Averaging Estimator for Cointegrated Vector Autoregressive Models

Yundong Tu¹ Yanping Yi²

¹Guanghua School of Management, Peking University

²School of Economics, Shanghai University of Finance and Economics

IMS of NUS, May 22, 2014 Work in Progress

Outline

Introduction and Motivation

- 2 Model and Estimation
 - Unconstrained Estimation
 - Constrained Estimation
 - Pretesting Estimation
 - Mallow's Selection Estimation
 - Mallow's Averaging Estimation
- 3 Asymptotic Comparison
- 4 Finite Sample Properties
- 5 Conclusion and Future Directions





- 2 Model and Estimation
- 3 Asymptotic Comparison
- 4 Finite Sample Properties
- 6 Conclusion and Future Directions



We are interested in forecasting the future value of non-stationary data, such as GDP and interest rate.

Why Non-stationary Data?

- Many financial and macroeconomic variables are found to be non-stationary. Examples include quarterly GDP, consumer price index, income, consumption, foreign exchange rate, interest rate, etc.
- Most non-stationary economic variables can be characterized by a unit root process.
- There are various existing test procedures for the existence of unit roots, such as Dickey and Fuller (1981), Phillips and Perron (1988), Elliott et al. (1996), Elliott and Stock (2001), among others.

Practical Approach to Multiple Non-stationary Time Series

 (i) Estimate VAR in levels / Unconstrained estimation: ignore whether the data are stationary or non-stationary, and simply fit and estimate VAR in levels. We consider an equivalent representation for a Bivariate VAR(p) model.

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \Pi \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + \sum_{j=1}^{p-1} \Pi_j \begin{bmatrix} \Delta Y_{t-j} \\ \Delta X_{t-j} \end{bmatrix} + a_0 + \epsilon_t, \quad (1)$$

where $\Delta Y_t = Y_t - Y_{t-1}$, Π has rank $0 \le r_0 \le 2$, ϵ_t is a bivariate error vector with mean 0 and nonsingular covariance matrix Ω .

Practical Approach to Multiple Non-stationary Time Series

Estimate VAR in levels / Unconstrained estimation: estimate model (1) directly. This method is pretty robust if one is only interested in forecasting Y_t .

$$\widetilde{\mu}_{t} = \widetilde{C}_{0} + \widetilde{C}_{1}Y_{t-1} + \widetilde{C}_{2}X_{t-1} + \sum_{j=1}^{p-1}\widetilde{\Pi}_{j}^{(Y)'} \begin{bmatrix} \Delta Y_{t-j} \\ \Delta X_{t-j} \end{bmatrix}, \quad (2)$$

where $\tilde{C}_0, \tilde{C}_1, \tilde{C}_2, \tilde{\Pi}_j^{(Y)} (j = 1, \dots, p-1)$ are obtained by least-square estimation.

Practical Approach to Multiple Nonstationary Time Series II

(ii) Estimate VAR in difference / Constrained estimation : difference non-stationary data before estimating VAR , i.e. estimate model (1) imposing the restriction $\Pi = 0$.

$$\hat{\mu}_t = \widehat{C}_0 + \sum_{j=1}^{p-1} \widehat{\Pi}_j^{(Y)\prime} \begin{bmatrix} \Delta Y_{t-j} \\ \Delta X_{t-j} \end{bmatrix}, \qquad (3)$$

If they are indeed co-integrated, we miss the information contained in the equilibrium relationship between Y_t and X_t , which matters in forecasting the short run dynamics.

Cointegration

Cointegration, coined by Granger (1983) and Engle and Granger (1987), was formalized to describe the phenomenon that linear combinations of unit root processes are stationary. Economic and financial theories often imply equilibrium relationships (cointegration) between time series variables.

- The permanent income model implies cointegration between consumption and income (Davison, Hendry, Srba, Yeo 1978)
- Purchasing power parity implies cointegration between the nominal exchange rate and foreign and domestic prices.
- Arbitrage arguments imply cointegration between spot and futures prices, and spot and forward prices, and bid and ask prices.

• • • •

Cointegration

Figure 1: Weekly U.S. interest rates of 3-month Treasury bill and 6-month Treasury bill (December 12, 1958 to August 6, 2004, measured in percentage):



10/38

Cointegration

- The two interest rates move in unison. We expect this long run equilibrium will help forecast the future value of interest rates.
- Engle and Granger (1987) provided a two-step procedure to test the null of no cointegration against the alternative that there exists one. Johansen (1988) proposed the maximum likelihood approach to test the cointegration relationship.

The VECM Representatoin for a Bivariate Cointegrated System

When Y_t and X_t are cointegrated, i.e. Π has rank $r_0 = 1$

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \begin{bmatrix} \alpha_y \\ \alpha_x \end{bmatrix} [Y_{t-1} - \beta X_{t-1} - c_0] + \sum_{j=1}^{p-1} \prod_j \begin{bmatrix} \Delta Y_{t-j} \\ \Delta X_{t-j} \end{bmatrix} + \epsilon_t,$$
(4)

where $[1 - \beta]'$ is the cointegrating vector, α_y and α_x are adjustment coefficients in this error correction model. The cointegration relationship appears explicitly in the estimation/forecast equation.

Practical Approach to Multiple Nonstationary Time Series III

- (iii) Pretesting estimation : test each individual series for unit roots and test for possible cointegration relationships. And then model (4) can be estimated via a two-step procedure.
 - **(**) estimate the cointegration coefficients $\widehat{\beta}$ and $\widehat{c_0}$
 - egress ΔY_t on Y_{t-1} βX_{t-1} c₀ and the lagged difference of Y_t and X_t to obtain an estimate of α_y.
 - **3** depending on the significance of $\hat{\alpha}_y$, one choose either unconstrained estimation or constrained estimation

Remark: If we know the number and location of cointegration relationships, the two-step procedure is equivalent to estimating VAR in levels.

Motivations

- If the reaction of Y_t to their moving out of long-run equilibrium is quite weak/slow, VAR in difference might improve the small-sample performance of estimates, and also the accuracy of forecast.
- That motivates us to use a weighted average of VAR in levels and VAR in difference.
- Question: how should we choose the weight?
- Answer : the choice is based on the Mallows criterion in our paper.

AMSE, AFR, Mallows criteria

• Asymptotic mean-squared error (AMSE) :

$$\lim_{n\to\infty}\sum_{t=1}^{n}E(\widehat{\mu}_t-E(\Delta Y_t|\mathcal{F}_{t-1})^2.$$

• Asymptotic forecast risk (AFR):

$$\lim_{n\to\infty} nE(\widehat{\mu}_{n+1}-E(\Delta Y_{n+1}|\mathcal{F}_n)^2).$$

 Mallows criterion : a penalized sum of squared errors, designed to be approximately unbiased estimate(up to a constant) for the in-sample AMSE.

Our Approach: Model Averaging based on Mallows criteria

Given the nature of the pretesting, we propose to adopt the averaging approach to multiple nonstationary time series.

- Forecast Combination was introduced by Bates and Granger (1969), and have been found useful in improving the forecast accuracy, as noted in Stock and Watson (1999, 2004, 2005) and references therein.
- Model averaging based on Mallows criteria has been proposed by Hansen (2007, 2008), and has been extended to models with structural breaks in Hansen (2009), to autoregression models with a near unit root in Hansen (2010), to heteroskedastic regression in Hansen and Racine (2012).
- In this paper we adopt the local asymptotic framework that was developed in Hjort and Claeskens (2003) and later used in Hansen (2010).

Contribution of This Paper

This paper contributes to the literature in the following aspects.

- We propose to use model averaging estimation for cointegrated VAR models.
- The AMSE and AFR of unconstrained estimation, constrained estimation, pretesting estimation, Mallows selected estimation, and Mallows averaging estimation are derived and shown to depend on the strength of the cointegration signal, the number and location of unit roots in the multivariate non-stationary system.

Contribution of This Paper

- The asymptotic comparison based on AMSE and AFR favors our proposed averaging estimator and cautions against the pretesting estimator
- Finite sample comparison also confirms our asymptotic results.

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Outline



2 Model and Estimation

- 3 Asymptotic Comparison
- ④ Finite Sample Properties
- 5 Conclusion and Future Directions

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Local to Error Correction Asymptotic Framework

We evaluate the forecast performance of different estimators under a local-to-error-correction framework.

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \begin{bmatrix} \alpha_y \\ \alpha_x \end{bmatrix} [Y_{t-1} - \beta X_{t-1} - c_0] + \sum_{j=1}^{p-1} \prod_j \begin{bmatrix} \Delta Y_{t-j} \\ \Delta X_{t-j} \end{bmatrix} + \epsilon_t$$
$$= \begin{bmatrix} \frac{c_y}{\sqrt{n}} \\ c_x \end{bmatrix} [Y_{t-1} - \beta X_{t-1} - c_0] + \sum_{j=1}^{p-1} \prod_j \begin{bmatrix} \Delta Y_{t-j} \\ \Delta X_{t-j} \end{bmatrix} + \epsilon_t,$$

<□ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Unconstrained Estimation

We show in the paper that

$$\frac{AMSE}{\Omega_{11}} = b^* + 1 + [2(p-1)+1],$$

where $b^* = E\left(\frac{\left[\int_0^1 W(t)dW(t) - W(1)\int_0^1 W(t)dt\right]^2}{\int_0^1 W^2(t)dt - \left[\int_0^1 W(t)dt\right]^2}\right).$

$$\frac{AFR}{\Omega_{11}} = E \left[T_1 + W(1) \right]^2 + \left[2(p-1) + 1 \right],$$

$$T_1 \equiv \left(W(1) - \int_0^1 W(t) dt \right) \times \frac{\int_0^1 W(t) dW(t) - W(1) \int_0^1 W(t) dt}{\int_0^1 W^2(t) dt - \left[\int_0^1 W(t) dt \right]^2}.$$

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Constrained Estimation

We also show that

• • $\frac{AMSE}{\Omega_{11}} = c_y^2 \frac{Var(\beta'_0 Z_{t-1}^*)}{\Omega_{11}} + [2(p-1)+1].$ • • $\frac{AFR}{\Omega_{11}} = c_y^2 \frac{Var(\beta'_0 Z_{t-1}^*)}{\Omega_{11}} + [2(p-1)+1].$

β'₀Z^{*}_{t-1} is the projection residual of (Y_{t-1} − βX_{t-1}) on a constant, and the lagged difference series.

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Pretesting Estimation

• The pretest estimator is defined based on the t-ratio

$$t_n = \frac{\hat{\alpha}_y}{se(\hat{\alpha}_y)},$$

where $\hat{\alpha}_y$ is an OLS estimate of α_y via the regression of ΔY_t on $Y_{t-1} - \hat{\beta} X_{t-1}$, a constant, and the lagged difference series

• Let *c*^{*} be the corresponding critical value. The pretesting estimator is

$$\hat{\mu}_t^{p} = \hat{\mu}_t \mathbb{1}(|t_n| \leq c^*) + \widetilde{\mu}_t \mathbb{1}(|t_n| > c^*).$$

• AMSE and AFR of the pretesting estimator can be easily derived.

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Mallow's selection estimation

• The optimal Mallows criterion for the unconstrained model is

$$M_0 = n \widetilde{\Omega}_{11} + 2 \widetilde{\Omega}_{11} \left(2p + b^*
ight)$$

 $E(M_0) - n \Omega_{11} \rightarrow \mathsf{AMSE}$ of the unconstrained estimator

where $\widetilde{\Omega}_{11}$ is the estimate of Ω_{11} from the unconstrained model.

• The optimal Mallows criterion for the constrained model is

$$M_1 = n\widehat{\Omega}_{11} + 2\widetilde{\Omega}_{11}[2(p-1)+1]$$

 $E(M_1) - n\Omega_{11} \rightarrow \text{AMSE}$ of the constrained estimator

where $\widehat{\Omega}_{11}$ is the estimate of Ω_{11} from the constrained model.

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Mallow's selection estimation

• Mallows selection chooses the model with the smallest criteria. Therefore, if

$$M_0 < M_1,$$

or equivalently

$$M_n \equiv \frac{n(\widehat{\Omega}_{11} - \widetilde{\Omega}_{11})}{\widetilde{\Omega}_{11}} > 2 + 2b^*,$$

the unconstrained model will be chosen. Otherwise, the constrained model will be chosen.

• Mallow's selection estimator :

$$\hat{\mu}_t^m = \widetilde{\mu}_t I(M_n > 2 + 2b^*) + \hat{\mu}_t I(M_n \le 2 + 2b^*).$$

• AMSE and AFR of the Mallow's selection estimator can be easily derived.

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Mallow's Averaging Estimation

• The optimal Mallows criterion for a weighted estimator $\hat{\mu}_t(w) = w\hat{\mu}_t + (1-w)\tilde{\mu}_t$ is

$$\begin{split} M_w &= n \widehat{\Omega}_{11}(w) + 2 \widetilde{\Omega}_{11}[2(p-1)+1] + 2(1-w)(1+b^*) \widetilde{\Omega}_{11} \\ & E(M_w) - n \Omega_{11} \rightarrow \mathsf{AMSE} \text{ of the weighted estimator} \end{split}$$

where $\widehat{\Omega}_{11}(w)$ is the estimate of Ω_{11} using the weighted estimator.

• Mallows weight \hat{w} is chosen by minimizing M_w ,

Unconstrained Estimation Constrained Estimation Pretesting Estimation Mallow's Selection Estimation Mallow's Averaging Estimation

Mallow's Averaging Estimation

 The Mallows averaging estimator is the weighted average of the constrained estimator and the unconstrained estimator using the Mallows weight ŵ,

$$\begin{aligned} \hat{\mu}_t^{ma} &= \hat{w}\hat{\mu}_t + (1-\hat{w})\widetilde{\mu}_t \\ &= \begin{cases} \frac{1+b^*}{M_n}\hat{\mu}_t + (1-\frac{1+b^*}{M_n})\widetilde{\mu}_t & \text{if } M_n > 1+b^* \\ \hat{\mu}_t & \text{otherwise} \end{cases} \end{aligned}$$

• AMSE and AFR of the Mallows averaging estimator can be derived using this expression.





2 Model and Estimation



- 4 Finite Sample Properties
- 5 Conclusion and Future Directions

Model Setup for Monte Carlo Simulations

We plot the AMSE and AFR of various estimators via simulations. Data generating process :

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \begin{bmatrix} \frac{c_y}{\sqrt{n}} \\ c_x \end{bmatrix} [Y_{t-1} - \beta X_{t-1}] + \epsilon_t, \quad (5)$$

(日) (同) (三) (三)

29 / 38

We set $\beta = 0.5$, $c_x = 0.5$, and let c_y take values from a grid formed by 100 points between -3 to 3, and $var(\epsilon_t) = l_2$.



Figure 2: AMSE





Figure 3: AFR



< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

Outline

- Introduction and Motivation
- 2 Model and Estimation
- 3 Asymptotic Comparison
- 4 Finite Sample Properties
- 5 Conclusion and Future Directions

Finite Sample Properties

We exam the finite sample performance , with the same choice of parameters as we did in the asymptotic performance. We replicate the process 10,000 times for T = 2000. The one-step-ahead forecast error are computed in each replication and the average over 10,000 replications are reported along the values of c_y in figure 4.

Mean Squared One-step-ahead Forecast Error

Figure 4: Mean Squared Forecast Error



・ロ ・ ・ (日 ・ ・ 三 ・ ・ 三 ・ つ へ ()
34 / 38

Outline

- 1 Introduction and Motivation
- 2 Model and Estimation
- 3 Asymptotic Comparison
- ④ Finite Sample Properties
- **5** Conclusion and Future Directions

Working in Progress

- K > 2 non-stationary time series are considered
- 0 < r < K cointegration relationships exist
- How will the Mallow's averaging estimator improve forecast performance?

Conclusion

In this paper, we have

- We propose to use model averaging estimation for cointegrated VAR models.
- The AMSE and AFR of various estimators are derived and shown to depend on the strength of the cointegration signal, the number and location of unit roots in the multivariate non-stationary system.
- Both asymptotic comparison and finite sample comparison favor our proposed averaging estimator and cautions against the pretesting estimator

Future directions

There are many directions to proceed in the future.

- the local-to-cointegration framework
- to incorporate the local-to-unity series in the model

(日) (同) (三) (三)

3

38 / 38