Factor Augmented Forecasting Subject to Structural Breaks in the Factor Structure

Xu Han*, Ze Yu Zhong†

*City University of Hong Kong, †Monash University

December 2024

We acknowledge the financial support of the Australian Research Council under Grants LP160101038, DP210101440, and DE200100693.

Introduction

Factor augmented forecasts are the *de facto* benchmark, Stock and Watson (2002, 2012)

- Extract factors from large datasets to use as predictors and maintain parsimony

Introduction

Factor augmented forecasts are the *de facto* benchmark, Stock and Watson (2002, 2012)

- Extract factors from large datasets to use as predictors and maintain parsimony

Challenge: instability in (high-dimensional) time series

- Great Moderation, Global Financial Crisis, COVID-19

Introduction

Factor augmented forecasts are the *de facto* benchmark, Stock and Watson (2002, 2012)

- Extract factors from large datasets to use as predictors and maintain parsimony

Challenge: instability in (high-dimensional) time series

- Great Moderation, Global Financial Crisis, COVID-19

Questions we answer

- How does a break in the factor structure affect factor-augmentated forecasts?
- Can we design a factor estimator that utilizes pre- and post-break data?
- Can we combine the these forecasts together regardless of break size?

Literature

Forecasting with observed regressors with structural changes

- Pesaran et al. (2006, 2013)

Literature

Forecasting with observed regressors with structural changes

- Pesaran et al. (2006, 2013)

Factor Models with Structural Changes:

- Bai et al. (2020), Baltagi et al. (2021), Chen et al. (2014), Duan et al. (2022), and Han and Inoue (2015)

Literature

Forecasting with observed regressors with structural changes

- Pesaran et al. (2006, 2013)

Factor Models with Structural Changes:

- Bai et al. (2020), Baltagi et al. (2021), Chen et al. (2014), Duan et al. (2022), and Han and Inoue (2015)

Literature on factor-augmented forecasting under instability is limited

- Small breaks: full sample PCA factors are still consistent, Bates et al. (2013)
- Large breaks:
    - Equivalent to model with constant loadings but possibly more (pseudo) factors
    - Generally does not distinguish between rotation and shift type breaks
- Other approaches such as local-PCA amount to continuous breaks in forecasting eqn, Fu et al., 2023

Literature

Forecasting with observed regressors with structural changes

- Pesaran et al. (2006, 2013)

Factor Models with Structural Changes:

- Bai et al. (2020), Baltagi et al. (2021), Chen et al. (2014), Duan et al. (2022), and Han and Inoue (2015)

Literature on factor-augmented forecasting under instability is limited

- Small breaks: full sample PCA factors are still consistent, Bates et al. (2013)
- Large breaks:
  - Equivalent to model with constant loadings but possibly more (pseudo) factors
  - Generally does not distinguish between rotation and shift type breaks
- Other approaches such as local-PCA amount to continuous breaks in forecasting eqn, Fu et al., 2023

Empirical results are generally mixed

- Out of sample: Banerjee et al. (2008), Corradi and Swanson (2014), and Massacci and Kapetanios (2024)
- In sample: Massacci (2019) and Stock and Watson (2009)

Contributions

Propose new “rotated factor” estimator

- Minimize the L2 distance between the pre and post-break loading matrices
- Robust to large shift breaks, unlike full-sample PC
- Outperforms split sample PCA when rotational break is small

Contributions

Propose new "rotated factor" estimator

- Minimize the L2 distance between the pre and post-break loading matrices
- Robust to large shift breaks, unlike full-sample PC
- Outperforms split sample PCA when rotational break is small

Establish asymptotic properties of the three competing factor estimators

- Out-of-sample (OOS) rankings of MSFEs under different types of breaks and magnitudes

Contributions

Propose new "rotated factor" estimator

- Minimize the L2 distance between the pre and post-break loading matrices
- Robust to large shift breaks, unlike full-sample PC
- Outperforms split sample PCA when rotational break is small

Establish asymptotic properties of the three competing factor estimators

- Out-of-sample (OOS) rankings of MSFEs under different types of breaks and magnitudes

Combine factor-based forecasts using cross-validation

- Automatic data-adaptive weights - no need to know/estimate break sizes

Contributions

Propose new "rotated factor" estimator

- Minimize the L2 distance between the pre and post-break loading matrices
- Robust to large shift breaks, unlike full-sample PC
- Outperforms split sample PCA when rotational break is small

Establish asymptotic properties of the three competing factor estimators

- Out-of-sample (OOS) rankings of MSFEs under different types of breaks and magnitudes

Combine factor-based forecasts using cross-validation

- Automatic data-adaptive weights - no need to know/estimate break sizes

Simulation and Empirical results show improved forecasting performance compared to existing methods

Model & Break Setup

Direct factor augmented forecast h-step ahead forecast ($t = 1, \ldots, T, i = 1, \ldots, N$):

$$y_{t+h} = \gamma_0 + \beta(L)^\intercal f_t + \gamma(L) y_t + \eta_{t+h}, \tag{1}$$

$$x_{it} = \begin{cases} \lambda_{1,it}^\intercal f_t + e_{it}, & t = 1, \ldots, \lfloor \pi T \rfloor, \\ \lambda_{2,it}^\intercal f_t + e_{it}, & t = \lfloor \pi T \rfloor + 1, \ldots, T. \end{cases} \tag{2}$$

- $\gamma(L)$ and $\beta(L)$ are p and q (unknown) finite-order lag polynomials
- f_t are r dimensional unobserved factors estimated via principal components on x_{it}
- Assume break is only in factor loadings
- r and π treated as known (can be consistently estimated/averaged over)

Model & Break Setup

Direct factor augmented forecast h-step ahead forecast ($t = 1, \ldots, T, i = 1, \ldots, N$):

$$y_{t+h} = \gamma_0 + \beta(L)^\intercal f_t + \gamma(L) y_t + \eta_{t+h}, \tag{1}$$

$$x_{it} = \begin{cases} \lambda_{1,it}^\intercal f_t + e_{it}, & t = 1, \ldots, \lfloor \pi T \rfloor, \\ \lambda_{2,it}^\intercal f_t + e_{it}, & t = \lfloor \pi T \rfloor + 1, \ldots, T. \end{cases} \tag{2}$$

- $\gamma(L)$ and $\beta(L)$ are p and q (unknown) finite-order lag polynomials
- f_t are r dimensional unobserved factors estimated via principal components on x_{it}
- Assume break is only in factor loadings
- r and π treated as known (can be consistently estimated/averaged over)

Stacking x_{it} into $T \times N$ matrices:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_1 \Lambda_1^\intercal + e_1 \\ F_2 \Lambda_2^\intercal + e_2 \end{bmatrix}. \tag{3}$$

Model - Break Decomposition of Λ_2

Decompose Λ_2 via a *projection* as:

$$\Lambda_2 = \Lambda_1 \underbrace{Z}_{\text{rotation}} + \underbrace{W}_{\text{shift}},$$

where $Z = E(\Lambda_1^\intercal \Lambda_1)^{-1} E(\Lambda_1^\intercal \Lambda_2)$ so that $E(\Lambda_1^\intercal W) = 0$.
Pure shift break if $Z = I_r$; pure rotation break if $W = 0$.

Model - Break Decomposition of Λ_2

Decompose Λ_2 via a *projection* as:

$$\Lambda_2 = \Lambda_1 \underbrace{Z}_{\text{rotation}} + \underbrace{W}_{\text{shift}},$$

where $Z = E(\Lambda_1^\intercal \Lambda_1)^{-1} E(\Lambda_1^\intercal \Lambda_2)$ so that $E(\Lambda_1^\intercal W) = 0$.
Pure shift break if $Z = I_r$; pure rotation break if $W = 0$.
Allow for different magnitude of both components with $\alpha, \nu \in [0, 1]$:

$$\begin{aligned} Z &= I_r + \frac{R}{N^{1-\nu}}, \\ W &= \frac{D}{N^{(1-\alpha)/2}}. \end{aligned} \tag{4}$$

Equation (4) implies $\Lambda_1^\intercal W/N = O_p\left(N^{\alpha/2-1}\right)$.

- Fraction of series breaking: $w_i \neq 0$ for $i = 1, \ldots, N_1$ with $N_1 \propto N^\alpha$, $\frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \lambda_{1i}^\intercal w_i = O_p(1)$.

Model - Equivalent Representation

Substituting $\Lambda_2 = \Lambda_1 Z + W$ into Equation (3):

$$\begin{aligned} X &= \begin{bmatrix} F_1 & 0 \\ F_2 Z^\intercal & F_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^\intercal \\ W^\intercal \end{bmatrix} + e \\ &= \begin{bmatrix} G_r & G_p \end{bmatrix} \begin{bmatrix} \Lambda_1^\intercal \\ W^\intercal \end{bmatrix} + e \\ X &= G\Xi^\intercal + e, \end{aligned} \tag{5}$$

where G denotes *pseudo* factors, with dimension larger than r when $W \neq 0$.

Model - Equivalent Representation

Substituting $\Lambda_2 = \Lambda_1 Z + W$ into Equation (3):

$$
\begin{aligned}
X &= \begin{bmatrix} F_1 & 0 \\ F_2 Z^\intercal & F_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^\intercal \\ W^\intercal \end{bmatrix} + e \\
&= \begin{bmatrix} G_r & G_p \end{bmatrix} \begin{bmatrix} \Lambda_1^\intercal \\ W^\intercal \end{bmatrix} + e \\
X &= G\Xi^\intercal + e, \qquad (5)
\end{aligned}
$$

where G denotes *pseudo* factors, with dimension larger than r when $W \neq 0$.

Changing r

Focus on square non-singular Z, which implies non-changing r.
Changing r can be accommodated with an $r_1 \times r_2$ "rectangular" Z.

Estimation: Three Factor Estimators

Pseudo Factors $\tilde{F}_P$: first r full-sample principal components

Estimation: Three Factor Estimators

Pseudo Factors $\tilde{F}_P$: first r full-sample principal components

Split-sample Factors $\tilde{F}_S = [\tilde{F}_1^\intercal, \tilde{F}_2^\intercal]^\intercal$

- $\tilde{F}_1$ and $\tilde{F}_2$ via split-sample PCA, robust to types of breaks
- Needs break in forecasting eqn because $\tilde{F}_1$ and $\tilde{F}_2$ estimate $F_1 H_1$ and $F_2 H_2$, $H_1 \neq H_2$

Estimation: Three Factor Estimators

Pseudo Factors $\tilde{F}_P$: first r full-sample principal components

Split-sample Factors $\tilde{F}_S = [\tilde{F}_1^\intercal, \tilde{F}_2^\intercal]^\intercal$

- $\tilde{F}_1$ and $\tilde{F}_2$ via split-sample PCA, robust to types of breaks
- Needs break in forecasting eqn because $\tilde{F}_1$ and $\tilde{F}_2$ estimate $F_1 H_1$ and $F_2 H_2$, $H_1 \neq H_2$

Rotated Factors $\tilde{F}_R$

- Compute $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ given $\tilde{F}_1$ and $\tilde{F}_2$ via LS

$$\tilde{Z} = \left(\tilde{\Lambda}_1^\intercal \tilde{\Lambda}_1\right)^{-1} \tilde{\Lambda}_1^\intercal \tilde{\Lambda}_2,$$
$$\tilde{F}_R = [\tilde{F}_1^\intercal, \tilde{Z}\tilde{F}_2^\intercal]^\intercal.$$

Estimation: Three Factor Estimators

Pseudo Factors $\tilde{F}_P$: first r full-sample principal components
Split-sample Factors $\tilde{F}_S = [\tilde{F}_1^\intercal, \tilde{F}_2^\intercal]^\intercal$

- $\tilde{F}_1$ and $\tilde{F}_2$ via split-sample PCA, robust to types of breaks
- Needs break in forecasting eqn because $\tilde{F}_1$ and $\tilde{F}_2$ estimate $F_1 H_1$ and $F_2 H_2$, $H_1 \neq H_2$

Rotated Factors $\tilde{F}_R$

- Compute $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ given $\tilde{F}_1$ and $\tilde{F}_2$ via LS

$$\tilde{Z} = \left(\tilde{\Lambda}_1^\intercal \tilde{\Lambda}_1\right)^{-1} \tilde{\Lambda}_1^\intercal \tilde{\Lambda}_2,$$
$$\tilde{F}_R = [\tilde{F}_1^\intercal, \tilde{Z}\tilde{F}_2^\intercal]^\intercal.$$

Rotated Factors aim to not need a break in forecasting equation

Assumptions - Factor Space

Assumption 1. $E\|f_t\|^4 < \infty$, $E\left(f_t f_t^\intercal\right) = \Sigma_F$ for some $\Sigma_F > 0$.
Assumption 2. $E\|\lambda_{1i}\|^4 \leq M$, $\|\Lambda_1^\intercal \Lambda_1 / N - \Sigma_{\Lambda_1}\| \xrightarrow{p} 0$ and $E\|w_i\|^4 \leq M$, $\|W^\intercal W / N^\alpha - \Sigma_W\| \xrightarrow{p} 0$, and $\|R\| \leq M$
Assumption 3. Moments of idiosyncratic errors in Bai (2003).
Assumption 4. $\{\lambda_{m,i}\}$, $\{f_t\}$ and $\{e_{it}\}$ are mutually independent groups for $m = 1, 2$.
Assumption 5. Weak serial and cross sectional correlation in errors.
Assumption 6. Subsample version of Assumption F in Bai (2003).
Assumption 7. Both $(\Sigma_{\Lambda_1}\Sigma_F)$ and $(\Sigma_{\Lambda_2}\Sigma_F)$ have distinct eigenvalues.
Assumption 8. Break fraction π is bounded away from 0 and 1.

Assumptions - Factor Space

Assumption 1. $E\|f_t\|^4 < \infty$, $E\left(f_t f_t^\intercal\right) = \Sigma_F$ for some $\Sigma_F > 0$.
Assumption 2. $E\|\lambda_{1i}\|^4 \leq M$, $\|\Lambda_1^\intercal \Lambda_1/N - \Sigma_{\Lambda_1}\| \xrightarrow{p} 0$ and $E\|w_i\|^4 \leq M$, $\|W^\intercal W/N^\alpha - \Sigma_W\| \xrightarrow{p} 0$, and $\|R\| \leq M$
Assumption 3. Moments of idiosyncratic errors in Bai (2003).
Assumption 4. $\{\lambda_{m,i}\}$, $\{f_t\}$ and $\{e_{it}\}$ are mutually independent groups for $m = 1, 2$.
Assumption 5. Weak serial and cross sectional correlation in errors.
Assumption 6. Subsample version of Assumption F in Bai (2003).
Assumption 7. Both $(\Sigma_{\Lambda_1}\Sigma_F)$ and $(\Sigma_{\Lambda_2}\Sigma_F)$ have distinct eigenvalues.
Assumption 8. Break fraction π is bounded away from 0 and 1.
Notation: $\delta_{NT} = \min(\sqrt{T}, \sqrt{N})$.

Theorem 1. Under Assumptions 1 to 8, as $N, T \to \infty$,

1. Pseudo Factors satisfy:

$$
\begin{aligned}
T^{-1}\left\|\tilde{F}_P - G_r H_G\right\|^2 &= O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{2\alpha-2}\right) && \text{for} \quad \alpha < 1, \\
T^{-1}\left\|\tilde{F}_P - F H_G\right\|^2 &= O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{2\alpha-2}\right) + O_p\left(N^{2\nu-2}\right) && \text{for} \quad \alpha < 1, \\
T^{-1}\left\|\tilde{F}_P - G H_{\Xi,r}\right\|^2 &= O_p\left(\delta_{NT}^{-2}\right) && \text{for} \quad \alpha = 1;
\end{aligned}
$$

Theorem 1. Under Assumptions 1 to 8, as $N, T \to \infty$,

1. Pseudo Factors satisfy:

$$T^{-1}\left\|\tilde{F}_P - G_r H_G\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{2\alpha-2}\right) \qquad \text{for} \quad \alpha < 1,$$

$$T^{-1}\left\|\tilde{F}_P - F H_G\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{2\alpha-2}\right) + O_p\left(N^{2\nu-2}\right) \qquad \text{for} \quad \alpha < 1,$$

$$T^{-1}\left\|\tilde{F}_P - G H_{\Xi,r}\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) \qquad \text{for} \quad \alpha = 1;$$

2. Split Sample Factors satisfy:

$$T^{-1}\left\|\tilde{F}_\iota - F_\iota H_\iota\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) \quad \text{for} \quad \iota = 1, 2;$$

Theorem 1. Under Assumptions 1 to 8, as $N, T \to \infty$,

1. Pseudo Factors satisfy:

$$T^{-1}\left\|\tilde{F}_P - G_r H_G\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{2\alpha-2}\right) \qquad \text{for} \quad \alpha < 1,$$

$$T^{-1}\left\|\tilde{F}_P - F H_G\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{2\alpha-2}\right) + O_p\left(N^{2\nu-2}\right) \qquad \text{for} \quad \alpha < 1,$$

$$T^{-1}\left\|\tilde{F}_P - G H_{\Xi,r}\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) \qquad \text{for} \quad \alpha = 1;$$

2. Split Sample Factors satisfy:

$$T^{-1}\left\|\tilde{F}_\iota - F_\iota H_\iota\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) \quad \text{for} \quad \iota = 1, 2;$$

3. Rotated Factors satisfy:

$$T^{-1}\left\|\tilde{F}_R - G_r H_1\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{\alpha-2}\right),$$

$$T^{-1}\left\|\tilde{F}_R - F H_1\right\|^2 = O_p\left(\delta_{NT}^{-2}\right) + O_p\left(N^{\alpha-2}\right) + O_p\left(N^{2\nu-2}\right).$$

Out of Sample Forecasts - Notation and Assumptions

Notation: collect all forecasting regressors into c_t:

$$y_{t+h} = c_t^{\intercal}\theta + \eta_{t+h}. \tag{6}$$

Compute forecasts as $\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P$, $\tilde{c}_{S,T}^{\intercal}\widehat{\theta}_S$ and $\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R$.

Out of Sample Forecasts - Notation and Assumptions

Notation: collect all forecasting regressors into c_t:

$$y_{t+h} = c_t^{\intercal}\theta + \eta_{t+h}. \tag{6}$$

Compute forecasts as $\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P$, $\tilde{c}_{S,T}^{\intercal}\widehat{\theta}_S$ and $\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R$.
Let $\mathcal{F}_t = \sigma(y_t, f_t, x_{1t}, x_{2t}, \ldots, y_{t-1}, f_{t-1}, x_{1,t-1}, x_{2,t-2}, \ldots)$.
Assumption 9.

1. $\mathbb{E}\left(\eta_{t+h}|\mathcal{F}_t\right) = 0$.
2. $(c_t^{\intercal}, \eta_{t+h}, e_{1t}, \ldots, e_{Nt})$ is piecewise strictly stationary and ergodic.
3. $\mathbb{E}\|c_t\|^4 \leq M$, $\mathbb{E}\eta_t^4 \leq M$, and $\frac{1}{T}\left(c_t c_t^{\intercal}\right) \xrightarrow{p} \Sigma_{CC} > 0$.
4. $\frac{1}{\sqrt{T}}\sum_{t=1-h}^{T-h} c_t\eta_{t+h} \xrightarrow{d} N(0, \Omega_{CC,\eta})$, where $\Omega_{CC,\eta} = \text{plim}\,\frac{1}{T}\sum_{t=1}^{T-h}\eta_{t+h}^2 c_t c_t^{\intercal} > 0$.

Out of Sample Forecasts

Theorem 2. Under Assumptions 1 to 9, if $N \propto T$ and $N, T \to \infty$, then

1. For small shift breaks $\alpha < 0.5$, $\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P - \tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R = o_p(N^{-1/2})$,
2. For small rotational breaks $\nu < 0.5$,

$$E\left(\mathrm{plim}_{NT}\, N\left\|\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R - c_T^{\intercal}\theta\right\|^2\right) < E\left(\mathrm{plim}_{NT}\, N\left\|\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P - c_T^{\intercal}\theta\right\|^2\right), \qquad \text{for} \quad \alpha = 0.5,$$

$$\left\|\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R - c_T^{\intercal}\theta\right\|^2 / \left\|\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P - c_T^{\intercal}\theta\right\|^2 \xrightarrow{p} 0, \qquad \text{for} \quad \alpha > 0.5.$$

Key results:

- Rotated factors are asymptotically equivalent to pseudo factors when $\alpha < 0.5$.
- Rotated factors *weakly dominate* pseudo factors when $\nu < 0.5$.

Out of Sample Forecasts

Theorem 2.

1. For moderate rotational breaks $\nu = 0.5$,

$$\left\|\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P - c_T^{\intercal}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R - c_T^{\intercal}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{S,T}^{\intercal}\widehat{\theta}_S - c_T^{\intercal}\theta\right\|^2, \qquad \text{for} \quad \alpha \leq 0.5,$$

$$\left\|\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R - c_T^{\intercal}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{S,T}^{\intercal}\widehat{\theta}_S - c_T^{\intercal}\theta\right\|^2,$$

$$\left\|\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P - c_T^{\intercal}\theta\right\|^2 / max\left[\left\|\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R - c_T^{\intercal}\theta\right\|^2, \left\|\tilde{c}_{S,T}^{\intercal}\widehat{\theta}_S - c_T^{\intercal}\theta\right\|^2\right] \xrightarrow{p} \infty \qquad \text{for} \quad \alpha > 0.5,$$

2. For large rotational breaks $\nu > 0.5$,

$$\left\|\tilde{c}_{S,T}^{\intercal}\widehat{\theta}_S - c_T^{\intercal}\theta\right\|^2 / min\left[\left\|\tilde{c}_{R,T}^{\intercal}\widehat{\theta}_R - c_T^{\intercal}\theta\right\|^2, \left\|\tilde{c}_{P,T}^{\intercal}\widehat{\theta}_P - c_T^{\intercal}\theta\right\|^2\right] \xrightarrow{p} 0.$$

Out of Sample Forecasts - Summary

	$\nu < 0.5$	$\nu = 0.5$	$\nu > 0.5$
$\alpha < 0.5$			
$\alpha = 0.5$	R		
$0.5 < \alpha < 1$			S
$\alpha = 1$			

Table 1: Summary of Theorem 2.

- Yellow region represents rotated factors are the best, orange represents the split sample factors are the best, white represents no dominating method.
- Red box represents region where rotated factors dominate the pseudo factors, blue box represents where the rotated factors are equivalent to pseudo factors.

Out of Sample Forecasts - Summary

	$\nu < 0.5$	$\nu = 0.5$	$\nu > 0.5$
$\alpha < 0.5$			
$\alpha = 0.5$	R		
$0.5 < \alpha < 1$			S
$\alpha = 1$			

Table 1: Summary of Theorem 2.

- Yellow region represents rotated factors are the best, orange represents the split sample factors are the best, white represents no dominating method.
- Red box represents region where rotated factors dominate the pseudo factors, blue box represents where the rotated factors are equivalent to pseudo factors.
- Generally, α and ν are not known and difficult to estimate $\Rightarrow$ *forecast combination*.

Forecast Combination

Allow unknown lag structure by defining $c_t = (1, f_t^\intercal, \ldots, f_{t-q_{max}}^\intercal, y_t, \ldots, y_{t-p_{max}})$:

$$y_{t+h} = \gamma_0 + \beta(L)^\intercal f_t + \gamma(L) y_t + \eta_{t+h} \tag{7}$$

$$y_{t+h} = c_t^\intercal \theta + \eta_{t+h}. \tag{8}$$

Three different factor estimates $\tilde{F}_P, \tilde{F}_S, \tilde{F}_R$, each with $\mathcal{M}$ approximating models. Total of $3 \times \mathcal{M}$ models, each $\tilde{c}_t(m)$ for $m = 1, \ldots, 3\mathcal{M}$ specifies subset of regressors.

Forecast Combination

Allow unknown lag structure by defining $c_t = (1, f_t^\intercal, \ldots, f_{t-q_{max}}^\intercal, y_t, \ldots, y_{t-p_{max}})$:

$$y_{t+h} = \gamma_0 + \beta(L)^\intercal f_t + \gamma(L) y_t + \eta_{t+h} \tag{7}$$

$$y_{t+h} = c_t^\intercal \theta + \eta_{t+h}. \tag{8}$$

Three different factor estimates $\tilde{F}_P, \tilde{F}_S, \tilde{F}_R$, each with $\mathcal{M}$ approximating models. Total of $3 \times \mathcal{M}$ models, each $\tilde{c}_t(m)$ for $m = 1, \ldots, 3\mathcal{M}$ specifies subset of regressors.

Remark.

- Assume that y_{t+h} is generated by f_t - implicitly assumes that Z rotation is not part of factors
- If Z is interpreted to be change in factors, y_{t+h} should be generated by g_t, rotational break ν does not matter
- Model averaging/selection automatically takes care of this

Forecast Model Estimation

OLS estimate $\widehat{\theta}(m) = \left(\tilde{C}(m)^\intercal \tilde{C}(m)\right)^{-1} \tilde{C}(m)^\intercal y$. Corresponding conditional forecast for mth model, and combined forecast are respectively:

$$\widehat{y}_{T+h|T}(m) = \tilde{c}_T(m)^\intercal \widehat{\theta}(m), \tag{9}$$

$$\widehat{y}_{T+h|T}(w) = \sum_{m=1}^{3\mathcal{M}} w(m)\widehat{y}_{T+h|T}, \tag{10}$$

where $w(m), m = 1, \ldots, 3\mathcal{M}$ are forecast weights s.t. $\sum_{m=1}^{3\mathcal{M}} w(m) = 1$.

Forecast Model Estimation

OLS estimate $\widehat{\theta}(m) = \left(\tilde{C}(m)^{\intercal}\tilde{C}(m)\right)^{-1}\tilde{C}(m)^{\intercal}y$. Corresponding conditional forecast for mth model, and combined forecast are respectively:

$$\widehat{y}_{T+h|T}(m) = \tilde{c}_T(m)^{\intercal}\widehat{\theta}(m), \tag{9}$$

$$\widehat{y}_{T+h|T}(w) = \sum_{m=1}^{3\mathcal{M}} w(m)\widehat{y}_{T+h|T}, \tag{10}$$

where $w(m), m = 1, \ldots, 3\mathcal{M}$ are forecast weights s.t. $\sum_{m=1}^{3\mathcal{M}} w(m) = 1$.
Usually use Mallows and full sample Cross Validation, (Cheng & Hansen, 2015)

- Requires data to be strictly stationary
- In-sample squared loss in general not good estimate for MSFE

Forecast Model Estimation

OLS estimate $\widehat{\theta}(m) = \left(\tilde{C}(m)^\intercal \tilde{C}(m)\right)^{-1} \tilde{C}(m)^\intercal y$. Corresponding conditional forecast for mth model, and combined forecast are respectively:

$$\widehat{y}_{T+h|T}(m) = \tilde{c}_T(m)^\intercal \widehat{\theta}(m), \tag{9}$$

$$\widehat{y}_{T+h|T}(w) = \sum_{m=1}^{3\mathcal{M}} w(m) \widehat{y}_{T+h|T}, \tag{10}$$

where $w(m), m = 1, \ldots, 3\mathcal{M}$ are forecast weights s.t. $\sum_{m=1}^{3\mathcal{M}} w(m) = 1$.
Usually use Mallows and full sample Cross Validation, (Cheng & Hansen, 2015)

- Requires data to be strictly stationary
- In-sample squared loss in general not good estimate for MSFE

$\Rightarrow$ Solution: use post break leave-h-out cross validation residuals.

Leave-h-out Cross Validation Criteria

Leave-h-out CV criterion for forecast selection:

$$CV_{h,T}(m) = \frac{1}{\lfloor (1-\pi)T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \tilde{\eta}_{t+h,h}(m)^2, \tag{11}$$

$$\widehat{m} = argmin_{1 \leq m \leq 3\mathcal{M}} CV_{h,T}(m). \tag{12}$$

Leave-h-out cross validation criterion for forecast combination:

$$CV_{h,T}(w) = \frac{1}{\lfloor (1-\pi)T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \left(\sum_{m=1}^{3\mathcal{M}} w(m)\tilde{\eta}_{t,h}(m) \right)^2, \tag{13}$$

$$\widehat{w} = \underset{w}{argmin} CV_{h,T}(w). \tag{14}$$

Theorem 3. Under Assumptions 1 to 9, we have for any $h \geq 1$, fixed $\mathcal{M}$ and w, as $N, T \to \infty$,

$$CV_{h,T}(w) = \tilde{L}_{T_2}(w) + \frac{1}{T_2}\eta_{(2)}^{\intercal}\eta_{(2)} + \frac{2}{\sqrt{T_2}}\tilde{r}_{1T}(w),$$

where $\tilde{r}_{1T}(w) \to \xi_1(w)$ and $E\xi_1(w) = 0$.
Cross validation criterion is asymptotically unbiased estimate of post break mean squared loss, and therefore MSFE, plus σ^2.

Simulation Study - Design

Generate break as $\Lambda_2 = \Lambda_1 Z + W$, $\Lambda_1 \sim N(0, I_r)$,

$$Z = I_r + \frac{R}{N^{1-\nu}}; \quad R \sim N(0_r, I_r), \tag{15}$$

$$W = \frac{1.5 \times D}{N^{(1-\alpha)/2}}; \quad D \sim MVN(0_r, I_r). \tag{16}$$

Approximate factor model with structural break:

$$x_{it} = \begin{cases} \lambda_{1i}^{\intercal} f_t + \sqrt{\theta} e_{it}, & t = 1, \ldots, \lfloor \pi T \rfloor, \\ \lambda_{2i}^{\intercal} f_t + \sqrt{\theta} e_{it}, & t = \lfloor \pi T \rfloor + 1, \ldots, T, \end{cases} \tag{17}$$

$$f_{k,t} = \rho f_{k,t-1} + u_{it}, \tag{18}$$

$$e_{it} = \alpha e_{i,t-1} + v_{it}, \tag{19}$$

$\rho = 0.7$ for serial correlation in factors, $\alpha = \beta = 0.3$ for mild serial and cross sectional correlation.

Simulation Study - Design

Regression equation:

$$y_{t+h} = \beta_1 f_t + \beta_2 f_t + \beta_3 f_t + \eta_{t+h}, \tag{20}$$

$$\eta_{t+h} = \sum_{j=1}^{h-1} \kappa^j \epsilon_{t+h-j}, \quad \epsilon_t \sim N(0,1), \kappa \in \{0.1, 0.5, 0.9\}. \tag{21}$$

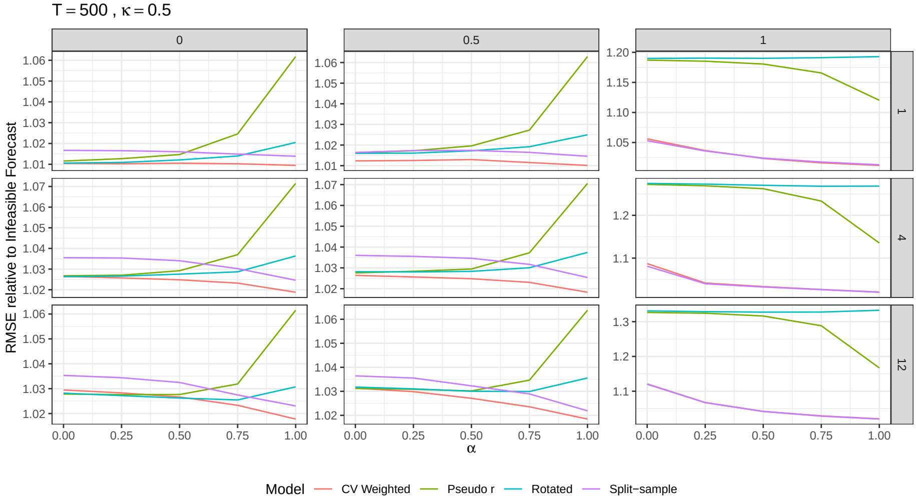
Candidate regressors:

$$\mathcal{C}_t = (1, \tilde{f}_t^\intercal, \ldots, \tilde{f}_{t-q_{max}}^\intercal, y_t, \ldots, y_{t-p_{max}}), \tag{22}$$

$\tilde{f}_t$ estimated using $\tilde{F}_P$ whole sample pseudo factors, $\tilde{F}_R$ split sample rotated factors and $\tilde{F}_S$ split sample factors.

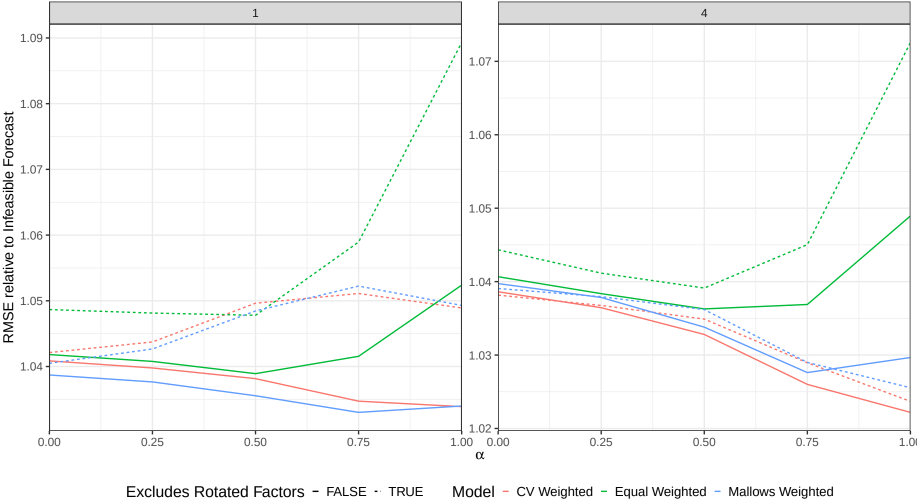
- Leave-h-out cross validation averaging
- Compare with Mallows model averaging, Bayesian model averaging, and equal weights

Simulation Study - Factor Estimator Results



Relative MSFE for each factor estimator, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$.

Simulation Study - Value of Rotated Factors



Relative MSFE for each factor estimator, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$.

Empirical Study

FRED-MD of McCracken and Ng (2016)

- 1984 March - 2020 February, Global Financial Crisis break of 2008 November, **stock_disentangling_2012-1**; Cheng et al. (2016)
- 2008 December - 2024 September, COVID19 break of 2020 March, Ng (2021)

Methodology

- Whole sample pseudo factors, rotated factors, and split sample factors
- Average over up to $r = 5$ factors, lag structure kept $p = 1, q = 3$

DFM-5 benchmark

- Direct forecast with 3 lags of y_t, augmented with 5 factors
- Very difficult to beat, even with more complex shrinkage methods

Results reported as percentiles of *RMSE* relative to DFM-5 benchmark

Global Financial Crisis Subsample

Table 2: Distributions of relative RMSEs by forecasting method, relative to DFM-5, h = 1, 2, 4, FREDMD Global Financial Crisis Subsample (1984 March - 2020 February, 2008 November Break), outlier adjusted, include = 99. No. of asterisks denote ranking.

Percentile	h = 1			h = 2			h = 3		
Model	0.250	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750
CV Weighted	0.982*	0.996**	1.006	0.984**	0.996**	1.005***	0.982*	0.996***	1.005
Equal Weighted	0.983**	0.995*	1.004**	0.983*	0.995*	1.003*	0.983**	0.994**	1.002*
Mallows Weighted	0.988	0.998	1.005***	0.988	0.998	1.005***	0.984***	0.997	1.005
Pseudo r	0.995	1.000	1.003*	0.996	1.000	1.003*	0.994	1.000	1.002*
Rotated	0.986***	0.996**	1.005***	0.985***	0.996**	1.009	0.985	0.993*	1.003***
Split-sample	0.994	1.010	1.032	0.989	1.005	1.027	0.991	1.008	1.028

COVID19 Subsample

Table 3: Distributions of relative RMSEs by forecasting method, relative to DFM-5, h = 1, 2, 4, FREDMD COVID-19 Subsample (2008 December - 2024 September, 2020 March Break), outlier adjusted, include = 99. No. of asterisks denote ranking.

Percentile	h = 1			h = 2			h = 3		
Model	0.250	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750
CV Weighted	0.912*	0.970**	1.011**	0.946	0.988	1.008**	0.935***	0.990***	1.010**
Equal Weighted	0.924***	0.985***	1.022	0.937**	0.976*	1.012	0.933**	0.982*	1.011***
Mallows Weighted	0.955	0.993	1.023	0.938***	0.981**	1.007*	0.963	0.994	1.019
Pseudo r	0.974	0.997	1.004*	0.981	1.000	1.011***	0.962	0.994	1.003*
Rotated	0.920**	0.969*	1.015***	0.924*	0.983***	1.021	0.923*	0.984**	1.016
Split-sample	0.989	1.049	1.116	0.965	1.009	1.080	0.990	1.047	1.115

Conclusion

Study the effects of structural breaks in the factor structure on factor augmented forecasting.

- Propose new rotated factor estimator, and derive it and other factor estimators asymptotic properties
- Detailed MSFE rankings of factor estimators
- CV can automatically choose and weight different estimators

Conclusion

Study the effects of structural breaks in the factor structure on factor augmented forecasting.

- Propose new rotated factor estimator, and derive it and other factor estimators asymptotic properties
- Detailed MSFE rankings of factor estimators
- CV can automatically choose and weight different estimators

Monte Carlo demonstrates the effectiveness of the proposed methods

Conclusion

Study the effects of structural breaks in the factor structure on factor augmented forecasting.

- Propose new rotated factor estimator, and derive it and other factor estimators asymptotic properties
- Detailed MSFE rankings of factor estimators
- CV can automatically choose and weight different estimators

Monte Carlo demonstrates the effectiveness of the proposed methods
Empirical results show that allowing for breaks in the forecasting equation directly does not work well, and better estimates of factors tends to work better

Thank You!