Time-varying forecasts by variational approximation of sequential Bayesian inference

HUI “Fox” LING * † ¶ and DOUGLAS B. STONE ‡ § ¶

* Corresponding author, Email: Fox.Ling@AllianzGI.com

† Allianz Global Investors U.S. LLC, 1633 Broadway, New York, NY 10019, USA

‡ Allianz Global Investors U.S. LLC, 600 West Broadway, San Diego, CA 92101, USA

§ (Current) GLP Risk Management, 18575 Locksley Street, San Diego, CA 92128, USA

¶ The authors truly appreciate the reviewers for their critiques and suggestions.

Version: April 10, 2015

Time-varying forecasts by variational approximation of sequential Bayesian inference

Methods developed for making time-varying forecasts in economic and financial analysis include (a) equal-weighted moving-window (or rolling) regression, (b) time-weighted (e.g. exponentially weighted) regression, (c) the Kalman filter and (d) adaptive Kalman filters. This paper developed a new method based on variational approximation of sequential Bayesian inference (VASB). Concepts and notions of the sequential Bayesian analysis and the variational approximation of an intractable posterior are simple and straightforward. Our VASB algorithm is not complicated and is easy to code. For a regression on multiple time-series, the regression coefficients, standard errors, prediction and residual error are time-varying and are estimated jointly at every time step. For a single time-series (e.g. price returns of an asset), its mean and variance are time-varying and are predicted jointly at every time step. The VASB algorithm performs better than the rolling and time-weighted statistics or regressions and the Kalman filter in terms of higher predictive power and stronger robustness. Derivations of the VASB algorithm are presented in the appendices.

Keywords: Time-varying forecasts; Time-varying regression; Time-varying variance; State space modeling of time-series; Sequential

Bayesian inference; Variational Bayes; Kalman filter; Bayesian filtering.

JEL Classification: C11, C22, C32, C53.

1. Introduction

Economic time-series predictions, asset allocation strategies, and many other analytical tasks in economics and finance require time-varying forecasts. Multivariate time-series analysis is a major quantitative and statistical research area with linear regression among the most popular approaches. Almost all financial market characteristics are time-varying due to continuous social evolutions, economic developments, technological progresses, regulatory changes, and financial innovations. Two widely used time-varying techniques are (a) rolling or moving-window regression – always using the latest information within a time window of fixed width and dropping the earlier data when the new ones become available; and (b) imposing exponential time-weighting of a fixed half-life in order to emphasize the recent data over the earlier ones. The size of the rolling window and the length of the half-life are among key parameters to be evaluated and decided empirically. In these conventional techniques, there is not a separate equation to explicitly govern the dynamic evolution or process of the unknown time-varying regression coefficients.

When applying a state space model to a time-varying regression, a separate equation dedicated to the dynamics of the regression coefficients is introduced along with the original regression equation. Linear state space model and its classic solution, the Kalman filter (KF), were developed in the early 1960's to meet engineering challenges such as automatic navigation, optimal control, and continuous tracking (Kalman 1960, Harvey 1991, Simon 2006). The basic state space model is expressed by two equations: a state transition equation, $x_t = F_t\, x_{t-1} + u_t$, and a measurement equation, $y_t = H_t\, x_t + v_t$, where the subscript t is a time index. When applying the model to a time-varying regression, the regression equation becomes the measurement equation for y_t. The scalar y_t is the dependent variable while the row vector H_t contains values of the independent variables at time t. The column vector x_t is the unknown regression coefficients to be solved for and the scalar v_t is the regression residual. The evolution of the regression coefficient vector x_t is explicitly described by state transition matrix F_t and error u_t. The Kalman filter is the solution to the linear Gaussian state space model. The filter estimates the state x_t recursively when the new values of H_t and new observation y_t become available. The KF is optimal when u_t and v_t are of known zero-mean normal distributions with Q_t and R_t as their covariance matrices, respectively (Simon 2006).

The Kalman filter has been examined and utilized by economic and financial researchers as a theoretically sound approach to time-varying estimations, predictions and regressions for at least four decades. Since the discussion by Belsley (1973) on KF as an optimal solution to a time-varying regression, researchers have published numerous studies about the state space models for economic and financial time-series analyses. Alvarez and Levinson (2007), Andersen and Benzoni (2009), Date and Ponomareva (2011), Javaheri *et al.* (2003), Mamaysky *et al.* (2008), Pasricha (2006), and Stock and Watson (2005, 2006) presented or summarized applications of linear and non-linear KFs in economics and finance. The KF applications include hedge fund risk factor exposures, alphas and betas of mutual funds, stochastic volatility models, interest rate models, exchange rate risk premiums, term structures of commodity prices, demand for international reserves, dynamic factor models, time-varying parameters for forecasts, unobserved components models, and Markov switching models.

The Kalman filter requires that, in addition to the data H_t and y_t, the matrices F_t, Q_t and R_t must be known *a priori*. In many real world applications, however, one or both of the covariance matrices, Q_t and R_t, are unknown and time-varying. Various researchers have developed adaptive Kalman filters (AKFs) to deal with these problems. Das and Ghoshal (2010) applied an AKF to estimate the CAPM beta which varies in time. Their AKF was proposed by Ding *et al.* (2006 a, 2006 b, 2007) for GPS/INS navigation data integration. At each time t, R_t is revised by the current data while Q_t is updated by rescaling the entire matrix. For many applications having two or more state variables, simply rescaling the entire matrix Q_t may not be realistic. Koopman and Bos (2004) suggested an adaptive state space model in which the two stochastic errors u_t and v_t are linked by the same stochastic disturbance ϵ_t: $u_t = U\epsilon_t$ and $v_t = V\epsilon_t$, where the covariance of ϵ_t is time-varying. Therefore, the task of estimating separate Q_t and R_t is reduced to estimating only one time-varying covariance of ϵ_t. Other AKFs with unknown and time-varying covariance have been proposed in various engineering fields. A common limitation of the AKFs is that one of the Q_t and R_t is assumed to be known, or is a known function of the other. In a typical economic or financial application, however, both Q_t and R_t are fully unknown and are time-varying.

A recent development in the state space modeling is to employ a sequential Bayesian approach or Bayesian filtering (Särkkä 2013). The new estimates of the hidden variables such as x_t are obtained as the posterior of a Bayesian inference in which the previous estimates made at earlier time t-1 form the prior and the current observation is incorporated in the likelihood. When the state transition and measurement equations are linear and the random errors are Gaussian with known covariance matrices Q_t and R_t, the exact posterior of x_t can be analytically expressed as Gaussian with its moments the same as those formulated by the Kalman filter (Šmídl and Quinn 2006 a). The sequential Bayesian posteriors in almost all other cases are not analytically tractable and, therefore, have to be estimated approximately. A stochastic approximation is to use a rather large random sample of the joint distribution in the model to construct the posterior numerically. A well-known example is particle filter (PF) with each random realization of the sample called a particle. They are already applied to financial time-series analysis and econometrics (Johannes and Polson 2009, Lopes and Tsay 2011). The PFs are computer-intensive and it is not easy to explain their numerical results.

An intractable joint posterior may be approximated analytically by some simplified distributions. One of the methods is variational approximation (VA): a joint posterior is approximated by a factorized distribution with all of the approximating factors (simpler distributions) to be identified and optimized iteratively (Bishop 2006, Tzikas *et al.* 2008, Grimmer 2011). Beal (2003) created a variational Bayesian algorithm to estimate the unknown x_t, $F_t = F$, $H_t = H$, and $R_t = R$ jointly for a linear dynamic system with a state space representation. Šmídl and Quinn (2006 a, 2006 b, 2008) developed a range of variational Bayesian filtering approaches to, for examples, a hidden Markov model with an unknown non-stationary transition matrix, and a non-linear state space model where a combination of PF and VA is applied. Särkkä and Nummenmaa (2009) and Särkkä and Hartikainen (2013) proposed variational Bayesian filters for linear and nonlinear state space models of two stochastic unknowns, Gaussian x_t and inverse-gamma R_t, with Q_t as a known input.

In a typical time-varying analysis in economics and finance, both of the state transition error u_t and the measurement error v_t are of distributions with unknown stochastic parameters Q_t and R_t whose distributions need to be estimated as well. We did not find published works to estimate the distributions of x_t, Q_t and R_t jointly using a Bayesian approach with an analytic approximation. This may be a reason why the existing state space modeling methods lack of wide proven applicability and deserved popularity in the analyses of financial markets.

As detailed in Appendix A of this paper, following the work of two-random-variable variational Bayes by Särkkä and Nummenmaa (2009) and Särkkä and Hartikainen (2013) in which x_t and R_t are unknown (with Q_t as a known input), applying general Bayesian filtering framework and techniques (Särkkä 2013, Özkana *et al.* 2013) to a linear state space model where both Q_t and R_t are among the unknown time-varying stochastic variables to be estimated, and making the standard variational Bayesian approximation (Bishop 2006, Tzikas *et al.* 2008, Grimmer 2011) to the joint distribution of the random variables, this paper developed a three-random-variable VASB (variational approximation of sequential Bayesian inference) algorithm to estimate the time-varying distributions of x_t, Q_t and R_t jointly.

The remainder of the paper is organized as follows. Section 2, "Time-varying regression and state space model", introduces and compares the equations of a multivariate time-varying regression and a state space model. Section 3, "VASB algorithm: a solution to the state space model", presents a short series of formulas of the VASB algorithm. Section 4, "State space solutions for multivariate regression and univariate moments", discusses setups of the state space models for multivariate regression and time-series moments. Section 5, "Error reduction target", discusses a mechanism to prevent both over- and under-fitting when real market data are involved. Section 6, "Case 1: simulations", compares the VASB with OLS (ordinary least-squares) and KF in some simulations of U.S. equity investment strategies. Section 7, "Case 2: examples of market forecasts", discusses advantages of the VASB over the rolling and time-weighted OLS and the KF when making actual forecasts in real market settings. Appendix A details our derivations of the VASB algorithm. Appendix B explains the use of error reduction target to prevent over- and under-fitting when both Q_t and R_t need to be estimated with real market data.

2. Time-varying regression and state space model

A multivariate time-varying linear regression on time-series of time t can be formulated as

$$r(t) = b_0(t) + b_1(t)\, z_1(t) + \cdots + b_{m-1}(t)\, z_{m-1}(t) + e(t), \qquad (1)$$

where $r(t)$ is the dependent variable, $z_j(t)$, $j = 1, 2, \cdots, m-1$, are $m-1$ independent variables, $b_j(t)$, $j = 0, 1, \cdots, m-1$, are m time-varying regression coefficients to be solved for, and $e(t)$ is the regression residual of unobservable variance $\sigma^2(t)$. The equally and exponentially weighted (rolling) regressions are popular approaches to the time-varying regression. These two simple "time-varying" techniques are, however, lack of sound theoretical foundation. A linear state space model provides a rigorous time-varying framework to the problem.

The two basic equations of a linear state space model can be formalized as

$$x_t = F_t\, x_{t-1} + a_t + u_t, \qquad (2)$$

$$y_t = H_t\, x_t + c_t + v_t, \qquad (3)$$

where Eq. (2) is the state transition equation and Eq. (3) the measurement equation in the same form as the multivariate regression Eq. (1). In these equations, t is the time index, $t = 1, 2, \cdots$, x_t an unknown $m \times 1$ hidden state vector to be estimated, F_t a known $m \times m$ state transition matrix, a_t a known $m \times 1$ state input vector, u_t an unobservable $m \times 1$ state transition error vector of zero mean and unobservable $m \times m$ covariance matrix Q_t and independent of $u_{t-1}, u_{t-2}, \cdots, v_t, v_{t-1}, \cdots$, y_t an observed $n \times 1$ measurement vector, H_t a known $n \times m$

measurement or observation matrix, c_t a known $n \times 1$ measurement input vector, and v_t an unobservable $n \times 1$ measurement error vector of zero mean and unobservable $n \times n$ covariance matrix R_t and independent of $u_t, u_{t-1}, \cdots, v_{t-1}, v_{t-2}, \cdots$.

To apply the state space model Eqs. (2) and (3) to the multivariate regression Eq. (1), following relations link them together: $y_t = r(t)$, $H_t = [1, z_1(t), z_2(t), \cdots, z_{m-1}(t)]$, $x_t = [b_0(t), b_1(t), b_2(t), \cdots, b_{m-1}(t)]^T$, $a_t = [0, \cdots, 0]^T$, $c_t = 0$, and

$$F_t = f_0\, I_m,\ \ 0 < f_0 \le 1, \tag{4}$$

where the superscript T denotes the transposition of a matrix and I_m is the $m \times m$ identity matrix. For the state space model Eqs. (2) and (3) defined by the regression Eq. (1) with the above set of relations, both Q_t and R_t are unknown and stochastic.

3. VASB algorithm: a solution to the state space model

Whether or not the unknown Q_t and R_t themselves are of explicit interest in a particular application, they have to be estimated in order to estimate the state x_t. This paper developed a VASB (variational approximation of sequential Bayesian inference) algorithm to estimate the three stochastic unknowns, x_t, Q_t and R_t, jointly.

Additional assumptions are needed in order to present the VASB algorithm. The error vectors in Eqs (2) and (3) are assumed to be Gaussian, $u_t \sim N(u_t; 0, Q_t)$ and $v_t \sim N(v_t; 0, R_t)$. A normal distribution is the most appropriate to be assumed because it has the maximum entropy (Rao 1973) among all possible distributions if only the mean and covariance are given and no other information is available. Therefore, if the initial state x_0 is Gaussian, the distribution of the state x_t estimated based on all information available at time t is a multivariate normal distribution, $x_t \sim N(x_t; x_{t|t}, P_{t|t})$. The state transition equation Eq. (2) indicates that the expected value of x_t can be predicted by

$$x_{t|t-1} = F_t\, x_{t-1|t-1} + a_t. \tag{5}$$

According to the measurement equation Eq. (3), the observation y_t can be predicted by

$$y_{t|t-1} = H_t\, x_{t|t-1} + c_t, \tag{6}$$

subject to a measurement prediction error

$$e_{t|t-1} = y_t - y_{t|t-1}. \tag{7}$$

The estimated expected values of Q_t and R_t are denoted as $Q_{t|t}$ and $R_{t|t}$, respectively. Our VASB algorithm assumes that the transition matrix F_t, such as Eq. (4), and the three expected values, $P_{t|t}$, $Q_{t|t}$ and $R_{t|t}$, are diagonal matrices.

Given initial values $x_{0|0}$, $P_{0|0}$, $Q_{0|0}$ and $R_{0|0}$, the VASB algorithm can be summarized by the following procedure at each time step $t = 1, 2, \cdots$. The detailed derivations of the VASB, including dynamic Bayesian modeling assumptions, are discussed in the Appendix A. The beginning block of the VASB contains the three statements Eqs. (5), (6) and (7) to make the data preparation for time step t. The middle block of the VASB is a variational approximation iteration of length L ($L = 10$, for example) to initialize and to update two diagonal covariance matrices $P_{t|t}^{(k)}$ and $R_{t|t}^{(k)}$, $k = 0, 1, \cdots, L$, for the time step t:

$$P_{t|t}^{(0)} = P_{t|t-1} = P_{t-1|t-1}, \tag{8}$$

$$R_{t|t}^{(0)} = R_{t|t-1} = R_{t-1|t-1}, \tag{9}$$

$$S_t^{(k)} = H_t\, P_{t|t}^{(k)} H_t^T + R_{t|t}^{(k)}, \tag{10}$$

$$K_t^{(k)} = P_{t|t}^{(k)} H_t^T (S_t^{(k)})^{-1}, \tag{11}$$

$$M_t^{(k)} = I_n - H_t\, K_t^{(k)}, \tag{12}$$

$$P_{t|t}^{(k)} = diag(\, P_{t|t}^{(0)} + \\ + K_t^{(k-1)} \,(\, e_{t|t-1}\, e_{t|t-1}^T - S_t^{(k-1)} \,)\, (K_t^{(k-1)})^T \,/\, T_0 \,), \tag{13}$$

$$R_{t|t}^{(k)} = diag(\, R_{t|t}^{(0)} + \\ + M_t^{(k-1)} \,(\, e_{t|t-1}\, e_{t|t-1}^T - S_t^{(k-1)} \,)\, (M_t^{(k-1)})^T \,/\, T_0 \,), \tag{14}$$

where the superscript -1 indicates the inverse of a matrix. The function $diag(\cdot)$ converts a square matrix into a diagonal one by setting all of the off-diagonal elements to 0. According to (A.17) in the Appendix A, the starting value $P_{t|t}^{(0)}$ by Eq. (8) for $t = 1$ can be replaced by $P_{1|1}^{(0)} = diag(\, F_1\, P_{0|0}\, F_1^T \,) + Q_{0|0}$ if the initial value $Q_{0|0} \neq 0$. The $m \times n$ matrix $K_t^{(k)}$ is a "Kalman gain" (Simon 2006). The value $1/T_0$ can be regarded as the weight of the last data point in updating $P_{t|t}^{(k)}$ and $R_{t|t}^{(k)}$. A reasonable example of T_0 could be $T_0 = 10$. The difference matrix, $e_{t|t-1} e_{t|t-1}^T - S_t^{(k)}$, in Eqs. (13) and (14) is the observed data to update the covariance $P_{t|t}^{(k)}$ and $R_{t|t}^{(k)}$. The ending block of the VASB is to calculate the results for the time step t:

$$x_{t|t} = x_{t|t-1} + K_t^{(L)}\, e_{t|t-1}, \tag{15}$$

$$y_{t|t} = H_t\, x_{t|t} + c_t, \tag{16}$$

$$P_{t|t} = P_{t|t}^{(L)} - diag(\, K_t^{(L)} S_t^{(L)} (K_t^{(L)})^T \,), \tag{17}$$

$$Q_{t|t} = q_{nne}(\, P_{t|t} - diag(\, F_t\, P_{t-1|t-1}\, F_t^T \,)\,), \tag{18}$$

$$R_{t|t} = R_{t|t}^{(L)}. \tag{19}$$

The function $q_{nne}(\cdot)$ in Eq. (18) is to make the elements of a matrix nonnegative, element by element: $q_{nne}(A_{ij}) = (A_{ij} + |A_{ij}|)/2$.

The VASB algorithm Eqs. (5) to (19) uses the last estimates, $x_{t-1|t-1}$, $P_{t-1|t-1}$ and $R_{t-1|t-1}$, plus the current information, F_t, H_t, a_t, c_t and y_t, to make the new estimates, $x_{t|t}$, $P_{t|t}$,

$Q_{t|t}$ and $R_{t|t}$. We coded and utilized the VASB algorithm with MATLAB, a high-level computer language of technical computing (www.mathworks.com).

4. State space solutions for multivariate regression and univariate moments

When applying the state space model Eqs. (2) and (3) to a time-varying analysis problem, we will explicitly define the observation y_t and measurement matrix H_t, and make the following VASB application assumptions, constraints or actions: defining the state transition matrix F_t by Eq. (4), using a column of 1 in the matrix H_t for estimating the time-varying conditional mean of y_t, assigning 0s to the state and measurement inputs a_t and c_t, assuming and constraining the random error covariance matrices $P_{t|t}$, $Q_{t|t}$ and $R_{t|t}$ as diagonal, and estimating the initial values $x_{0|0}$, $P_{0|0}$, $Q_{0|0}$ and $R_{0|0}$. For a multivariate regression, $x_{0|0}$, $P_{0|0}$ and $R_{0|0}$ are initial regression coefficients, standard error variances and residual error variance. Then the VASB algorithm Eqs. (5) to (19) is utilized to get the solutions to the time-varying analysis problem.

To solve the multivariate time-varying regression Eq. (1), a state space model Eqs. (2) and (3) can be applied by defining

$$y_t = r(t), \tag{20}$$

$$H_t = [\ 1, z_1(t), z_2(t), \cdots, z_{m-1}(t)\ ], \tag{21}$$

along with our VASB application assumptions, constraints and actions. After obtaining the VASB results $x_{t|t-1}$, $x_{t|t}$, $y_{t|t-1}$, $y_{t|t}$, $P_{t|t}$ and $R_{t|t}$ by Eqs. (5) to (19), the m predicted time-varying regression coefficients $b_j(t|t-1)$, $j = 0, 1, \cdots, m-1$, and the predicted conditional expectation $r(t|t-1)$ are:

$$[\ b_0(t|t-1), b_1(t|t-1), \cdots, b_{m-1}(t|t-1)\ ] = x_{t|t-1}^T,$$

$$r(t|t-1) = y_{t|t-1},$$

while the m estimated coefficients $b_j(t|t)$ and the estimated conditional expectation $r(t|t)$ are:

$$[\, b_0(t|t), b_1(t|t), \cdots, b_{m-1}(t|t) \,] = x_{t|t}^T,$$

$$r(t|t) = y_{t|t}.$$

The standard error variance is $P_{t|t}$ and the residual error variance is $R_{t|t}$. The VASB method estimates the regression coefficients, standard errors, prediction and residual error jointly at every time step t.

The traditional concept of time-varying regression treats the regression coefficients $b_j(t)$ and residual error $\sigma^2(t)$ as non-random parameters to be solved for. When applying the state space model and the VASB solution Eqs. (2) to (19) to the regression Eq. (1), the regression coefficients, standard error variances and residual error variance are represented by the hidden state vector x_t and hidden (diagonal) matrices P_t and R_t, which are random variables whose distributions are estimated jointly. A special case is a stochastic local level model (Koopman and Bos 2004, Zivot and Wang 2006),

$$r(t) = b_0(t) + e(t), \tag{22}$$

where the hidden random variable $b_0(t)$ represents the stochastic local level (the time-varying mean) and the error $e(t) \sim N(0, \sigma^2(t))$. A state space model Eqs. (2) and (3) to be used to predict and estimate the time-varying mean and variance of the time-series $r(t)$ in Eq. (22) can be defined simply by

$$y_t = r(t), \tag{23}$$

$$H_t = 1, \tag{24}$$

along with our VASB application assumptions, constraints and actions. With the VASB results $x_{t|t-1}$, $x_{t|t}$, $y_{t|t-1}$, $y_{t|t}$, $P_{t|t}$, $Q_{t|t}$ and $R_{t|t}$ by Eqs. (5) to (19), the predicted time-varying mean and variance of $r(t)$ are

$$\mu(t|t-1) = y_{t|t-1} = x_{t|t-1} = F\, x_{t-1|t-1}, \tag{25}$$

$$\sigma^2(t|t-1) = F\, P_{t-1|t-1}\, F^T + Q_{t-1|t-1} + R_{t-1|t-1}, \tag{26}$$

while the estimated mean and variance are $\mu(t|t) = y_{t|t} = x_{t|t}$ and $\sigma^2(t|t) = P_{t|t} + R_{t|t}$. The VASB method predicts (and estimates) the stochastic local level Eq. (25) and stochastic volatility Eq. (26) jointly at every time step t.

5. Error reduction target

When applying the VASB algorithm to real financial market data which often deviate from normal distributions, the process of sequentially updating $x_{t|t}$, $P_{t|t}$ and $R_{t|t}$ may exhibit a problematic tendency: the ratio $tr(P_{t|t})/tr(R_{t|t})$ may become higher and higher or lower and lower, here $tr(\cdot)$ is the trace of a square matrix. This suggests a positive feedback loop for over- or under-fitting. By the nature of a Bayesian update, the overall estimation error associated with the posterior is smaller than the overall prediction error associated with the prior. At a time step t, an overly large error reduction causes over-fitting while an unduly small reduction causes under-fitting.

According to the Appendix B, rescaling the two initial values $P_{t|t}^{(0)}$ and $R_{t|t}^{(0)}$ in Eqs. (8) and (9) before making the VA iteration Eqs. (10) to (14) can make the VASB aiming at a specified level of error reduction and, therefore, preventing both over- and under-fitting. Assume that an error reduction target g, $0 < g < 1$, represents a desired variance ratio of the lowered estimation error to the original prediction error. When the state space model Eqs. (2) and (3) is defined by Eqs. (20) and (21) or (23) and (24), i.e. when H_t is a $1 \times m$ matrix, this initial value rescaling adjustment can be accomplished by inserting between the VASB Eqs. (9) and (10) the following two rescaling statements derived by the Appendix B:

$$P_{t|t}^{(0)} = [\,(\,1 - g^{1/2}\,)\,S_t^{(0)} \,/\, (\,S_t^{(0)} - R_{t|t}^{(0)}\,)\,]\,P_{t|t}^{(0)}, \tag{27}$$

$$R_{t|t}^{(0)} = [\,g^{1/2}\,S_t^{(0)} \,/\, R_{t|t}^{(0)}\,]\,R_{t|t}^{(0)}, \tag{28}$$

where the values in the brackets, [·], are rescaling coefficients, and $S_t^{(0)} = H_t\,P_{t|t}^{(0)} H_t^T + R_{t|t}^{(0)}$ by Eq. (10). The covariance matrix $P_{t-1|t-1}$ in Eq. (18) should be rescaled by Eq. (27) as well. An error reduction target $g = 0.9$ is a sensible example.

The error reduction target g controls the relative allocation of the randomness between $P_{t|t}$ and $R_{t|t}$. It serves a similar function as the signal-to-noise ratio q in the adaptive state space model suggested by Koopman and Bos (2004); as the multiplier λ in the Hodrick–Prescott filter (Hodrick and Prescott 1997); and as the smoothness parameter or noise ratio λ in the flexible least squares (FLS) method discussed by Markov *et al.* (2006). Unlike q and λ, the range of g is clearly limited: $0 < g < 1$.

6. Case 1: simulations

Let's at first examine some simulations to compare the estimates and forecasts made by the VASB algorithm, the rolling and time-weighted OLS analysis and the Kalman filter when "true" values of the hidden state x_t are already known. To make the simulations relevant to financial and investment analysis, they are carried out based on monthly data of U.S. equity market indices.

6.1. Data

The monthly data are price returns of Dow Jones Industrial (DJI) Index and total returns of Russell Indices from January 1979 to June 2013, obtained from FactSet (www.factset.com). The first 4½ years are used to estimate the initial results of or initial values for the OLS, KF and

VASB. The next 15 years from July 1983 to June 1998 (Period I) serve as in-sample data for model parameter selection. The last 15 years from July 1998 to June 2013 (Period II) can function as out-of-sample data. The entire 30 years of Periods I and II are combined together for some sample statistics of the estimates and forecasts.

6.2. "True" values of the hidden state

In the simulations, the independent variables include a constant, monthly returns of Russell 3000, $r_{R3K}(t)$, differential returns of Russell 3000 Growth minus Russell 3000 Value, $r_{R3G-R3V}(t)$, and those of Russell 1000 minus Russell 2000, $r_{R1K-R2K}(t)$. The Russell 1000, 2000 and 3000 represent U.S. equity markets of large-, small- and all-capitalization stocks. The simulated response $y_t = y(t)$ is a linear combination of the independent variables plus a random disturbance,

$$y_t = y(t) = H_t\, x_t + v_t, \tag{29}$$

where the independent variables,

$$H_t = H(t) = [\ 1,\ r_{R3K}(t),\ r_{R3G-R3V}(t),\ r_{R1K-R2K}(t)\ ], \tag{30}$$

the known time-varying state (or the linear combination coefficients),

$$x_t = x(t) = [\ b_0(t), b_{R3K}(t), b_{R3G-R3V}(t), b_{R1K-R2K}(t)\ ]^T, \tag{31}$$

and $v_t = v(t)$ is the random disturbance. An element $b_j(t)$ of x_t can be set as time-varying or constant 0. The disturbances v_t are random samples from normal distributions of zero-mean and time-varying variance,

$$v_t \sim N(\ 0,\ \sigma^2_{DJI-R3K}(t; T_v)\ ), \tag{32}$$

where $\sigma^2_{DJI-R3K}(t; T_v)$ is trailing T_v-month variance of differential returns of DJI minus Russell 3000, $r_{DJI-R3K}(t)$, and $12 \le T_v \le 36$.

Two U.S. equity investment strategies are simulated with known state x_t. An "all-capitalization market timing strategy" is with the true value of time-varying state,

$$x_t = x(t) = [\,0, (1 + \tfrac{1}{4}\,\sin\tfrac{2\pi}{T_C} t), 0, 0\,]^T, \tag{33}$$

where $T_C = 60$, the nonzero time-varying coefficient $b_{R3K}(t)$ experiences a full cycle in every 60 months. A "time-varying small-capitalization growth style market neutral strategy" is with the true value of state,

$$x_t = x(t) = [\,0, 0, (1 + \tfrac{1}{4}\,\sin\tfrac{2\pi}{T_C} t), (-1 + \tfrac{1}{4}\,\cos\tfrac{2\pi}{T_C} t)\,]^T. \tag{34}$$

With each known time-varying state x_t by Eq. (33) or (34), the observation generating process Eq. (29) is applied 100 times. Each time, the value of parameter T_v in Eq. (32) is randomly chosen from $\{12, 13, \cdots, 36\}$.

Then the estimated state $x_{t|t}$, estimated response $y_{t|t} = H_t\, x_{t|t}$, forecasted state $x_{t|t-1} = x_{t-1|t-1}$ and forecasted response $y_{t|t-1} = H_t\, x_{t|t-1}$ are obtained by the rolling OLS regression, the Kalman filter and the VASB algorithm, with their parameters taking a range of values. The bias and dispersion of the estimated and forecasted states and responses are measured by sample mean and standard deviation of monthly differences $\Delta x_{t|t} = x_{t|t} - x_t$, $\Delta y_{t|t} = y_{t|t} - y_t$, $\Delta x_{t|t-1} = x_{t|t-1} - x_t$ and $\Delta y_{t|t-1} = y_{t|t-1} - y_t$ over the sample of 36,000 monthly data obtained from the 100 simulations with 360 monthly estimates or forecasts (in the Periods I and II) in each simulation.

6.3. Solutions by OLS, KF and VASB

The biases and dispersions of the response $y_{t|t}$ and state $x_{t|t}$ estimated by the rolling T_r-month OLS regressions in the simulation of "all-capitalization market timing strategy" with Eq. (33) are shown in the top panel of Table 1.1, where $T_r = \{18, 24, \cdots, 42\}$. The first data column RMSE (root mean square error) is RMS of $y_{t|t} - y_t$. The remaining columns are sample mean and standard deviation of individual elements of $x_{t|t} - x_t$. The biases and dispersions of the response $y_{t|t-1}$ and state $x_{t|t-1}$ forecasted by the same rolling T_r-month OLS regressions in the same simulation are shown in the top panel of Table 1.2. The first data column RMSE is the RMS of $y_{t|t-1} - y_t$. The remaining columns are the sample mean and standard deviation of elements of $x_{t-1|t-1} - x_t$. The biases and dispersions of the estimates and forecasts in the simulation of "time-varying small-capitalization growth style market neutral strategy" with Eq. (34) are shown in Tables 2.1 and 2.2.

The predicted and estimated states and responses, $x_{t|t-1}$, $x_{t|t}$, $y_{t|t-1}$ and $y_{t|t}$, for the time-varying regression Eq. (29) can be computed by the Kalman filter (Simon 2006), assuming the transition matrix $F_t = I_m$. The transition and measurement error covariance matrices, Q_t and R_t, are among the known inputs required by the KF which can be implemented by Eqs. (5), (6), (7), (A.17), (10), (11), (15), (16) and (17), with $L = k = 0$. The published adaptive KFs that we have read estimate only one of Q_t and R_t independently (Ding *et al.* 2006 b and 2007, Koopman and Bos 2004, Särkkä and Hartikainen 2013, Särkkä and Nummenmaa 2009, Šmídl and Quinn 2006 a). Fortunately, the rolling OLS can estimate both Q_t and R_t for the KF. At each time t, a trailing OLS regression estimates the variance $P_{t|t}^{OLS}$ of the regression coefficients and $R_{t|t}^{OLS}$ of the residual error. The biases and dispersions of the estimates or forecasts by the KF of time-varying $Q_t = Q_{t|t}^{OLS} = q_{nne}\left(P_{t|t}^{OLS} - P_{t-1|t-1}^{OLS}\right)$ based on Eq. (18) and $R_t = R_{t|t}^{OLS}$ according to Eq. (19), corresponding to $T_r = \{18, 24, \cdots, 42\}$, are shown in the second panel of the Tables 1.1, 1.2, 2.1 and 2.2, where the trivial initial values $x_{0|0} = 0$ and $P_{0|0} = 0$ are applied.

The transition error u_t and measurement error v_t for the KF can be treated as stationary processes of constant variances $Q = 0$ by a stationary Eq. (18) with $F_t = I_m$ and $R = \langle R_{t|t}^{OLS} \rangle$, the sample mean of $R_{t|t}^{OLS}$ over the in-sample Period I. The biases and dispersions of the estimates or forecasts by the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$, corresponding to $T_r = \{18, 24, \cdots, 42\}$, are shown in the third panel of the Tables 1.1, 1.2, 2.1 and 2.2, where $x_{0|0} = 0$ and $P_{0|0} = 0$.

The biases and dispersions of the estimates or forecasts by the VASB of $F_t = I_m$ and various levels of error reduction target g are shown in the bottom panel of the Tables 1.1, 1.2, 2.1 and 2.2, where $g = \{0.6, 0.7, 0.8, 0.94, n/a\}$. The value $g = n/a$ means not to make the adjustment by Eqs. (27) and (28). The initial values for the VASB are $x_{0|0} = 0$, $P_{0|0} = 0$, and $Q_{0|0}$ and $R_{0|0}$ are estimated by the 36-month OLS regression. The VASB parameters T_0 in Eqs. (13) and (14) and L in Eqs. (15), (17) and (19) are set to $T_0 = 6$ and $L = 5$.

6.4. Estimates vs. forecasts

Interesting contrasts between the estimates and forecasts can be observed based on the biases and dispersions. The Tables 1.1 and 2.1 indicate an apparent tradeoff between estimating the state by $x_{t|t}$ with lower standard deviation of $x_{t|t} - x_t$ (StDev) and fitting the response by $y_{t|t}$ with lower RMS of $y_{t|t} - y_t$ (RMSE): a model having lower StDev is typically associated with a higher RMSE. The Tables 1.2 and 2.2, however, indicate a similarity between forecasting the state by $x_{t|t-1} = x_{t-1|t-1}$ with lower StDev of $x_{t|t-1} - x_t$ and forecasting the response by $y_{t|t-1} = H_t\, x_{t-1|t-1}$ with lower RMS of $y_{t|t-1} - y_t$: a model having lower StDev is typically associated with a lower RMSE. A comparison between the Tables 1.1 and 1.2 or 2.1 and 2.2 reveals that the estimated and forecasted states, $x_{t|t}$ and $x_{t|t-1}$, have very similar biases and dispersions; while the estimated and forecasted responses, $y_{t|t}$ and $y_{t|t-1}$, are obviously different: better estimates of lower RMSEs often result in worse forecasts of higher RMSEs. These are outcomes of a basic observation in the financial and investment analysis: a better estimate may come from an over-fitting, while a better forecast needs to avoid both over- and under-fitting.

An estimator (or forecaster) achieving smaller StDev (with similar or smaller bias) and smaller RMSE can be regarded as a better estimator (or forecaster). The KF of time-varying $Q_t = Q_{t|t}^{OLS}$ and $R_t = R_{t|t}^{OLS}$ is a better estimator and forecaster than the rolling OLS. The KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$ is a better forecaster than the KF of time-varying Q_t and R_t because the lack of time-variability in Q and R reduces the chance of over-fitting. In this simulation study, the VASB of a lower g (to prevent under-fitting) is as a good estimator as the KF of time-varying Q_t and R_t, while the VASB of a higher g (to prevent over-fitting) is as a good forecaster as the KF of constant Q and R.

7. Case 2: examples of market forecasts

Now let's compare the VASB algorithm with the rolling or time-weighted OLS analysis and the Kalman filter by using them to make actual forecasts in real financial market settings.

7.1. Forecasting DJI by Russell indices

When predicting monthly returns of Dow Jones Industrial Index using the most recent monthly returns of Russell Indices, the VASB is able to make better forecasts than the rolling OLS and the KF of Q_t and R_t estimated by OLS. The predictability is measured by information ratio (IR) and information coefficient (IC) of the forecasts.

As in the simulation study, the monthly price returns of DJI index and the total returns of Russell indices from January 1979 to June 2013 are utilized. The first 4½ years are used to estimate the initial results of or initial values for the OLS, KF and VASB. The next 15 years from July 1983 to June 1998 (Period I) serve as the in-sample data for model selection. The last

15 years from July 1998 to June 2013 (Period II) are set aside as the out-of-sample data to evaluate the selected forecasting models. These two time periods of equal length, as well as the price or total returns of the indices, are chosen by our data availability, not by any analysis on the conditions of the markets or on the relative performances of the OLS, KF and VASB forecasts.

Our multivariate regression model forecasting the DJI returns is based on the following reasoning. According to the investment theories (e.g. arbitrage pricing theory) and observations (e.g. trend following practices), the DJI returns $r_{DJI}(t)$ may be linearly explained by (a) concurrent returns of Russell 3000, $r_{R3K}(t)$, differential returns of Russell 3000 Growth minus Russell 3000 Value, $r_{R3G-R3V}(t)$, and those of Russell 1000 minus Russell 2000, $r_{R1K-R2K}(t)$, to account for the equity markets and styles; (b) lagged DJI returns, $r_{DJI}(t-1) = r_{DJI-R3K}(t-1) + r_{R3K}(t-1)$, due to a certain level of serial correlation; (c) lagged DJI return residuals uncorrelated with the Russell indices: possible residual reversal behavior; and (d) time-varying means of these returns, $\mu_{DJI-R3K}(t|t)$, $\mu_{R3K}(t|t)$, $\mu_{R3G-R3v}(t|t)$ and $\mu_{R1K-R2K}(t|t)$, representing the most recent trends in the equity markets. Therefore, the predictable component of $r_{DJI}(t)$ may be forecasted 1-month ahead by a multivariate time-varying regression in a form of Eq. (29) as,

$$y_t = H_t\, x_t + v_t, \tag{35}$$

where y_t is a scalar and H_t a 1×9 matrix:

$$y_t = r_{DJI}(t), \tag{36}$$

$$\begin{aligned} H_t = [\,1,\ & r_{DJI-R3K}(t-1),\ r_{R3K}(t-1),\ r_{R3G-R3V}(t-1), \\ & r_{R1K-R2K}(t-1),\ \mu_{DJI-R3K}(t-1|t-1),\ \mu_{R3K}(t-1|t-1), \\ & \mu_{R3G-R3V}(t-1|t-1),\ \mu_{R1K-R2K}(t-1|t-1)\,]. \end{aligned} \tag{37}$$

The time-varying mean of an independent variable, $\mu_j(t-1|t-1)$ in H_t, can be estimated by time-weighted average,

$$\mu_j(t|t) = (1 - w_0)\,\mu_j(t-1|t-1) + w_0\, r_j(t), \tag{38}$$

where $0 < w_0 < 1$. After solving for the regression coefficients $x_{t|t}$ by the rolling OLS, the KF or the VASB, the 1-month ahead forecast is either of the following two predictions:

$$r_{DJI}(t+1|t) = y_{t+1|t} = H_{t+1} x_{t+1|t} = H_{t+1}\, F_{t+1}\, x_{t|t}, \tag{39}$$

$$r_{DJI}(t+1|t) = y_{t|t} = H_t\, x_{t|t}, \tag{40}$$

where the elements of the estimated state vector $x_{t|t}$ can be denoted as

$$\begin{aligned} x_{t|t} = [\, & b_0(t), b^{(r)}_{DJI-R3K}(t), b^{(r)}_{R3K}(t), b^{(r)}_{R3G-R3V}(t), b^{(r)}_{R1K-R2K}(t), \\ & b^{(\mu)}_{DJI-R3K}(t), b^{(\mu)}_{R3K}(t), b^{(\mu)}_{R3G-R3V}(t), b^{(\mu)}_{R1K-R2K}(t)\,]^T. \end{aligned} \tag{41}$$

The predictability of the forecasts by Eq. (39) or (40) is commonly measured by performances of actionable investment decisions based on the forecasts. According to Grinold and Kahn (1992), the information ratio (IR) and information coefficient (IC) are two effective measures. If residual return relative to a benchmark is referred to as alpha, IR is the ratio of annualized alpha to annualized residual risk and IR is proportional to the t-statistic of the time-varying alpha, while IC is the correlation between the forecasted and the subsequently realized alphas. Assume that the realized monthly returns of the DJI index is denoted by $r(t)$, and the 1-month ahead forecasts by $r(t|t-1)$. An actionable investment decision in our case is to take long or short positions proportional to the forecasts and, therefore, the long/short investment return in the month t is proportional to the product $r(t|t-1)\, r(t)$. The average investment performance over a given time period $t \in [t_1, t_2]$ is,

$$f(t_1, t_2) = (t_2 - t_1 + 1)^{-1} \sum_{t=t_1}^{t_2} r(t)\, r(t|t-1). \tag{42}$$

The statistical significance of the forecasts $r(t|t-1)$ can be measured by the t-statistic of the performance $f(t_1, t_2)$. The effectiveness of the forecasts can be measured by the normalized value of the performance,

$$\rho(t_1, t_2) = [\, \textstyle\sum_{t=t_1}^{t_2} r(t)\, r(t|t-1)\,] \times$$
$$\times\, [\, \textstyle\sum_{t=t_1}^{t_2} r(t)^2\,]^{-1/2}\, [\, \textstyle\sum_{t=t_1}^{t_2} r(t|t-1)^2\,]^{-1/2}. \qquad (43)$$

Following the notions by Grinold and Kahn (1992), in the context of the long/short investment with cash as the benchmark, the performance $f(t_1, t_2)$ by Eq. (42) is the alpha, the t-statistic of $f(t_1, t_2)$ is proportional to the IR, and the normalized performance $\rho(t_1, t_2)$ by Eq. (43) is the IC. A higher t-statistic (IR) and/or ρ (IC) indicate a higher predictive power of the forecasts.

7.2. Results of DJI forecasts

Various rolling OLS regressions, Kalman filters and VASB models similar to those tested in the simulation Section 6 are utilized to forecast the monthly DJI returns.

The performances of forecasting DJI returns by rolling T_r-month OLS regression, $T_r = \{12, 18, \cdots, 42\}$, are shown in Table 3.1, where the upper panel is for the $y_{t+1|t}$ by Eq. (39), the lower panel for the $y_{t|t}$ by Eq. (40), the left panel for the in-sample Period I, and the right panel for the out-of-sample Period II. The 4 data columns for each time period are the t-statistic (proportional to IR) of $f(t_1, t_2)$ by Eq. (42), the level of $\rho(t_1, t_2)$ (or IC) by Eq. (43), and the RMSE and standard deviation (StDev) of the monthly forecasts $y_{t+1|t}$ or $y_{t|t}$. The time-varying means $\mu_j(t|t)$ in H_t of Eq. (37) are calculated by the time-weighted averages Eq. (38) with $w_0 = 1/6$. The best in-sample forecasting models identified by the in-sample IR or IC, the "$y_{t+1|t}$ by 42-Mon OLS" and "$y_{t|t}$ by 24- or 30-Mon OLS", are not the best out-of-sample forecasting models measured by the out-of-sample IR or IC.

Sample statistics of the 9 time-varying regression coefficients, i.e. the elements of $x_{t|t}$ as in Eq. (41), by the rolling 36-month OLS are shown in Table 3.2, where the upper panel is for the 180-month Period I, and the lower panel for the 180-month Period II. In each period, the sample distribution of t-statistic of a rolling regression coefficient is summarized by the 25th, 50th and 75th percentiles. The p-value of a regression coefficient is of the null hypothesis H_0 that the expected value of the coefficient is 0. Here the H_0 cannot be rejected for the coefficients of growth-value spreads $r_{R3G-R3V}(t)$ and $\mu_{R3G-R3V}(t|t)$. The sample mean, standard deviation and serial correlation of the monthly regression coefficients are also shown in the Table 3.2 to indicate the average effect of an independent variable on the forecasts.

The performances of forecasting DJI returns by the KF of time-varying Q_t and R_t are shown in Table 4, where $Q_t = Q_{t|t}^{OLS} = q_{nne}(P_{t|t}^{OLS} - P_{t-1|t-1}^{OLS})$ based on Eq. (18), $R_t = R_{t|t}^{OLS}$ according to Eq. (19), and $x_{0|0} = 0$ and $P_{0|0} = 0$. A comparison of the IR and IC between the Tables 4 and 3.1 reveals that the performances of DJI forecasts by these KF estimates are not as good as those by the rolling OLS analysis. According to the discussions in Section 6, over-fitting by the KF of time-varying Q_t and R_t could be the primary reason for the poor forecasts.

The performances of forecasting DJI returns by the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$ are shown in Table 5.1, where $Q = 0$ by a stationary Eq. (18) with $F_t = I_m$, and $\langle R_{t|t}^{OLS} \rangle$ is the sample mean of $R_{t|t}^{OLS}$ over the in-sample Period I. The best forecasting models in the in-sample Period I, the "$y_{t+1|t}$ and $y_{t|t}$ by KF of $Q = 0$ and $R = \langle R_{t|t}^{(42M\ OLS)} \rangle$", are also the best forecasting models in the out-of-sample Period II. A comparison of the IR and IC between the Tables 5.1 and 3.1 demonstrates that the out-of-sample performances of DJI forecasts by these KF estimates are consistently better than those by the rolling OLS analysis. Sample statistics of the 9 time-varying elements of the state $x_{t|t}$ estimated by the KF of $Q = 0$ and $R = \langle R_{t|t}^{(42M\ OLS)} \rangle$ are shown in Table 5.2. According to the mean, standard deviation and t-statistic, the variables

$r_{R3K}(t)$ and $r_{R3G-R3V}(t)$ may have weaker effects on the forecasts.

The performances of forecasting DJI returns by the VASB algorithm Eqs. (5) to (19) with $F_t = I_m$ are shown in Table 6.1, where the error reduction target $g = \{0.84, 0.86, \cdots, 0.94, n/a\}$. The value $g = n/a$ means not to make the adjustment by Eqs. (27) and (28). The initial values for the VASB are $x_{0|0} = 0$, $P_{0|0} = 0$, and $Q_{0|0}$ and $R_{0|0}$ are estimated by the 36-month OLS regression. The VASB parameters T_0 in Eqs. (13) and (14) and L in Eqs. (15), (17) and (19) are set to $T_0 = 6$ and $L = 5$. The best forecasting models in the in-sample Period I are the "$y_{t+1|t}$ and $y_{t|t}$ by VASB of $g = 0.94$ and $g = n/a$". The $y_{t+1|t}$ by $g = n/a$ is the best forecasting model in the out-of-sample Period II, while the $y_{t|t}$ by $g = 0.94$ is the worst in Period II. Sample statistics of the 9 time-varying elements of the state $x_{t|t}$ estimated by the VASB of $g = 0.94$ are shown in Table 6.2. According to the t-statistic, p-value, mean and standard deviation, the differential returns $r_{R3G-R3V}(t)$ and $r_{R1K-R2K}(t)$ may have no or weaker effects on the forecasts.

Table 7 compares the selected (i.e. the best in-ample) and the best out-of-sample (whether or not being selected) OLS, KF and VASB forecasting models collected from the Tables 3.1, 5.1 and 6.1. The selected VASB of $g = 0.94$ and $g = n/a$ models have higher or at least similar predictive power relative to the best rolling OLS or the best KF of constant Q and R models with $y_{t+1|t}$ or $y_{t|t}$ forecasts in the in-sample Period I or out-of-sample Period II.

The impact of the assumption about the transition matrix F_t on the performances of DJI forecasts is reviewed by Table 8. When applying the VASB of $g = 0.94$ or $g = n/a$ as in the Table 6.1, the use of $f_0 < 1$ in Eq. (4) will reduce the predictive power of the in-sample and out-of-sample forecasts.

The forecasts of DJI returns shown in the Tables 3.1, 5.1 and 6.1 (by the OLS, KF and

VASB) are all based on exactly the same set of independent variables H_t in Eq. (37), in which the time-varying means, $\mu_j(t|t)$, are estimated by the time-weighted averages Eq. (38) with $w_0 = 1/6$. When these time-varying means, $\mu_j(t|t)$, are estimated by the VASB stochastic local level model Eqs. (23) to (26) with $F_t = F_\mu = 1$ and $g = g_\mu = 0.6$ instead, the performances of forecasting DJI returns by the VASB are further improved, as shown in Table 9 (vs. the Table 7). The VASB of $g = 0.94$ and $g = n/a$ are again the best forecasting models in the in-sample Period I, and are the best or good forecasting models in the out-of-sample period II. Both the VASB stochastic local level solutions for the time-varying means $\mu_j(t|t)$ in Eq. (37) and the VASB state space solutions for the time-varying regression Eqs. (35) to (37) contribute to the increase in the predictive power of forecasting the DJI returns.

7.3. Forecasting time-varying mean and variance

When predicting time-varying mean and variance of monthly returns of 19 financial market indices listed in Table 10, the VASB algorithm is able to make more accurate forecasts than the rolling and time-weighted statistics in a majority of cases.

Monthly returns of the 19 indices from January 1990 to June 2013 are obtained from Morningstar EnCorr database (corporate.morningstar.com). The first 4 years are used to estimate the initial results of or initial values for the calculations. The remaining 19½ years or 234 months from January 1994 to June 2013 is the period for model evaluation.

For each of the 19 index return time-series $r_j(t)$, $j = 1, 2, \cdots, 19$, the 1-month ahead forecasts of time-varying mean and variance are denoted as $\mu_j(t|t-1;\theta)$ and $\sigma_j^2(t|t-1;\theta)$, where θ is a method-specific set of parameters. The time-varying accuracy of the forecasted mean and variance can be evaluated jointly using an l-month log-likelihood (Litterman and

Winkelmann 1998, Patton 2011). For the j-th index, the l-month log-likelihood of the forecasts is:

$$\ln L_j(t;\theta) = (-1/2)\sum_{k=t-l+1}^{t}\{\ln 2\pi + \ln\sigma_j^2(k|k-1;\theta) + \\ + [\,r_j(k) - \mu_j(k|k-1;\theta)\,]^2 / \sigma_j^2(k|k-1;\theta)\,\}. \quad (44)$$

A higher level of the log-likelihood indicates more accurate forecasts (Litterman and Winkelmann 1998). Patton (2011) proved that the log-likelihood belongs to a family of robust loss functions for volatility forecasts. Over a given time period, the average of the l-month log-likelihood can serve as a measure of the overall accuracy of the forecasts. In our example, we use the 234-month average of 12-month log-likelihood over the 19½ years of evaluation period.

For the j-th index $r_j(t)$, denote the 12-month log-likelihood Eq. (44) of the best rolling forecasts as $\ln L_j(t;\theta_j^R)$, that of the beset time-weighted forecasts as $\ln L_j(t;\theta_j^W)$ and that of the best VASB forecasts as $\ln L_j(t;\theta_j^B)$, and denote difference in the 12-month log-likelihood of the best time-weighted vs. the best rolling forecasts as $\Delta\ln L_j(t;\theta_j^W,\theta_j^R) = \ln L_j(t;\theta_j^W) - \ln L_j(t;\theta_j^R)$, that of the best VASB vs. the best rolling forecasts as $\Delta\ln L_j(t;\theta_j^B,\theta_j^R) = \ln L_j(t;\theta_j^B) - \ln L_j(t;\theta_j^R)$ and that of the best VASB vs. the best time-weighted forecasts as $\Delta\ln L_j(t;\theta_j^B,\theta_j^W) = \ln L_j(t;\theta_j^B) - \ln L_j(t;\theta_j^W)$. Then relative accuracy between a pair of the best forecasts can be measured by average and p-value of the 234 monthly differences in 12-month log-likelihood. A positive average difference indicates that the forecasts by the first method are more accurate, while a small p-value of the differences ($p < 0.05$, for example) indicates that the average difference is statistically significant.

7.4. Results of forecasting mean and variance

The 1-month ahead forecasts of the time-varying mean and variance of the j-th index $r_j(t)$ by the traditional rolling statistics are,

$$\mu_j(t|t-1;T_m,T_v) = \mu_j(t-1|t-1;T_m,T_v) =$$
$$= T_m^{-1} \textstyle\sum_{k=t-T_m}^{t-1} r_j(k), \qquad (45)$$

$$\sigma_j^2(t|t-1;T_m,T_v) = \sigma_j^2(t-1|t-1;T_m,T_v) =$$
$$= (T_v-1)^{-1} \textstyle\sum_{k=t-T_v}^{t-1} [\, r_j(k) - \mu_j(k|k-1;T_m,T_v)\,]^2, \qquad (46)$$

where the parameter set $\theta = (T_m, T_v)$ with T_m is the window size for forecasting the mean and T_v the window size for the variance. The best rolling forecasts with the highest 234-month average of 12-month log-likelihood can be denoted as $\theta_j^R = (T_{m,j}^R, T_{v,j}^R)$. The first 3 data columns of Table 11 show the parameters and the average log-likelihood of the best rolling forecasts for each of the 19 indices, where the candidate values of the parameters are $\{T_m, T_v\} = \{6, 12, \cdots, 48\}$.

The 1-month ahead forecasts of the time-varying mean and variance of $r_j(t)$ by the traditional time-weighted statistics are,

$$\mu_j(t|t-1;T_m,T_v) = \mu_j(t-1|t-1;T_m,T_v) =$$
$$= T_m^{-1}\, r_j(t-1) + (1-T_m^{-1})\, \mu_j(t-2|t-2;T_m,T_v),$$

$$\sigma_j^2(t|t-1;T_m,T_v) = \sigma_j^2(t-1|t-1;T_m,T_v) =$$
$$= T_v^{-1}\, [\, r_j(t-1) - \mu_j(t-1|t-2;T_m,T_v)\,]^2 +$$
$$+ (1-T_v^{-1})\, \sigma_j^2(t-2|t-2;T_m,T_v),$$

where the parameter set $\theta = (T_m, T_v)$ with $1/T_m$ is the weight of the last data point for forecasting the mean and $1/T_v$ the weight for the variance. The best time-weighted forecasts with the highest 234-month average of 12-month log-likelihood can be denoted as $\theta_j^W = (T_{m,j}^W, T_{v,j}^W)$. The middle 3 data columns of the Table 11 show the parameters and the average log-likelihood of the best time-weighted forecasts, where the candidate values of the parameters are $\{T_m, T_v\} = \{6, 12, \cdots, 48\}$.

The 1-month ahead forecasts of the time-varying mean and variance of $r_j(t)$ by the VASB solutions to stochastic local level model Eqs. (23) to (26) are of three parameters: the state transition coefficient F in Eqs. (25), (26) and (18), the error reduction target g in Eqs. (27) and (28), and the parameter T_0 in Eqs. (13) and (14). The rolling forecasts Eqs. (45) and (46) indicate that $1/T_m$ equals to the ratio of the variance of forecasted mean over the forecasted variance, and $1/T_v$ is the weight of data to estimate the variance. In the VASB solutions to the stochastic local level model, the variance of forecasted mean in Eq. (25) is $FP_{t-1|t-1}F^T + Q_{t-1|t-1} \approx P_{t-1|t-1}$ if the random error processes are relatively stationary, and the forecasted variance $\sigma_j^2(t|t-1;\theta) = FP_{t-1|t-1}F^T + Q_{t-1|t-1} + R_{t-1|t-1} \approx P_{t-1|t-1} + R_{t-1|t-1} = S_t^{(0)}$ by Eqs. (26), (24) and (10). According to the discussion about (B.8) in the Appendix B, if the fraction $1/T_m$ is a desired level for the variance ratio $P_{t-1|t-1}/S_t^{(0)}$, the error reduction target g is a function of T_m as

$$g = [1 - (1/T_m)]^2. \qquad (47)$$

The VASB Eqs. (13) and (14) show that the parameter $1/T_0$ serves as the weight of data to estimate the variance $P_{t|t} + R_{t|t}$, i.e. T_0 and T_v are equivalent to each other. With these parameter relations, the VASB forecasts of the time-varying mean and variance are,

$$\mu_j(t|t-1;F,T_m,T_v) = x_{t|t-1} = F\,x_{t-1|t-1},$$

$$\sigma_j^2(t|t-1;F,T_m,T_v) = F\,P_{t-1|t-1}F^T + Q_{t-1|t-1} + R_{t-1|t-1},$$

where the parameter set $\theta = (F, T_m, T_v)$. The best VASB forecasts with the highest 234-month average of 12-month log-likelihood can be denoted as $\theta_j^B = (F_j^B, T_{m,j}^B, T_{v,j}^B)$. The last 4 data columns of the Table 11 show the parameters and the average log-likelihood of the best VASB forecasts, where the candidate values of the parameters are $F = \{0.90, 0.92, \cdots, 1.00\}$, $g = \{[1 - (1/T_m)]^2 \colon T_m = 6, 12, \cdots, 48, n/a\}$ and $T_0 = \{T_v \colon T_v = 6, 12, \cdots, 48\}$. The value $T_m = n/a$ means not to make the adjustment by Eqs. (27) and (28). The initial values $x_{0|0}$, $P_{0|0}$, $Q_{0|0}$ and $R_{0|0}$ for the VASB are arithmetic averages of the estimates made in 25 different initial time windows $t \in [1, t_2]$, $t_2 = \{6, 7, \cdots, 30\}$.

To compare the best time-weighted vs. the best rolling forecasts, the first 3 data columns of Table 12 show the 234-month average of the 12-month log-likelihood difference $\Delta \ln L_j(t; \theta_j^W, \theta_j^R)$, the p-value of the null hypothesis for the difference, and whether the difference is statistically significant, i.e. whether the alternative hypothesis H_a can be considered by $p < 0.05$. The best time-weighted forecasts are more accurate for 16 of the 19 indices, and significantly more accurate for 12 of the indices. To compare the best VASB vs. the best rolling forecasts, the middle 3 data columns of the Table 12 show the average, p-value and H_a statistics of the log-likelihood difference $\Delta \ln L_j(t; \theta_j^B, \theta_j^R)$. The best VASB forecasts are more accurate for 18 of the 19 indices, and significantly more accurate for 13 of the indices. To compare the best VASB vs. the best time-weighted forecasts, the last 3 data columns of the Table 12 show the average, p-value and H_a statistics of the log-likelihood difference $\Delta \ln L_j(t; \theta_j^B, \theta_j^W)$. The best VASB forecasts are more accurate for 18 of the 19 indices, and significantly more accurate for 13 of the indices.

8. Conclusion

Successful time-varying forecasts are essential in many economic and financial analysis tasks. This paper developed a VASB (variational approximation of sequential Bayesian inference) algorithm for making the time-varying forecasts. The VASB, as an adaptive solution to the linear Gaussian state space models, outperforms the moving-window and time-weighted OLS analysis and performs better or as good as the Kalman filter in terms of higher predictive power. The VASB is more general and robust than the KF and adaptive KFs by not requiring *a priori* knowledge of any of the random error covariance matrices. Such requirements are not realistic in a typical economic or financial research. We expect the VASB algorithm to serve as a useful solution to the state space modeling in general and to the economic and financial forecasting in particular.

Reference

Alvarez, M. and Levinson, M., Hedge Fund Risk Modeling - Version 1.2. MSCI Barra, April 2007.

Andersen, T.G. and Benzoni, L., Stochastic Volatility. WP 2009-04, Federal Reserve Bank of Chicago, 2009.

Belsley, D.A., The Applicability of the Kalman Filter in the Determination of Systematic Parameter Variation. *Annals of Economic and Social Measurement*, 1973, **2** (5), 531–535.

Beal, M.J., Variational Algorithms for Approximate Bayesian Inference. A Thesis Submitted for the Degree of Doctor of Philosophy, University of London, 2003.

Bishop, C.M., *Pattern Recognition and Machine Learning*. 2006 (Springer Science+Business Media, LLC: New York).

Das, A. and Ghoshal, T.K., Market Risk Beta Estimation using Adaptive Kalman Filter. *International Journal of Engineering Science and Technology*, 2010, **2** (7), 1923–1934.

Date, P. and K. Ponomareva, K., Linear and non-linear filtering in mathematical finance: a review. *IMA Journal of Management Mathematics*, 2011, **22**, 195–211.

Ding, W., Wang, J. and Rizos, C., Stochastic modelling strategies in GPS/INS data fusion process. IGNSS Symposium, Gold Coast, Australia, 17-21 July 2006 (2006 a).

Ding, W., Wang, J. and Rizos, C., Improving Covariance Based Adaptive Estimation for GPS/INS Integration. IAIN/GNSS, Jeju, Korea, 18-20 October 2006 (2006 b).

Ding, W., Wang, J. and Rizos, C., Improving Adaptive Kalman Estimation in GPS/INS Integration. *Journal of Navigation*, 2007, **60**, 517–529.

Grimmer, J., An Introduction to Bayesian Inference via Variational Approximations. *Political Analysis*. 2011, **19**, 32–47.

Grinold, R.C., and Kahn, R.N., Information Analysis. *The Journal of Portfolio Management*, 1992, **18** (3), 14–21.

Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*. 1991 (Cambridge University Press: Cambridge, UK).

Hodrick, R.J. and Prescott, E.C., Postwar U.S. Business Cycle: an Empirical Investigation. *Journal of Money, Credit and Banking*, 1997, **29** (1), 1–16.

Javaheri, A., Lautier, D. and Galli, A., Filtering in Finance. *Wilmott Magazine*, 2003, May/June, 2–18.

Johannes, M. and Polson, N., Particle Filtering. In *Handbook of Financial Time Series*, edited by T.G. Andersen, R.A. Davis, J.-P. Kreiß and T.V. Mikosch, pp. 1015–1029, 2009 (Springer-Verlag: Berlin).

Kalman, R.E., A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 1960, **82** (1), 35–45.

Koopman, S.J. and Bos, C.S., State Space Models with a Common Stochastic Variance. *Journal of Business and Economic Statistics*, 2004, **22** (3), 346–357.

Litterman, R. and Winkelmann, K., Estimating Covariance Matrices. Working Paper, Goldman Sachs Group, Inc., 1998.

Lopes, H.F. and Tsay, R.S., Particle Filters and Bayesian Inference in Financial Econometrics. *Journal of Forecasting*, 2011, **30**, 168–209.

Mamaysky, H., Spiege, M. and Zhang, H., Estimating the Dynamics of Mutual Fund Alphas and Betas. *Review of Financial Studies*, 2008, **21** (1), 233–264.

Markov, M., Muchnik, I., Mottl, V. and Krasotkina, O., Dynamic Analysis of Hedge Funds. Third IASTED Conference on Financial Engineering and Applications, October 2006.

Muirhead, R.J., *Aspects of Multivariate Statistical Theory*. 2005 (John Wiley & Sons, Inc.: Hoboken).

Özkana, E., Šmídl, V., Saha, S., Lundquist, C. and Gustafsson, F., Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters, *Automatica*, 2013, **49**, 1566–1575.

Pasricha, G.K., Kalman Filter and its Economic Applications. Paper No. 22734, Munich Personal RePEc Archive (MPRA), 2006.

Patton, A.J., Volatility Forecast Comparison using Imperfect Volatility Proxies. *Journal of Econometrics*, 2011, **160** (1), 246–256.

Rao, C.R., *Linear Statistical Inference and its Applications, Second Edition*, 1973 (John Wiley & Sons, Inc.: New York, NY).

Särkkä, S., *Bayesian Filtering and Smoothing*, 2013, (Cambridge University Press: Cambridge, UK).

Särkkä, S. and Hartikainen, J., Non-linear noise adaptive Kalman filtering via variational Bayes, Paper presented at the IEEE International Workshop on Machine Learning for Signal Processing, Southampton, UK, 22-25 September, 2013.

Särkkä, S. and Nummenmaa, A., Recursive Noise Adaptive Kalman Filtering by Variational

Bayesian Approximations. *IEEE Transactions on Automatic Control*, 2009, **54** (3), 596–600.

Schön, T.B., Karlsson, R. and Gustafsson, F., The Marginalized Particle Filter: Analysis, Applications and Generalizations. *ESAIM: Proceedings*, 2007, **19**, 53–64.

Simon, D., *Optimal State Estimation*. 2006 (John Wiley & Sons, Inc.: Hoboken, NJ).

Šmídl, V. and Quinn, A., *The Variational Bayes Method in Signal Processing*. 2006 (Springer-Verlag: Berlin) (2006 a).

Šmídl, V. and Quinn, A., The Variational Bayes Approximation in Bayesian Filtering. Proceedings of the IEEE International Conference 2006 on Acoustics, Speech, and Signal Processing, 2006 (2006 b).

Šmídl, V. and Quinn, A., Variational Bayesian Filtering. *IEEE Transactions on Signal Processing*, 2008, **56** (11), 5020–5030.

Stock, J.H. and Watson, M.W., Implications of dynamic factor models for VAR analysis. Working Paper No. 11467, NBER, July 2005.

Stock, J.H. and Watson, M.W., Forecasting with Many Predictors. In *Handbook of Economic Forecasting*, *Volume 1*, edited by G. Elliott, C.W.J. Granger and A.G. Timmermann, pp. 515–554, 2006 (North-Holland: Amsterdam, Netherlands).

Tzikas, D.G., Likas, A.C. and Galatsanos, N.P., The Variational Approximation for Bayesian Inference. *IEEE Signal Processing Magazine*. November 2008, 131-146.

Witkovský, V., Computing the Distribution of a Linear Combination of Inverted Gamma Variables, *Kybernetika*, 2001, 37 (1), 79–90.

Zivot, E. and Wang, J., *Modeling Financial Time Series with S-PLUS, 2nd Ed*. 2006 (Springer-

Verlag: New York).

Appendix A. Derivations of the VASB algorithm

This appendix presents step-by-step derivations of our VASB (variational approximation of sequential Bayesian inference) algorithm Eqs. (5) to (19) for solving the state space model Eqs. (2) and (3) defined by the time-varying regression Eq. (1) and the assumptions and conditions.

A.1. Multivariate normal distribution and product of inverse-gamma distributions

Probability density function of an $m \times 1$ random vector x in a multivariate normal (or Gaussian) distribution is

$$p(x; \mu, \Sigma) = N(x; \mu, \Sigma) =$$
$$= (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp[-(1/2)\, (x-\mu)^T \Sigma^{-1} (x-\mu)] =$$
$$= (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp[-(1/2)\, tr(\Sigma^{-1}(x-\mu)(x-\mu)^T)], \qquad \text{(A.1)}$$

where μ is an $m \times 1$ expected value vector, Σ an $m \times m$ covariance matrix, $|\Sigma|$ the determinant of Σ, and the trace $tr(AB) = tr(BA)$. When only given the mean μ and covariance Σ, the normal distribution has the maximum entropy, i.e. is the most appropriate to be assumed, among all possible distributions on R^m (Rao 1973). Logarithm of the density function (A.1) is

$$\ln p(\, x;\ \mu, \Sigma\,) = \ln N(\, x;\ \mu, \Sigma\,) =$$
$$= (-1/2)\, [\, \ln|\Sigma| + tr(\Sigma^{-1} (x-\mu)\, (x-\mu)^T)\,] + const. \qquad \text{(A.2)}$$

When M and v are non-random matrix and vector, the expected value of a random matrix $(Mx + v)(Mx + v)^T$ is

$$E(Mx+v)(Mx+v)^T = M\Sigma M^T + (M\mu + v)(M\mu + v)^T, \qquad \text{(A.3)}$$

because of the definition $\Sigma = E(x-\mu)(x-\mu)^T$.

Assume that a univariate normal distribution ($m = 1$) is conditioned on variance $\Sigma = \sigma^2$, where σ^2 is also a random variable and its distribution can be refined by Bayes rule and data. It is well known that if the prior of σ^2 is an inverse-gamma distribution (a conjugate prior), the posterior of σ^2 is also inverse-gamma. The density function of $x = \sigma^2$ in an inverse-gamma distribution is

$$p(\,\sigma^2; \alpha, \beta\,) = p(\,x; \alpha, \beta\,) = IG(\,x; \alpha, \beta\,) =$$
$$= \Gamma(\alpha)^{-1}\,\beta^{\alpha}\,x^{-(\alpha+1)}\,\exp(\,-\beta/x\,), \qquad \text{(A.4)}$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters of $IG(x)$, and $\Gamma(\cdot)$ is gamma function. The expected values of x^{-1} and $(\ln x)$ are

$$E(\,x^{-1}\,) = \alpha\,/\,\beta, \qquad \text{(A.5)}$$

$$E(\,\ln x\,) = \ln\beta - \Gamma'(\alpha)\,/\,\Gamma(\alpha), \qquad \text{(A.6)}$$

where $\Gamma'(\alpha) = \partial\Gamma(\alpha)/\partial\alpha$.

When the likelihood of an observation y is Gaussian, $p(y|\sigma^2) = N(y; \mu, \sigma^2) \propto \sigma^{-1}\exp[-(y-\mu)^2/(2\sigma^2)]$, and the prior of the variance σ^2 is inverse-gamma, $p(\sigma^2) = IG(\sigma^2; \alpha_0, \beta_0) \propto (\sigma^2)^{-(\alpha_0+1)}\exp(-\beta_0/\sigma^2)$, then the posterior of σ^2 is, $p(\sigma^2|y) \propto p(y|\sigma^2)\,p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)}\exp(-\beta/\sigma^2)$, where the posterior shape and scale parameters are $\alpha = \alpha_0 + 1/2$ and, by (A.5), $\beta = E(\sigma^{-2})^{-1}\alpha = \beta_0 + (y-\mu)^2/2 = E_0(\sigma^{-2})^{-1}\alpha_0 + (y-\mu)^2/2$. The expected value of the posterior $E(\sigma^{-2})^{-1}$ is a weighted average of the expected value of the prior $E_0(\sigma^{-2})^{-1}$ and the data $(y-\mu)^2$. The prior shape parameter α_0 represents the weight, or strength, of the inverse-gamma prior.

If an unknown parameter β and three known ones, α_p, $\alpha_q \neq \alpha_p$ and β_q, are parameters of two inverse-gamma distributions, $p(x; \alpha_p, \beta) = IG(x; \alpha_p, \beta)$ and $q(x; \alpha_q, \beta_q) = IG(x; \alpha_q, \beta_q)$, and expected values $E_p(v) = \int_0^\infty v\,p(x; \alpha_p, \beta)\,dx$ and $E_q(v) = \int_0^\infty v\,q(x; \alpha_q, \beta_q)\,dx$, then both of the Kullback-Leibler divergences from $p(x)$ to $q(x)$, $KL(p \parallel q) = E_p(\ln[p(x; \alpha_p, \beta)/q(x; \alpha_q, \beta_q)]) \geq 0$, and the KL-divergence from $q(x)$ to $p(x)$, $KL(q \parallel p) = E_q(\ln[q(x; \alpha_q, \beta_q)/p(x; \alpha_p, \beta)]) \geq 0$, are minimized when the unknown scale parameter β satisfies the relation,

$$\alpha_p \,/\, \beta = E_p(x^{-1}) = \alpha_q \,/\, \beta_q = E_q(x^{-1}), \tag{A.7}$$

thanks to the expected values (A.5) and (A.6). The KL-divergence is the most utilized measure of difference between density functions of two distributions (Bishop 2006, Özkana *et al.* 2013).

If the $m \times m$ covariance matrix Σ in (A.1) is diagonal and its diagonal elements, $\Sigma_{jj} = \sigma_j^2$, are inverse-gamma variables and independent of one another, the joint distribution of $\{\sigma_j^2\}$ can be expressed as the product of the m inverse-gamma distributions (A.4),

$$p(\,\Sigma;\ \{\,\alpha_j,\ \beta_j\,\}\,) = \prod_{j=1}^{m} IG(\,\sigma_j^2;\ \alpha_j,\ \beta_j\,) =$$
$$= \prod_{j=1}^{m} [\,\Gamma(\alpha_j)^{-1}\,\beta_j^{\,\alpha_j}\,(\sigma_j^2)^{-(\alpha_i+1)}\exp(\,-\beta_j/\sigma_j^2\,)\,].$$

If there is no information on the strengths of individual σ_j^2, it is most appropriate to assume the same value for all shape parameters $\{\alpha_j\}$. When $\alpha_j = \alpha$ and B is a diagonal matrix with $B_{jj} = \beta_j$, the density function of the joint distribution becomes

$$p(\,\Sigma;\ \alpha, B\,) = IG_D(\,\Sigma;\ \alpha, B\,) =$$
$$= \Gamma(\alpha)^{-m}\,|B|^{\alpha}\,|\Sigma|^{-(\alpha+1)}\exp[\,-tr(\,B\,\Sigma^{-1}\,)\,]. \tag{A.8}$$

Since both of the two matrices B and Σ are diagonal, we will refer to the distribution $IG_D(\Sigma; \alpha, B)$ of Σ as a “diagonal inverse-gamma distribution” in this paper. Logarithm of the density function (A.8) is

$$\ln p(\,\Sigma;\ \alpha, B\,) = \ln IG_D(\,\Sigma;\ \alpha, B\,) =$$
$$= (-1)\,[\,(\alpha+1)\ \ln|\Sigma| + tr(\,B\,\Sigma^{-1}\,)\,] + const. \tag{A.9}$$

According to (A.5), the expected value of Σ^{-1} is

$$E(\,\Sigma^{-1}\,) = \alpha\,B^{-1}. \tag{A.10}$$

The scale parameter B can be expressed by the expected value as

$$B = E(\,\Sigma^{-1}\,)^{-1}\,\alpha. \tag{A.11}$$

The density function (A.8) or (A.9) is in a matrix form borrowed from inverse-Wishart distribution. A Wishart distribution is a multivariate generalization of gamma distribution (Muirhead 2005).

A.2. Sequential Bayesian inference

One of the latest progresses in dynamic state space modeling is Bayesian filtering (Särkkä 2013). Assuming Z_t denotes a set of unknown random variables, x_t, P_t, Q_t and R_t, to be estimated with the observations, y_t, $t = 1, 2, \cdots$, subject to a Markovian state space model, a Bayesian filter is to estimate the conditional distribution $p(Z_t|y_{1:t})$ based on the previously estimated one $p(Z_{t-1}|y_{1:t-1})$ and the current data y_t. If $p(y_t|Z_t)$ is the likelihood of y_t and the prior $p(Z_t|y_{1:t-1})$ is the dynamic prediction of Z_t, then the posterior,

$$p(\, Z_t \mid y_{1:t} \,) \propto p(\, y_t \mid Z_t \,)\, p(\, Z_t \mid y_{1:t-1} \,) \propto p(\, y_t,\, Z_t \mid y_{1:t-1} \,). \qquad \text{(A.12)}$$

The predictive prior $p(Z_t|y_{1:t-1})$ is calculated by combining the dynamic evolution $p(Z_t|Z_{t-1})$ and the previous posterior $p(Z_{t-1}|y_{1:t-1})$ using the Chapman-Kolmogorov equation, $p(Z_t|y_{1:t-1}) = \int p(Z_t|Z_{t-1})\, p(Z_{t-1}|y_{1:t-1})\, dZ_{t-1}$. In practice, the prediction may be obtained, estimated or assumed directly by analyzing the dynamic model without resorting to carrying out the integration. For the space model Eqs. (2) and (3) of unknown x_t and R_t (with known Q_t), Särkkä and Nummenmaa (2009) and Särkkä and Hartikainen (2013) proposed a two-random-variable variational Bayesian solution.

Based on the Gaussian errors and other assumptions imposed in the Sections 2 and 3, the measurement equation Eq. (3) gives rise to the likelihood of the observation y_t, which is Gaussian given the random parameters x_t and R_t,

$$p(\, y_t \mid x_t,\, R_t \,) = N(\, y_t;\, (\, H_t\, x_t + c_t \,),\, R_t \,). \qquad \text{(A.13)}$$

According to the state transition equation Eq. (2), and the fact that x_{t-1} is Gaussian, the predictive prior of x_t is Gaussian given the random parameter P_t,

$$p(\, x_t \mid P_t \,) = N(\, x_t;\; x_{t|t-1},\, P_t \,), \tag{A.14}$$

where the expected value $x_{t|t-1} = F_t\, x_{t-1|t-1} + a_t = F_t\, E(x_{t-1}) + a_t$.

The random variable P_t is a covariance matrix of a normal distribution, is assumed to be diagonal, needs to be updated by data through the Bayes rule, there is lack of information on relative prior strengths of different (diagonal) elements, and the prior strengths cannot be assumed rising (or falling) in a time-varying analysis. Comparing these features of P_t with the characteristics of inverse-gamma distribution described in the subsection A.1, it is appropriate to assume that the predictive prior of P_t is "diagonal inverse-gamma" by (A.8),

$$p(\, P_t \mid y_{1:t-1} \,) = IG_D(\, P_t;\; \alpha_{t|t-1},\, B_{t|t-1} \,), \tag{A.15}$$

with a constant shape parameter,

$$\alpha_{t|t-1} = \alpha_0 = T_0 \,/\, 2. \tag{A.16}$$

Therefore, the (previous) posterior of P_{t-1} must be "diagonal inverse-gamma" as well: $p(P_{t-1}|y_{1:t-1}) = IG_D(P_{t-1}; \alpha_{t-1|t-1}, B_{t-1|t-1})$. If the transition covariance Q_t is a non-random known input, the prior scale parameter simply is

$$B_{t|t-1} = (\, F_t\, P_{t-1|t-1}\, F_t^T + Q_t \,)\, \alpha_0, \tag{A.17}$$

based on Eqs. (2) and (A.11). When Q_t is a random unknown and the state error covariance P_t varies slowly relative to the state vector x_t, a technique utilizing the KL-divergence, as proposed by Özkana *et al.* (2013), can be used to optimize the prior scale parameter $B_{t|t-1}$. According to the discussion about (A.7) and (A.16), when

$$B_{t|t-1} = P_{t-1|t-1}\, T_0 \,/\, 2 = E(P_{t-1}^{-1})^{-1}\, T_0 \,/\, 2, \tag{A.18}$$

the KL-divergences between $p(P_t|y_{1:t-1})$ and $p(P_{t-1}|y_{1:t-1})$ are minimized, i.e. the predictive prior $p(P_t|y_{1:t-1})$ is the closest to the previous posterior $p(P_{t-1}|y_{1:t-1})$. A constant diagonal transition matrix F_t, such as Eq. (4), is a key condition for P_t to vary slowly in order to justify the use of scale parameter (A.18). Applying the same set of arguments about P_t to the measurement error covariance R_t in the likelihood (A.13), the predictive prior of R_t can be assumed as

$$p(\, R_t \mid y_{1:t-1}\,) = IG_D(\, R_t;\ \alpha_{t|t-1},\ C_{t|t-1}\,), \tag{A.19}$$

where the same shape parameter by (A.16) is applied because of no information on relative prior strengths between P_t and R_t, and the prior scale parameter is

$$C_{t|t-1} = R_{t-1|t-1}\, T_0\,/\,2 = E(R_{t-1}^{-1})^{-1}\, T_0\,/\,2. \tag{A.20}$$

As long as P_t varies slowly to justify the use of (A.18), a prediction of the unknown transition covariance Q_t is not needed for the Bayesian filtering. By Eq. (2), the conditional distribution of Q_t is

$$p(\, Q_t \mid y_{1:t}\,) = p(\, P_t - F_t\, P_{t-1}\, F_t^T \mid y_{1:t}\,). \tag{A.21}$$

If jumps in P_t are caused by a non-diagonal matrix F_t when Q_t is a random unknown, further research is needed on what should be an appropriate predictive prior of P_t and whether linear combination of inverse-gamma distributions (Witkovský 2001) should be involved.

All of the predictive priors of random variables x_t, P_t and R_t at time t are formed by the expected values of their previous posteriors $x_{t-1|t-1} = E(x_{t-1})$, $P_{t-1|t-1} = E(P_{t-1}^{-1})^{-1}$ and $R_{t-1|t-1} = E(R_{t-1}^{-1})^{-1}$ at time t-1. These priors can be referred to as “sequential priors” and the analysis can be referred to as “sequential Bayesian”. Based on the likelihood and prior formulas (A.13), (A.14), (A.15) and (A.19), the joint probability density function in (A.12) of the random variables y_t, x_t, P_t and R_t is:

$$\begin{aligned}
&p(\, y_t,\, Z_t \mid y_{1:t-1}\,) = p(\, y_t, x_t, P_t, R_t\,) \propto \\
&\propto p(\, y_t \mid x_t, R_t\,)\, p(\, x_t \mid P_t\,)\, p(\, P_t \mid y_{1:t-1}\,)\, p(\, R_t \mid y_{1:t-1}\,) = \\
&= N(\, y_t;\ (\, H_t\, x_t + c_t\,),\ R_t\,) \times N(\, x_t;\ x_{t|t-1},\ P_t\,) \times \\
&\quad \times IG_D(\, P_t;\ \alpha_0,\ B_{t|t-1}\,) \times IG_D(\, R_t;\ \alpha_0,\ C_{t|t-1}\,).
\end{aligned} \tag{A.22}$$

The sequential Bayesian inference is to evaluate the posterior $p(x_t, P_t, R_t \mid y_t)$, a three-random-variable solution to the state space model Eqs. (2) and (3), based on the joint distribution (A.22).

A.3. Variational approximation

The joint posterior $p(x_t, P_t, R_t \mid y_t)$ based on (A.22) is not analytically tractable. Several types of approximations can be applied to estimate an analytically intractable posterior. A stochastic approximation is a numerical method with Monte Carlo simulation and particle filters as its best known examples. An analytic approximation tries to derive a closed-form expression. A marginalized particle filter is a combination of the two (Schön *et al.* 2007, Özkana *et al.* 2013). The variational approximation (VA) is a well-developed analytic one.

The VA approach is to use a factorized distribution to approximate an intractable joint posterior (Bishop 2006, Tzikas *et al.* 2008, Grimmer 2011). Assume that the joint posterior $p(x_t, P_t, R_t \mid y_t)$ can be approximated by a factorized density function, $q(x_t, P_t, R_t) = q_x(x_t)\, q_P(P_t)\, q_R(R_t)$. Beyond the assumption of factorization, there is no need to identify the function form of $q(x_t, P_t, R_t)$ in advance. The form will come naturally in the VA process. The expected value of a random variable or function z with respect to the various approximating density functions can be denoted by "expectation operators",

$$E_x\, z = E_x(z) = \int z\, q_x(x_t)\, dx_t,$$

$$E_P\, z = E_P(z) = \int z\, q_P(P_t)\, dP_t,$$

$$E_R\, z = E_R(z) = \int z\, q_R(R_t)\, dR_t.$$

According to the VA method (Bishop 2006, Tzikas *et al.* 2008, Grimmer 2011), the optimal factors $q_x(x_t)$, $q_P(P_t)$ and $q_R(R_t)$ of the approximating density function are:

$$\ln q_x(x_t) = E_P\, E_R\, \ln p(\, y_t, x_t, P_t, R_t\,) + const, \tag{A.23}$$

$$\ln q_P(P_t) = E_x\, E_R\, \ln p(\, y_t, x_t, P_t, R_t\,) + const, \tag{A.24}$$

$$\ln q_R(R_t) = E_x\, E_P\, \ln p(\, y_t, x_t, P_t, R_t\,) + const. \tag{A.25}$$

These factor optimization equations are interrelated: each optimal factor is formulated in terms of all other optimal factors. A fortunate characteristic of the VA theory is that an iterative approach with reasonable initial values always converges to a local optimal result.

A.4. Function forms of the approximating factors

The actual forms of the approximating factors $q_x(x_t)$, $q_P(P_t)$ and $q_R(R_t)$ can be identified through the optimization equations (A.23) to (A.25) and the logarithm of the joint distribution $p(y_t, x_t, P_t, R_t)$ in (A.22):

$$\begin{aligned}
&\ln p(\, y_t,\, x_t,\, P_t,\, R_t\,) = \\
&= \ln N(\, y_t;\, (\, H_t\, x_t + c_t\,),\, R_t\,) + \ln N(\, x_t;\, x_{t|t-1},\, P_t\,) + \\
&\quad + \ln IG_D(\, P_t;\, \alpha_0,\, B_{t|t-1}\,) + \ln IG_D(\, R_t;\, \alpha_0,\, C_{t|t-1}\,) + const = \\
&= (-1/2)\,\{\, \ln|R_t| + tr(\, [y_t - (H_t x_t + c_t)][y_t - (H_t x_t + c_t)]^T R_t^{-1}\,)\,\} + \\
&\quad + (-1/2)\,[\, \ln|P_t| + tr(\, (x_t - x_{t|t-1})(x_t - x_{t|t-1})^T P_t^{-1}\,)\,] + \\
&\quad + (-1)\,[\, (\alpha_0 + 1)\ln|P_t| + tr(\, B_{t|t-1}\, P_t^{-1}\,)\,] + \\
&\quad + (-1)\,[\, (\alpha_0 + 1)\ln|R_t| + tr(\, C_{t|t-1}\, R_t^{-1}\,)\,] + const.
\end{aligned} \tag{A.26}$$

Let’s first examine the factor $q_x(x_t)$. Substituting (A.26) into (A.23) gives rise to

$$\begin{aligned}
&\ln q_x(x_t) = E_P\, E_R\, \ln p(\, y_t, x_t, P_t, R_t\,) + const = \\
&= (-1/2)\,\{\, tr(\, [\, y_t - (H_t x_t + c_t)\,]\,[\, y_t - (H_t x_t + c_t)\,]^T E_R(R_t^{-1})\,) + \\
&\quad + tr(\, (x_t - x_{t|t-1})\,(x_t - x_{t|t-1})^T E_P(P_t^{-1})\,)\,\} + const.
\end{aligned} \tag{A.27}$$

Comparing (A.27) with the normal distribution (A.2) yields

$$\begin{aligned}
&q_x(\, x_t\,) = \\
&= N(\, y_t;\, (H_t x_t + c_t),\, E_R(R_t^{-1})^{-1}\,) \times N(\, x_t;\, x_{t|t-1},\, E_P(P_t^{-1})^{-1}\,) \propto \\
&\propto \exp\ \{\, (-1/2)\, x_t^T\, [\, H_t^T\, E_R(R_t^{-1})\, H_t + E_P(P_t^{-1})\,]\, x_t + \\
&\quad + [\, (y_t - c_t)^T\, E_R(R_t^{-1})\, H_t + x_{t|t-1}^T\, E_P(P_t^{-1})\,]\, x_t\,\}.
\end{aligned} \tag{A.28}$$

where the expected values $E_R(R_t^{-1})$ and $E_P(P_t^{-1})$ are yet to be known. The approximating factor

$q_x(x_t)$ by (A.28) is a normal distribution

$$q_x(\, x_t \,) = N(\, x_t;\, \mu_t,\, \Sigma_t \,) \propto \exp\, (-\tfrac{1}{2}\, x_t^T\, \Sigma_t^{-1}\, x_t + \mu_t^T\, \Sigma_t^{-1}\, x_t \,), \qquad \text{(A.29)}$$

and its moments can be identified by comparing (A.28) and (A.29) as

$$\Sigma_t = [\, H_t^T\, E_R(R_t^{-1})\, H_t + E_P(P_t^{-1})\,]^{-1}, \qquad \text{(A.30)}$$

$$\mu_t = \Sigma_t\, [\, H_t^T\, E_R(R_t^{-1})\, (\, y_t - c_t \,) + E_P(P_t^{-1})\, x_{t|t-1}\,]. \qquad \text{(A.31)}$$

By applying the Woodbury matrix identity,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1},$$

and denoting

$$S_t = H_t\, E_P(P_t^{-1})^{-1}\, H_t^T + E_R(R_t^{-1})^{-1}, \qquad \text{(A.32)}$$

$$K_t = E_P(P_t^{-1})^{-1}\, H_t^T\, S_t^{-1}, \qquad \text{(A.33)}$$

the matrix parameter Σ_t (A.30) can be simplified as

$$\Sigma_t = E_P(P_t^{-1})^{-1} - K_t\, S_t\, K_t^T. \qquad \text{(A.34)}$$

The parameter Σ_t is the covariance of the approximating posterior $q_x(x_t)$ by (A.29) while the random variable P_t is the stochastic covariance of the prior $p(x_t|P_t)$ by (A.14). The matrix K_t by (A.33) is the Kalman gain (Simon 2006). Substituting (A.34) into (A.31) results in

$$\mu_t = x_{t|t-1} - K_t\, H_t\, x_{t|t-1} + E_P(P_t^{-1})^{-1}\, H_t^T\, [\, I -$$
$$- S_t^{-1}\, H_t\, E_P(P_t^{-1})^{-1}\, H_t^T\,]\, E_R(R_t^{-1})\, (y_t - c_t).$$

Because of the relation $H_t\, E_P(P_t^{-1})^{-1} H_t^T = S_t - E_R(R_t^{-1})^{-1}$ by rearranging the definition (A.32), the expected value

$$\mu_t = x_{t|t-1} - K_t\, H_t\, x_{t|t-1} + K_t\, (\, y_t - c_t \,).$$

Defining measurement prediction error,

$$e_{t|t-1} = y_t - (\, H_t\, x_{t|t-1} + c_t \,), \qquad \text{(A.35)}$$

the expression (A.31) of the expected value μ_t can be simplified as

$$\mu_t = x_{t|t-1} + K_t\, e_{t|t-1}. \tag{A.36}$$

The matrix S_t by (A.32) is the covariance of the measurement prediction error by (A.35) (Simon 2006).

Now let's figure out the actual forms of the other two factors $q_P(P_t)$ and $q_R(R_t)$. Substituting (A.26) into (A.24) yields

$$\begin{aligned}\ln q_P(P_t) = (-1)\,\{\, &(\alpha_0 + 1/2 + 1) \ln |P_t| + tr(\,[\,B_{t|t-1} + \\ &+ (1/2)\, E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T\,]\, P_t^{-1}\,)\,\} + \\ &+ const,\end{aligned} \tag{A.37}$$

where $B_{t|t-1}$ by (A.18) is diagonal. When the expected value $E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T$ is also diagonal either by data or by our diagonal constraint, the factor $q_P(P_t)$ is a "diagonal inverse-gamma distribution" by comparing (A.37) with (A.9):

$$q_P(\,P_t\,) = IG_D(\,P_t;\ \alpha_{t|t},\ B_{t|t}\,), \tag{A.38}$$

$$\alpha_{t|t} = \alpha_0 + 1/2 = (\,T_0 + 1\,)\,/\,2, \tag{A.39}$$

$$B_{t|t} = B_{t|t-1} + (1/2)\, diag(\,E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T\,). \tag{A.40}$$

The expected value $E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T$ will be evaluated later based on the Gaussian factor $q_x(x_t)$ of (A.29).

The actual form of the factor $q_R(R_t)$ can be worked out in a similar way. Substituting (A.26) into (A.25) yields

$$\begin{aligned}\ln q_R(R_t) = (-1)\,\{\, &(\alpha_0 + 1/2 + 1) \ln |R_t| + tr(\,[\,C_{t|t-1} + \\ &+ (1/2)\, E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T\,]\, R_t^{-1}\,)\,\} + \\ &+ const,\end{aligned} \tag{A.41}$$

where $C_{t|t-1}$ by (A.20) is diagonal. When the expected value $E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T$ is also diagonal either by data or by our diagonal constraint, the factor $q_R(R_t)$ is also a "diagonal inverse-gamma distribution" by comparing (A.41) with (A.9):

$$q_R(R_t) = IG_D(R_t;\ \alpha_{t|t},\ C_{t|t}), \tag{A.42}$$

$$C_{t|t} = C_{t|t-1} + \\ + (1/2)\ diag(E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T), \tag{A.43}$$

here the $\alpha_{t|t}$ of (A.39) applies. The expected value $E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T$ will be evaluated later based on $q_x(x_t)$ of (A.29) as well.

<u>A.5. Iteration of the variational approximation</u>

As the function forms of the approximating factors $q_x(x_t)$, $q_P(P_t)$ and $q_R(R_t)$ now identified, the four expected values, $E_P(P_t^{-1})$, $E_R(R_t^{-1})$, $E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T$ and $E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T$, in the density functions of these factors can now be evaluated.

The distribution forms of $q_P(P_t)$ and $q_R(R_t)$ in (A.38) and (A.42) and the expected value formula (A.10) lead to

$$E_P(P_t^{-1}) = \alpha_{t|t}\, B_{t|t}^{-1}, \tag{A.44}$$

$$E_R(R_t^{-1}) = \alpha_{t|t}\, C_{t|t}^{-1}. \tag{A.45}$$

But the matrix parameters $B_{t|t}$ by (A.40) needs $E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T$, while $C_{t|t}$ by (A.43) needs $E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T$. Because the factor $q_x(x_t)$ is a normal distribution by (A.29), (A.36) and (A.34), according to the expected values in (A.3) and (A.44), as well as the definition (A.35),

$$E_x(x_t - x_{t|t-1})(x_t - x_{t|t-1})^T = \Sigma_t + (\mu_t - x_{t|t-1})\,(\mu_t - x_{t|t-1})^T = \\ = E_P(P_t^{-1})^{-1} + K_t\,(e_{t|t-1}\, e_{t|t-1}^T - S_t)\, K_t^T. \tag{A.46}$$

The evaluation of $E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T$ needs a definition

$$M_t = I_n - H_t\, K_t, \tag{A.47}$$

and the relation

$$M_t\, S_t\, M_t^T = (\, I_n - H_t\, K_t\,)\, S_t\, (\, I_n - H_t\, K_t\,)^T =$$
$$= E_R(R_t^{-1})^{-1} - H_t\, E_P(P_t^{-1})^{-1}\, H_t^T + H_t\, K_t\, S_t\, (\, H_t\, K_t\,)^T. \tag{A.48}$$

According to (A.29), (A.36), (A.34), (A.3) and (A.48), the expected value

$$E_x(y_t - H_t x_t - c_t)(y_t - H_t x_t - c_t)^T =$$
$$= H_t\, \Sigma_t\, H_t^T + (\, y_t - H_t\, \mu_t - c_t\,)\, (\, y_t - H_t\, \mu_t - c_t\,)^T =$$
$$= H_t\, E_P(P_t^{-1})^{-1}\, H_t^T - H_t\, K_t\, S_t\, K_t^T\, H_t^T +$$
$$+ (\, e_{t|t-1} - H_t\, K_t\, e_{t|t-1}\,)\, (\, e_{t|t-1} - H_t\, K_t\, e_{t|t-1}\,)^T =$$
$$= E_R(R_t^{-1})^{-1} + M_t\, (\, e_{t|t-1}\, e_{t|t-1}^T - S_t\,)\, M_t^T. \tag{A.49}$$

Substituting (A.46) into (A.40) and then into (A.44) yields

$$E_P(P_t^{-1})^{-1} = B_{t|t} \,/\, \alpha_{t|t} =$$
$$= (\, B_{t|t-1} \,/\, \alpha_{t|t}\,) + [1/(2\, \alpha_{t|t})]\; diag(\, E_P(P_t^{-1})^{-1} +$$
$$+ K_t\, (\, e_{t|t-1}\, e_{t|t-1}^T - S_t\,)\, K_t^T\,).$$

The expected value $E_P(P_t^{-1})$ can be expressed by rearranging the above equation as

$$E_P(P_t^{-1})^{-1} = diag(\, 2\, B_{t|t-1} +$$
$$+ K_t\, (\, e_{t|t-1}\, e_{t|t-1}^T - S_t\,)\, K_t^T\,) \,/\, (\, 2\, \alpha_{t|t} - 1\,). \tag{A.50}$$

Similarly, substituting (A.49) into (A.43) and then into (A.45) yields

$$E_R(R_t^{-1})^{-1} = C_{t|t} \,/\, \alpha_{t|t} =$$
$$= (\, C_{t|t-1} \,/\, \alpha_{t|t}\,) + [1/(2\, \alpha_{t|t})]\; diag(\, E_R(R_t^{-1})^{-1} +$$
$$+ M_t\, (\, e_{t|t-1}\, e_{t|t-1}^T - S_t\,)\, M_t^T\,).$$

The expected value $E_R(R_t^{-1})$ can be expressed by rearranging the above equation as

$$E_R(R_t^{-1})^{-1} = diag(\, 2\, C_{t|t-1} +$$
$$+ M_t\, (\, e_{t|t-1}\, e_{t|t-1}^T - S_t\,)\, M_t^T\,) \,/\, (\, 2\, \alpha_{t|t} - 1\,). \tag{A.51}$$

The same difference matrix, $e_{t|t-1}e_{t|t-1}^T - S_t$, in both (A.50) and (A.51) is the current data to update the covariance matrices P_t and R_t.

When the two expected values $E_P(P_t^{-1})$ and $E_R(R_t^{-1})$ are fully defined by (A.50) and

(A.51), the approximating distribution $q_x(x_t)$ is fully defined by (A.29), (A.36) and (A.34); $q_P(P_t)$ is defined by (A.38), (A.39), (A.40) and (A.46); and $q_R(R_t)$ is defined by (A.42), (A.39), (A.43) and (A.49). Therefore, the converging iteration to solve for the three interrelated distributions (A.23), (A.24) and (A.25) can be simplified into an iteration to solve for the two interrelated expected values (A.50) and (A.51). With $k = 0, 1, \cdots, L$ as the iteration index, the two expected values $E_P(P_t^{-1})^{-1}$ and $E_R(R_t^{-1})^{-1}$ in (A.50), (A.51), (A.32) and (A.33) can be denoted as $P_{t|t}^{(k)}$ and $R_{t|t}^{(k)}$. The two initial values $P_{t|t}^{(0)}$ and $R_{t|t}^{(0)}$ of the iteration can be set by Eqs. (8) and (9) based on (A.18), (A.20), (A.16) and (A.10). The three definitions (A.32), (A.33) and (A.47) can be translated into the VASB Eqs. (10), (11) and (12). And then, the two interrelated equations (A.50) and (A.51) can be evaluated by the VASB Eqs. (13) and (14), because of the expressions (A.18), (A.20) and (A.39).

Assume that the VA iteration, Eqs. (10) to (14), ends at $k = L$. The expressions of μ_t in (A.36) and Σ_t in (A.34) of the approximating posterior $q_x(x_t)$ in (A.29) result in the VASB Eqs. (15) and (17). Taking expectation on both sides of the measurement equation Eq. (3) gives Eq. (16). The conditional distribution (A.21) leads to Eq. (18). Finally, Eq. (19) is self-explained.

Appendix B. Error reduction target

When applying the VASB algorithm Eqs. (5) to (19) to real financial markets, both covariance matrices Q_t and R_t are time-varying stochastic unknowns to be estimated. The process of sequentially updating $x_{t|t}$, $P_{t|t}$ and $R_{t|t}$ all together may become unstable because of the deviations of real market data from the assumed normal and inverse-gamma distributions. The ratio $tr(P_{t|t})/tr(R_{t|t})$ may become higher and higher, or lower and lower. Such observations suggest some positive feedback loops leading to over-fitting if the ratio $tr(P_{t|t})/tr(R_{t|t})$ is unusually large, or under-fitting if it is unusually small.

By the nature of a Bayesian update, the overall estimation error associated with the posterior is smaller than the overall prediction error associated with the prior. At a time step t, an overly large error reduction causes over-fitting while an unduly small reduction causes under-fitting. One technique to keep the Bayesian error reduction at a sensible level is to adjust the initial values $P_{t|t}^{(0)}$ and $R_{t|t}^{(0)}$ in Eqs. (8) and (9) before making the VA iteration Eqs. (10) to (14).

If a Bayesian inference Eq. (15) is made without making the VA iteration, the measurement estimation error $e_{t|t}^{(0)}$ can be expressed by the prediction error $e_{t|t-1}$ in (A.35) as

$$
\begin{aligned}
e_{t|t}^{(0)} &= y_t - [\, H_t\, (\, x_{t|t-1} + K_t^{(0)}\, e_{t|t-1}\,) + c_t\,] = \\
&= (\, I_n - H_t\, K_t^{(0)}\,)\, e_{t|t-1} = M_t^{(0)}\, e_{t|t-1}.
\end{aligned}
$$

The matrix

$$
\begin{aligned}
M_t^{(0)} &= I_n - H_t\, K_t^{(0)} = I_n - H_t\, P_{t|t}^{(0)}\, H_t^T\, (S_t^{(0)})^{-1} = \\
&= I_n - (\, S_t^{(0)} - R_{t|t}^{(0)}\,)\, (S_t^{(0)})^{-1} = R_{t|t}^{(0)}\, (S_t^{(0)})^{-1}, \qquad \text{(B.1)}
\end{aligned}
$$

is an error reduction matrix. The eigenvalues of the $n \times n$ symmetric matrix $(M_t^{(0)})^T M_t^{(0)}$ measure the degrees of error reduction in variance. Since the trace of the matrix equals to the sum of its eigenvalues,

$$
\begin{aligned}
g_t^{(0)} &= tr(\, (M_t^{(0)})^T\, M_t^{(0)}\,)\, /\, n = \\
&= tr(\, M_t^{(0)}\, (M_t^{(0)})^T\,)\, /\, n = tr(\, R_{t|t}^{(0)}\, (S_t^{(0)})^{-2}\, R_{t|t}^{(0)}\,)\, /\, n = \\
&= tr(\, R_{t|t}^{(0)}\, (\, H_t\, P_{t|t}^{(0)}\, H_t^T + R_{t|t}^{(0)}\,)^{-2}\, R_{t|t}^{(0)}\,)\, /\, n, \qquad \text{(B.2)}
\end{aligned}
$$

can serve as an "average error reduction ratio". When the value of $g_t^{(0)}$ is unusually small (or large), the VASB may be over-fitting (or under-fitting). The $g_t^{(0)}$ expression (B.2) suggests that rescaling the initial values $P_{t|t}^{(0)}$ and $R_{t|t}^{(0)}$, replacing them by the rescaled $p_t P_{t|t}^{(0)}$ and $r_t R_{t|t}^{(0)}$, where

$p_t > 0$ and $r_t > 0$, may be used to steer the error reduction to a desired level in order to prevent the VASB from both over- and under-fitting.

When an "error reduction target" g, $0 < g < 1$, is specified as an additional parameter for the VASB algorithm, the relative values of the two rescaling coefficients, p_t and r_t, can be determined by

$$tr(\, r_t\, R_{t|t}^{(0)}\, (\, p_t\, H_t\, P_{t|t}^{(0)} H_t^T + r_t\, R_{t|t}^{(0)}\,)^{-2}\, r_t\, R_{t|t}^{(0)}\,)\, /\, n = g, \tag{B.3}$$

which is to make the adjusted average error reduction ratio, $g_t^{(0)}$ by (B.2), equal to the specified error reduction target g. Since the unadjusted and adjusted predictive distributions for y_t, by Eqs. (6) and (10) before and after the initial value rescaling adjustment, are

$$p_U(\, y_t\,) = N(\, y_t;\, y_{t|t-1},\, H_t\, P_{t|t}^{(0)} H_t^T + R_{t|t}^{(0)}\,),$$

$$p_A(\, y_t\,) = N(\, y_t;\, y_{t|t-1},\, p_t\, H_t\, P_{t|t}^{(0)} H_t^T + r_t\, R_{t|t}^{(0)}\,),$$

the two rescaling coefficients, $p_t > 0$ and $r_t > 0$, can be optimized by minimizing the KL-divergence from $p_A(y_t)$ to $p_U(y_t)$, $KL(\, p_A(y_t) \parallel p_U(y_t)\,) \geq 0$, or minimizing that from $p_U(y_t)$ to $p_A(y_t)$, $KL(\, p_A(y_t) \parallel p_U(y_t)\,) \geq 0$, subject to the constraint (B.3).

When the observation y_t is a scalar, the dimension $n = 1$ and $H_t\, P_{t|t}^{(0)} H_t^T$ and $R_{t|t}^{(0)}$ in (B.3) are also scalars. The KL-divergences, $KL(\, p_A(y_t) \parallel p_U(y_t)\,) = 0$ and $KL(\, p_A(y_t) \parallel p_U(y_t)\,) = 0$, if

$$p_t\, H_t\, P_{t|t}^{(0)} H_t^T + r_t\, R_{t|t}^{(0)} = H_t\, P_{t|t}^{(0)} H_t^T + R_{t|t}^{(0)}. \tag{B.4}$$

The two equations (B.3) and (B.4) for the two rescaling coefficients p_t and r_t become

$$r_t^2\, R_{t|t}^{(0)}\, (S_t^{(0)})^{-2}\, R_{t|t}^{(0)} = g, \tag{B.5}$$

$$p_t\, H_t\, P_{t|t}^{(0)} H_t^T + r_t\, R_{t|t}^{(0)} = S_t^{(0)}. \tag{B.6}$$

The solutions p_t and r_t to the system (B.5) and (B.6) lead to the adjustment Eqs. (27) and (28), where g, $0 < g < 1$, is the error reduction target in variance.

For a stochastic local level model Eqs. (23) and (24), when a given fraction q is a desired level to be targeted by the variance ratio $P_{t|t}^{(L)}/S_t^{(L)}$ for a modeling reason, this "variance ratio target" q can serve as a modeling parameter via the error reduction target g. With the target q, the adjusted initial variance ratio can be set as

$$p_t\, P_{t|t}^{(0)} /\, S_t^{(0)} = q. \tag{B.7}$$

By applying Eqs. (24) and (B.7) to the condition (B.5) and (B.6), the error reduction target g can be expressed as a function of the variance ratio target q as:

$$g = (\, 1 - p_t\, P_{t|t}^{(0)} \,/\, S_t^{(0)}\,)^2 = (\, 1 - q\,)^2. \tag{B.8}$$

This parameter relation leads to Eq. (47).

Table 1.1. Biases and dispersions of the estimated response $y_{t|t}$ and state $x_{t|t}$ by solving the simulated time-varying regression Eqs. (29) to (31) using the rolling T_r-month OLS, the KF of time-varying $Q_t = Q_{t|t}^{OLS}$ and $R_t = R_{t|t}^{OLS}$, the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$, and the VASB of $F_t = I_m$ and various g. The response y_t is generated by the known x_t of Eq. (33) to simulate an "all-capitalization market timing strategy". The biases and dispersions of the estimates are measured by the RMS of monthly error $\Delta y_{t|t} = y_{t|t} - y_t$, and the mean and standard deviation of the elements of monthly difference $\Delta x_{t|t} = x_{t|t} - x_t$. For the estimates, a lower StDev of $\Delta x_{t|t}$ is typically associated with a higher RMSE. The KF of time-varying Q_t and R_t and the VASB can yield a good $y_{t|t}$ with a low RMSE.

Estimates	RMSE	$\boldsymbol{b_0(t)}$		$\boldsymbol{b_{R3K}(t)}$		$\boldsymbol{b_{R3G-R3V}(t)}$		$\boldsymbol{b_{R1K-R2K}(t)}$	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
True Value of x_t		0		$1 + a \sin \phi t$		0		0	
18-Mon OLS	1.431	.021	.449	-.009	.188	-.002	.210	-.001	.166
24-Mon OLS	1.531	.036	.382	-.014	.206	.003	.177	-.002	.143
30-Mon OLS	1.612	.049	.341	-.021	.227	.007	.158	-.004	.129
36-Mon OLS	1.671	.062	.315	-.028	.240	.010	.146	-.004	.122
42-Mon OLS	1.696	.073	.294	-.035	.241	.012	.135	-.006	.114
KF, $Q_{t\|t}^{18M}$, $R_{t\|t}^{18M}$	1.343	.019	.322	-.023	.168	-.018	.150	-.011	.125
KF, $Q_{t\|t}^{24M}$, $R_{t\|t}^{24M}$	1.441	.024	.271	-.034	.174	-.012	.129	-.018	.111
KF, $Q_{t\|t}^{30M}$, $R_{t\|t}^{30M}$	1.502	.019	.241	-.033	.174	-.016	.117	-.019	.099
KF, $Q_{t\|t}^{36M}$, $R_{t\|t}^{36M}$	1.537	.028	.218	-.035	.173	-.015	.109	-.019	.091
KF, $Q_{t\|t}^{42M}$, $R_{t\|t}^{42M}$	1.564	.030	.203	-.032	.173	-.014	.105	-.014	.086
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$	1.726	-.019	.159	-.010	.185	-.034	.091	.017	.061
KF, Q=0, $\langle R_{t\|t}^{24M} \rangle$	1.723	-.020	.149	-.009	.183	-.032	.087	.015	.055
KF, Q=0, $\langle R_{t\|t}^{30M} \rangle$	1.723	-.021	.145	-.008	.182	-.033	.086	.015	.054
KF, Q=0, $\langle R_{t\|t}^{36M} \rangle$	1.721	-.007	.145	-.012	.182	-.027	.084	.012	.053
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	1.720	.006	.141	-.014	.182	-.024	.083	.010	.053
VASB, $g = 0.6$	1.425	.000	.334	-.017	.167	-.011	.167	-.009	.136
VASB, $g = 0.7$	1.531	.002	.273	-.015	.172	-.010	.135	-.011	.113
VASB, $g = 0.8$	1.626	-.001	.212	-.009	.177	-.011	.106	-.009	.090
VASB, $g = 0.9$	1.715	-.010	.163	.012	.181	-.022	.091	.001	.069
VASB, $g = n/a$	1.790	-.019	.162	.052	.187	-.041	.094	-.007	.061

Table 1.2. Biases and dispersions of the forecasted response $y_{t|t-1}$ and state $x_{t|t-1}$ by solving the simulated time-varying regression Eqs. (29) to (31) using the rolling T_r-month OLS, the KF of time-varying $Q_t = Q_{t|t}^{OLS}$ and $R_t = R_{t|t}^{OLS}$, the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$, and the VASB of $F_t = I_m$ and various g. The response y_t is generated by the known x_t of Eq. (33) to simulate an "all-capitalization market timing strategy". The biases and dispersions of the forecasts are measured by the RMS of monthly error $\Delta y_{t|t-1} = y_{t|t-1} - y_t$, and the mean and standard deviation of the elements of monthly difference $\Delta x_{t|t-1} = x_{t|t-1} - x_t$. For the forecasts, a lower StDev of $\Delta x_{t|t-1}$ is typically associated with a lower RMSE. The KF of constant Q and R and the VASB can yield a good $y_{t|t-1}$ with a low RMSE.

Forecasts	RMSE	$\boldsymbol{b_0(t)}$		$\boldsymbol{b_{R3K}(t)}$		$\boldsymbol{b_{R3G-R3V}(t)}$		$\boldsymbol{b_{R1K-R2K}(t)}$	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
True Value of x_t		0		$1 + a\sin\phi t$		0		0	
18-Mon OLS	1.942	.022	.450	-.009	.200	-.002	.210	-.001	.166
24-Mon OLS	1.950	.036	.383	-.014	.218	.004	.177	-.002	.143
30-Mon OLS	1.951	.049	.342	-.021	.237	.007	.158	-.004	.129
36-Mon OLS	1.972	.062	.316	-.029	.247	.010	.146	-.004	.122
42-Mon OLS	1.964	.073	.294	-.035	.245	.012	.136	-.006	.114
KF, $Q_{t\|t}^{18M}$, $R_{t\|t}^{18M}$	1.843	.019	.322	-.023	.176	-.018	.150	-.011	.125
KF, $Q_{t\|t}^{24M}$, $R_{t\|t}^{24M}$	1.825	.024	.271	-.034	.180	-.012	.130	-.018	.111
KF, $Q_{t\|t}^{30M}$, $R_{t\|t}^{30M}$	1.817	.019	.242	-.033	.180	-.016	.117	-.019	.099
KF, $Q_{t\|t}^{36M}$, $R_{t\|t}^{36M}$	1.808	.028	.219	-.035	.178	-.015	.109	-.019	.091
KF, $Q_{t\|t}^{42M}$, $R_{t\|t}^{42M}$	1.801	.029	.203	-.032	.177	-.014	.106	-.014	.086
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$	1.798	-.020	.160	-.010	.186	-.034	.091	.017	.061
KF, Q=0, $\langle R_{t\|t}^{24M} \rangle$	1.793	-.021	.149	-.009	.184	-.032	.087	.015	.056
KF, Q=0, $\langle R_{t\|t}^{30M} \rangle$	1.792	-.022	.145	-.008	.183	-.033	.086	.015	.054
KF, Q=0, $\langle R_{t\|t}^{36M} \rangle$	1.789	-.007	.146	-.012	.183	-.028	.084	.012	.053
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	1.787	.006	.142	-.014	.183	-.025	.084	.011	.053
VASB, $g = 0.6$	1.854	.001	.334	-.017	.175	-.011	.167	-.009	.136
VASB, $g = 0.7$	1.827	.002	.273	-.015	.178	-.010	.135	-.011	.113
VASB, $g = 0.8$	1.804	.000	.212	-.009	.181	-.011	.106	-.009	.090
VASB, $g = 0.9$	1.796	-.010	.164	.012	.183	-.022	.091	.001	.069
VASB, $g = n/a$	1.829	-.019	.163	.052	.188	-.041	.095	-.007	.062

Table 2.1. Biases and dispersions of the estimated response $y_{t|t}$ and state $x_{t|t}$ by solving the simulated time-varying regression Eqs. (29) to (31) using the rolling T_r-month OLS, the KF of time-varying $Q_t = Q_{t|t}^{OLS}$ and $R_t = R_{t|t}^{OLS}$, the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$, and the VASB of $F_t = I_m$ and various g. The response y_t is generated by the known x_t of Eq. (34) to simulate a "time-varying small-capitalization growth style market neutral strategy". The biases and dispersions of the estimates are measured by the RMS of monthly error $\Delta y_{t|t} = y_{t|t} - y_t$, and the mean and standard deviation of the elements of monthly difference $\Delta x_{t|t} = x_{t|t} - x_t$. For the estimates, a lower StDev of $\Delta x_{t|t}$ is typically associated with a higher RMSE. The KF of time-varying Q_t and R_t and the VASB can yield a good $y_{t|t}$ with a low RMSE.

Estimates	RMSE	$\boldsymbol{b_0(t)}$		$\boldsymbol{b_{R3K}(t)}$		$\boldsymbol{b_{R3G-R3V}(t)}$		$\boldsymbol{b_{R1K-R2K}(t)}$	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
True Value of x_t		0		0		$1 + a\sin\phi t$		$-1 + a\cos\phi t$	
18-Mon OLS	1.406	-.001	.439	-.002	.124	-.004	.258	-.005	.218
24-Mon OLS	1.483	.008	.366	-.005	.101	.000	.258	-.009	.228
30-Mon OLS	1.538	.015	.326	-.006	.088	.001	.264	-.013	.239
36-Mon OLS	1.574	.021	.299	-.008	.080	.000	.263	-.014	.245
42-Mon OLS	1.591	.028	.278	-.010	.074	-.001	.255	-.014	.242
KF, $Q_{t\|t}^{18M}$, $R_{t\|t}^{18M}$	1.320	.008	.317	-.008	.083	.003	.217	-.016	.188
KF, $Q_{t\|t}^{24M}$, $R_{t\|t}^{24M}$	1.405	.022	.263	-.011	.070	.003	.207	-.025	.187
KF, $Q_{t\|t}^{30M}$, $R_{t\|t}^{30M}$	1.461	.044	.231	-.016	.063	.005	.203	-.029	.185
KF, $Q_{t\|t}^{36M}$, $R_{t\|t}^{36M}$	1.494	.057	.214	-.020	.058	.010	.202	-.031	.184
KF, $Q_{t\|t}^{42M}$, $R_{t\|t}^{42M}$	1.520	.069	.201	-.024	.055	.016	.202	-.030	.185
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$	1.659	.053	.156	-.013	.058	.024	.202	-.004	.186
KF, Q=0, $\langle R_{t\|t}^{24M} \rangle$	1.655	.087	.145	-.029	.053	.047	.198	-.015	.183
KF, Q=0, $\langle R_{t\|t}^{30M} \rangle$	1.654	.099	.141	-.034	.050	.051	.197	-.020	.183
KF, Q=0, $\langle R_{t\|t}^{36M} \rangle$	1.654	.092	.143	-.031	.048	.048	.197	-.019	.183
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	1.654	.095	.139	-.031	.048	.048	.197	-.019	.183
VASB, $g = 0.6$	1.390	.004	.335	-.009	.084	.012	.224	-.009	.193
VASB, $g = 0.7$	1.482	.002	.273	-.010	.071	.015	.211	-.011	.188
VASB, $g = 0.8$	1.564	.009	.215	-.015	.059	.025	.200	-.012	.186
VASB, $g = 0.9$	1.640	.034	.169	-.027	.054	.049	.197	-.015	.186
VASB, $g = n/a$	1.691	.081	.160	-.040	.056	.064	.200	-.026	.187

Table 2.2. Biases and dispersions of the forecasted response $y_{t|t-1}$ and state $x_{t|t-1}$ by solving the simulated time-varying regression Eqs. (29) to (31) using the rolling T_r-month OLS, the KF of time-varying $Q_t = Q_{t|t}^{OLS}$ and $R_t = R_{t|t}^{OLS}$, the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$, and the VASB of $F_t = I_m$ and various g. The response y_t is generated by the known x_t of Eq. (34) to simulate a "time-varying small-capitalization growth style market neutral strategy". The biases and dispersions of the forecasts are measured by the RMS of monthly error $\Delta y_{t|t-1} = y_{t|t-1} - y_t$, and the mean and standard deviation of the elements of monthly difference $\Delta x_{t|t-1} = x_{t|t-1} - x_t$. For the forecasts, a lower StDev of $\Delta x_{t|t-1}$ is typically associated with a lower RMSE. The KF of constant Q and R and the VASB can yield a good $y_{t|t-1}$ with a low RMSE.

Forecasts	RMSE	$\boldsymbol{b_0(t)}$		$\boldsymbol{b_{R3K}(t)}$		$\boldsymbol{b_{R3G-R3V}(t)}$		$\boldsymbol{b_{R1K-R2K}(t)}$	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
True Value of x_t		0		0		$1 + a \sin \phi t$		$-1 + a \cos \phi t$	
18-Mon OLS	1.895	-.001	.439	-.002	.124	-.004	.267	-.004	.228
24-Mon OLS	1.867	.007	.367	-.005	.101	.000	.267	-.009	.237
30-Mon OLS	1.845	.015	.326	-.006	.088	.001	.270	-.012	.247
36-Mon OLS	1.834	.021	.299	-.008	.080	.000	.268	-.013	.250
42-Mon OLS	1.813	.028	.278	-.010	.074	-.001	.257	-.014	.245
KF, $Q_{t\|t}^{18M}$, $R_{t\|t}^{18M}$	1.784	.009	.317	-.008	.083	.003	.223	-.016	.195
KF, $Q_{t\|t}^{24M}$, $R_{t\|t}^{24M}$	1.753	.022	.263	-.011	.070	.003	.213	-.025	.192
KF, $Q_{t\|t}^{30M}$, $R_{t\|t}^{30M}$	1.738	.044	.232	-.016	.063	.005	.207	-.029	.190
KF, $Q_{t\|t}^{36M}$, $R_{t\|t}^{36M}$	1.733	.057	.215	-.020	.058	.011	.205	-.031	.188
KF, $Q_{t\|t}^{42M}$, $R_{t\|t}^{42M}$	1.728	.069	.201	-.024	.055	.017	.205	-.030	.188
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$	1.716	.053	.156	-.013	.058	.024	.203	-.004	.187
KF, Q=0, $\langle R_{t\|t}^{24M} \rangle$	1.710	.087	.145	-.030	.053	.047	.199	-.015	.184
KF, Q=0, $\langle R_{t\|t}^{30M} \rangle$	1.708	.099	.142	-.034	.050	.051	.198	-.020	.184
KF, Q=0, $\langle R_{t\|t}^{36M} \rangle$	1.707	.092	.143	-.031	.048	.048	.198	-.019	.184
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	1.706	.095	.140	-.031	.048	.048	.198	-.019	.184
VASB, $g = 0.6$	1.806	.004	.335	-.009	.084	.012	.230	-.009	.200
VASB, $g = 0.7$	1.767	.003	.274	-.011	.071	.015	.216	-.010	.194
VASB, $g = 0.8$	1.736	.009	.215	-.015	.059	.026	.204	-.011	.190
VASB, $g = 0.9$	1.717	.034	.169	-.027	.055	.049	.198	-.015	.188
VASB, $g = n/a$	1.721	.081	.161	-.040	.057	.064	.201	-.026	.188

Table 3.1. Performances of forecasting DJI returns by applying the rolling T_r-month OLS regression to the multivariate time-varying regression Eqs. (35) to (37). The forecasting performances of $y_{t+1|t}$ of Eq. (39) (the upper panel) and $y_{t|t}$ of Eq. (40) (the lower panel) are measured by the t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts. The model(s) selected by the highest IR or IC in the in-sample period (7/1983-6/1998) can be evaluated by the IR and IC in the out-of-sample period (7/1998-6/2013). Also shown are RMSE and standard deviation of the monthly forecasts. The best in-sample models, "$y_{t+1|t}$ by 42-Mon OLS" and "$y_{t|t}$ by 24/36-Mon OLS", are not the best out-of-sample models.

Rolling OLS	**July 1983 – June 1998**				**July 1998 – June 2013**			
	***t* (IR)**	***ρ* (IC)**	RMSE	StDev	***t* (IR)**	***ρ* (IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by rolling OLS								
12-Mon OLS	-0.16	-0.019	14.59	13.62	0.51	0.040	10.13	9.23
18-Mon OLS	0.99	0.079	7.14	5.59	-1.10	-0.079	7.18	5.19
24-Mon OLS	1.37	0.115	5.93	4.09	0.44	0.036	5.88	3.94
30-Mon OLS	1.71	0.139	5.41	3.25	0.77	0.060	5.51	3.44
36-Mon OLS	2.02	0.167	5.09	2.77	0.37	0.033	5.56	3.38
42-Mon OLS	**2.32**	**0.188**	4.89	2.45	**0.43**	**0.039**	5.36	3.04
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by rolling OLS								
12-Mon OLS	1.24	0.098	5.58	3.70	0.90	0.083	5.75	3.94
18-Mon OLS	2.03	0.136	5.14	3.03	0.48	0.043	5.49	3.30
24-Mon OLS	2.75	**0.190**	4.84	2.62	**0.48**	**0.042**	5.36	3.08
30-Mon OLS	**2.76**	0.185	4.82	2.45	**0.56**	**0.050**	5.21	2.84
36-Mon OLS	2.50	0.172	4.75	2.18	0.74	0.067	5.11	2.71
42-Mon OLS	2.15	0.158	4.79	2.07	0.93	0.082	5.03	2.59

Table 3.2. Sample statistics of the elements of monthly state vector $x_{t|t}$ estimated by the rolling 36-month OLS in the Table 3.1 for the two 180-month periods. The "25% t" denotes the 25 percentile of the 180 monthly t-statistic of an element. The 180-month p-value of null hypothesis, sample mean, standard deviation and serial correlation of an element reflect the average effect of an independent variable on the forecast $r_{DJI}(t|t-1)$. According to the p-value, the growth-value spreads $r_{R3G-R3V}(t)$ and $\mu_{R3G-R3V}(t|t)$ may have no effects on the forecasts.

	25% t	**50% t**	**75% t**	**p-Val**	**Mean**	**StDev**	**SCorr**
In-sample Period I: July 1983 – June 1998							
$b_0(t)$	1.06	1.79	2.71	0.000	3.87	2.40	0.93
$b_{DJI-R3K}^{(r)}(t)$	-0.34	0.29	0.69	0.026	0.10	0.62	0.91
$b_{R3K}^{(r)}(t)$	-0.94	-0.63	0.42	0.000	-0.08	0.23	0.93
$b_{R3G-R3V}^{(r)}(t)$	-0.90	-0.02	0.40	***0.220***	0.05	0.55	0.96
$b_{R1K-R2K}^{(r)}(t)$	-1.25	-0.54	0.17	0.000	-0.20	0.34	0.94
$b_{DJI-R3K}^{(\mu)}(t)$	-2.03	-0.83	0.49	0.000	-1.87	4.64	0.95
$b_{R3K}^{(\mu)}(t)$	-1.43	-1.02	-0.55	0.000	-1.39	1.13	0.89
$b_{R3G-R3V}^{(\mu)}(t)$	-0.46	0.05	1.48	0.000	0.95	3.36	0.95
$b_{R1K-R2K}^{(\mu)}(t)$	-0.21	0.69	1.41	0.000	1.13	2.30	0.95
Out-of-sample Period II: July 1998 – June 2013							
$b_0(t)$	-0.03	0.77	1.56	0.000	1.63	2.49	0.98
$b_{DJI-R3K}^{(r)}(t)$	-0.67	0.38	2.08	0.000	0.60	1.24	0.97
$b_{R3K}^{(r)}(t)$	0.42	0.72	1.13	0.000	0.20	0.14	0.79
$b_{R3G-R3V}^{(r)}(t)$	-0.81	-0.22	0.60	***0.130***	-0.04	0.37	0.90
$b_{R1K-R2K}^{(r)}(t)$	-1.62	-0.80	0.31	0.000	-0.19	0.46	0.95
$b_{DJI-R3K}^{(\mu)}(t)$	-2.60	-1.18	-0.38	0.000	-6.17	5.91	0.97
$b_{R3K}^{(\mu)}(t)$	-2.33	-1.73	-1.24	0.000	-2.18	1.07	0.89
$b_{R3G-R3V}^{(\mu)}(t)$	-1.33	0.08	0.76	***0.244***	-0.22	2.56	0.93
$b_{R1K-R2K}^{(\mu)}(t)$	-0.58	0.51	1.98	0.000	1.56	4.25	0.98

Table 4. Performances of forecasting DJI returns by applying the Kalman filter of time-varying $Q_t = Q_{t|t}^{OLS}$ and $R_t = R_{t|t}^{OLS}$ to the multivariate time-varying regression Eqs. (35) to (37). The forecasting performances of $y_{t+1|t}$ of Eq. (39) (the upper panel) and $y_{t|t}$ of Eq. (40) (the lower panel) are measured by the t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts. The model(s) selected by the highest IR or IC in the in-sample period (7/1983-6/1998) can be evaluated by the IR and IC in the out-of-sample period (7/1998-6/2013). Also shown are RMSE and standard deviation of the monthly forecasts. The best in-sample models, "$y_{t+1|t}$ by KF of $Q_{t|t}^{(12M\ OLS)}$ and $R_{t|t}^{(12M\ OLS)}$" and "$y_{t|t}$ by KF of $Q_{t|t}^{(36M\ OLS)}$ and $R_{t|t}^{(36M\ OLS)}$", are not as good as the selected rolling OLS models in the Table 3.1.

Kalman filter	**July 1983 – June 1998**				**July 1998 – June 2013**			
	t (IR)	**ρ (IC)**	RMSE	StDev	**t (IR)**	**ρ (IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by Kalman filter								
KF, $Q_{t\|t}^{12M}$, $R_{t\|t}^{12M}$	**1.23**	**0.093**	5.85	4.11	**0.27**	**0.024**	6.39	4.62
KF, $Q_{t\|t}^{18M}$, $R_{t\|t}^{18M}$	0.64	0.047	5.52	3.34	-0.48	-0.043	6.22	4.08
KF, $Q_{t\|t}^{24M}$, $R_{t\|t}^{24M}$	0.52	0.042	5.18	2.57	-0.28	-0.022	5.50	3.01
KF, $Q_{t\|t}^{30M}$, $R_{t\|t}^{30M}$	0.48	0.040	4.98	2.21	-0.39	-0.032	5.39	2.77
KF, $Q_{t\|t}^{36M}$, $R_{t\|t}^{36M}$	1.00	0.080	4.88	2.28	-0.28	-0.024	5.20	2.44
KF, $Q_{t\|t}^{42M}$, $R_{t\|t}^{42M}$	1.00	0.083	4.71	1.78	-0.38	-0.033	5.12	2.22
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by Kalman filter								
KF, $Q_{t\|t}^{12M}$, $R_{t\|t}^{12M}$	0.58	0.048	5.99	4.13	0.80	0.070	6.10	4.44
KF, $Q_{t\|t}^{18M}$, $R_{t\|t}^{18M}$	0.71	0.053	5.58	3.48	0.76	0.069	5.84	4.02
KF, $Q_{t\|t}^{24M}$, $R_{t\|t}^{24M}$	0.75	0.051	5.23	2.81	0.87	0.076	5.40	3.30
KF, $Q_{t\|t}^{30M}$, $R_{t\|t}^{30M}$	1.08	0.074	5.01	2.48	0.80	0.071	5.20	2.91
KF, $Q_{t\|t}^{36M}$, $R_{t\|t}^{36M}$	**1.62**	**0.121**	4.82	2.37	**1.20**	**0.106**	4.97	2.60
KF, $Q_{t\|t}^{42M}$, $R_{t\|t}^{42M}$	1.42	0.097	4.71	1.88	1.17	0.102	4.89	2.35

Table 5.1. Performances of forecasting DJI returns by applying the Kalman filter of constant $Q = 0$ and $R = \langle R_{t|t}^{OLS} \rangle$ (sample mean of $R_{t|t}^{OLS}$ in the in-sample period) to the multivariate time-varying regression Eqs. (35) to (37). The forecasting performances of $y_{t+1|t}$ of Eq. (39) (the upper panel) and $y_{t|t}$ of Eq. (40) (the lower panel) are measured by the t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts. The model(s) selected by the highest IR or IC in the in-sample period (7/1983-6/1998) can be evaluated by the IR and IC in the out-of-sample period (7/1998-6/2013). Also shown are RMSE and standard deviation of the monthly forecasts. The best in-sample models, "$y_{t+1|t}$ and $y_{t|t}$ by KF of Q=0 and $\langle R_{t|t}^{(42M\ OLS)} \rangle$", are also the best out-of-sample models. These KF models of constant Q and R are consistently better than the rolling OLS models in the Table 3.1.

Kalman filter	**July 1983 – June 1998**				**July 1998 – June 2013**			
	***t* (IR)**	***ρ* (IC)**	RMSE	StDev	***t* (IR)**	***ρ* (IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by Kalman filter								
KF, Q=0, $\langle R_{t\|t}^{12M} \rangle$	1.32	0.124	4.63	1.70	1.31	0.126	4.75	1.56
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$	1.95	0.162	4.40	1.07	1.59	0.141	4.58	1.15
KF, Q=0, $\langle R_{t\|t}^{24M} \rangle$	1.99	0.167	4.39	1.03	1.54	0.138	4.58	1.12
KF, Q=0, $\langle R_{t\|t}^{30M} \rangle$	2.12	0.178	4.35	0.83	1.58	0.136	4.55	0.96
KF, Q=0, $\langle R_{t\|t}^{36M} \rangle$	2.19	0.182	4.34	0.78	1.61	0.137	4.54	0.92
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	**2.24**	**0.183**	4.33	0.73	**1.66**	**0.137**	4.53	0.87
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by Kalman filter								
KF, Q=0, $\langle R_{t\|t}^{12M} \rangle$	1.51	0.119	4.58	1.51	1.85	0.156	4.67	1.55
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$	2.08	0.158	4.40	1.03	2.09	0.164	4.54	1.18
KF, Q=0, $\langle R_{t\|t}^{24M} \rangle$	2.11	0.163	4.39	0.99	2.04	0.160	4.54	1.15
KF, Q=0, $\langle R_{t\|t}^{30M} \rangle$	2.24	0.174	4.36	0.83	2.11	0.160	4.51	1.00
KF, Q=0, $\langle R_{t\|t}^{36M} \rangle$	2.30	0.178	4.34	0.79	2.15	0.161	4.50	0.97
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	**2.34**	**0.180**	4.34	0.75	**2.20**	**0.163**	4.49	0.93

Table 5.2. Sample statistics of the elements of monthly state vector $x_{t|t}$ estimated by the KF of constant $Q = 0$ and $R = \langle R_{t|t}^{(42M\ OLS)} \rangle$ in the Table 5.1 for the two 180-month periods. The "$25\%\ t$" denotes the 25 percentile of the 180 monthly t-statistic of an element. The 180-month p-value of null hypothesis, sample mean, standard deviation and serial correlation of an element reflect the average effect of an independent variable on the forecast $r_{DJI}(t|t-1)$. According to the Mean, StDev and $25\%\ t$ or $75\%\ t$, the variables $r_{R3K}(t)$ and $r_{R3G-R3V}(t)$ may have weaker effects than others on the forecasts.

	25% *t*	**50% *t***	**75% *t***	***p*-Val**	**Mean**	**StDev**	**SCorr**
			In-sample Period I: July 1983 – June 1998				
$b_0(t)$	2.74	3.23	3.48	0.000	1.238	0.141	0.94
$b_{DJI-R3K}^{(r)}(t)$	-0.18	-0.05	0.07	0.000	-0.023	0.073	0.97
$b_{R3K}^{(r)}(t)$	-0.02	0.13	0.39	0.000	0.014	0.026	0.94
$b_{R3G-R3V}^{(r)}(t)$	-0.09	0.09	0.26	0.000	0.016	0.043	0.91
$b_{R1K-R2K}^{(r)}(t)$	-1.66	-1.56	-1.33	0.000	-0.224	0.047	0.94
$b_{DJI-R3K}^{(\mu)}(t)$	-1.43	-1.22	-0.79	0.000	-0.676	0.153	0.97
$b_{R3K}^{(\mu)}(t)$	-2.04	-1.80	-1.25	0.000	-0.357	0.065	0.94
$b_{R3G-R3V}^{(\mu)}(t)$	-0.26	-0.12	0.46	0.000	0.120	0.292	0.98
$b_{R1K-R2K}^{(\mu)}(t)$	1.26	1.36	1.57	0.000	0.663	0.152	0.95
			Out-of-sample Period II: July 1998 – June 2013				
$b_0(t)$	4.50	4.66	4.78	0.000	1.173	0.138	0.99
$b_{DJI-R3K}^{(r)}(t)$	-0.79	-0.71	-0.40	0.000	-0.099	0.036	0.94
$b_{R3K}^{(r)}(t)$	0.04	0.15	0.69	0.000	0.015	0.017	0.97
$b_{R3G-R3V}^{(r)}(t)$	0.25	0.40	0.57	0.000	0.025	0.030	0.95
$b_{R1K-R2K}^{(r)}(t)$	-2.61	-2.54	-2.18	0.000	-0.206	0.034	0.97
$b_{DJI-R3K}^{(\mu)}(t)$	-3.11	-2.78	-2.33	0.000	-1.168	0.159	0.99
$b_{R3K}^{(\mu)}(t)$	-2.59	-2.43	-2.20	0.000	-0.343	0.054	0.98
$b_{R3G-R3V}^{(\mu)}(t)$	-0.02	0.35	0.45	0.000	0.036	0.120	0.96
$b_{R1K-R2K}^{(\mu)}(t)$	3.09	3.39	3.88	0.000	0.920	0.108	0.97

Table 6.1. Performances of forecasting DJI returns by applying the VASB algorithm of $F_t = I_m$ and various g to the multivariate time-varying regression Eqs. (35) to (37). The forecasting performances of $y_{t+1|t}$ of Eq. (39) (the upper panel) and $y_{t|t}$ of Eq. (40) (the lower panel) are measured by the t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts. The model(s) selected by the highest IR or IC in the in-sample period (7/1983-6/1998) can be evaluated by the IR and IC in the out-of-sample period (7/1998-6/2013). Also shown are RMSE and standard deviation of the monthly forecasts. The best in-sample models, "$y_{t+1|t}$ and $y_{t|t}$ by VASB of $g = 0.94$ and $g = n/a$", are also the best out-of-sample model with $y_{t+1|t}$ but not the best with $y_{t|t}$. A comparison of the best VASB vs. the best OLS and KF is shown in the Table 7.

VASB, $F_t = I_m$	**July 1983 – June 1998**				**July 1998 – June 2013**			
	t (IR)	**ρ (IC)**	RMSE	StDev	**t (IR)**	**ρ (IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by VASB								
VASB, $g = 0.84$	2.30	0.196	4.32	0.74	1.14	0.089	4.58	0.88
VASB, $g = 0.86$	2.38	0.200	4.31	0.69	1.18	0.092	4.57	0.82
VASB, $g = 0.88$	2.47	0.204	4.30	0.64	1.24	0.096	4.55	0.75
VASB, $g = 0.90$	2.57	0.208	4.30	0.59	1.31	0.102	4.54	0.68
VASB, $g = 0.92$	2.67	0.212	4.29	0.54	1.42	0.109	4.52	0.62
VASB, $g = 0.94$	**2.78**	0.215	4.28	0.48	**1.56**	**0.116**	4.50	0.56
VASB, $g = n/a$	2.70	**0.218**	4.29	0.71	**1.70**	**0.136**	4.52	0.74
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by VASB								
VASB, $g = 0.84$	2.23	0.195	4.33	0.73	2.71	0.217	4.41	0.90
VASB, $g = 0.86$	2.29	0.199	4.32	0.68	2.72	0.217	4.41	0.84
VASB, $g = 0.88$	2.36	0.203	4.31	0.63	2.71	0.215	4.41	0.77
VASB, $g = 0.90$	2.44	0.208	4.30	0.58	2.66	0.209	4.41	0.70
VASB, $g = 0.92$	2.54	0.212	4.29	0.53	2.56	0.198	4.42	0.64
VASB, $g = 0.94$	2.66	**0.217**	4.28	0.47	**2.41**	**0.182**	4.44	0.57
VASB, $g = n/a$	**2.70**	0.209	4.30	0.69	**2.63**	**0.204**	4.43	0.75

Table 6.2. Sample statistics of the elements of monthly state vector $x_{t|t}$ estimated by the VASB of $g = 0.94$ in the Table 6.1 for the two 180-month periods. The "25% t" denotes the 25 percentile of the 180 monthly t-statistic of an element. The 180-month p-value of null hypothesis, sample mean, standard deviation and serial correlation of an element reflect the average effect of an independent variable on the forecast $r_{DJI}(t|t-1)$. According to the 25% t or 75% t, p-value, Mean and StDev, the spreads $r_{R3G-R3V}(t)$ and $r_{R1K-R2K}(t)$ may have no or weaker effects relative to others on the forecasts.

	25% t	**50% t**	**75% t**	**p-Val**	**Mean**	**StDev**	**SCorr**
In-sample Period I: July 1983 – June 1998							
$b_0(t)$	0.75	1.65	2.78	0.000	0.511	0.232	0.99
$b_{DJI-R3K}^{(r)}(t)$	-0.21	-0.12	-0.06	0.000	-0.016	0.015	0.97
$b_{R3K}^{(r)}(t)$	-0.60	-0.13	-0.03	0.000	-0.014	0.015	0.99
$b_{R3G-R3V}^{(r)}(t)$	-0.28	-0.06	-0.02	0.000	-0.013	0.015	0.98
$b_{R1K-R2K}^{(r)}(t)$	-0.17	-0.07	-0.03	0.000	-0.007	0.007	0.93
$b_{DJI-R3K}^{(\mu)}(t)$	-0.57	-0.36	-0.19	0.000	-0.150	0.084	0.98
$b_{R3K}^{(\mu)}(t)$	0.55	0.91	1.26	0.000	0.182	0.056	0.96
$b_{R3G-R3V}^{(\mu)}(t)$	-0.19	-0.12	0.04	0.000	-0.021	0.051	0.96
$b_{R1K-R2K}^{(\mu)}(t)$	0.44	0.64	1.13	0.000	0.246	0.119	0.99
Out-of-sample Period II: July 1998 – June 2013							
$b_0(t)$	2.42	3.02	3.96	0.000	0.793	0.050	0.93
$b_{DJI-R3K}^{(r)}(t)$	-0.65	-0.45	-0.25	0.000	-0.057	0.017	0.96
$b_{R3K}^{(r)}(t)$	-0.80	-0.56	-0.40	0.000	-0.033	0.008	0.92
$b_{R3G-R3V}^{(r)}(t)$	-0.22	0.09	0.22	***0.639***	0.001	0.027	0.99
$b_{R1K-R2K}^{(r)}(t)$	-0.50	-0.36	-0.17	0.000	-0.025	0.014	0.97
$b_{DJI-R3K}^{(\mu)}(t)$	-1.01	-0.79	-0.65	0.000	-0.388	0.101	0.98
$b_{R3K}^{(\mu)}(t)$	0.68	1.00	1.34	0.000	0.178	0.039	0.96
$b_{R3G-R3V}^{(\mu)}(t)$	-0.05	0.21	0.44	0.000	0.075	0.098	0.98
$b_{R1K-R2K}^{(\mu)}(t)$	1.31	1.63	2.09	0.000	0.572	0.054	0.95

Table 7. Performances of forecasting DJI returns by applying the best VASB (in the Table 6.1) to the multivariate time-varying regression Eqs. (35) to (37) vs. the best rolling OLS (in the Table 3.1) and the best KF of constant Q and R (in the Table 5.1). The forecasting performances of $y_{t+1|t}$ of Eq. (39) (the upper panel) and $y_{t|t}$ of Eq. (40) (the lower panel) are measured by the t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts. The model(s) selected by the highest IR or IC in the in-sample period (7/1983-6/1998) can be evaluated by the IR and IC in the out-of-sample period (7/1998-6/2013). Also shown are RMSE and standard deviation of the monthly forecasts. The models not showing their in-sample results are those not being selected but having the better out-of-sample performance(s) than the selected ones. The selected VASB models performed better than, or at least similar to, the best OLS or the best KF models.

VASB, $F_t = I_m$	**July 1983 – June 1998**				**July 1998 – June 2013**			
	t (IR)	**ρ (IC)**	RMSE	StDev	**t (IR)**	**ρ (IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by OLS, KF and VASB								
42-Mon OLS	2.32	0.188	4.89	2.45	0.43	0.039	5.36	3.04
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$					1.59	**0.141**	4.58	1.15
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	2.24	0.183	4.33	0.73	1.66	0.137	4.53	0.87
VASB, $g = 0.94$	**2.78**	0.215	4.28	0.48	1.56	0.116	4.50	0.56
VASB, $g = n/a$	2.70	**0.218**	4.29	0.71	**1.70**	0.136	4.52	0.74
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by OLS, KF and VASB								
24-Mon OLS	2.75	0.190	4.84	2.62	0.48	0.042	5.36	3.08
30-Mon OLS	**2.76**	0.185	4.82	2.45	0.56	0.050	5.21	2.84
KF, Q=0, $\langle R_{t\|t}^{18M} \rangle$					2.09	0.164	4.54	1.18
KF, Q=0, $\langle R_{t\|t}^{42M} \rangle$	2.34	0.180	4.34	0.75	2.20	0.163	4.49	0.93
VASB, $g = 0.86$					**2.72**	**0.217**	4.41	0.84
VASB, $g = 0.94$	2.66	**0.217**	4.28	0.47	2.41	0.182	4.44	0.57
VASB, $g = n/a$	2.70	0.209	4.30	0.69	2.63	0.204	4.43	0.75

Table 8. Performances of forecasting DJI returns by applying the VASB algorithm of $g = 0.94$ or $g = n/a$ and various $F_t = f_0\, I_m$ to the multivariate time-varying regression Eqs. (35) to (37), the same way as the Table 6.1. The forecasting models achieving higher t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts have higher predictive power. Also shown are RMSE and standard deviation of the monthly forecasts. The VASB of $F_t = I_m$ is better than the VASB of $F_t = f_0\, I_m$, $f_0 < 1$.

VASB, $F_t = f_0\, I_m$	**July 1983 – June 1998**				**July 1998 – June 2013**			
	$\boldsymbol{t}$ **(IR)**	$\boldsymbol{\rho}$ **(IC)**	RMSE	StDev	$\boldsymbol{t}$ **(IR)**	$\boldsymbol{\rho}$ **(IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by VASB of $g = 0.94$								
VASB, $F_t = 0.90\, I_m$	1.19	0.101	4.37	0.15	-0.31	-0.027	4.52	0.14
VASB, $F_t = 0.92\, I_m$	1.28	0.111	4.37	0.17	-0.20	-0.018	4.52	0.16
VASB, $F_t = 0.94\, I_m$	1.40	0.123	4.36	0.19	-0.04	-0.003	4.51	0.18
VASB, $F_t = 0.96\, I_m$	1.58	0.140	4.35	0.22	0.20	0.018	4.51	0.22
VASB, $F_t = 0.98\, I_m$	1.94	0.166	4.33	0.28	0.62	0.053	4.50	0.29
VASB, $\boldsymbol{F_t = 1.00\, I_m}$	**2.78**	**0.215**	4.28	0.48	**1.56**	**0.116**	4.50	0.56
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by VASB of $g = 0.94$								
VASB, $F_t = 0.90\, I_m$	1.34	0.123	4.36	0.17	0.53	0.044	4.50	0.17
VASB, $F_t = 0.92\, I_m$	1.44	0.135	4.36	0.18	0.58	0.049	4.50	0.18
VASB, $F_t = 0.94\, I_m$	1.57	0.150	4.35	0.20	0.67	0.057	4.50	0.20
VASB, $F_t = 0.96\, I_m$	1.74	0.166	4.34	0.23	0.84	0.073	4.50	0.24
VASB, $F_t = 0.98\, I_m$	2.02	0.186	4.32	0.28	1.27	0.110	4.48	0.31
VASB, $\boldsymbol{F_t = 1.00\, I_m}$	**2.66**	**0.217**	4.28	0.47	**2.41**	**0.182**	4.44	0.57
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by VASB of $g = n/a$								
VASB, $F_t = 0.90\, I_m$	0.78	0.057	4.38	0.24	-1.40	-0.130	4.53	0.15
VASB, $F_t = 0.92\, I_m$	0.85	0.065	4.38	0.26	-1.24	-0.114	4.53	0.17
VASB, $F_t = 0.94\, I_m$	0.94	0.076	4.37	0.29	-0.98	-0.088	4.53	0.19
VASB, $F_t = 0.96\, I_m$	1.10	0.094	4.37	0.33	-0.56	-0.050	4.53	0.22
VASB, $F_t = 0.98\, I_m$	1.46	0.129	4.35	0.39	0.18	0.015	4.52	0.28
VASB, $\boldsymbol{F_t = 1.00\, I_m}$	**2.70**	**0.218**	4.29	0.71	**1.70**	**0.136**	4.52	0.74
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by VASB of $g = n/a$								
VASB, $F_t = 0.90\, I_m$	1.35	0.096	4.37	0.32	-0.34	-0.027	4.52	0.22
VASB, $F_t = 0.92\, I_m$	1.49	0.112	4.36	0.33	-0.25	-0.020	4.52	0.23
VASB, $F_t = 0.94\, I_m$	1.63	0.130	4.35	0.34	-0.09	-0.008	4.52	0.24
VASB, $F_t = 0.96\, I_m$	1.79	0.152	4.34	0.37	0.20	0.017	4.51	0.27
VASB, $F_t = 0.98\, I_m$	2.06	0.180	4.32	0.41	0.91	0.081	4.49	0.33
VASB, $\boldsymbol{F_t = 1.00\, I_m}$	**2.70**	**0.209**	4.30	0.69	**2.63**	**0.204**	4.43	0.75

Table 9. Performances of forecasting DJI returns by applying the VASB algorithm of $F_t = I_m$ and various g to the multivariate time-varying regression Eqs. (35) to (37). The time-varying means of Russell indices are estimated by the "VASB stochastic local levels", instead of the time-weighted averages as those for the Tables 3.1, 4, 5.1, 6.1 and 7. The forecasting models achieving higher t-statistic (proportional to the information ratio IR) or ρ (the information coefficient IC) of the monthly forecasts have higher predictive power. Also shown are RMSE and standard deviation of the monthly forecasts. The best in-sample (7/1983-6/1998) models, "$y_{t+1|t}$ and $y_{t|t}$ by VASB of $g = 0.94$ and $g = n/a$", are also the best out-of-sample (7/1998-6/2013) models with $y_{t+1|t}$ but not the best with $y_{t|t}$. A comparison with the Tables 6.1 and 7 indicates that the use of VASB stochastic local levels results in a higher predictive power than the use of time-weighted averages.

VASB, $F_t = I_m$	**July 1983 – June 1998**				**July 1998 – June 2013**			
	***t* (IR)**	**ρ (IC)**	RMSE	StDev	***t* (IR)**	**ρ (IC)**	RMSE	StDev
Forecasting DJI returns with $y_{t+1\|t}$ of Eq. (39) estimated by VASB								
VASB, $g = 0.84$	2.45	0.200	4.32	0.70	1.44	0.108	4.56	0.83
VASB, $g = 0.86$	2.54	0.205	4.31	0.65	1.45	0.109	4.55	0.77
VASB, $g = 0.88$	2.64	0.210	4.30	0.60	1.47	0.111	4.54	0.70
VASB, $g = 0.90$	2.75	0.215	4.29	0.55	1.52	0.113	4.53	0.62
VASB, $g = 0.92$	2.87	0.219	4.28	0.50	1.60	0.117	4.52	0.55
VASB, $\boldsymbol{g = 0.94}$	**2.99**	0.223	4.28	0.45	**1.71**	**0.122**	4.50	0.48
VASB, $\boldsymbol{g = n/a}$	2.77	**0.225**	4.28	0.62	**1.96**	**0.150**	4.49	0.67
Forecasting DJI returns with $y_{t\|t}$ of Eq. (40) estimated by VASB								
VASB, $g = 0.84$	2.34	0.197	4.33	0.69	2.84	0.224	4.41	0.86
VASB, $g = 0.86$	2.42	0.202	4.31	0.64	2.83	0.222	4.41	0.80
VASB, $g = 0.88$	2.51	0.207	4.30	0.60	2.79	0.217	4.41	0.72
VASB, $g = 0.90$	2.62	0.213	4.29	0.55	2.72	0.209	4.42	0.65
VASB, $g = 0.92$	2.75	0.220	4.28	0.50	2.61	0.198	4.43	0.57
VASB, $\boldsymbol{g = 0.94}$	**2.91**	**0.227**	4.27	0.45	**2.46**	**0.182**	4.44	0.50
VASB, $g = n/a$	2.63	0.210	4.29	0.61	2.80	0.211	4.42	0.70

Table 10. Nineteen financial market indices. Monthly returns of the 19 indices from January 1990 through June 2013 are obtained from Morningstar EnCorr database.

Symbol	Index Name
USAgg	Barclays US Agg Bond
EAFE	MSCI EAFE GR
SP500	IA SBBI S&P 500
Gold	London Fix Gold PM
REITs	FTSE NAREIT All REITs
SmallStk	IA SBBI US Small Stock
LTCorp	IA SBBI US LT Corp
LTGovt	IA SBBI US LT Govt
Yen	Japanese Yen
Franc	Swiss Franc
Sterling	Sterling
Euro	Euro
EmgMkt	MSCI EM GR
GSCI	S&P GSCI
Austr$	Australian Dollar
Can$	Canadian Dollar
HiYld	BofAML US HY Master II
GlobAgg	Barclays Global Aggregate
GlobTreas	Barclays Global Treasury

Table 11. The best rolling, time-weighted and VASB forecasts of time-varying mean and variance of monthly returns for each of the 19 financial market indices. The parameter $1/T_m$ functions as the weight of data in forecasting the mean, $1/T_v$ as the weight of data for forecasting the variance, and F_t is the state transition coefficient of the VASB model. The log-likelihood value (LL) is the 234-month average of 12-month log-likelihood of the forecasts. A higher (i.e. less negative) LL value indicates more accurate forecasts, the same statistical concept as that underlying the classic maximum likelihood estimation (MLE).

Index	**Best Rolling**			**Best Time-Wghtd**			**Best VASB**			
	$T^R_{m,j}$	$T^R_{v,j}$	LL	$T^W_{m,j}$	$T^W_{v,j}$	LL	F^B_j	$T^B_{m,j}$	$T^B_{v,j}$	LL
USAgg	36	18	-18.1	48	48	-18.0	1.00	48	48	-18.0
EAFE	36	12	-36.1	48	12	-35.8	0.96	18	6	-35.8
SP500	36	12	-34.3	48	6	-34.2	0.98	30	6	-34.1
Gold	48	42	-35.2	48	30	-35.2	0.98	48	24	-35.2
REITs	48	12	-36.5	48	12	-36.5	1.00	n/a	12	-36.4
SmallStk	48	18	-38.4	48	12	-38.5	0.98	30	6	-38.4
LTCorp	36	24	-29.3	48	12	-28.9	1.00	n/a	12	-28.8
LTGovt	36	12	-30.8	48	36	-30.6	1.00	n/a	42	-30.5
Yen	48	48	-31.3	48	30	-31.1	0.90	48	48	-30.9
Franc	48	48	-31.2	48	24	-31.1	0.90	n/a	30	-31.0
Sterling	42	48	-28.3	48	24	-28.2	0.90	n/a	24	-28.1
Euro	48	24	-30.1	48	12	-29.9	0.90	n/a	12	-29.8
EmgMkt	36	12	-40.8	36	12	-40.5	0.90	18	6	-40.4
GSCI	48	36	-39.0	48	18	-38.8	1.00	n/a	12	-38.8
Austr$	48	18	-31.8	48	12	-31.6	0.90	42	6	-31.5
Can$	48	18	-26.2	48	12	-26.2	0.94	n/a	12	-26.1
HiYld	18	18	-28.1	6	12	-27.6	0.90	24	6	-27.0
GlobAgg	42	24	-22.7	48	18	-22.7	1.00	n/a	24	-22.7
GlobTreas	42	24	-24.8	48	18	-24.8	1.00	n/a	24	-24.8

Table 12. Comparing the accuracy of forecasted time-varying mean and variance of monthly returns of the 19 financial market indices by the best rolling, the best time-weighted and the best VASB models. The measures ΔLL is the 234-month average of difference of 12-month log-likelihood, and p-value is of the null hypothesis for the difference of log-likelihood. The best time-weighted forecasts are more accurate than the best rolling ones (with positive ΔLL) for 16 indices, and significantly more accurate (with small p-value favoring alternative hypothesis H_a) for 12 indices. The best VASB forecasts are more accurate than the best rolling ones (with positive ΔLL) for 18 indices and significantly more accurate (with small p-Val favoring H_a) for 13 indices. The best VASB forecasts are more accurate than the best time-weighted ones (with positive ΔLL) for 18 indices, and significantly more accurate (with small p-Val favoring H_a) for 13 indices.

Index	TimeWtd vs. Rolling			VASB vs. Rolling			VASB vs. TimeWtd		
	ΔLL	p-Val	H_a	ΔLL	p-Val	H_a	ΔLL	p-Val	H_a
USAgg	0.113	0.025	Yes	0.138	0.007	Yes	0.026	0.004	Yes
EAFE	0.233	0.003	Yes	0.321	0.000	Yes	0.088	0.070	
SP500	0.073	0.269		0.165	0.001	Yes	0.092	0.024	Yes
Gold	-0.008	0.846		-0.024	0.591		-0.016	0.341	
REITs	0.031	0.691		0.111	0.255		0.080	0.004	Yes
SmallStk	-0.041	0.259		0.002	0.971		0.043	0.293	
LTCorp	0.486	0.000	Yes	0.564	0.000	Yes	0.078	0.006	Yes
LTGovt	0.189	0.010	Yes	0.296	0.000	Yes	0.107	0.000	Yes
Yen	0.193	0.000	Yes	0.385	0.000	Yes	0.192	0.000	Yes
Franc	0.083	0.015	Yes	0.178	0.000	Yes	0.095	0.000	Yes
Sterling	0.179	0.000	Yes	0.282	0.000	Yes	0.103	0.000	Yes
Euro	0.207	0.003	Yes	0.289	0.001	Yes	0.082	0.001	Yes
EmgMkt	0.345	0.000	Yes	0.459	0.000	Yes	0.115	0.035	Yes
GSCI	0.122	0.007	Yes	0.168	0.015	Yes	0.045	0.199	
Austr$	0.138	0.001	Yes	0.245	0.000	Yes	0.107	0.010	Yes
Can$	-0.016	0.783		0.048	0.437		0.063	0.009	Yes
HiYld	0.550	0.001	Yes	1.105	0.000	Yes	0.555	0.000	Yes
GlobAgg	0.032	0.330		0.040	0.380		0.007	0.767	
GlobTreas	0.027	0.429		0.047	0.310		0.019	0.384	