OPTIMAL FORECASTING UNDER PARAMETER INSTABILITY

Yu Bai
Monash University

November 1, 2023

MOTIVATING EXAMPLE

- Predictive regression under parameter instability

$$y_{t+1} = X_t'\theta_t + \varepsilon_{t+1}, \quad t = 1, 2, \cdots, T-1. \tag{1}$$

- Under mean squared error (MSE) loss: $L(y_{T+1}, \hat{y}_{T+1|T}) = (y_{T+1} - \hat{y}_{T+1|T})^2$, the optimal forecast is $\hat{y}_{T+1|T} = X_T'\hat{\theta}_T$.
- Rolling window forecast scheme:

$$\hat{\theta}_T = \Big(\sum_{t=T-R_0+1}^{T-1} X_t X_t'\Big)^{-1}\Big(\sum_{t=T-R_0+1}^{T-1} X_t y_{t+1}\Big), \tag{2}$$

where R_0 is the window size.

▸ Return to main slide

MOTIVATING EXAMPLE

- (2) can be written more generally as

$$\hat{\theta}_{b,T} = \Big(\sum_{t=1}^{T-1} k_{tT} X_t X_t'\Big)^{-1} \Big(\sum_{t=1}^{T-1} k_{tT} X_t y_{t+1}\Big), \tag{3}$$

  where

  - $k_{tT} = K\big((t-T)/(Tb)\big)$ is the weighting function;
  - $b = b_T > 0$ is a tuning parameter satisfying $b \to 0$, $Tb \to \infty$ as $T \to \infty$.

- If $K(u) = \mathbb{1}_{\{-1<u<0\}}$, (3) becomes (2) with $R_0 = \lfloor Tb \rfloor$.

RESEARCH QUESTION

(1) What types of time variation are allowable for using estimator like (3)?

(2) How to select the tuning parameter b optimally?

(3) Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?

RESEARCH QUESTION

1. What types of time variation are allowable for using estimator like (3)?
2. How to select the tuning parameter b optimally?
3. Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?

RESEARCH QUESTION

1. What types of time variation are allowable for using estimator like (3)?
2. How to select the tuning parameter b optimally?
3. Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?

ESTIMATION UNDER PARAMETER INSTABILITY

THE ESTIMATOR

- y_{t+h}: target
- X_t: predictors
- $\hat{y}_{t+h|t}(\theta)$: forecast
- $\ell_t(\theta) = L(y_{t+h}, \hat{y}_{t+h|t}(\theta))$: loss function
- Parameter estimates:

$$\hat{\theta}_{K,b,T} = \arg\min_{\theta \in \Theta} \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \ell_t(\theta), \tag{4}$$

  where

  - $k_{tT} = K\big((t-T)/(Tb)\big)$, $K(\cdot)$ is a weighting function;
  - $b = b_T > 0$ is the tuning parameter satisfying $b \to 0$, $Tb \to \infty$ as $T \to \infty$.

ON CONSISTENCY

- We adopt the framework of locally stationary: Karmakar et al. (2022, JoE), Dahlhaus et al. (2019, Bernoulli), etc..
- We assume that

$$\theta_{t,T} = \theta\left(t/T\right) = \theta(u), \quad \theta(\cdot) : (0, 1] \longrightarrow \Theta.$$

- What are the minimal conditions on $\theta(\cdot)$ to ensure that $\hat{\theta}_{K,b,T} \xrightarrow{p} \theta_1$?

ON CONSISTENCY

- Hölder-type continuity condition:

$$|\theta_\ell(t/T) - \theta_\ell(s/T)| \leqslant c_\ell \left(\frac{|t-s|}{T}\right)^\gamma, \quad t, s = 1, 2, \cdots, T,$$

  for each $\ell = 1, 2, \cdots, k$ where $0 < \gamma \leqslant 1$ and c_ℓ is a positive bounded constant.

Example

1. Abrupt structural change: $\theta_\ell(\cdot) = a_T \mathbb{1}_{\{t/T > e\}}$, where $e \in (0, 1]$ and $a_T = o(1)$ as $T \to \infty$;
2. Smooth structural change: $\theta_\ell(\cdot)$ is twice continuously differentiable;
3. Realization of persistent bounded stochastic processes: $\theta_{\ell,t} = \frac{1}{\sqrt{T}} v_t$, where $(1-L)^{d-1} v_t \overset{i.i.d.}{\sim} \mathcal{N}$.

ON CONSISTENCY

- Hölder-type continuity condition:

$$|\theta_\ell(t/T) - \theta_\ell(s/T)| \leqslant c_\ell \left(\frac{|t-s|}{T}\right)^\gamma, \quad t, s = 1, 2, \cdots, T,$$

for each $\ell = 1, 2, \cdots, k$ where $0 < \gamma \leqslant 1$ and c_ℓ is a positive bounded constant.

Example

1. Abrupt structural change: $\theta_\ell(\cdot) = a_T \mathbb{1}_{\{t/T>e\}}$, where $e \in (0, 1]$ and $a_T = o(1)$ as $T \to \infty$;
2. Smooth structural change: $\theta_\ell(\cdot)$ is twice continuously differentiable;
3. Realization of persistent bounded stochastic processes: $\theta_{\ell,t} = \frac{1}{\sqrt{T}} v_t$, where $(1-L)^{d-1} v_t \overset{i.i.d.}{\sim} \mathcal{N}$.

ON CONSISTENCY

- Hölder-type continuity condition:

$$|\theta_\ell(t/T) - \theta_\ell(s/T)| \leqslant c_\ell \left(\frac{|t-s|}{T}\right)^\gamma, \quad t, s = 1, 2, \cdots, T,$$

  for each $\ell = 1, 2, \cdots, k$ where $0 < \gamma \leqslant 1$ and c_ℓ is a positive bounded constant.

Example

1. Abrupt structural change: $\theta_\ell(\cdot) = a_T \mathbb{1}_{\{t/T>e\}}$, where $e \in (0, 1]$ and $a_T = o(1)$ as $T \to \infty$;
2. Smooth structural change: $\theta_\ell(\cdot)$ is twice continuously differentiable;
3. Realization of persistent bounded stochastic processes: $\theta_{\ell,t} = \frac{1}{\sqrt{T}} v_t$, where $(1-L)^{d-1} v_t \overset{i.i.d.}{\sim} \mathcal{N}$.

ON CONSISTENCY

- It can be shown that

$$\|\hat{\theta}_{K,b,T} - \theta_1\| = O_p\left((Tb)^{-1/2} + b^{\gamma}\right).$$

- Easier to estimate if γ is large.

OUT-OF-SAMPLE FORECASTING

END-OF-SAMPLE RISK

- Two inputs $\Rightarrow$ K and b
- End-of-sample risk:

$$E_T\left(\ell_{T+h}(\hat{\theta}_{K,b,T})\right) \approx R_T^1 + R_T^2 + R_T^3,$$

where

$$\begin{aligned}
R_T^1 &= E_T\left(\ell_{T+h}(\theta_1)\right) \\
R_T^2 &= E_T\left(\frac{\partial \ell_{T+h}(\theta_1)}{\partial \theta'}\right)(\hat{\theta}_{K,b,T} - \theta_1) \\
R_T^3 &= \frac{1}{2}(\hat{\theta}_{K,b,T} - \theta_1)' E_T\left(\frac{\partial^2 \ell_{T+h}(\overline{\theta}_1)}{\partial \theta \partial \theta'}\right)(\hat{\theta}_{K,b,T} - \theta_1),
\end{aligned}$$

and $\overline{\theta}_1$ lies between $\hat{\theta}_{K,b,T}$ and θ_1.

DECOMPOSITION

- R_T^1: does not involve parameter estimates
- R_T^2: drops out if $E_T\left(\frac{\partial \ell_{T+h}(\theta_1)}{\partial \theta'}\right) = 0$
  - ε_{t+h} are uncorrelated ▸ Back to example
- Minimizing the conditional expected loss is equivalent to minimize R_T^3.
- Define the regret risk Hirano and Wright (2017, ECTA):

$$R_T(K, b) = (\hat{\theta}_{K,b,T} - \theta_1)' \, E_T\left(\frac{\partial^2 \ell_{T+h}(\bar{\theta}_1)}{\partial \theta \partial \theta'}\right) (\hat{\theta}_{K,b,T} - \theta_1). \qquad (5)$$

SELECTION OF THE TUNING PARAMETER b

- Select b by minimizing R_T^3:

$$\hat{b} := \arg\min_{b \in I_T} \, (\hat{\theta}_{b,T} - \theta_1)' \, \omega_T(\overline{\theta}_1) \, (\hat{\theta}_{b,T} - \theta_1). \tag{6}$$

where $I_T = [\underline{b}, \overline{b}]$ is the candidate choice set of b.

Theorem

Under certain regularity conditions, the optimal tuning parameter $\hat{b}$ obtained by minimizing (6) is of order $T^{-\frac{1}{2\gamma+1}}$ in probability for some $0 < \gamma \leqslant 1$.

SELECTION OF THE TUNING PARAMETER b

- (6) is not feasible since it involves θ_1.
- If $\theta(\cdot)$ is twice continuously differentiable, we can approximate $\theta(1)$ by

$$\theta(t/T) \approx \theta + \theta' \left(\frac{t-T}{T}\right) + \frac{\theta''}{2}\left(\frac{t-T}{T}\right)^2, \tag{7}$$

where $\theta = \theta_1$, $\theta' = \theta_1^{(1)}$ and $\theta'' = \theta^{(2)}(c)$, where c lies between 1 and t/T.

▸ More on example

SELECTION OF THE TUNING PARAMETER b

- Then, the local-linear estimator is defined by the minimizer of

$$\min_{(\theta,\theta')\in\Theta\times\Theta'} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \ell_t\Big(\theta + \theta'(t/T - 1)\Big), \quad (8)$$

  where
  - $\tilde{k}_{tT} = K\left(\frac{t-T}{T\tilde{b}}\right)$;
  - $\tilde{b}$ is such that $\tilde{b} \to 0$ and $T\tilde{b} \to \infty$ as $T \to \infty$.

SELECTION OF THE TUNING PARAMETER b

- This leads to the following feasible selection criteria:

$$\hat{b} := \arg\min_{b \in I_T} (\hat{\theta}_{b,T} - \tilde{\theta}_T)' \, \omega_T\left(\tilde{\theta}_T\right) (\hat{\theta}_{b,T} - \tilde{\theta}_T). \tag{9}$$

  where
  - $\tilde{\theta}_T$: first $k \times 1$ elements of the minimizer of (8).

Theorem

Under certain regularity conditions, choosing $\hat{b}$ by (9) *is asymptotically optimal in the sense that*

$$(\hat{\theta}_{b,T} - \tilde{\theta}_T)' \, \omega_T\left(\tilde{\theta}_T\right) (\hat{\theta}_{b,T} - \tilde{\theta}_T) \asymp \inf_{b \in I_T} (\hat{\theta}_{b,T} - \theta_1)' \, \omega_T\left(\theta_1\right) (\hat{\theta}_{b,T} - \theta_1)$$

where $\tilde{\theta}_T$ is the local linear estimator from (8) *with tuning parameter $\tilde{b}$.*

IMPLICATION ON THE CHOICE OF WEIGHTING FUNCTION

- Typical choices of weighting function:

$$K_1(u) = \mathbb{1}_{\{-1<u<0\}}, \quad K_2(u) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \mathbb{1}_{\{u<0\}},$$
$$K_3(u) = \frac{3}{2}(1-u^2)\mathbb{1}_{\{-1<u<0\}}.$$

- All data are used for $K_2(u)$, but not for $K_1(u)$ and $K_3(u)$. ▸ G I
  - Kapetanios et al. (2019, JAE), Dendramis et al. (2020, JRSSa): find $K_2(u)$ is the best
  - Farmer et al. (2023, JF): recommend to use $K_3(u)$
- What types of weighting function shall we choose?

IMPLICATION ON THE CHOICE OF WEIGHTING FUNCTION

- (4) admits the following decomposition:

$$\hat{\theta}_{K,b,T} - \theta_1 = -H_{1,T}^{-1}(\theta_1)\Big(\underbrace{S_{1,T}}_{\text{variance}} + \underbrace{B_{2,T}}_{\text{bias}}\Big),$$

where

$$H_{1,T}(\theta_1) = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2 \ell_t\big(\theta_1\big)}{\partial\theta\partial\theta'}, \quad S_{1,T} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial \ell_t(\theta(t/T))}{\partial\theta},$$

$$B_{2,T} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2 \ell_t(\overline{\theta}_1)}{\partial\theta\partial\theta'}\big(\theta_1 - \theta(t/T)\big),$$

and $\overline{\theta}_1$ lies between $\hat{\theta}_{K,b,T}$ and θ_1.

(i) If $T^{1/2}b^{1/2+\gamma} \to 0$, we have

$$Tb \cdot R_T(K,b) \xrightarrow{d} \phi_{0,K}\Sigma_1^{1/2}Z'\omega_T(\theta_1)Z\Sigma_1^{1/2},$$

where $\phi_{0,K} = \int_{\mathcal{C}} K^2(u)du$, $Z \sim \mathcal{N}(0, I_k)$ and Σ_1 is defined as in Lemma C1;

(ii) If $T^{1/2}b^{1/2+\gamma} \to \infty$, we have

$$b^{-2\gamma} \cdot R_T(K,b) \xrightarrow{p} \mu_{\gamma,K}^2\mathcal{C}'\omega_T(\theta_1)\mathcal{C},$$

where $\mu_{\gamma,K} = \int u^{\gamma}K(u)du$ and $\mathcal{C} = (c_1, \cdots, c_{\bar{k}})'$ is a collection of Hölder constant;

IMPLICATION ON THE CHOICE OF WEIGHTING FUNCTION

(i)

(ii)

(iii) If $T^{1/2}b^{1/2} \asymp b^{-\gamma}$, we have

$$Tb \cdot \left(R_T(K, b) + b^{2\gamma} \mu_{\gamma,K}^2 \mathcal{C}' \omega_T(\theta_1) \mathcal{C} \right) \xrightarrow{d} \phi_{0,K} \Sigma_1^{1/2} Z' \omega_T(\theta_1) Z \Sigma_1^{1/2},$$

where $\mu_{\gamma,K}$, $\mathcal{C}$ and $\phi_{0,K}$ are defined as in (i) and (ii).

WHAT HAVE WE LEARNED?

- Reflects the usual bias-variance trade-off:
  - When variance dominates $T^{1/2}b^{1/2+\gamma} \to 0$, choose a weighting function which has smallest $\phi_{0,K}$;
  - Otherwise, $\mu_{\gamma,K}$ also plays a role.
- Assume $\gamma = 1$:
  - $\Rightarrow$ May fall into cases (ii) and (iii), but at the slowest rate

MONTE CARLO EXPERIMENTS

SUMMARY OF THE RESULTS

- We consider DGPs used in Pesaran and Timmermann (2007, JoE) and Inoue et al. (2017, JoE).
- Types of time variation include all considered in ▸ Example
- We find that our methods are useful: results are robust under various types of structural change.
- Using all data and downweighting them ($K_2(u)$) is generally preferred.

APPLICATION: BOND RETURN PREDICTABILITY

TARGET

- (log) Yield of an *n*-year bond:

$$y_t^{(n)} = -\frac{1}{n} p_t^{(n)},$$

  where
  - $p_t^{(n)}$ is the log price of the *n*-year zero-coupon bond at time *t*.
- Holding-period return:

$$r_{t+12}^{(n)} = p_{t+12}^{(n-1)} - p_t^{(n)}.$$

- The excess return is

$$rx_{t+12}^{(n)} = r_{t+12}^{(n)} - y_t^{(1)},$$

  where
  - $y_t^{(1)}$ is the one-year risk-free rate.

PREDICTIVE REGRESSIONS

(i) Fama-Bliss (FB) univariate

$$rx_{t+12}^{(n)} = \alpha + \beta\, fs_t^{(n)} + \varepsilon_{t+12};$$

(ii) Cochrane-Piazzesi (CP) univariate

$$rx_{t+12}^{(n)} = \alpha + \beta CP_t + \varepsilon_{t+12};$$

(iii) Fama-Bliss and Cochrane-Piazzesi predictors

$$rx_{t+12}^{(n)} = \alpha + \beta_1\, fs_t^{(n)} + \beta_2 CP_t + \varepsilon_{t+12}.$$

▸ More details

DATA

- Bond markets:
  - United States (Liu and Wu (2021, JFE))
  - Canada (Bank of Canada)
  - United Kingdom (Bank of England)
  - Japan (Ministry of Finance Japan)
- Sample period: 1986M1 – 2022M12
- Maturity up to 5 years
- $n = 2, 3, 4, 5$

FORECAST EVALUATION

- Benchmark: 3 PCs from global yield curve
- Starts from 2000M1
- MSE loss:

$$R(K,b) = (\hat{\theta}_{K,b,T} - \theta_1)' \left(X_T X_T'\right) (\hat{\theta}_{K,b,T} - \theta_1) \qquad (10)$$

- Set $b = cT^{-1/3}$, c ranges from 1 to 7 with a course grid of width 0.1
- $\tilde{b} = 1.06T^{-1/5}$
- Weighting functions:

$$K_1(u) = \mathbb{1}_{\{-1<u<0\}}, \quad K_2(u) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \mathbb{1}_{\{u<0\}},$$
$$K_3(u) = \frac{3}{2}(1-u^2) \mathbb{1}_{\{-1<u<0\}}.$$

RESULTS

‣ details

- VERY PROMISING: sizable and sometimes significant improvement over the benchmark forecasts
  - particularly when $K_3(u)$ is used with optimal tuning parameter selection
- Japan: $K_2(u)$ is better, but differences are small
- Non-local estimator ?
  - Not useful, particularly for Canada

CONCLUSION

CONCLUSION

- What types of time variation are allowable for using estimator like (3)?
  - A: Hölder-type continuity condition
- How to select the tuning parameter b optimally?
  - A: minimizing regret risk, asymptotic optimality
- Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?
  - A: No, properties of TVP, rolling window selection outperformed

CONCLUSION

- What types of time variation are allowable for using estimator like (3)?
  - A: Hölder-type continuity condition
- How to select the tuning parameter b optimally?
  - A: minimizing regret risk, asymptotic optimality
- Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?
  - A: No, properties of TVP, rolling window selection outperformed

CONCLUSION

- What types of time variation are allowable for using estimator like (3)?
  - A: Hölder-type continuity condition
- How to select the tuning parameter b optimally?
  - A: minimizing regret risk, asymptotic optimality
- Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?
  - A: No, properties of TVP, rolling window selection outperformed

CONCLUSION

- What types of time variation are allowable for using estimator like (3)?
  - A: Hölder-type continuity condition
- How to select the tuning parameter b optimally?
  - A: minimizing regret risk, asymptotic optimality
- Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?
  - A: No, properties of TVP, rolling window selection outperformed

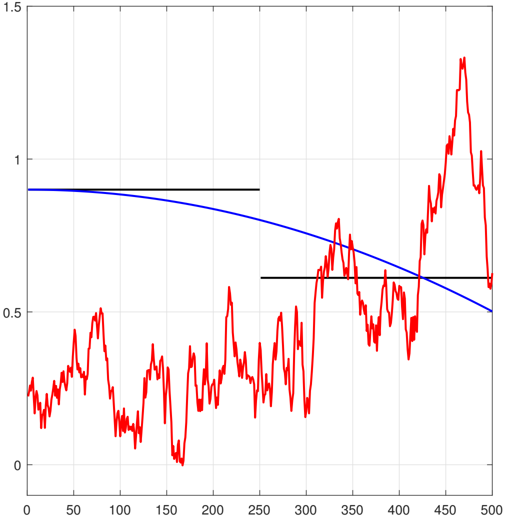
CONCLUSION

- What types of time variation are allowable for using estimator like (3)?
  - A: Hölder-type continuity condition
- How to select the tuning parameter b optimally?
  - A: minimizing regret risk, asymptotic optimality
- Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?
  - A: No, properties of TVP, rolling window selection outperformed

CONCLUSION

- What types of time variation are allowable for using estimator like (3)?
  - A: Hölder-type continuity condition
- How to select the tuning parameter b optimally?
  - A: minimizing regret risk, asymptotic optimality
- Is the weighting function $K(u) = \mathbb{1}_{\{-1<u<0\}}$ always the best choice?
  - A: No, properties of TVP, rolling window selection outperformed

APPENDIX SLIDES



Notes: black line: $\theta(t/T) = 0.9 - \frac{1}{T^{0.2}}\mathbb{1}(t \geqslant 0.5T + 1)$; blue line: $\theta(t/T) = 0.9 - 0.4(t/T)^2$; red line: $\theta(t/T)$ is a realization from the process $\frac{1}{\sqrt{T}}v_t$, where $\Delta v_t \overset{i.i.d.}{\sim} (0, 1)$.

- $\overline{\theta}_{\ell,t}$ satisfies:

$$|\overline{\theta}_{\ell,t} - \overline{\theta}_{\ell,s}| \leqslant \xi_{\ell,ts} \left(\frac{|t-s|}{T}\right)^{\gamma}$$

  where

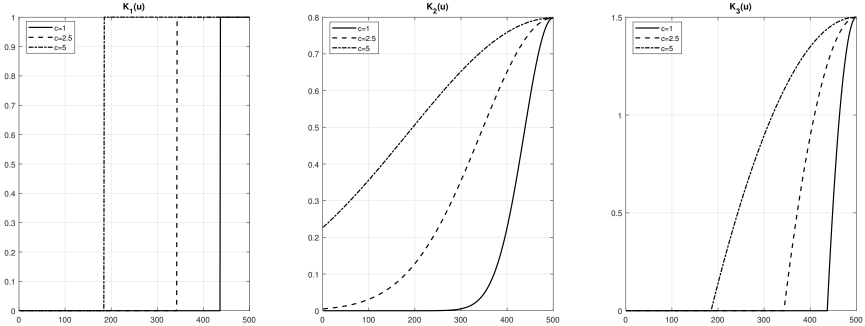
  - $\xi_{\ell,ts}$ has a thin-tailed distribution: $\mathbb{P}\left(|\xi_{\ell,ts}| > \omega\right) \leqslant \exp\left(-c_0|\omega|^{\alpha}\right)$, $\omega > 0$, for some $c_0 > 0$, $\alpha > 0$

Return to main slide

Example

suppose that $\theta(t/T)$ is a realization of a bounded random walk process: $\frac{1}{\sqrt{T}}v_t$, where $\Delta v_t \overset{i.i.d.}{\sim} \mathcal{N}$. Simple algebra gives $\theta(t/T) = \sqrt{\frac{t}{T}}\frac{1}{\sqrt{t}}v_t$. We know that $\frac{1}{\sqrt{t}}v_t = O_p(1)$, this implies that $\theta(t/T) = C_t\sqrt{\frac{t}{T}}$, where C_t is a positive bounded constant.

Return to main slide



Notes: Shape of the weighting function with $T = 500$, $b = cT^{-1/3}$ with c equal to 1,2.5 and 5.

▸ Return to main slide

- The Fama-Bliss (FB) forward spreads are given by

$$fs_t^{(n)} = f_t^{(n)} - y_t^{(1)} = p_t^{(n-1)} - p_t^{(n)} - y_t^{(1)}.$$

- The Cochrane-Piazzesi (CP) factor is constructed as the linear combination of forward rates:

$$CP_t = \hat{\gamma}' \boldsymbol{f}_t,$$

where

  - $\boldsymbol{f}_t = (y_t^{(1)}, f_t^{(2)}, f_t^{(3)}, f_t^{(4)}, f_t^{(5)})'$;
  - The coefficient vector $\hat{\gamma}$ is estimated from a predictive regression of $\frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)}$ on $[1\ \ \boldsymbol{f}_t']'$.

▸ Return to main slide

Table 1: Out-of-sample forecasting performance on bond returns: United States

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
			USA - 2 years						USA - 3 years		
PC-yields	1.592					PC-yields	6.046				
FB	1.047	1.150	1.103	0.958	0.852	FB	0.979	1.038	0.967	0.922	0.743
CP	1.113	1.122	0.949	1.005	0.744	CP	1.106	1.075	0.899	0.965	0.705
FB+CP	1.107	0.964	0.876	0.919	0.652	FB+CP	1.116	0.903	0.780	0.882	0.578*
			USA - 4 years						USA - 5 years		
PC-yields	11.836					PC-yields	18.670				
FB	0.960	0.943	0.884	0.905	0.709	FB	0.941	0.872	0.875	0.900	0.707
CP	1.101	1.037	0.863	0.943	0.708*	CP	1.099	1.025	0.862	0.941	0.738*
FB+CP	1.099	0.778	0.694	0.841	0.518*	FB+CP	1.075	0.751	0.693	0.861	0.535*

Table 2: Out-of-sample forecasting performance on bond returns: Canada

	Non-local	R = 60	Opt-R	Opt-G	Opt-E		Non-local	R = 60	Opt-R	Opt-G	Opt-E
	Canada - 2 years						Canada - 3 years				
PC-yields	1.171					PC-yields	3.534				
FB	1.011	0.920	0.953	0.826	0.726	FB	1.029	0.868	0.905	0.859	0.706
CP	1.051	0.888	0.908	0.809	0.744	CP	1.094	0.907	0.898	0.852	0.757
FB+CP	1.034	0.861	0.931	0.798	0.687	FB+CP	1.096	0.813	0.852	0.826	0.642
	Canada - 4 years						Canada - 5 years				
PC-yields	6.545					PC-yields	10.133				
FB	1.033	0.860	0.887	0.892	0.707	FB	1.032	0.873	0.899	0.929	0.730
CP	1.129	0.911	0.859	0.882	0.758	CP	1.165	0.931	0.864	0.914	0.781
FB+CP	1.137	0.822	0.847	0.861	0.661	FB+CP	1.149	0.843	0.867	0.895	0.682

Table 3: Out-of-sample forecasting performance on bond returns: UK

	Non-local	R = 60	Opt-R	Opt-G	Opt-E		Non-local	R = 60	Opt-R	Opt-G	Opt-E
	UK - 2 years						UK - 3 years				
PC-yields	1.415					PC-yields	4.378				
FB	0.807	0.821	0.907	0.790	0.648	FB	0.897	0.897	1.057	0.866	0.769
CP	0.923	0.764	0.704	0.646	0.593	CP	1.041	0.839	0.769	0.729	0.650
FB+CP	0.921	0.688	0.669	0.645	0.514	FB+CP	1.050	0.751	0.745	0.724	0.591
	UK - 4 years						UK - 5 years				
PC-yields	8.224					PC-yields	12.962				
FB	0.949	0.942	1.042	0.897	0.884	FB	0.980	0.983	1.028	0.923	0.936
CP	1.087	0.884	0.811	0.782	0.691*	CP	1.097	0.916	0.850	0.813	0.727
FB+CP	1.092	0.797	0.780	0.770	0.638	FB+CP	1.075	0.835	0.811	0.789	0.669

Table 4: Out-of-sample forecasting performance on bond returns: Japan

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	Japan - 2 years						Japan - 3 years				
PC-yields	0.333					PC-yields	1.146				
FB	0.222	0.105*	0.115*	0.099*	0.097*	FB	0.244	0.148*	0.167*	0.145*	0.150*
CP	0.582	0.102*	0.112*	0.093*	0.098*	CP	0.677	0.155*	0.164*	0.140*	0.144*
FB+CP	0.610	0.101*	0.134*	0.091*	0.094*	FB+CP	0.679	0.149*	0.179*	0.140*	0.141*
	Japan - 4 years						Japan - 5 years				
PC-yields	2.517					PC-yields	4.050				
FB	0.246	0.197*	0.186*	0.186*	0.165*	FB	0.291	0.267*	0.243*	0.247*	0.219*
CP	0.817	0.181*	0.182*	0.162*	0.160*	CP	0.902	0.220*	0.223*	0.196*	0.190*
FB+CP	0.772	0.186*	0.189*	0.162*	0.168*	FB+CP	0.871	0.182*	0.187*	0.167*	0.169*

‣ Return to main slide