H}\boldsymbol{W}_{T+j-1|T}\boldsymbol{H}' + \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}'.
\end{aligned}$$

Note that $\boldsymbol{m}_{T|T} = \boldsymbol{x}_T$ and $\boldsymbol{W}_{T|T} = \boldsymbol{O}$. We follow the usual practice in time series analysis and do not attempt to incorporate the effects of estimation error into the distributions.

3.5 Model selection

Automatic model selection is an important feature of forecasting frameworks. One approach is to evaluate the forecasting accuracy of each model over a section of the data that has been withheld from the estimation process. This, however, is not reliable with small samples. A second option is to devote the entire sample to the estimation process and use an information criterion for model selection. A study by Billah *et al.* (2006) suggests that the Akaike's information criterion (AIC) is the best of the common information criteria in a forecasting context. Letting M designate the number of unknown parameters the multivariate AIC is specified as

$$\text{AIC} = -2L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}_0}) + 2M,$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{x}_0}$ denote the maximum likelihood estimates. Although the AIC is used to choose appropriate lag lengths in VARIMA models, it is well known that it cannot be used to decide between different levels of differencing; other approaches, such as the Dickey–Fuller test, must be used instead. By conditioning on the seed states in the state space approach, it can be established that this problem with the AIC disappears. An advantage of the state space approach based on the conditional likelihood (3.5) is that the AIC can be used to assess state space models with different implied orders of differencing, without recourse to unit root tests.

4 Application

To gauge the forecasting capacity of the VISTS framework and to compare it with commonly used alternatives, we applied it to the monthly exchange rate time series of the UK pound (UKP) and US dollar (USD) against the Australia dollar (AUD) (see Figure 1). The expectation was that changes in economic conditions in Australia could

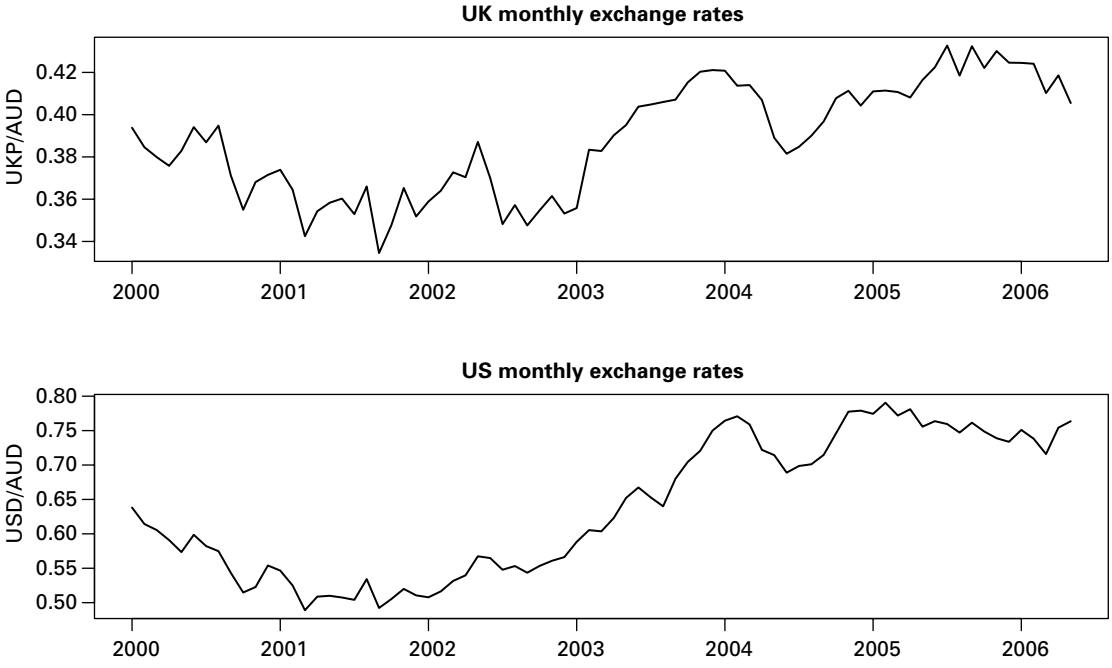

Figure 1 Monthly exchange rates

affect both exchange rates simultaneously, and so create interdependencies between them. The data comprised 77 observations spanning the period January 2000 to May 2006. The natural logarithm of each series was taken before the models were fitted to the first 60 observations.

Their forecasting performances were evaluated on the 17 withheld observations using the mean absolute scaled error (MASE) of Hyndman and Koehler (2006)

$$\text{MASE}_{T+j} = \frac{1}{h}\sum_{j=1}^{h}\left(\frac{|e_{T+j|T}|}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}\right),$$

where h denotes the length of the maximum forecast horizon and $j = 1, 2, \ldots, h$. Essentially, the MASE is the average of the absolute scaled forecast error over horizons 1 to h, where the scaling factor is the average absolute within-sample first difference.

The MASE has two important properties: one, it is scale invariant; and two, it is numerically stable. The second of these properties makes this forecasting accuracy measure particularly attractive, differentiating it from other commonly used measures such as the mean absolute percentage error and the root mean square error. It is because of these two properties that the MASE can be averaged across series and compared across datasets, as we do in the following section.

Table 1 Forecasting accuracy (MASE) of innovations state space models; the value in bold is the minimum value

Model	Univariate	Multivariate
Local level	12.68	15.00
Local trend	11.87	32.91
Damped local trend	13.00	**11.40**

The first exercise with the data was geared to determining whether the exploitation of series interdependencies in the multivariate approach could lead to better forecasts. Common cases of the innovations state space model, in both univariate and multivariate forms, were fitted to the log-transformed data and used to generate predictions. Table 1 shows the results in terms of the MASE. The results are a little mixed in terms of the benefits of exploiting inter-series dependencies. Nevertheless, the multivariate damped local trend model had the lowest measure of all.

A second exercise was to compare the innovations approaches with other common methods. RWs, e.g., typically outperform traditional economic models (Meese and Rognoff, 1983), and so the challenge, in part, is to see whether the VISTS approach can do better. Evidence that this may be possible is provided by Clarida *et al.* (2003), who illustrate the importance of inter-series dependencies when modelling four exchange rates simultaneously using a vector error correction mechanism.

The results shown in Table 2 indicate that the RW works well. Its performance is similar to that of the local level models. It is not as good, however, as the multivariate innovations damped trend model.

As was stated previously, this new innovations approach is an alternative to the MSOE form proposed by Harvey (1989). The results in Table 2 indicate that the innovations approach outperformed the multiple disturbance state space approaches. In theory, for a restricted range of parameter values, the approaches are known to be equivalent. However, the innovations models can take parameter values that have no counterpart in the uncorrelated MSOE models. Some of the estimated parameter values turned out to belong in this region. This illustrates the point that the innovations framework can be more flexible than its MSOE counterpart.

Table 2 Comparison of approaches: MASE over the holdout period covering 17 observations. The value in bold is the minimum value

	Multivariate MSOE	Univariate innovations	Multivariate innovations
Random walk		16.64	
Local level	17.39	12.68	15.00
Local trend	58.85	11.87	32.91
Damped local trend	26.47	13.00	**11.40**
VAR(1)			15.71
VAR(2)			14.26
VAR(3)			17.15

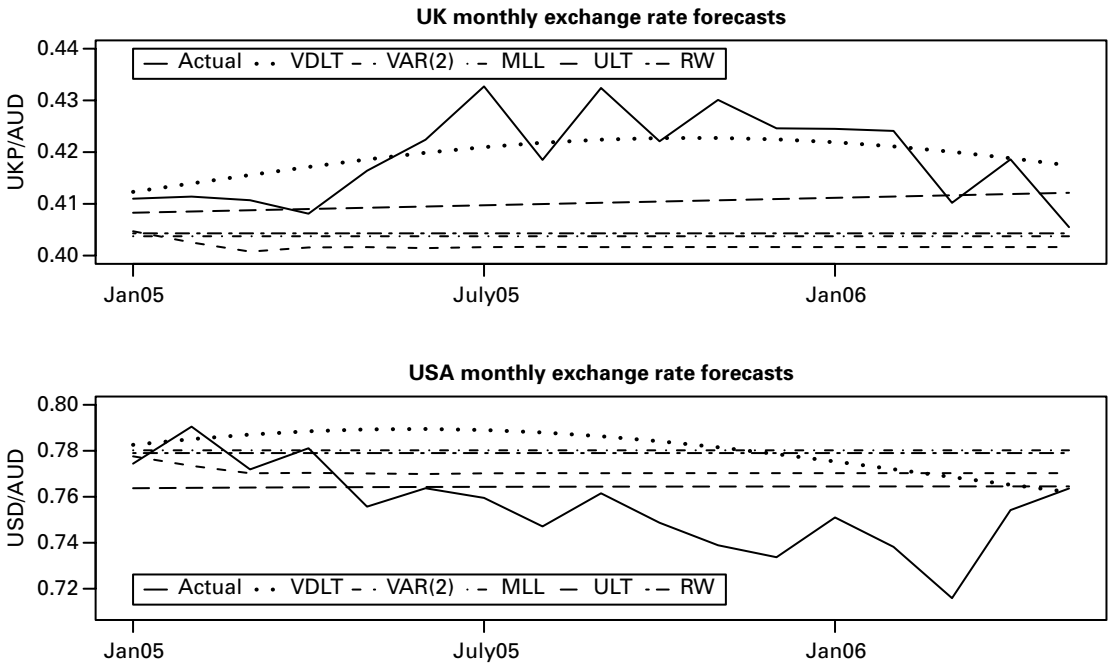

Figure 2 Predicted monthly exchange rates

The VAR is an alternative model formulation, equivalent to a VARIMA(p,0,0) model, where p is the number of autoregressive 'lags'. The VARs considered were restricted to have one, two or three lags. The maximum number of lags was set to three because this corresponds to the number of unknowns in the VDLT model. As prior testing of the series revealed that the log of the series was non-stationary, the VAR models were fitted to the first difference of the logs. The results indicate that the VISTS approach forecasts better than the VAR approach for this particular set of data.

Figure 2 displays the monthly Australian exchange rates for the UKP and the USD for the period of the holdout sample. The observations are indicated by the solid line in both panels. The forecasts from the vector damped local trend (VDLT), (differenced) VAR(2), univariate local trend (ULT), conventional local level (MLL) models, and the random walk (RW), are also displayed. The trajectory of the VDLT forecast is particularly impressive, closely following the 17 monthly observations of the UKP/AUD exchange rate from December 2004 onwards. The forecast of the USD/AUD exchange rates is marginally inferior at some horizons; however, overall the VDLT model produces the most accurate forecasts (see Table 2).

When fitting a family of models, it is a common practice to use the AIC to select a model. The VAR(3) had the lowest AIC value among the VARs; however, it produced forecasts inferior to the VAR(1) and VAR(2). On the other hand, the

Table 3 AIC of the fitted VISTS models

Model	AIC
Vector local level	−13.608
Vector local trend	−13.628
VDLT	−13.763

damped trend had the lowest AIC among the VISTS models and produced the most accurate forecasts (see Tables 2 and 3). Therefore, we can conclude that the AIC correctly anticipated the forecasting capacities of the VISTS models. This is consistent with the findings of Billah *et al.* (2006) on the relationship between information criteria and the forecasting performances of various forms of exponential smoothing.

5 Forecasting experiment

5.1 Design

The analysis of a single (albeit multivariate) time series leaves open the possibility that the better performance of the VISTS model was simply due to chance. Consequently, we undertake an extensive forecasting study, similar to that of the M3-competition (Makridakis and Hibon, 2000), to examine the robustness of the approach on a large number of series.

Some have criticized the M3-competition as being unrealistic. For example, Armstrong (2001) argues that domain knowledge is often used by forecasters when predicting future observations. Automatic forecasting exercises of this nature preclude any information of this type by design. Furthermore, the gauging of relative forecast performance using only numbers and not statistical tests is also cited as a limitation (Granger, 2001). Nevertheless, automatic forecasting is widely used in practice, and the relative performance of forecasting methods under such conditions is of genuine interest. Therefore, we proceed with the experiment, paying particular attention to the following four aspects:

1. The advantages of a multivariate versus a univariate framework.
2. The accuracy of the VISTS specification versus the MSOE structural time series specifications.
3. The predictive ability of the VISTS framework compared to that of the most common alternative VAR.
4. A general comparison combining aspects 1 and 3 using an automatic model selection procedure.

The forecasting accuracy of all models is evaluated once using 1000 different datasets. The variables and their starting dates are randomly chosen from the Watson (2003) macroeconomic database. The Watson (2003) database comprises eight groups which

Table 4 Index of time series models included in forecasting experiment

Model	Description
VLL	Vector local level model
VLT	Vector local trend model
VDLT	Vector damped local trend model
VAR	Vector autoregression
TLL	MSOE multivariate local level model
TLT	MSOE multivariate local trend model
TDLT	MSOE multivariate damped local trend model
UISTS	Univariate innovations structural time series models

can be loosely considered to represent different economic sectors. The number of variables in each group ranges from 13 to 27. All variables are real and non-seasonal, with observations from January 1959 to December 1998. The variables are of varying orders of integration (see Watson, 2003). Every dataset comprises two randomly chosen variables from different economic sectors. The starting date is randomly chosen, where the only restriction is that there must be enough observations to fit and evaluate out-of-sample forecasts up to 12 horizons.

Two sizes of estimation sample are considered, 30 and 100. These sample sizes were chosen because they resemble small and large sample sizes that occur in practice. All variables are standardized by dividing by the standard deviation of their first difference.

The forecasts of the three VISTS models are compared to three alternatives: VAR models, MSOE structural time series models and univariate innovations structural time series (UISTS) models. All three approaches have been established for some time and are commonly used.

The multivariate form of the ARIMA methodology is confined to VAR (Lütkepohl, 2005), with at most three lags fitted to the first difference of the data. An upper limit of three lags was set because this corresponds to the largest VISTS model: the damped local trend model. The optimal lag length is determined using the multivariate form of the AIC.

The number of MSOE structural time series models available to the forecaster is quite large; however, here it is strictly limited to the equivalent of the vector local level, vector local trend and vector damped local trend models. These models are denoted as the TLL, TLT and TDLT, respectively. Table 4 lists the full set of alternatives considered.

As before, we average the MASE over the two series for each dataset, and over each horizon. The maximum horizon considered is 12.

5.2 Results

5.2.1 Multivariate versus UISTS models

In this first section, a comparison is made between the univariate and multivariate innovations structural time series models. The objective of this comparison is to

Table 5 Percentage of first ranks, sample size 30. The largest proportion in each row is in bold

Horizon	VLL	VLT	VDLT	ULL	ULT	UDLT
1	15.6	**21.3**	19.7	13.3	15.9	14.2
2	13.7	**22.3**	19.4	13.0	16.5	15.7
3	12.6	**23.7**	20.4	12.1	16.5	14.7
4	11.1	**23.5**	19.7	10.8	19.0	15.9
5	11.1	**22.3**	20.2	9.4	20.2	16.8
6	9.9	21.8	20.5	9.8	**21.9**	16.1
7	11.0	**22.1**	18.9	9.2	21.5	17.3
8	10.4	**22.4**	17.9	9.2	21.4	18.7
9	10.3	**22.4**	18.0	9.3	21.4	18.6
10	9.4	**22.5**	18.1	9.5	21.6	18.9
11	9.4	**22.4**	18.2	9.5	21.5	19.0
12	9.0	**22.1**	19.1	9.6	21.5	18.8
Average	11.1	**22.4**	19.2	10.4	19.9	17.1

determine whether there is any advantage in extending the univariate form into a multivariate specification.

Tables 5 and 6 display the percentage of times that each model produces the most accurate forecasts over each horizon. The largest percentage in each row is bolded. For example, the vector local trend (VLT) model produced the most accurate forecasts 12 steps ahead 22.1% of the time in the small sample application.

In general, the VLT model appears to be the most accurate, registering the highest average in both tables. The second most accurate model depends on which sample size is being considered. In the large sample case, the VDLT model is the second most accurate (as indicated by the average), whereas it is the third most accurate model in the small sample size context, after the univariate and vector local trend models. The VLL model appears to produce the least accurate predictions in general.

Table 6 Percentage of first ranks, sample size 100. The largest proportion in each row is in bold

Horizon	VLL	VLT	VDLT	ULL	ULT	UDLT
1	13.5	**24.7**	19.9	13.6	16.2	12.1
2	11.9	22.7	**23.3**	12.9	18.3	12.3
3	10.1	23.9	**24.1**	9.6	19.7	12.6
4	9.1	23.8	**23.4**	9.3	21.1	13.6
5	8.4	22.5	**24.2**	8.7	21.2	15.0
6	9.0	22.3	22.3	8.9	**22.7**	14.9
7	8.2	21.7	21.4	9.5	**23.1**	16.1
8	8.0	21.5	21.2	9.3	**23.4**	16.6
9	8.0	21.4	21.5	8.7	**23.5**	16.9
10	8.5	20.8	21.4	8.6	**23.1**	17.7
11	8.1	20.4	21.3	9.4	**23.1**	17.7
12	7.7	20.7	21.5	9.4	**22.6**	18.1
Average	9.2	**22.2**	22.1	9.8	21.5	15.3

Table 7 Percentage of first places, small sample size. The largest proportions for each row are in bold

Horizon	VLL	VLT	VDLT	TLL	TLT	TDLT
1	12.4	17.4	16.8	14.1	**25.3**	14.0
2	12.1	19.6	18.3	13.1	**21.9**	15.7
3	11.1	**20.8**	19.1	12.7	20.2	16.1
4	10.9	**21.5**	18.8	11.3	20.4	17.3
5	10.3	**21.1**	19.9	10.4	**21.1**	17.2
6	10.2	20.9	20.6	9.9	**21.2**	17.2
7	10.4	**21.6**	20.3	9.3	20.6	17.8
8	9.9	**21.6**	20.4	9.3	20.4	18.4
9	9.6	**21.9**	19.9	9.6	20.8	18.2
10	9.2	**21.6**	19.9	9.1	21.4	18.8
11	8.9	**21.6**	20.8	9.5	21.1	18.1
12	8.8	**21.8**	21.0	8.9	21.3	18.2
Average	10.3	21.0	19.7	10.6	**21.3**	17.3

In summary, the results in Tables 5 and 6 indicate that the VISTS models generated the most accurate forecasts for over 50% of the comparisons. Therefore, it can be concluded that there is a significant advantage in extending the univariate formulation to a multivariate specification.

5.2.2 VISTS versus the MSOE structural approach

In this section, the VISTS framework is compared to the MSOE form of the structural time series model.

Tables 7 and 8 display the percentages of first places for each model over horizons 1 to 12. The maximum percentage in each row is bolded. As is consistent with

Table 8 Percentage of first places, large sample size. The largest proportions for each row are in bold

Horizon	VLL	VLT	VDLT	TLL	TLT	TDLT
1	9.5	22.0	18.5	14.1	**24.0**	11.9
2	9.7	**23.2**	20.4	13.1	21.0	13.4
3	7.3	**26.9**	21.3	11.5	20.7	12.3
4	7.3	**26.8**	23.2	10.6	20.0	12.2
5	7.7	**25.7**	21.7	10.4	21.1	13.4
6	7.7	**26.5**	21.7	10.3	20.6	13.3
7	7.6	**25.8**	22.8	10.0	20.7	13.1
8	7.5	**25.7**	22.4	9.9	20.4	14.1
9	7.1	**25.7**	23.2	9.8	20.1	14.1
10	7.0	**25.5**	22.9	9.9	20.8	13.9
11	7.1	**25.5**	22.9	10.2	20.1	14.2
12	7.1	**26.7**	23.2	9.6	19.3	14.1
Average	7.7	**25.5**	22.0	10.8	20.7	13.3

the previous findings, the VLT model appears to have produced the most accurate forecasts more often than any of the alternatives in the small sample size context (but not the highest average in the large sample case). The VDLT model was second to the VLT model in the large sample case. The VISTS models that are characterized by a local trend outperformed their MSOE structural form equivalents over all horizons except the first in the context of the large sample size.

Again, it is apparent that all the models considered experience some degree of success. Moreover, the innovations models appear to perform slightly better than their MSOE counterparts.

In summary, the VISTS approach is arguably a simpler and more flexible form than the MSOE structural time series model and has been shown to produce accurate forecasts on a regular basis. Therefore, the evidence suggests that the VISTS models provide a useful alternative for those wanting to employ a structural time series approach.

5.2.3 Comparison with VAR

The objective of this comparison is to gauge the forecasting accuracy of the VISTS framework against what is arguably the most popular multivariate time series modelling tool, the VAR approach. Both methods include an automatic model selection procedure using the AIC.

Table 9 displays the relative predictive performance of VAR and VISTS models using an automatic model selection procedure. The VAR framework was fitted to the first differences of the data. The maximum order permitted for the VAR was three lags, as this reflects the number of unknown parameters in the VDLT model. The overall results are quite similar, with the VISTS slightly outperforming the VAR in the large sample, and vice versa in the small sample.

Table 9 Percentage of first places

	Small sample		Large sample	
Horizon	VISTS	VAR	VISTS	VAR
1	48.3	51.7	51.1	48.9
2	48.7	51.3	52.1	47.9
3	49.4	50.6	53.8	46.2
4	48.5	51.5	53.9	46.1
5	48.0	52.0	52.6	47.4
6	47.4	52.6	51.7	48.3
7	47.4	52.6	52.0	48.0
8	47.6	52.4	51.7	48.3
9	47.8	52.2	53.0	47.0
10	47.8	52.2	53.9	46.1
11	48.3	51.7	54.0	46.0
12	47.5	52.5	53.6	46.4
Average	48.1	51.9	52.8	47.2

Table 10 Percentage of first places

	Small sample			Large sample		
Horizon	VISTS	UISTS	VAR	VISTS	UISTS	VAR
1	29.9	35.7	34.4	31.9	33.5	34.6
2	31.0	37.4	31.7	33.5	34.5	32.2
3	31.4	38.3	30.3	34.3	35.5	30.2
4	30.7	38.9	30.4	33.5	36.3	30.2
5	30.0	39.3	30.7	31.2	38.6	30.2
6	29.8	39.3	30.9	30.6	38.3	31.1
7	30.9	37.9	31.2	31.5	37.5	31.0
8	31.0	37.6	31.4	32.4	37.2	30.4
9	31.5	38.0	30.5	32.3	38.4	29.3
10	30.7	38.4	30.9	32.8	39.3	27.9
11	30.1	39.6	30.3	33.0	38.8	28.2
12	30.1	39.1	30.8	32.9	38.8	28.3
Average	30.6	38.3	31.1	32.5	37.2	30.3

5.2.4 Automatic model selection

This final comparison considers all the alternatives except the MSOE structural time series models. The objective is to determine which (if any) of the approaches that include an automatic model selection feature produced a relatively higher degree of forecasting accuracy.

Table 10 displays the percentage of times that each alternative is most accurate. It is clear that all three approaches perform relatively well.

In particular, the results show that the univariate innovations structural time series approach outperformed the multivariate alternatives across all horizons. The relative performances of the VISTS and VAR approaches are consistent with the findings in the previous section: VISTS performed relatively better when fitted to the larger sample.

In general, the results both here and in the previous sections show that the VISTS framework produces more accurate forecasts than the other forecasting tools considered. For each comparison, the VISTS framework produced the most accurate forecasts on a consistent basis (approximately 50% of the time).

6 Conclusion

We have introduced and evaluated the vector version of the innovations state space model as a mechanism for forecasting related macroeconomic time series. It was compared with the vector autoregressive and conventional MSOE state space approaches that are typically used in multiple time series studies. By conditioning on seed states rather than an initial run of series values, it was shown that maximum likelihood estimates could be obtained with a simple recursion reminiscent of exponential

smoothing. Some preliminary empirical studies have indicated that VISTS is a robust approach to forecasting and performs better than the conventional approaches in a significant proportion of cases.

Future work will be directed to further developing the framework to explicitly exploit co-movements in time series reminiscent of those that have been the focus of the co-integration literature. Other work will be directed to generalizing the univariate framework in Hyndman *et al*. (2002), so that seasonal effects, non-linear relationships and heteroscedastic error processes are accommodated.

References

Anderson B and Moore JB (1979) *Optimal filtering*. New Jersey: Prentice-Hall.

Ansley CF and Kohn R (1985) Estimation, filtering and smoothing in state space models with incompletely specified conditions. *Annals of Statistics*, **13**, 1286–316.

Armstrong JS (2001) Should we redesign forecasting competitions? *International Journal of Forecasting*, **17**, 542–45.

Athanasopoulos G and Vahid F (2008) VARMA versus VAR for macroeconomic forecasting. *Journal of Economic and Business Statistics*, **26**, 237–52.

Billah B, King M, Snyder R and Koehler A (2006) Exponential smoothing model selection for forecasting. *International Journal of Forecasting*, **22**, 239–47.

Box GEP and Jenkins G (1970) *Time series analysis: forecasting and control*. San Francisco: Holden-Day.

Brockwell PJ and Davis RA (1991) *Time series: theory and methods*, 2nd edn. New York: Springer-Verlag.

Brown RG (1959) *Statistical forecasting for inventory control*. New York: McGraw-Hill.

Clarida R, Sarno L, Taylor M and Valente G (2003) The out-of-sample success of term structure models as exchange rate predictors: one step along. *Journal of International Economics*, **60**, 61–83.

de Jong P (1991) The diffuse Kalman Filter. *Annals of Statistics*, **19**, 1073–83.

Dickey D and Fuller W (1979) Distribution of the estimates for autoregressive time series with a unit root. *Journal of American Statistical Association*, **74**, 427–31.

Durbin J and Koopman S (2001) *Time series analysis by state space method*. Oxford: Oxford University Press.

Enns PG, Machak JA, Spivey A and Wrobleski WJ (1982) Forecasting applications of an adaptive multiple exponential smoothing model. *Management Science*, **28**, 1035–44.

Gardner ES and McKenzie E (1985) Forecasting trends in time series. *Management Science*, **31**, 1237–46.

Granger C (2001) Comments on the M3 forecast evaluation and a comparison with a study by Stock and Watson. *International Journal of Forecastings*, **17**, 565–67.

Granger CW and Newbold P (1986) *Forecasting economic time series*, 2nd edn. New York: Academic Press.

Hamilton JD (1994) *Time series analysis*. Princeton: Princeton University Press.

Hannan EJ (1969) The identification of vector mixed Autoregressive-Moving Average systems. *Biometrica* **56**, 223–25.

Hannan EJ and Deistler M (1988) *The statistical theory of linear systems*. New York: John Wiley & Sons.

Harvey A (1986) Analysis and generalisation of a multivariate exponential smoothing model. *Management Science*, **32**, 374–80.

Harvey A and Koopman S (1997) Multivariate structural time series model. In Heij C,

Schumacher H and Hanzon B (eds) *System dynamics in economic and financial models*. Chichester: John Wiley and Sons, 269–96.

Harvey AC (1989) *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

Hyndman R and Koehler AB (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting*, **22**, 679–88.

Hyndman RJ (2001) It's time to move from 'what' to 'why'. *International Journal of Forecasting*, **17**, 567–70.

Hyndman RJ, Koehler AB, Ord JK and Snyder RD (2008) *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer.

Hyndman RJ, Koehler AB, Snyder RD and Grose S (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, **18**, 439–54.

Jones RH (1966) Exponential smoothing for multivariate time series. *Journal of the Royal Statistical Society, Series B*, **28**, 241–51.

Kwiatkowski D, Phillips PCB, Schmidt P and Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**, 159–78.

Lütkepohl H (2005) *New introduction to multiple time series analysis*. Berlin: Springer-Verlag.

Makridakis M and Hibon M (2000) The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–76.

Meese R and Rognoff K (1983) Empirical exchange rate model of the seventies. *Journal of International Economics*, **14**, 3–24.

Ord JK, Koehler AB and Snyder RD (1997) Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, **92**, 1621–29.

Pfeffermann D and Allon J (1989) Multivariate exponential smoothing: method and practice. *International Journal of Forecasting*, **5**, 83–98.

Sims C (1980) Macroeconomics and reality. *Econometrica*, **48**, 1–49.

Snyder RD (1985) Recursive estimation of dynamic linear statistical models. *Journal of the Royal Statistical Society, Series B*, **47**, 272–76.

Tiao G and Tsay R (1989) Model specification in multivariate time series. *Journal of the Royal Statistical Society, Series B*, **51**, 157–213.

Watson M (2003) Macroeconomic forecasting using many predictors. In Dewatripont M, Hansen P and Turnovsky SJ eds. *Advances in economics and econometrics, theory and applications, eighth world congress of the econometric society*, **3**. USA: Cambridge University Press, 87–115.