Multi-step ahead predictions in a normal BVAR(p) model using MC sampling

Jan Šindelář

Department of Adaptive Systems, UTIA CAS

Department of Probability and Mathematical Statistics, Charles University

e-mail: jenik.sindelar@gmail.com

- Motivation
- Modelling using BVAR(p) process
- Predictions
- Conclusion
- Possible future research

- Motivation
- Modelling using BVAR(p) process
- Predictions
- Conclusion
- Possible future research

- Motivation
- Modelling using BVAR(p) process
- Predictions
- Conclusion
- Possible future research

- Motivation
- Modelling using BVAR(p) process
- Predictions
- Conclusion
- Possible future research

- Motivation
- Modelling using BVAR(p) process
- Predictions
- Conclusion
- Possible future research

Let's have a probability space $(\Omega, \mathcal{F}, p)$ equipped with a filtration $\mathcal{F}_t$, where $t \in \mathcal{T} = \{1, \dots, T\}$.
We have to solve a general optimization task

$$\max_{\pi_t, \dots, \pi_{t+h-1}} \mathbb{E}[U(W_{t+h})|\mathcal{F}_t] \tag{1}$$

where $\pi_t : \Omega \to \mathcal{U}$ are policies ($\mathcal{F}_t$ measurable functions) and U is a given concave utility function – can be solved by Dynamic Programming or some of its approximations.

Wealth

$$W_{t+1} = W_t + [Y_{t+1} - Y_t] \sum_{k=1}^{t} \pi_k - K|\pi_t| \tag{2}$$

where Y_t is the market price and K are transaction costs.

Or sometimes we use a simple approximation of the general utility through Taylor expansion around the mean value of wealth

$$U(W_{t+h}) \approx U(\bar{W}_{t+h}) + \dot{U}(\bar{W}_{t+h})[W_{t+h} - \bar{W}_{t+h}] + \frac{1}{2}\ddot{U}(\bar{W}_{t+h})[W_{t+h} - \bar{W}_{t+h}]^2 \tag{3}$$

where $\bar{W}_{t+h} = \mathbb{E}[W_{t+h}|\mathcal{F}_t]$. By applying the mean value operator we get

$$\mathbb{E}[U(W_{t+h})|\mathcal{F}_t] \approx U(\bar{W}_{t+h}) + \dot{U}(\bar{W}_{t+h})\underbrace{\mathbb{E}\left[(W_{t+h} - \bar{W}_{t+h})|\mathcal{F}_t\right]}_{0} + \frac{1}{2}\ddot{U}(\bar{W}_{t+h})\underbrace{\mathbb{E}\left[(W_{t+h} - \bar{W}_{t+h}]^2|\mathcal{F}_t\right]}_{\mathrm{var}[W_{t+h}|\mathcal{F}_t]} \tag{4}$$

– mean-variance optimization

Remark: We can use a different measure of risk to solve a mean-risk optimization problem.

We need to model the evolution of the price.

VAR(p) process:

$$\mathbf{D}_{t+1} = \mathbf{A}_1 \cdot \mathbf{D}_t + \mathbf{A}_2 \cdot \mathbf{D}_{t-1} + \cdots + \mathbf{A}_p \cdot \mathbf{D}_{t+1-p} + \mathbf{\Sigma} \cdot \mathbf{e}_{t+1} \qquad (5)$$

where $D_{1;t} = Y_t, \forall t \in \mathcal{T}$.

Rewritten into recursive, Markovian form:

$$\underbrace{\begin{bmatrix} \mathbf{D}_{t+1} \\ \mathbf{D}_t \\ \mathbf{D}_{t-1} \\ \vdots \\ \mathbf{D}_{t+1-p} \\ 1 \end{bmatrix}}_{\mathbf{\Phi}_{t+1}} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & & \mathbf{A}_{p-1} & \mathbf{A}_p & \mathbf{c} \\ \mathbf{I} & \mathbf{0} & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & & \\ \mathbf{0} & \mathbf{0} & & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{D}_t \\ \mathbf{D}_{t-1} \\ \mathbf{D}_{t-2} \\ \vdots \\ \mathbf{D}_{t-p} \\ 1 \end{bmatrix}}_{\mathbf{\Phi}_t} + \underbrace{\begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \ldots & \mathbf{0} \\ \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{bmatrix} \mathbf{e}_{t+1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}}_{\mathbf{e}_{t+1}}$$

- $\mathcal{L}(\mathbf{e}_{t+1}) = \mathcal{N}(0, 1), \forall t \in \mathcal{T}, \forall h \neq 0$ $\mathbf{e}_{t+h}$ independent of $\mathbf{e}_t$
- $\mathbf{\Theta} = \mathbf{A}, \mathbf{\Sigma}$ are parameters not measurable $\mathcal{F}_t$ – Bayesian viewpoint, generally time dependent

All random variables are absolutely continuous with respect to the underlying Lebesgue measure – densities exist. Parameters $\boldsymbol{\Theta}$ are difficult to model – not observable. We suppose

$$f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\ldots,\boldsymbol{\theta}_p,\mathcal{F}_t) = f(\boldsymbol{\theta}|\mathcal{F}_t) \quad \forall t \in \mathcal{T} \tag{6}$$

- The model is assimilated using Bayesian parameter estimation
- Prior distribution is chosen in a conjugate form – in the AR model case the Gauss-Inverse-Wishart distribution
- The model is made adaptive using exponential forgetting between data updating steps

$$f(\theta_t = \theta|\mathcal{F}_t) = f(\theta_{t-1} = \theta|\mathcal{F}_t)^{\lambda} \tag{7}$$

- The structure of the model can be estimated using Maximum Likelihood hypothesis testing.

All random variables are absolutely continuous with respect to the underlying Lebesgue measure – densities exist. Parameters $\boldsymbol{\Theta}$ are difficult to model – not observable. We suppose

$$f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\ldots,\boldsymbol{\theta}_p,\mathcal{F}_t) = f(\boldsymbol{\theta}|\mathcal{F}_t) \quad \forall t \in \mathcal{T} \tag{6}$$

- The model is assimilated using Bayesian parameter estimation
- Prior distribution is chosen in a conjugate form – in the AR model case the Gauss-Inverse-Wishart distribution
- The model is made adaptive using exponential forgetting between data updating steps

$$f(\theta_t = \theta|\mathcal{F}_t) = f(\theta_{t-1} = \theta|\mathcal{F}_t)^{\lambda} \tag{7}$$

- The structure of the model can be estimated using Maximum Likelihood hypothesis testing.

All random variables are absolutely continuous with respect to the underlying Lebesgue measure – densities exist. Parameters $\boldsymbol{\Theta}$ are difficult to model – not observable. We suppose

$$f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\ldots,\boldsymbol{\theta}_p,\mathcal{F}_t) = f(\boldsymbol{\theta}|\mathcal{F}_t) \quad \forall t \in \mathcal{T} \tag{6}$$

- The model is assimilated using Bayesian parameter estimation
- Prior distribution is chosen in a conjugate form – in the AR model case the Gauss-Inverse-Wishart distribution
- The model is made adaptive using exponential forgetting between data updating steps

$$f(\theta_t = \theta|\mathcal{F}_t) = f(\theta_{t-1} = \theta|\mathcal{F}_t)^{\lambda} \tag{7}$$

- The structure of the model can be estimated using Maximum Likelihood hypothesis testing.

All random variables are absolutely continuous with respect to the underlying Lebesgue measure – densities exist. Parameters $\boldsymbol{\Theta}$ are difficult to model – not observable. We suppose

$$f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\ldots,\boldsymbol{\theta}_p,\mathcal{F}_t) = f(\boldsymbol{\theta}|\mathcal{F}_t) \quad \forall t \in \mathcal{T} \tag{6}$$

- The model is assimilated using Bayesian parameter estimation
- Prior distribution is chosen in a conjugate form – in the AR model case the Gauss-Inverse-Wishart distribution
- The model is made adaptive using exponential forgetting between data updating steps

$$f(\boldsymbol{\theta}_t = \boldsymbol{\theta}|\mathcal{F}_t) = f(\boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}|\mathcal{F}_t)^{\lambda} \tag{7}$$

- The structure of the model can be estimated using Maximum Likelihood hypothesis testing.

All random variables are absolutely continuous with respect to the underlying Lebesgue measure – densities exist. Parameters $\boldsymbol{\Theta}$ are difficult to model – not observable. We suppose

$$f(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1},\ldots,\boldsymbol{\theta}_p,\mathcal{F}_t) = f(\boldsymbol{\theta}|\mathcal{F}_t) \quad \forall t \in \mathcal{T} \tag{6}$$

- The model is assimilated using Bayesian parameter estimation
- Prior distribution is chosen in a conjugate form – in the AR model case the Gauss-Inverse-Wishart distribution
- The model is made adaptive using exponential forgetting between data updating steps

$$f(\boldsymbol{\theta}_t = \boldsymbol{\theta}|\mathcal{F}_t) = f(\boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}|\mathcal{F}_t)^{\lambda} \tag{7}$$

- The structure of the model can be estimated using Maximum Likelihood hypothesis testing.

Suppose we learn the parameters perfectly - fixed,known. From equation (5) we get a full distribution prediction

$$\mathbf{\Phi}_{t+h} = \mathbf{A}^h \mathbf{\Phi}_t + \sum_{i=0}^{h-1} \mathbf{A}^i \mathbf{\Sigma} \mathbf{e}_{t+h-i} \tag{8}$$

where $\mathbf{A}^0 \equiv \mathbf{I}$ and we can also compute first two moments this random variable

$$\begin{aligned} \boldsymbol{\mu}_h = \mathbb{E}\left[\mathbf{\Phi}_{t+h} | \mathcal{F}_t, \boldsymbol{\theta}\right] &= \mathbf{A}^h \mathbf{\Phi}_t \\ \mathbf{R}_h = \operatorname{cov}\left[\mathbf{\Phi}_{t+h} | \mathcal{F}_t, \boldsymbol{\theta}\right] &= \sum_{i=0}^{h-1} \left[\mathbf{A}^i \mathbf{\Sigma} \left(\mathbf{A}^i \mathbf{\Sigma}\right)'\right] \end{aligned} \tag{9}$$

and if $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ – specifies the distribution fully.

Bayesian setup $\Rightarrow$ parameters unknown $\Rightarrow$ infinite number of recursions in (5)

$$\downarrow$$

Instead we draw Monte Carlo samples $\boldsymbol{\theta}_i, i \in 1, 2, \ldots, N$ from distribution $f(\boldsymbol{\theta}|\mathcal{F}_t)$

$$\downarrow$$

We use averaging instead of integration to get

$$f_N(\phi_{t+h}|\phi_t) = \frac{1}{N} \sum_{i=1}^{N} f(\phi_{t+h}|\phi_t, \boldsymbol{\theta}_i) \tag{10}$$

should converge pointwise to general solution for $N \to \infty$

First two moments:

$$\mathbb{E}_N\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]$$

$$\mathrm{cov}_N\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t\right] = \mathbb{E}_N\left[\left(\mathbf{\Phi}_{t+h}-\mathbb{E}_N\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t\right]\right)\left(\mathbf{\Phi}_{t+h}-\mathbb{E}_N\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t\right]\right)'|\mathcal{F}_t\right] =$$

$$\frac{1}{N}\sum_i \mathbf{R}_{h;i} + \frac{1}{N}\sum_i\left(\mathbb{E}\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]\right)\left(\mathbb{E}\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]\right)' -$$

$$-\frac{1}{N^2}\left(\sum_i\mathbb{E}\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]\right)\left(\sum_j\mathbb{E}\left[\mathbf{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\right)'$$

where $\mathbf{R}_{h;i} = \mathrm{cov}(\mathbf{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i)$.

Example

Forgetting factor $\lambda = 0.98$, parameters time-dependent, 2 data channels, 50 Monte Carlo steps

$$\begin{aligned} x_{t+1} &= 0.85x_t - 0.95x_{t-1} + 0.6y_t + 1.4e_{1;t+1} \\ y_{t+1} &= ay_t - 0.7y_{t-1} - c + 0.9e_{2;t+1} \end{aligned}$$

where

$$\begin{aligned} a &= \sin\left(\frac{t}{3}\right) \\ c &= \frac{1}{2}\sin\left(\frac{t}{5}\right) + \frac{1}{5}\cos\left(\frac{t}{2}\right) \end{aligned}$$

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

- Better structure estimation
- Optimization
  - Jan Zeman
  - Mean-variance optimization
  - use of the Monte Carlo
  - Properties of the Bellman function
  - Portfolio optimization instead of a single asset case
  - Derivative pricing

Thank you for your attention.