Variational Bayesian filtering solutions to popular quantitative models

H. Fox Ling, PhD, CFA
H.Fox.Ling@gmail.com
January 2018

Content

- Bayesian filtering framework
- Solution 1, stochastic factor model
- Solution 2, time-series regression
- Solution 3, vector autoregressive model
- Reference
- Appendix: factor model comparison

VBF

- **Bayesian filtering framework**
- Solution 1, stochastic factor model
- Solution 2, time-series regression
- Solution 3, vector autoregressive model
- Reference
- Appendix: factor model comparison

Background (I)

- Many time-series models have both
  - frequentist and probabilistic solutions
- Examples of popular time-series models
  - Simple and multivariate time-varying regression
  - Local level, or unobserved component, model
  - Vector autoregressive (VAR) model
  - Statistical factor model (SFM)
- Popular frequentist solutions
  - Trailing (rolling, moving) window analysis
  - Time-weighted analysis (e.g. exponentially decayed weights)

Background (II)

- Popular frequentist solutions (continue)
    - Principal component analysis (PCA, trailing or time-weighted) for SFM
- Some probabilistic filtering solutions
    - Kalman filter (KF, when both transition noise and residual error are known inputs)
    - Adaptive KF (if either transition noise or residual error is unknown)
    - Bayesian filter (BF, if both transition noise and residual error are unknowns)
    - Particle filter (if BF posterior is approximated by stochastic sampling)

Background (III)

- Some probabilistic filtering solutions (continue)
    - Variational BF (if BF posterior is approximated by factorized distribution)
- Variational Bayesian approximation (VBA)
    - Approximating posterior by factorized distribution
    - One of the useful tools in machine learning
- Frequentist vs. Bayesian?
    - SAS (Dec 2017): "Bayesian Analysis: Advantages and Disadvantages“, http://documentation.sas.com/?docsetId=statug&docsetTarget=statug_introbayes_sect015.htm&docsetVersion=14.3&locale=en

Background (IV)

- This presentation
  - Framework and model-specific formulas of
  - variational Bayesian filtering solutions
  - to these popular time-series models

State Space Model

- State space model
  - Observed data vector time-series: y_t, $t = 1, 2, \cdots$
  - Unknown hidden state vector time-series: Z_t
  - Equations of state space representation

    distribution of data: $p(\, y_t \mid Z_t \,)$

    dynamics of states: $p(\, Z_t \mid Z_{t-1} \,)$
- Filtering
  - Estimating hidden states Z_t by all of available data $y_{1:t}$
- A probabilistic filtering [1]
  - Estimating conditional distribution $p(\, Z_t \mid y_{1:t} \,)$ by

    last estimate $p(\, Z_{t-1} \mid y_{1:t-1} \,)$ and current data y_t

Bayesian Filtering

- Chapman-Kolmogorov equation
  - Predicted states: $p(\, Z_t \mid y_{1:t-1} \,)$
  - $p(\, Z_t \mid y_{1:t-1} \,) = \int p(\, Z_t \mid Z_{t-1} \,)\, p(\, Z_{t-1} \mid y_{1:t-1} \,)\, dZ_{t-1}$
  - In practice, prediction may be obtained, estimated or assumed directly by dynamics without integration
- Bayesian filtering [1] at time t
  - Likelihood (data distribution): $p(\, y_t \mid Z_t \,)$
  - Prior (last prediction of states): $p(\, Z_t \mid y_{1:t-1} \,)$
  - Posterior (new estimate of states): $p(\, Z_t \mid y_{1:t} \,) \propto$

    $\propto p(\, y_t \mid Z_t \,)\, p(\, Z_t \mid y_{1:t-1} \,) \propto p(\, y_t, Z_t \mid y_{1:t-1} \,)$
- Most of actual posteriors, $p(\, Z_t \mid y_{1:t} \,)$, are intractable

Approximation of Posterior

- Stochastic approximation
  - Approximating posterior by a numerical distribution using stochastic sampling
  - Approximated posterior → exact posterior when sample size → ∞
  - Example: Markov Chain Monte Carlo (MCMC) method
  - Algorithm: particle filter (each numeric pass is a particle)
  - Model size may not be too large
- Analytic approximation
  - Approximating posterior by a simpler tractable distribution
  - Example: variational Bayesian approximation (VBA)

Variational Bayesian Filter (I)

- Variational Bayesian approximation (VBA) [2, 3, 4, 5]
  - The only assumption: approximating posterior by factorized, i.e. separable, distribution
  - Developed long ago in quantum mechanics
  - Now used in science, engineering and machine learning
  - On its way to economics and finance
  - Model size can be very large
- Variational Bayesian (VB) filtering
  - Grouping hidden states as $Z_t^T = [\, Z_{1,t}^T, Z_{2,t}^T, \cdots, Z_{k,t}^T \,]$
  - Approximating posterior $p(Z_t|y_{1:t})$ by a factorized distribution $q(Z_t) = \prod_{j=1}^{k} q_j(Z_{j,t})$

Variational Bayesian Filter (II)

- Advantages of VBA [2, 3, 4, 5]
  - No need to assume specific form of $q_j(Z_{j,t})$
  - Optimal forms of $\{q_j(Z_{j,t})\}$ will be identified one by one
  - Optimal distributions $\{q_j(Z_{j,t})\}$ can be solved iteratively
  - VBA iteration always converges with appropriate initial values
  - Optimal $q(Z_t)$ approximates true local optima
- Major weakness of VBA [2, 3, 4, 5]
  - Factorized approximation $q(Z_t)$ with $k \geq 2$ will likely not to converge to true optima

VBF

- Bayesian filtering framework
- **Solution 1, stochastic factor model**
- Solution 2, time-series regression
- Solution 3, vector autoregressive model
- Reference
- Appendix: factor model comparison

Factor Model Overview (I)

- Stationary factor model of vector time-series y_t

$$y_t = X f_t + e_t$$

- Time-varying factor model of data vector y_t

$$y_t = X_t f_t + e_t$$

- Symbols and dimensions
  - t = time index, $t = 1, 2, \cdots$
  - y_t = n-vector for n observed time-series, $n \gg 1$
  - f_t = m-vector for m factor scores, $m \ll n$
  - X_t = $n \times m$ matrix of factor loadings, sensitivities of n time-series to m factor scores
  - e_t = n-vector for n error time-series

Factor Model Overview (II)

- Fundamental factor models

  $$y_t = X_t f_t + e_t$$

  - known factor loadings X_t, unknown factor scores f_t
  - estimation: cross-sectional regression

- Models with known factor time-series

  $$y_t = X_t f_t + e_t$$

  - unknown loadings X_t, known factors f_t
  - estimation: time-series regressions

- Statistical factor models (only data y_t is known)

  $$y_t = X_t f_t + e_t$$

  - unknown loadings X_t, unknown factors f_t
  - estimation: numerous approaches

Estimations of FMs (I)

- Statistical factor model of data vector y_t

  $$y_t = X_t f_t + e_t$$

- Frequentist modeling
  - Random variables: f_t, e_t
  - Deterministic parameters: X_t, $\mathrm{cov}(f_t)$, $\mathrm{cov}(e_t)$
  - Estimating optimal values of the variables and parameters given data y_t
- Bayesian modeling
  - Random variables: X_t, f_t, e_t
  - Random variables: $Cov(vec(X_t))$, $Cov(f_t)$, $Cov(e_t)$
  - Estimating optimal conditional joint distribution of the random variables given data y_t

Estimations of FMs (II)

- Stationary statistical factor models

  $$y_t = X\, f_t + e_t$$

  - unknown constant loadings X, unknown factors f_t
- Estimation methods
  - PCA, asymptotic PCA, PCA augmented with Bayesian analysis, thresholding principal orthogonal complements
  - maximum likelihood (ML) factor analysis, expectation-maximization (EM) factor analysis, Bayesian factor analysis, ML-EM-Bayesian factor model, variational Bayesian factor analysis (VBFA), multi-step VBFA
  - variational Bayesian PCA, MCMC Bayesian PCA
  - dynamic factor model (DFM)

Estimations of FMs (III)

- Time-varying statistical factor models

  $$y_t = X_t\, f_t + e_t$$

  - unknown time-varying loadings X_t, unknown factors f_t
- Estimation methods
  - *by rolling-window, a local stationarity approximation*: all of the approaches to stationary statistical factor models
  - *by sequential Bayes, with stochastic approximation, for low-dimensional FMs*: MCMC Bayesian factor analysis, Gibbs sampling for DFM, dynamic latent factors and time-varying sparse loadings with MCMC solutions
  - *by sequential Bayes, with analytic approximation, for high-dimensional FMs*: variational Bayesian filtering [6]

Bayesian Modeling (I)

- The simplest distributions to be assumed
  - for random variables in time-varying statistical factor models, $y_t = X_t f_t + e_t$
- Multivariate normal distributions
  - Maximum entropy (i.e. the most appropriate assumption) given only mean and variance-covariance
  - $f_t \sim N(E(f_t), P_t)$
  - $X_{j-col,t} \sim N(E(X_{j-col,t}), U_{j,t}),\ \ j = 1, 2, \cdots, m$
  - $e_t \sim N(0, R_t)$
  - Elements of f_t, X_t and e_t independent of each other
  - P_t, $\{U_{j,t}\}$ and R_t: diagonal random matrixes

Bayesian Modeling (II)

- Inverse-gamma distribution
  - Conjugate prior of variance of normal distribution
  - Diagonal elements of P_t, $\{U_{j,t}\}$ and R_t: independent inverse-gamma variables
- A "diagonal inverse-gamma" distribution $IG_D(\cdot)$
  - = Product of independent inverse-gamma distributions $IG(\cdot)$
  - Expression: $IG_D(P_t; \alpha_t, B_{P,t}) = \prod_{j=1}^{m} IG((P_t)_{jj}; \alpha_t, (B_{P,t})_{jj})$
- "Diagonal inverse-gamma" distributions [6]
  - $P_t \sim IG_D(\alpha_t, B_{P,t})$
  - $U_{j,t} \sim IG_D(\alpha_t, B_{U,j,t}),\ \ j = 1, 2, \cdots, m$
  - $R_t \sim IG_D(\alpha_t, B_{R,t})$

Bayesian Modeling (III)

- "Diagonal inverse-gamma" distributions (continue)
  - α_t: shape parameter common to all variables
  - $B_{P,t}$, $\{B_{U,j,t}\}$ and $B_{R,t}$: diagonal matrixes of scale parameters
- Bayesian modeling
  - Joint distribution = Likelihood × Priors,

$$p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) =$$

$$= p(y_t|X_t, f_t, R_t)p(f_t|P_t)p(P_t)p(X_t|\{U_{j,t}\})p(\{U_{j,t}\})p(R_t)$$

- Bayesian solution
  - Posterior, joint conditional distribution:

$$p(\, f_t, P_t, X_t, \{U_{j,t}\}, R_t \mid y_t \,)$$

Sequential Bayes (I)

- Time-varying statistical factor models

  $$y_t = X_t\, f_t + e_t$$

  - Random variables: f_t, P_t, X_t, $\{U_{j,t}\}$, R_t
  - Estimated means: $f_{t|t}$, $P_{t|t}$, $X_{t|t}$, $\{U_{j,t|t}\}$, $R_{t|t}$
  - Predicted means: $f_{t+1|t}$, $P_{t+1|t}$, $X_{t+1|t}$, $\{U_{j,t+1|t}\}$, $R_{t+1|t}$
- Predicting "slow-changing" variables by random walks
  - $P_{t|t-1} = P_{t-1|t-1}$
  - $X_{t|t-1} = X_{t-1|t-1}$
  - $U_{j,t|t-1} = U_{j,t-1|t-1}$
  - $R_{t|t-1} = R_{t-1|t-1}$

Sequential Bayes (II)

- Predicting of "fast-changing" variable by long-term mean
  - $f_{t|t-1} = 0$
- Likelihood at time t
  - $p(y_t | X_t, f_t, R_t) = N(y_t; X_t f_t, R_t)$
- "Sequential priors" at time t
  - Expected value of a prior = Predicted mean of the variable
  - $p(f_t | P_t) = N(f_t; f_{t|t-1}, P_t)$
  - $p(P_t) = IG_D(P_t; \alpha_{t|t-1}, B_{P,t|t-1})$
  - $p(X_t | \{U_{j,t}\}) = \prod_{j=1}^{m} N(X_{j-col,t}; X_{j-col,t|t-1}, U_{j,t})$
  - $p(\{U_{j,t}\}) = \prod_{j=1}^{m} IG_D(U_{j,t}; \alpha_{t|t-1}, B_{U,j,t|t-1})$
  - $p(R_t) = IG_D(R_t; \alpha_{t|t-1}, B_{R,t|t-1})$

Variational Bayes (I)

- Variational Bayesian approximation (VBA) of posterior

$$p(f_t, P_t, X_t, \{U_{j,t}\}, R_t | y_t) \approx q(f_t, P_t, X_t, \{U_{j,t}\}, R_t) =$$
$$= \prod_{j=1}^{m} q_{f,j}(f_{j,t}) \times \prod_{j=1}^{m} q_{P,j}(P_{j,t}) \times \prod_{j=1}^{m} q_{X,j}(X_{j-col,t}) \times$$
$$\times \prod_{j=1}^{m} q_{U,j}(U_{j,t}) \times q_R(R_t)$$

- "Expectation operators" [6], expectation w.r.t. $q_j(\theta_j)$
  - $E_{f,j}\, z = E_{f,j}(z) = \int z\, q_{f,j}(f_{j,t})\, df_{j,t}$
  - $E_{P,j}\, z = E_{P,j}(z) = \int z\, q_{P,j}(P_{j,t})\, dP_{j,t}$
  - $E_{X,j}\, z = E_{X,j}(z) = \int z\, q_{X,j}(X_{j-col,t})\, dX_{j-col,t}$
  - $E_{U,j}\, z = E_{U,j}(z) = \int z\, q_{U,j}(U_{j,t})\, dU_{j,t}$
  - $E_R\, z = E_R(z) = \int z\, q_R(R_t)\, dR_t$

Variational Bayes (II)

- Combined expectation operators [6]
  - $E_f\, z = \left(\prod_{j=1}^{m} E_{f,j}\right) z = \int z\, \prod_{j=1}^{m} q_{f,j}(f_{j,t})\, df_{j,t}$
  - $E_P\, z = \left(\prod_{j=1}^{m} E_{P,j}\right) z = \int z\, \prod_{j=1}^{m} q_{P,j}(P_{j,t})\, dP_{j,t}$
  - $E_X\, z = \left(\prod_{j=1}^{m} E_{X,j}\right) z = \int z\, \prod_{j=1}^{m} q_{X,j}(X_{j-col,t})\, dX_{j-col,t}$
  - $E_U\, z = \left(\prod_{j=1}^{m} E_{U,j}\right) z = \int z\, \prod_{j=1}^{m} q_{U,j}(U_{j,t})\, dU_{j,t}$
- With known Joint distribution = Likelihood × Priors,

$$p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) =$$

$$= p(y_t|X_t, f_t, R_t)p(f_t|P_t)p(P_t)p(X_t|\{U_{j,t}\})p(\{U_{j,t}\})p(R_t)$$

optimal approximating distributions $q_j(\theta_j)$ are the solutions to a set of VBA optimization equations [2, 3, 4, 5, 6]

Variational Bayes (III)

- VBA optimization equations, for $j = 1, 2, \cdots, m$
  - $\ln q_{f,j}(f_{j,t}) = const +$
    $$+ \left(\prod_{k \neq j} E_{f,k}\right) E_P\, E_X\, E_U\, E_R\, \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$$
  - $\ln q_{P,j}(P_{j,t}) = const +$
    $$+ E_f\left(\prod_{k \neq j} E_{P,k}\right) E_X\, E_U\, E_R\, \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$$
  - $\ln q_{X,j}(X_{j-col,t}) = const +$
    $$+ E_f\, E_P\left(\prod_{k \neq j} E_{X,k}\right) E_U\, E_R\, \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$$
  - $\ln q_{U,j}(U_{j,t}) = const +$
    $$+ E_f\, E_P\, E_X\left(\prod_{k \neq j} E_{U,k}\right) E_R\, \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$$

Variational Bayes (IV)

- VBA optimization equations (continue)
  - $\ln q_R(R_t) = const +$

$$+ E_f\, E_P\, E_X\, E_U\, \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$$

- The VBA optimization equations
  - interrelated: each optimal solution $q_j(\theta_j)$ is expressed in terms of all others
  - able to identify function forms of $q_j(\theta_j)$, $j = 1, 2, \cdots, k$
  - can be solved iteratively: estimating each $q_j(\theta_j)$ with the latest estimates of all others
  - with reasonable initial distributions, the iterative solutions always converge to local optima
  - the sequential priors offer reasonable initial distributions

Variational Bayes Filtering (I)

- VBA iteration at time t
  - Using executable iteration (iteration index $v = 1, 2, \cdots, L$) to solve the set of VBA optimization equations
  - Expected values, $f_{t|t}$, $P_{t|t}$, $X_{t|t}$, $\{U_{j,t|t}\}$, $R_{t|t}$, of random variables to be estimated iteratively
- Initial values for the VBA iteration at time t
  - factors: $f_{t|t}^{(0)} = 0$; $P_{t|t}^{(0)} = P_{t-1|t-1}$
  - loadings: $X_{t|t}^{(0)} = X_{t-1|t-1}$; $U_{j,t|t}^{(0)} = U_{j,t-1|t-1}$, $j = 1, \cdots, m$
  - errors: $R_{t|t}^{(0)} = R_{t-1|t-1}$
  - additional: $Q_{t|t}^{(0)} = P_{t|t}^{(0)}$; $V_{j,t|t}^{(0)} = U_{j,t|t}^{(0)}$, $j = 1, \cdots, m$

Variational Bayes Filtering (II)

- A parameter for VB filtering [6]
  - Residual-to-specific variance ratio target (or RSVRT), h, $0 \ll h < 1$
  - To prevent filter from over- and under-fitting
- Rescaling the diagonal $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ [6]
  - for $j = 1, \cdots, m$
  - $W_{t|t}^{(0)} = \sum_{j=1}^{m} P_{j,t|t}^{(0)} U_{j,t|t}^{(0)} + R_{t|t}^{(0)}$
  - $U_{j,t|t}^{(0)} = \{ (W_{t|t}^{(0)} - R_{t|t}^{(0)})^{-1} [(1-h) W_{t|t}^{(0)}] \} U_{j,t|t}^{(0)}$
  - $R_{t|t}^{(0)} = [(R_{t|t}^{(0)})^{-1} (h W_{t|t}^{(0)})] R_{t|t}^{(0)}$
  - $V_{j,t|t}^{(0)} = U_{j,t|t}^{(0)}$

Variational Bayes Filtering (III)

- At each VBA iterate $v \geq 1$ at time t
  - The first group of updates: updating factors one by one, for $j = 1, 2, \cdots, m$,
  - $Q_{j,t|t}^{(v)} = \{ (P_{j,t|t}^{(v-1)})^{-1} + $
    $+ tr((R_{t|t}^{(v-1)})^{-1} [V_{j,t|t}^{(v-1)} + X_{j-col,t|t}^{(v-1)} (X_{j-col,t|t}^{(v-1)})^T]) \}^{-1}$
  - $f_{j,t|t}^{(v)} = Q_{j,t|t}^{(v)} \{ (P_{j,t|t}^{(v-1)})^{-1} f_{j,t|t}^{(0)} +$
    $+ tr((R_{t|t}^{(v-1)})^{-1}[y_t - \sum_{k \neq j} f_{k,t|t}^{(v-1)} X_{k-col,t|t}^{(v-1)}](X_{j-col,t|t}^{(v-1)})^T) \}$
  - $P_{j,t|t}^{(v)} = (T_0 + 1)^{-1} [T_0 P_{j,t|t}^{(0)} + Q_{j,t|t}^{(v)} + (f_{j,t|t}^{(v)} - f_{j,t|t}^{(0)})^2]$

Variational Bayes Filtering (IV)

- At each VBA iterate $v \geq 1$ at time t (continue)
  - The second group of updates: updating loadings one by one, for $j = 1, 2, \cdots, m$,
  - $V_{j,t|t}^{(v)} = U_{j,t|t}^{(v-1)} -$
    $$- U_{j,t|t}^{(v-1)} \{ U_{j,t|t}^{(v-1)} + R_{t|t}^{(v-1)} / [Q_{j,t|t}^{(v)} + (f_{j,t|t}^{(v)})^2] \}^{-1} U_{j,t|t}^{(v-1)}$$
  - $X_{j-col,t|t}^{(v)} = V_{j,t|t}^{(v)} \{ (U_{j,t|t}^{(v-1)})^{-1} X_{j-col,t|t}^{(0)} +$
    $$+ (R_{t|t}^{(v-1)})^{-1} f_{j,t|t}^{(v)} [y_t - \sum_{k \neq j} f_{k,t|t}^{(v)} X_{k-col,t|t}^{(v-1)}] \}$$
  - $U_{j,t|t}^{(v)} = (T_0 + 1)^{-1} diag(T_0 U_{j,t|t}^{(0)} + V_{j,t|t}^{(v)} +$
    $$+ (X_{j-col,t|t}^{(v)} - X_{j-col,t|t}^{(0)}) (X_{j-col,t|t}^{(v)} - X_{j-col,t|t}^{(0)})^T)$$

Variational Bayes Filtering (V)

- At each VBA iterate $v \geq 1$ at time t (continue)
  - The last update: updating variance of errors
  - $R_{t|t}^{(v)} = (T_0 + 1)^{-1}\, diag(\, T_0\, R_{t|t}^{(0)} +$
    $+ \sum_{j=1}^{m} [Q_{j,t|t}^{(v)} V_{j,t|t}^{(v)} + Q_{j,t|t}^{(v)} X_{j-col,t|t}^{(v)} (X_{j-col,t|t}^{(v)})^T + (f_{j,t|t}^{(v)})^2 V_{j,t|t}^{(v)}] +$
    $+ (\, y_t - X_{t|t}^{(v)} f_{t|t}^{(v)}\,)\,(\, y_t - X_{t|t}^{(v)} f_{t|t}^{(v)}\,)^T\,)$
- VB filtering estimates as of time t
  - factors: $f_{t|t} = f_{t|t}^{(L)}$; $P_{t|t} = P_{t|t}^{(L)}$
  - loadings: $X_{t|t} = X_{t|t}^{(L)}$; $U_{j,t|t} = U_{j,t|t}^{(L)}$, $j = 1, 2, \cdots, m$
  - errors: $R_{t|t} = R_{t|t}^{(L)}$

Cov Matrixes of Portfolios (I)

- For factor model, $y_t = X_t\, f_t + e_t$, expected or optimal values estimated
  - by VB filtering: $f_{t|t}$, $P_{t|t}$, $X_{t|t}$, $\{U_{j,t|t}\}$, $R_{t|t}$
  - by rolling PCA: $f_{t|t}$, $P_{t|t}$, $X_{t|t}$, $R_{t|t}$
- Applications of factor models
  - Factor-based forecasts of time-varying variance-covariance matrix of time-series
  - Variance-covariance-based forecasts of time-varying volatilities of asset portfolios
- Predicted variance-covariance matrix
  - $C_{t+1|t} = E_{t+1|t}(\, y_{t+1}\, y_{t+1}^T\,)$

Cov Matrixes of Portfolios (II)

- Variance-covariance matrix forecasted by VB filtering

$$C_{t+1|t} = C_{t+1|t}^{(X)} + C_{t+1|t}^{(Spec)}$$

  - loadings-based var-cov: $C_{t+1|t}^{(X)} = X_{t|t}\, P_{t|t}\, (X_{t|t})^T$
  - specific variance: $C_{t+1|t}^{(Spec)} = \sum_{j=1}^{m} P_{j,t|t}\, U_{j,t|t} + R_{t|t}$

- Portfolio of weights vector w
  - forecasted portfolio variance

$$(\sigma_{t+1|t}^{(w)})^2 = (\sigma_{t+1|t}^{(X|w)})^2 + (\sigma_{t+1|t}^{(Spec|w)})^2$$

  - loadings-based component: $(\sigma_{t+1|t}^{(X|w)})^2 = w_t^T\, C_{t+1|t}^{(X)}\, w_t$
  - specific component: $(\sigma_{t+1|t}^{(Spec|w)})^2 = w_t^T\, C_{t+1|t}^{(Spec)}\, w_t$

VBF

- Bayesian filtering framework
- Solution 1, stochastic factor model
- **Solution 2, time-series regression**
- Solution 3, vector autoregressive model
- Reference
- Appendix: factor model comparison

State Space Model (I)

- A linear state space model

$$x_t = F_t\, x_{t-1} + a_t + u_t$$

$$y_t = H_t\, x_t + c_t + v_t$$

- Notations and assumptions [7]
  - t = time index, $t = 1, 2, \cdots$
  - x_t = unknown $m \times 1$ hidden state vector
  - $Var_t(x_t) = P_t$ with unknown diagonal P_t
  - F_t = known $m \times m$ state transition matrix
  - a_t = known $m \times 1$ state input vector
  - u_t = unobservable $m \times 1$ state transition error vector
  - $E_t(u_t) = 0$

State Space Model (II)

- Notations and assumptions (continue)
  - $Var_t(u_t) = Q_t$ with unknown diagonal Q_t
  - $Cov_t(u_t, u_{t-\tau}) = 0$ for $\tau \geq 1$
  - y_t = observed $n \times 1$ measurement / data vector
  - H_t = known $n \times m$ measurement / observation matrix
  - c_t = known $n \times 1$ measurement input vector
  - v_t = unobservable $n \times 1$ measurement error vector
  - $E_t(v_t) = 0$
  - $Var_t(v_t) = R_t$ with unknown diagonal R_t
  - $Cov_t(v_t, v_{t-\tau}) = 0$ for $\tau \geq 1$
  - $Cov_t(u_t, v_\tau) = Cov_\tau(u_t, v_\tau) = 0$

Time-Series Models (I)

- Simple or multivariate time-varying regression
  - $n = 1,\ m \geq 2$
  - $F_t = I_m,\ a_t = 0$
  - $(H_t)_{j=1} = 1,\ c_t = 0$
  - response variable: y_t
  - predictor variable(s): $(H_t)_{j=2:m}$
  - regression coefficient(s): $(x_t)_{i=2:m}$; est.: $(x_{t|t})_{i=2:m}$
  - regression intercept: $(x_t)_{i=1}$; est.: $(x_{t|t})_{i=1}$
  - standard variance: P_t; est.: $P_{t|t}$
  - residual variance: R_t; est.: $R_{t|t}$

Time-Series Models (II)

- Stochastic local level model
  - $n = 1,\ m = 1$
  - $F_t = H_t = 1,\ a_t = c_t = 0$
  - time-series: y_t
  - local level of time-series: x_t; est.: $x_{t|t}$
  - standard variance: P_t; est.: $P_{t|t}$
  - residual variance: R_t; est.: $R_{t|t}$
  - local variance, estimated: $P_{t|t} + R_{t|t}$

Variational Bayes Filtering (I)

- A parameter for VB filtering [7]
  - Error reduction target, g, $0 < g < 1$
  - To prevent filter from over- and under-fitting
- Initial values for VB filtering at time t
  - $x_{t|t-1} = F_t\, x_{t-1|t-1} + a_t$
  - $y_{t|t-1} = H_t\, x_{t|t-1} + c_t$
  - $e_{t|t-1} = y_t - y_{t|t-1}$
- Initial values for VBA iteration
  - $P_{t|t}^{(0)} = P_{t|t-1} = P_{t-1|t-1}$
  - $R_{t|t}^{(0)} = R_{t|t-1} = R_{t-1|t-1}$

Variational Bayes Filtering (II)

- Initial values for VBA iteration (continue)
  - $S_t^{(0)} = H_t\, P_{t|t}^{(0)} H_t^T + R_{t|t}^{(0)}$
- Initial value adjustment, given parameter g [7]
  - $P_{t|t}^{(0)} = [\,(\,1 - g^{1/2}\,)\, S_t^{(0)} \,/\, (\,S_t^{(0)} - R_{t|t}^{(0)}\,)\,]\, P_{t|t}^{(0)}$
  - $R_{t|t}^{(0)} = [\, g^{1/2}\, S_t^{(0)} \,/\, R_{t|t}^{(0)}\,]\, R_{t|t}^{(0)}$
- VBA iteration [7], for $k = 1, 2, \cdots, L$
  - $S_t^{(k)} = H_t\, P_{t|t}^{(k)} H_t^T + R_{t|t}^{(k)}$
  - $K_t^{(k)} = P_{t|t}^{(k)} H_t^T (S_t^{(k)})^{-1}$
  - $M_t^{(k)} = I_n - H_t\, K_t^{(k)}$

Variational Bayes Filtering (III)

- VBA iteration, for $k = 1, 2, \cdots, L$ (continue)
  - $P_{t|t}^{(k)} = diag(\, P_{t|t}^{(0)} +$
    $+ K_t^{(k-1)} (\, e_{t|t-1}\, e_{t|t-1}^T - S_t^{(k-1)})\, (K_t^{(k-1)})^T \,/\, T_0 \,)$
  - $R_{t|t}^{(k)} = diag(\, R_{t|t}^{(0)} +$
    $+ M_t^{(k-1)} (\, e_{t|t-1}\, e_{t|t-1}^T - S_t^{(k-1)})\, (M_t^{(k-1)})^T \,/\, T_0 \,)$
- Estimates of VB filtering
  - $x_{t|t} = x_{t|t-1} + K_t^{(L)}\, e_{t|t-1}$
  - $y_{t|t} = H_t\, x_{t|t} + c_t$
  - $P_{t|t} = P_{t|t}^{(L)}$

Variational Bayes Filtering (IV)

- Estimates of VB filtering (continue)
  - $R_{t|t} = R_{t|t}^{(L)}$
  - $Q_{t|t} = q_{nne}(\, P_{t|t} - diag(\, F_t\, P_{t-1|t-1}\, F_t^T\,)\,)$

    here $q_{nne}(A_{ij}) = (A_{ij} + |A_{ij}|)/2$
- VB filtering estimates can be used for
  - Forecasting future values of time-series by (multiple) predictors
  - Estimating variance / covariance (matrix) of time-series

VBF

- Bayesian filtering framework
- Solution 1, stochastic factor model
- Solution 2, time-series regression
- **Solution 3, vector autoregressive model**
- Reference
- Appendix: factor model comparison

State Space Model (I)

- Vector time-series and data generating process (DGP)
  - Observed data vector time-series : y_t
  - Assumed DGP: vector autoregressive (VAR) model
- State space representation of VAR model

$$c_t = c_{t-1} + u_{c,t}$$

$$A_{k,t} = A_{k,t-1} + U_{k,t}, \quad k = 1, 2, \cdots, p$$

$$y_t = c_t + \sum_{k=1}^{p} A_{k,t}\, y_{t-k} + v_t$$

- Notations and assumptions [8]
  - t = time index, $t = 1, 2, \cdots$
  - y_t = $n \times 1$ vector of observed data

State Space Model (II)

- Notations and assumptions (continue)
  - c_t = $n \times 1$ vector of unobservable additive component
  - $u_{c,t}$ = $n \times 1$ vector of random transition noise
  - $A_{k,t}$ = $n \times n$ matrix of unobservable VAR coefficients
  - $U_{k,t}$ = $n \times n$ matrix of random transition noise
  - $u_t = vec(\,[\,u_{c,t}, U_{1,t}, U_{2,t}, \cdots, U_{p,t}\,]\,)$
  - $E_t(u_t) = 0$
  - $Var_t(u_t) = Q_t$ with unknown diagonal Q_t
  - $Cov_t(u_t, u_{t-\tau}) = 0$ for $\tau \geq 1$
  - v_t = $n \times 1$ vector of unobservable random innovations
  - $E_t(v_t) = 0$

State Space Model (III)

- Notations and assumptions (continue)
  - $Var_t(v_t) = R_t$ with unknown diagonal R_t
  - $Cov_t(v_t, v_{t-\tau}) = 0$ for $\tau \geq 1$
  - $Cov_t(u_t, v_\tau) = Cov_\tau(u_t, v_\tau) = 0$
  - k = number of time lags
  - p = order of VAR model
- VAR relation with time-varying estimates

$$c_{t|t} = c_{t-1|t-1} + u_{c,t|t}$$

$$A_{k,t|t} = A_{k,t-1|t-1} + U_{k,t|t}, \quad k = 1, 2, \cdots, p$$

$$y_t = c_{t|t} + \sum_{k=1}^{p} A_{k,t|t}\, y_{t-k} + v_{t|t}$$

Variational Bayes Filtering (I)

- A parameter for VB filtering [8]
  - Error reduction target, g, $0 < g < 1$
  - To prevent filter from over- and under-fitting
- Notations
  - for $j = 1, 2, \cdots, 1 + np$
  - $X_{t-1|t-1} = [\, c_{t-1|t-1}, A_{1,t-1|t-1}, A_{2,t-1|t-1}, \cdots, A_{p,t-1|t-1} \,]$
  - $X_{j,t-1|t-1} = (X_{t-1|t-1})_{j-col}$
  - $P_{j,t-1|t-1} = Cov_{t-1}(X_{j,t-1|t-1})$
  - $z_t = [\, 1, vec(\, [\, y_{t-1}, y_{t-2}, \cdots, y_{t-p} \,] \,)^T \,]^T$
  - $z_{j,t} = (z_t)_j$

Variational Bayes Filtering (II)

- Initial values for VB filtering at time t
  - $X_{t|t-1} = X_{t-1|t-1}$
  - $v_{t|t-1} = y_t - X_{t|t-1}\, z_t$
  - $P_{j,t|t-1} = P_{j,t-1|t-1}, \quad j = 1, 2, \cdots, 1 + np$
  - $R_{t|t-1} = R_{t-1|t-1}$
- Initial values for VBA iteration
  - $P_{j,t|t}^{(0)} = P_{j,t|t-1}$
  - $R_{t|t}^{(0)} = R_{t|t-1}$

Variational Bayes Filtering (III)

- Estimating initial value rescaling coefficients [8], $a_{P,t}$ and $a_{R,t}$, given parameter g, and $J = 1 + np$
  - $\Sigma_{Una,t} = \sum_{j=1}^{J} z_{j,t}^2 \, P_{j,t|t-1} + R_{t|t-1}$
  - $\Sigma_{Adj,t} = \Sigma_{Adj,t}(a_{P,t}, a_{R,t}) = a_{P,t} \sum_{j=1}^{J} z_{j,t}^2 \, P_{j,t|t-1} + a_{R,t} \, R_{t|t-1}$
  - $\lambda_{i,t} = (R_{t|t-1}^{-1} \sum_{j=1}^{J} z_{j,t}^2 \, P_{j,t|t-1})_{i,i}$
  - $\min_{a_{P,t}, a_{R,t}} \; tr(\Sigma_{Adj,t}^{-1} \, \Sigma_{Una,t} + \Sigma_{Una,t}^{-1} \, \Sigma_{Adj,t})$

    $\text{s.t.} \;\; \sum_{i=1}^{n} (1 + (a_{P,t} / a_{R,t}) \, \lambda_{i,t})^{-2} = ng$
- Initial value adjustment, given g, $a_{P,t}$ and $a_{R,t}$
  - $P_{j,t|t}^{(0)} = a_{P,t} \, P_{j,t|t}^{(0)}, \quad j = 1, 2, \cdots, 1 + np$
  - $R_{t|t}^{(0)} = a_{R,t} \, R_{t|t}^{(0)}$

Variational Bayes Filtering (IV)

- VBA iteration [8], for $k = 1, 2, \cdots, L$
  - for $j = 1, 2, \cdots, J$, and $J = 1 + np$
  - $S_t^{(k)} = \sum_{j=1}^{J} z_{j,t}^2 \, P_{j,t|t}^{(k)} + R_{t|t}^{(k)}$
  - $K_{j,t}^{(k)} = z_{j,t} \, P_{j,t|t}^{(k)} \, (S_t^{(k)})^{-1}$
  - $M_t^{(k)} = I_n - \sum_{j=1}^{J} z_{j,t} \, K_{j,t}^{(k)}$
  - $P_{j,t|t}^{(k)} = diag(\, P_{j,t|t}^{(0)} +$

    $+ K_{j,t}^{(k-1)} (\, v_{t|t-1} \, v_{t|t-1}^T - S_t^{(k-1)}) \, K_{j,t}^{(k-1)} / \, T_0 \,)$
  - $R_{t|t}^{(k)} = diag(\, R_{t|t}^{(0)} +$

    $+ M_t^{(k-1)} (\, v_{t|t-1} \, v_{t|t-1}^T - S_t^{(k-1)}) \, M_t^{(k-1)} / \, T_0 \,)$

Variational Bayes Filtering (V)

- Estimates of VB filtering
  - for $j = 1, 2, \cdots, 1 + np$
  - $X_{j,t|t} = X_{j,t|t-1} + K_{j,t}^{(L)}\, v_{t|t-1}$
  - $X_{t|t} = [\, X_{1,t|t}, X_{2,t|t}, \ \cdots, X_{J,t|t} \,]$
  - $[\, c_{t|t}, A_{1,t|t}, A_{2,t|t}, \ \cdots, A_{p,t|t} \,] = X_{t|t}$
  - $P_{j,t|t} = P_{j,t|t}^{(L)}$
  - $R_{t|t} = R_{t|t}^{(L)}$
- VB filtering estimates can be used for
  - Forecasting future values of time-series by VAR modeling
  - Estimating variance / covariance of time-series

VBF

- Bayesian filtering framework
- Solution 1, stochastic factor model
- Solution 2, time-series regression
- Solution 3, vector autoregressive model
- **Reference**
- Appendix: factor model comparison

Literature (I)

- [1] Särkkä, S., *Bayesian Filtering and Smoothing*, 2013, (Cambridge University Press: Cambridge, UK).
- [2] Bishop, C.M., *Pattern Recognition and Machine Learning*, 2006 (Springer Science+Business Media, LLC: New York).
- [3] Tzikas, D.G., Likas, A.C. and Galatsanos, N.P., The variational approximation for Bayesian inference: life after the EM algorithm. *IEEE Signal Processing Magazine*, November 2008, 131–146.
- [4] Ormerod, J.T. and Wand, M.P., Explaining Variational Approximations. *The American Statistician*, 2010, **64** (2), 140–153.
- [5] Grimmer, J., An introduction to Bayesian inference via variational approximations. *Political Analysis*, 2011, **19**, 32–47.

Literature (II)

- [6] Ling, H.F. and Franzen, C., Online learning of time-varying
- stochastic factor structure by variational sequential Bayesian factor analysis, *Quantitative Finance*, 2017, **17** (8), 1277–1304.
- [7] Ling, H.F. and Stone, D.B., Time-varying forecasts by variational approximation of sequential Bayesian inference. *Quantitative Finance*, 2016, **16** (1), 43–67.
- [8] Ling, H.F., Time-varying stochastic vector autoregressive model estimated by variational sequential Bayesian algorithm, *technical document*.

VBF

- Bayesian filtering framework
- Solution 1, stochastic factor model
- Solution 2, time-series regression
- Solution 3, vector autoregressive model
- Reference
- Appendix: factor model comparison

Factor Model Evaluation

- Factor model evaluation by accuracies of
  - factor-based forecasts of time-varying variance-covariance matrix of time-series
  - variance-covariance-based forecasts of time-varying volatilities of asset portfolios

Difference of 2 Cov Matrixes

- Difference of two distributions p and q
  - the best known measure:
    Kullback-Leibler divergence (KLD)
  - KLD is not symmetric and not a “distance”
  - symmetrized KL distance:
    KL distance = (KLD from p to q + KLD from q to p) / 2
- Difference of two variance-covariance matrixes
  - (when they represent two multivariate normal distributions)
  - can be measured by the symmetrized KL distance
  - smaller KL distance: smaller difference between the two variance-covariance matrixes

Accuracy of Portfolio Risk (I)

- Bias statistic

  = standard deviation of realized portfolio returns standardized by forecasted portfolio mean and variance

  - simple and popular
  - if less (or greater) than 1: variance is over- (or under-) predicted
  - over- and under-predictions may be cancelled out

- Log-likelihood

  = logarithm of likelihood calculated by forecasted portfolio mean and variance and realized portfolio return

  - popular and proven optimal
  - higher log-likelihood: more accurate forecasts

Accuracy of Portfolio Risk (II)

- Q-statistic
  - closely related to the log-likelihood
  - proven optimal
  - smaller Q-statistic: more accurate forecasts
- Volatility reduction

  = realized portfolio volatility reduction by variance minimization based on the forecasted variance-covariance matrix
  - practically important and popular
  - larger volatility reduction: more accurate forecasts
- The four measures
  - largely agree with each other

Estimated Cov Matrixes (I)

- Simulated data
  - 200 simulated daily time-series of 9 years in length
  - known daily data generating model (DGM): daily factor model of 10 factors
  - the daily DGM: 10-factor rolling 65-day PCA on 9-year daily returns of 200 Russell Top Index stocks
- Tested values
  - known daily variance-covariance matrixes by the DGM
  - daily variance-covariance estimated by 5-, 10- and 15-factor rolling PCA of various window size T_{MW}
  - daily variance-covariance estimated by 5-, 10- and 15-factor VB filtering of various RSVRT h

Estimated Cov Matrixes (II)

- Test statistic
  - 8-Year and annual averages of symmetrized KL distances
  - between the known daily variance-covariance matrix and those estimated by rolling PCA and VB filtering
- Results
  - under-specification (model m < “true” $m = 10$) results in more accurate forecasts
  - the best rolling PCA is of $T_{MW} = 65$
  - the best VB filtering is of $h = 0.94$
- Conclusion
  - Variance-covariance matrixes estimated by VB filtering are more accurate than those by rolling PCA

Global Stocks and Portfolios

- Global stock data
  - 807 stocks chosen from (1610 stocks in) MSCI World Index on 01/31/2014
  - all stocks have complete 25-year history of monthly total returns in USD from Jan 1989 to Dec 2013
- Random portfolios
  - 1000 long-only portfolios; 1000 long/short portfolios

Portfolio Volatility (I)

- Tested values
  - portfolio volatilities forecasted by 10-factor rolling PCA of various window size T_{MW}
  - portfolio volatilities forecasted by 10-factor VB filtering of various RSVRT h
- Test statistic
  - 240-month and 60-month values of
  - bias statistic, $BS^{(PCA)}$, $BS^{(VSB)}$
  - log-likelihood, $LL^{(PCA)}$, $LL^{(VSB)}$
  - Q-statistic, $QS^{(PCA)}$, $QS^{(VSB)}$
  - volatility-reduction, $VM^{(PCA)}$, $VM^{(VSB)}$

Portfolio Volatility (II)

- Results
  - Rolling PCA of $T_{MW} = 37$ can be regarded as the best
  - VB filtering of $h = 0.86$ can be regarded as the best
  - (For forecasting volatilities of both long-only and long/short portfolios)

VB Filtering vs. Rolling PCA (I)

- Model comparison with 1000 long-only portfolios
  - VB filtering can make more accurate portfolio volatility forecasts than rolling PCA in
  - 11 or 14 out of the 16 rolling 60-month periods and
  - the entire 240-month period
- Conclusion
  - variances of long-only portfolios are overwhelmingly dominated by loadings-based (or systematic) component
  - implication: VB filtering can make more accurate forecasts in factor-based variance-covariance than rolling PCA

VB Filtering vs. Rolling PCA (II)

- Model comparison with 1000 long/short portfolios
  - VB filtering can make more accurate portfolio volatility forecasts than rolling PCA in
  - all of the 16 rolling 60-month periods and
  - the entire 240-month period
- Conclusion
  - variances of long/short portfolios are dominated by specific component
  - implication: VB filtering can make more accurate forecasts in specific variance as well than rolling PCA

Variational Bayes Filter Solutions

- Questions ?
- Discussions

- **Thank you very much !**