

# Dynamic Factor Model with infinite-dimensional factor space: forecasting

Mario Forni, Università di Modena and RECent

Alessandro Giovannelli, Università di Roma, Tor Vergata

Marco Lippi, Einaudi Institute for Economics and Finance, Roma

Stefano Soccorsi, Università di Roma, La Sapienza

2015 NBER-NSF Time Series Conference

Vienna

September 25-26

# Outline

## Introduction

### Prediction. Finite-dimensional factor space

SW (time domain)

FHLR (frequency domain)

### Prediction. Infinite-dimensional factor space (frequency domain)

FHLZ. Blockwise autoregressive representation and the static form

## Empirical results

# Outline

## Introduction

## Prediction. Finite-dimensional factor space

SW (time domain)

FHLR (frequency domain)

## Prediction. Infinite-dimensional factor space (frequency domain)

FHLZ. Blockwise autoregressive representation and the static form

## Empirical results

# Outline

## Introduction

## Prediction. Finite-dimensional factor space

SW (time domain)

FHLR (frequency domain)

## Prediction. Infinite-dimensional factor space (frequency domain)

FHLZ. Blockwise autoregressive representation and the static form

## Empirical results

# Outline

## Introduction

## Prediction. Finite-dimensional factor space

- SW (time domain)

- FHLR (frequency domain)

## Prediction. Infinite-dimensional factor space (frequency domain)

- FHLZ. Blockwise autoregressive representation and the static form

## Empirical results

# General form of DFM

## General form of the Large-Dimensional Dynamic Factor Model

$$x_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt} + \xi_{it}, \quad i \in \mathbb{N}, \quad (1)$$

where

$$c_{if}(L) = c_{if,0} + c_{if,1}L + \cdots + c_{if,s_1}L^{s_1} \quad \text{and} \quad d_{if}(L) = d_{if,0} + d_{if,1}L + \cdots + d_{if,s_2}L^{s_2},$$

$\mathbf{u}_t = (u_{1t} \ u_{2t} \ \cdots \ u_{qt})'$  is a  $q$ -dimensional white noise, the variables  $\xi_{it}$  are idiosyncratic components.

# Finite dimension

If the common components

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}, \quad i \in \mathbb{N},$$

span a **finite-dimensional vector space**, then (1) can be rewritten in the **static form**

$$x_{it} = \lambda_{i1} F_{1t} + \lambda_{i2} F_{2t} + \cdots + \lambda_{ir} F_{rt} + \xi_{it}. \quad (2)$$

For example, if

$$\chi_{it} = \frac{a_{i1}}{1 - \alpha L} u_{1t} + a_{i2} u_{2t} + a_{i3} u_{2,t-1}, \quad (3)$$

then the model has a **static representation** with  $r = 3$ ,

$$\lambda_{ij} = a_{ij}, \quad F_{1t} = (1 - \alpha L)^{-1} u_{1t}, \quad F_{2t} = u_{2t}, \quad F_{3t} = u_{2,t-1}.$$

# Finite dimension

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it}. \quad (2)$$

The finite-dimension assumption or, equivalently, the existence of the static representation (2), has been almost universally adopted in the recent literature (see in particular Stock and Watson, 2002a,b, Bai and Ng, 2002).

# Finite dimension

Under the **finite-dimension assumption**, the factors  $F_{jt}$  and the loadings  $\lambda_{ij}$  can be consistently estimated, as the number of variables and the number of observations for each variable tend to infinity, by means of the **principal components of the observables**  $x_{it}$ . The estimated factors are then used as predictors in forecasting the variables  $x_{it}$ .

# Infinite dimension

A motivation for studying the general model (1), without restrictions on the dimension of the factor space, as argued in FHLZ (2015), is that the finite-dimension assumption (2) rules out models as simple as

$$x_{it} = \frac{a_i}{1 - \alpha_i L} u_t + \xi_{it}, \quad (4)$$

i.e.

$$x_{it} = a_i(u_t + \alpha_i u_{t-1} + \alpha_i^2 u_{t-2} + \cdots) + \xi_{it},$$

for  $i \in \mathbb{N}$ , where  $u_t$  is a scalar white noise and the coefficients  $\alpha_i$  are drawn from, say, the uniform distribution between  $-0.8$  and  $0.8$ . For, unless  $\alpha_i$  takes a finite number of values as  $i \in \mathbb{N}$ , the stochastic variables  $\chi_{it} = a_i(1 - \alpha_i L)^{-1} u_t$  span an **infinite-dimensional space**.

# Infinite dimension

On the other hand, as pointed out in Forni et al. (2015a, Journal of Econometrics), when a dataset is given, with finite  $n$  (number of variables) and  $T$  (number of observations), **the static method might perform well even under misspecification**, i.e. even if the data were generated by a model not fulfilling the finite-dimension assumption.

In Forni et al. (2015b) the static and the dynamic methods have been applied to simulated data in several Monte Carlo experiments. A very short summary of the results is that (i) when the data are generated by infinite-dimensional models which generalize (4), the estimation of impulse-response functions and predictions obtained by the dynamic method are by far better than those obtained by the static method; (ii) when the data are generated by (2), still the dynamic method performs slightly better.

THE PRESENT PAPER COMPARES THESE METHODS USING MONTHLY MACROECONOMIC US DATA 1959-2014 (THE STOCK-WATSON DATASET).

# Previous literature comparing different factor models

- Boivin and Ng (2005), International Journal of Central Banking
- D'Agostino and Giannone (2012) Oxford Bulletin of Economics and Statistics
- Schumacher (2007), Journal of Forecasting
- Luciani (2014), International Journal of Forecasting

# Prediction: Stock and Watson

Starting with (2) the forecasting equation is obtained projecting  $x_{1,t+1}$  on the space spanned by

$$\mathbf{F}_t, \mathbf{F}_{t-1}, \dots; x_{1t}, x_{1,t-1}, \dots$$

Thus

$$x_{1,t+1|t} = \alpha(L)\mathbf{F}_t + \beta(L)x_{1t}, \quad (5)$$

where lagged values of  $x_{1t}$  account for possible autocorrelation (predictability) of the idiosyncratic component. Equation (5) is often used without lags for  $\mathbf{F}_t$  or without the term including  $x_{1,t}$ :

$$x_{1,t+1|t} = \alpha\mathbf{F}_t + \beta(L)x_{1t}, \quad x_{1,t+1|t} = \alpha\mathbf{F}_t.$$

We refer to this predictor as SW, from Stock and Watson (2002).

# Prediction: FHLR

FHLR. (i) The procedure starts with the estimation of the spectral density matrix of the  $x$ 's, call it  $\Sigma^x(\theta)$ .

(ii)  $q$  is determined by the Hallin and Liska Criterion.

(iii) The spectral density of the common components is estimated using the first  $q$  Brillinger principal components of the  $x$ 's (by means of the frequency domain eigenvectors of  $\Sigma^x(\theta)$ ), call it  $\Sigma^\chi(\theta)$ . (iv)  $\Sigma^\chi(\theta)$  and  $\Sigma^\xi(\theta)$  are used to compute the autocovariance matrices of the common and idiosyncratic components:

$$\Gamma_k^\chi, \quad \Gamma_k^\xi, \quad k \in \mathbb{Z},$$

$$\Gamma_k^\chi = \int_{-\pi}^{\pi} e^{ik\theta} \Sigma^\chi(\theta) d\theta, \quad \Gamma_k^\xi = \int_{-\pi}^{\pi} e^{ik\theta} \Sigma^\xi(\theta) d\theta$$

# Prediction: FHLR

(iv) **Under the finite-dimension assumption**, the covariances  $\Gamma_0^x$  and  $\Gamma_0^\xi$  are employed to estimate a basis in the factor space by means of generalized principal components (the estimated variance of the idiosyncratic is taken into account):

$$G_{1t}, G_{2t}, \dots, G_{rt}.$$

The covariances  $\Gamma_1^x$  and  $\Gamma_1^\xi$  are employed to project  $\chi_{1,t+1}$  on the factors:

$$\chi_{1,t+1|t} = \gamma \mathbf{G}_t$$

(a dynamically more complex version, allowing for lags of the factors, as in SW, can be obtained in the same way). Lastly, the idiosyncratic component is predicted by means of a univariate model. This predictor is based on Forni, Hallin, Lippi and Reichlin (2000, 2005). We refer to it as FHLR.

# Prediction: FHLZ

Let us go back to model (1), reported here for convenience:

$$x_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt} + \xi_{it}, \quad i \in \mathbb{N},$$

where

$$c_{if}(L) = c_{if,0} + c_{if,1}L + \cdots + c_{if,s_1}L^{s_1} \quad \text{and} \quad d_{if}(L) = d_{if,0} + d_{if,1}L + \cdots + d_{if,s_2}L^{s_2},$$

with common components

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt}.$$

# Prediction: FHLZ. Preliminary

We start in the same way as in FHLR (2005), but now we have rates.

(i) We start with an estimated lag-window spectral density  $\hat{\Sigma}_n^x(\theta)$ . We assume that

$$\hat{\Sigma}_{ij}^x(\theta) \rightarrow \Sigma_{ij}^x(\theta)$$

in probability for  $T \rightarrow \infty$ , uniformly in  $\theta$ ,  $i$  and  $j$ , with rate  $1/\sqrt{\rho_T}$ , where  $\rho_T = T^{2/3}/\log T$  (see Wu and Zaffaroni, 2015). This is the price paid to non parametric estimation of the spectrum of the  $x$ 's.

(ii) Then we obtain an estimator for the spectral density of the  $\chi$ 's,  $\hat{\Sigma}_n^\chi(\theta)$  (principal components in the frequency domain), and prove that

$$\hat{\Sigma}_{ij,n}^\chi(\theta) \rightarrow \Sigma_{ij}^\chi(\theta)$$

as  $T$  and  $n$  tend to infinity, in probability at the rate  $\max(n^{-1/2}, \rho_T^{-1/2})$ .

# Prediction: FHLZ. Preliminary

(iii) Then we prove that

$$\hat{\gamma}_{js,k,n}^{\chi} = \int_{-\pi}^{\pi} \hat{\sigma}_{js,n}^{\chi}(\theta) e^{ik\theta} d\theta \rightarrow \int_{-\pi}^{\pi} \sigma_{js,n}^{\chi}(\theta) e^{ik\theta} d\theta = \gamma_{js,k}^{\chi}$$

in probability as  $T$  and  $n$  tend to infinity, at the same rate as above.

**Here we move away from FHLR in that we do not assume finite dimension of the factor space.**

# FHLZ. Autoregressive representation for the common components

We have the vector

$$\chi_t = (\chi_{1t} \ \chi_{2t} \ \cdots \ \chi_{nt}, \ \cdots)',$$

and we have estimated (we know) its spectral density and therefore its covariance function.

In principle we can estimate a VAR for  $\chi_t$ . You may object: but if you could not estimate a VAR for  $\mathbf{x}_t$  because of the large dimension, how do you think you can do it for  $\chi_t$ ?

The answer is: unlike  $\mathbf{x}_t$ ,  $\chi_t$  is a singular vector, a large dimensional vector driven by a small number of shocks.

# FHLZ. Consequences of the singularity of $\chi_t$

Singular vectors have special properties, recently studied by Anderson and Deistler in a number of papers. An example. Let  $n = 2$  and  $q = 1$ :

$$\begin{aligned} y_{1t} &= \alpha_1 w_t + \beta_1 w_{t-1} \\ y_{2t} &= \alpha_2 w_t + \beta_2 w_{t-1} \end{aligned}$$

We see that

$$w_t = \frac{1}{\alpha_1 \beta_2 - \alpha_2 \beta_1} (\beta_2 y_{1t} - \beta_1 y_{2t}),$$

so that

$$\begin{pmatrix} 1 - \delta \beta_1 \beta_2 L & \delta \beta_1^2 L \\ -\delta \beta_2^2 & 1 + \delta \beta_1 \beta_2 L \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} w_t$$

where  $\delta = 1/(\alpha_1 \beta_2 - \alpha_2 \beta_1)$ . Note that the autoregressive representation exists if and only if  $\alpha_1 \beta_2 - \alpha_2 \beta_1 \neq 0$ , thus generically.

# FHLZ. Consequences of the singularity of $\chi_t$

Suppose now that the dimension of the vector is 4, with  $q = 1$ :

$$y_{it} = \alpha_i w_t + \beta_i w_{t-1}, \quad i = 1, 2, 3, 4$$

You see that this vector has, generically, the autoregressive representation

$$\begin{pmatrix} 1 - a_{11}L & a_{12}L & 0 & 0 \\ a_{21}L & 1 - a_{22}L & 0 & 0 \\ 0 & 0 & 1 - a_{31}L & a_{32}L \\ 0 & 0 & a_{41}L & 1 - a_{42}L \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} w_t$$

# FHLZ: Additional assumption

## A. Current and past values of any $(q + 1)$ -tuple

$$(\chi_{i_1,t} \chi_{i_2,t} \cdots \chi_{i_{q+1},t})$$

**span the same space spanned by current and past values of all the  $\chi_{it}$ .**

Let me illustrate this by an example. Assume  $q = 1$  and

$$\begin{aligned}\chi_{1t} &= u_{t-1} \\ \chi_{2t} &= u_{t-1} \\ \chi_{jt} &= u_t, \text{ for } j \geq 3\end{aligned}$$

Then Assumption A does not hold. Thus (ii) rules out cases like this. On the other hand, if

$$\chi_{it} = \alpha_i u_t + \beta_i u_{t-1},$$

then generically A holds.

# FHLZ: AR representation for the common components

Assume for convenience that  $n = (q + 1)m$ . Under Assumptions (i) and (ii) we obtain the representation

$$\begin{pmatrix} A^1(L) & 0 & \cdots & 0 \\ 0 & A^2(L) & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & A^m(L) \end{pmatrix} \begin{pmatrix} \chi_t^1 \\ \chi_t^2 \\ \vdots \\ \chi_t^m \end{pmatrix} = \begin{pmatrix} R^1 \\ R^2 \\ \vdots \\ R^m \end{pmatrix} \mathbf{v}_t$$

$$A_n(L)\chi_{nt} = R_n\mathbf{v}_t$$

where the vectors  $\chi_t^k$  are non overlapping  $(q + 1)$ -dimensional selections, the blocks  $A^k(L)$  are  $(q + 1) \times (q + 1)$ , the matrix  $R_n$  is  $n \times q$  and  $\mathbf{v}_t$  is a  $q$  dimensional white noise. The matrices  $A^k(L)$  are minimum order.

# FHLZ: The model in static form

Start again with

$$A_n(L)\chi_{nt} = R_n \mathbf{v}_t.$$

Using  $\mathbf{x}_{nt} = \chi_{nt} + \xi_{nt}$ ,

$$A_n(L)(\mathbf{x}_{nt} - \xi_{nt}) = R_n \mathbf{v}_t,$$

that is:

$$A_n(L)\mathbf{x}_{nt} = \mathbf{z}_{nt} = R_n \mathbf{v}_t + A_n(L)\xi_{nt},$$

$$\mathbf{z}_{nt} = R_n \mathbf{v}_t + \Xi_{nt}.$$

Under mild assumptions we can prove that  $\Xi_{nt}$  is idiosyncratic. Thus the original dynamic model has been transformed into a **static** one. We use the spectral density of the  $\chi$ 's to obtain the matrices  $A_n(L)$ , thus  $\mathbf{z}_{nt}$ . Then use the static model to estimate  $\mathbf{v}_t$ .

# FHLZ: The forecasting equation

Inverting the autoregressive equation

$$A^k(L)\chi_t^k = R^k v_t$$

we get

$$\chi_t^k = [A^k(L)]^{-1} R^k v_t = B_0^k v_t + B_1^k v_{t-1} + \dots$$

This is used to predict  $\chi_{t+h}^k$ :

$$\chi_{t+h|t}^k = B_h^k v_t + B_{h+1}^k v_{t-1} + \dots$$

# FHLZ: Reordering the panel

Under our assumptions, the particular grouping into  $(q + 1)$ -dimensional subvectors of  $\chi_{nt}$  does not matter in population. **However, in practice the grouping makes a difference.** We deal with this problem by **randomizing** the groupings and **averaging** over the results.

# FHLZ. Forecasting equation

**SW:**  $x_{1,t+h|t} = \alpha(L)\mathbf{F}_t + \beta(L)x_{1t}, \quad x_{1,t+h|t} = \alpha\mathbf{F}_t$

**FHLR:**  $x_{1,t+h|t} = \alpha(L)\mathbf{G}_t + \beta(L)x_{1t}, \quad x_{1,t+h|t} = \alpha\mathbf{G}_t$

**FHLZ:** We use  $\hat{A}^k(L)$  and  $\hat{u}_t$ :

$$x_{t+h|t}^1 = \hat{A}^k(L)^{-1} \hat{R}^k E[\hat{v}_{t+h}] = \hat{B}_h \hat{v}_t + \hat{B}^{h+1} \hat{v}_{t-1} + \dots$$

# DATA

The dataset consists of 115 U.S. macroeconomic and financial time series observed at monthly frequency between January 1959 and September 2014. See the paper for details and the transformations necessary to achieve stationarity. We concentrate on inflation  $(1 - L^{12})\text{CPI}_t$  and industrial production  $\text{IP}_t$ .

We use the subsample up to December 1984 to calibrate the forecasting equations, the remaining sample for the comparison. Estimation and pseudo out-of-sample forecasting are based on a ten-years rolling window.

# Parameters. Calibration

- SW:**
- (i) Number of static factors,
  - (ii) Include lagged factors in the forecasting equation or not,
  - (iii) Include lagged  $x$  in the forecasting equation or not.

- FHLR:** (i), (ii), (iii), plus
- (a) Number of dynamic factors,
  - (b) Window and kernel in the spectral estimation.

- FHLZ:** (a) and (b), plus
- ( $\alpha$ ) Degree of the polynomials  $A^k(L)$ ,
  - ( $\beta$ ) Number of permutations of the panel.

# Average results

h	IP				CPI		SW	AR
	FHLZ	FHLR	SW	AR	FHLZ	FHLR		
1	0.949	0.910	0.926	1.000	0.958	1.066	1.085	1.000
3	1.048	0.944	0.902	1.000	0.891	1.001	1.066	1.000
6	0.981	0.893	1.020	1.000	0.887	1.013	1.146	1.000
12	0.906	0.840	0.987	1.000	0.972	1.120	0.985	1.000

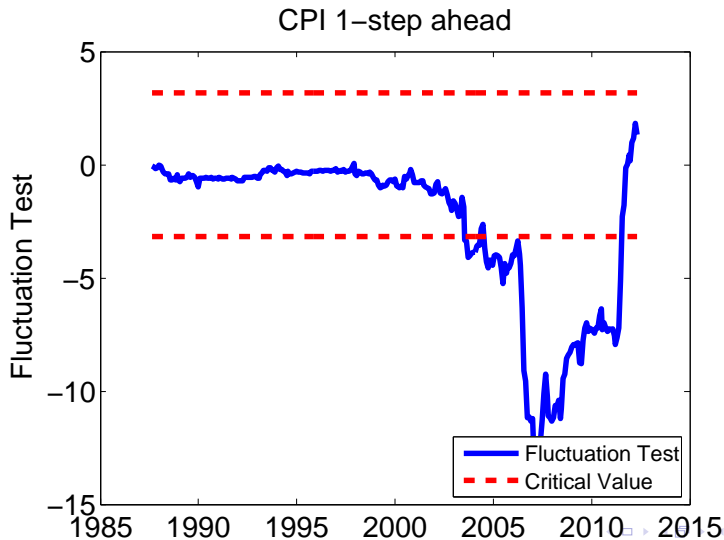
# Testing the relative forecasting performance

We use the test by Giacomini and Rossi to test the Null of equal performance at each point in time of two forecast methods, whose prediction errors are denoted by  $\hat{\epsilon}_{t+h}$  and  $\hat{\eta}_{t+h}$ :

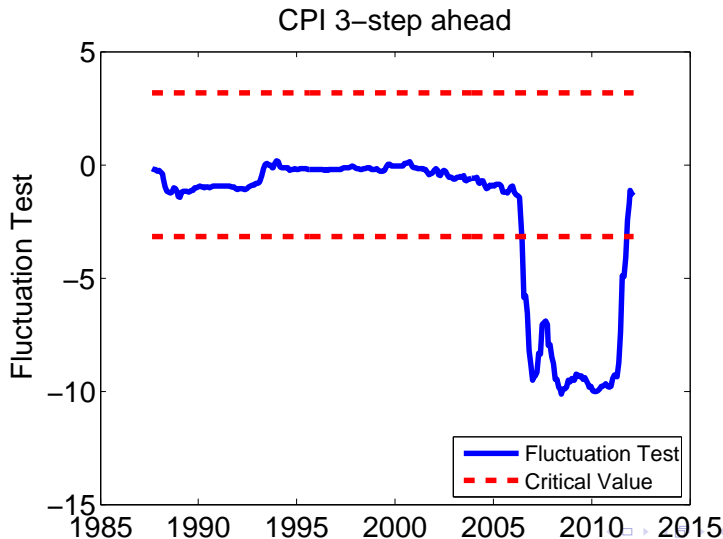
$$F_{t,m} = \frac{1}{\hat{\sigma} m^{1/2}} \left( \sum_{j=t-m/2}^{t+m/2} \hat{\epsilon}_{j+h}^2 - \sum_{j=t-m/2}^{t+m/2} \hat{\eta}_{j+h}^2 \right)$$

where between brackets we have the difference between squared prediction errors smoothed over a rolling window of size  $m + 1$ . This is normalized using  $m$  and  $\hat{\sigma}$  is a Heteroskedasticity and Autocorrelation Consistent estimator of the variance of  $\epsilon^2 - \eta^2$ . G. and R. compute critical values that, if crossed, signal that one method outperforms the other.

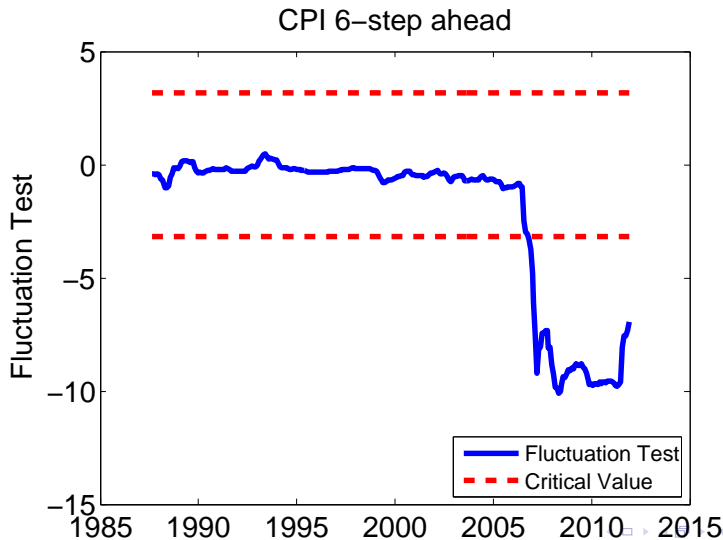
# FHLZ vs SW



# FHLZ vs SW

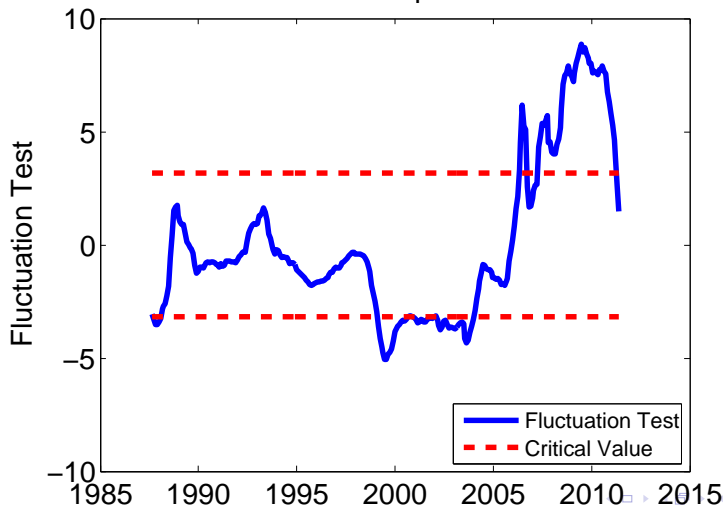


# FHLZ vs SW



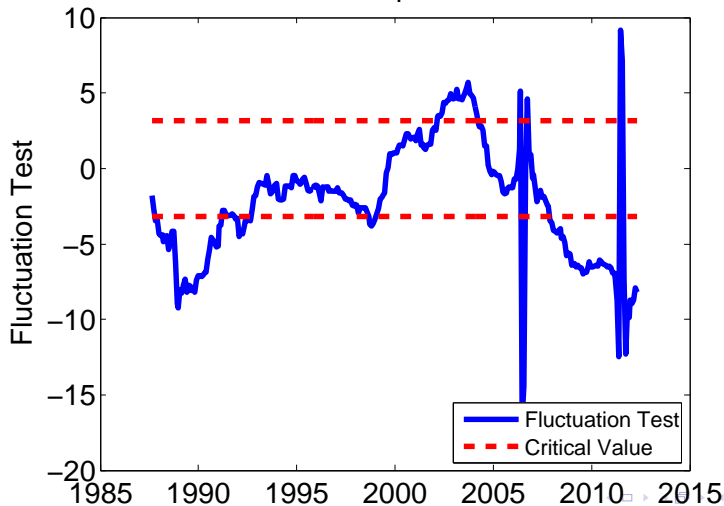
# FHLZ vs SW

## CPI 12-step ahead



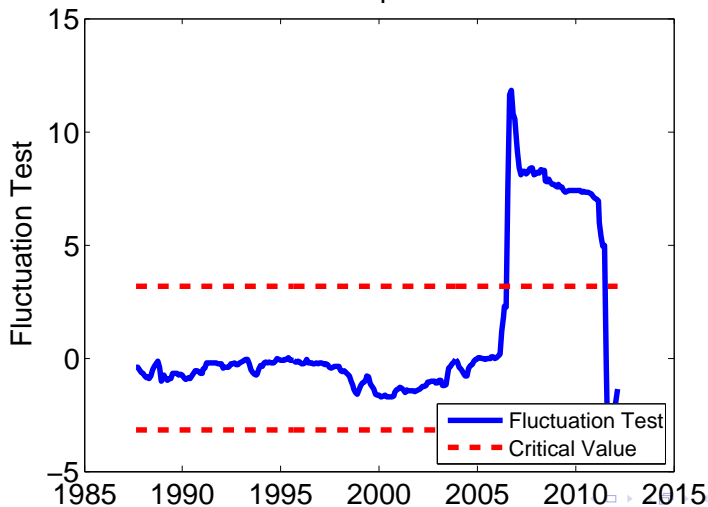
# FHLR vs SW

IP 1-step ahead



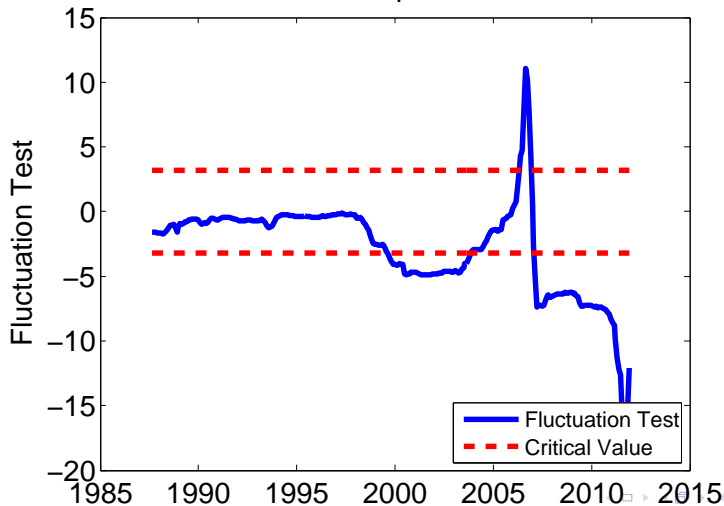
# FHLR vs SW

IP 3-step ahead



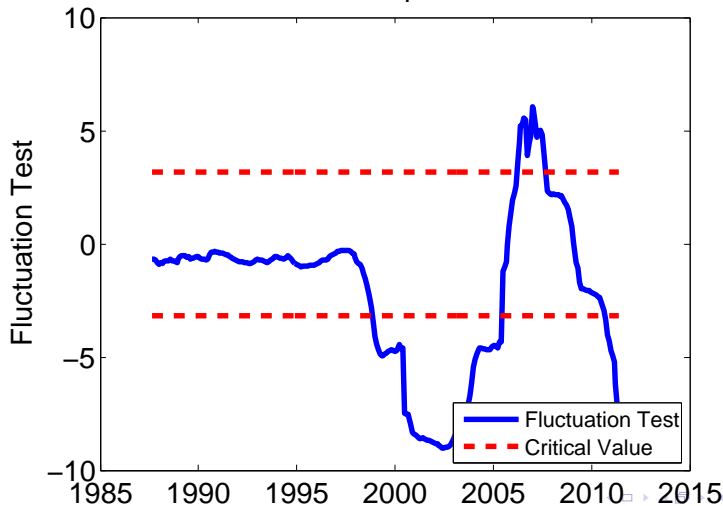
# FHLR vs SW

IP 6-step ahead



# FHLR vs SW

IP 12-step ahead



# Conclusion

Good results of frequency-domain methods as compared to time domain. In particular, FHLZ very good with nominal variables; FHLR very good with real variables.

Potential for improving on FHLZ: we obtain many predictors by reordering the variables. We averaged over them and also made some attempts with LASSO etc., with some good results. Still to be explored systematically.