Robust Bond Risk Premia

Michael D. Bauer[1] James D. Hamilton[2]

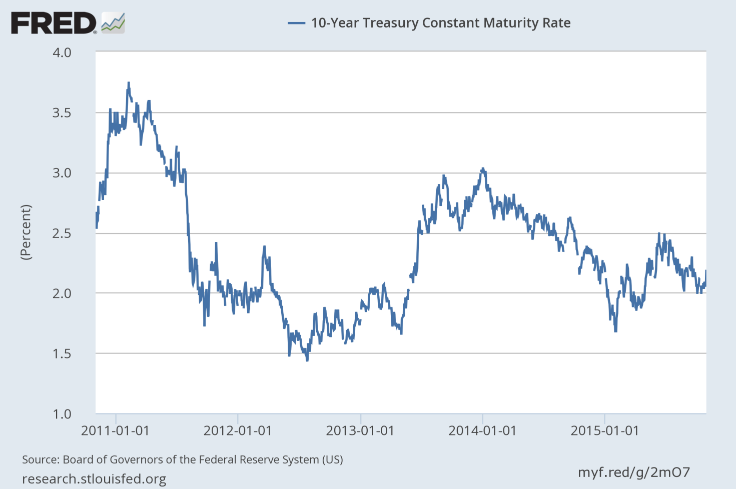
[1]Federal Reserve Bank of San Francisco

[2]University of California, San Diego

November 5, 2015

FRBSF-BoC Conference on Fixed Income Markets

Is 10-year yield around 2% the new normal?



Understanding long-term interest rates

Long-term rate = expected short-term rates + term premium

- Are expected future rates only 2%?
  - Real rate near zero for a decade?
  - Fed won't hit its 2% inflation target?
- Or is it the term premium?
  - LSAP produced negative term premium?
  - Flight to safety?

Can distinguish expectation component from term premium if we have correct model to forecast interest rates.

What variables predict interest rates and bond returns?

- Yield on any security at time t is a function of state vector z_t.
- Under standard assumptions (e.g., Duffee, 2013) we should be able to back out z_t from yields.
- Three principal components (level, slope, and curvature) summarize almost all information in the cross-section of the yield curve.

Spanning hypothesis

Level, slope, and curvature are all that are needed to predict bond yields and excess returns.

- This is much weaker than expectations hypothesis.

Evidence against spanning hypothesis

Several recent studies find that variables *in addition to level/slope/curvature* help predict future bond returns.

Study	**Proposed predictors**
Joslin, Priebsch and Singleton (2014)	inflation and output
Ludvigson and Ng (2009, 2010)	factors from macro data sets
Cochrane and Piazzesi (2005)	4th and 5th PC
Greenwood and Vayanos (2014)	maturity structure of Treasury debt
Cooper and Priestley (2008)	output gap

Predictive regressions

Evidence in these studies comes from regressions of common form:

y_{t+h} = yield or bond return
x_{1t} = summary of yield curve
x_{2t} = proposed predictors

$$y_{t+h} = \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+h}$$

$$H_0: \quad \beta_2 = 0$$

Studies find:

- big increase in R^2 when x_{2t} added to regression
- very low p-value for test of H_0

Our paper

- We document serious small-sample problems caused by serially correlated predictors and correlation between x_{1t} and lagged u_{t+h}.
- We revisit the evidence in these studies and find z_t only needs to include level and slope of the yield curve.

Econometrics of testing the spanning hypothesis

$$y_{t+h} = \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+h}$$

Two problems have not previously been recognized:

1. Spurious increase in R^2 when x_{2t} added
   - Overlapping returns ($h > 1$) and persistent x_{2t} increase small-sample mean and variance of ΔR^2 even though $\beta_2 = 0$
2. "Standard error bias" if x_{1t} is not strictly exogenous
   - HAC standard errors too small, so conventional tests of $\beta_2 = 0$ reject too often
   - Separate issue from "Stambaugh bias" in $\hat{\beta}_1$

Source of standard error bias

$$y_{t+h} = x_{1t}'\beta_1 + x_{2t}'\beta_2 + u_{t+h}$$

OLS estimate $\hat{\beta}_2$ could be obtained as follows:

1. Regress x_{2t} on x_{1t}
2. Regress y_{t+h} on x_{1t}
3. Regress residuals $\tilde{y}_{t+h}$ on residuals $\tilde{x}_{2t}$.

- Under usual asymptotics the intermediate regression (1) is irrelevant
- But if regressors are highly persistent (1) is like a spurious regression and residuals $\tilde{x}_{2t}$ differ significantly from true x_{2t}

Simple example

$$x_{1t} \text{ and } x_{2t} \text{ scalars}$$

$$y_{t+1} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_{t+1}$$

$$x_{i,t+1} = \rho_i x_{it} + \varepsilon_{i,t+1} \qquad \rho_1, \rho_2 \text{ near } 1$$

$$\beta_1 = \rho_1, \quad \beta_0 = \beta_2 = 0$$

$$E\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ u_t \end{bmatrix}\begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} & u_t \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \delta\sigma_1\sigma_u \\ 0 & \sigma_2^2 & 0 \\ \delta\sigma_1\sigma_u & 0 & \sigma_u^2 \end{bmatrix}$$

- If $\delta \neq 0$ then x_{1t} is not strictly exogenous.

t-test under local-to-unity asymptotics

- Asymptotic distribution of t-statistic:

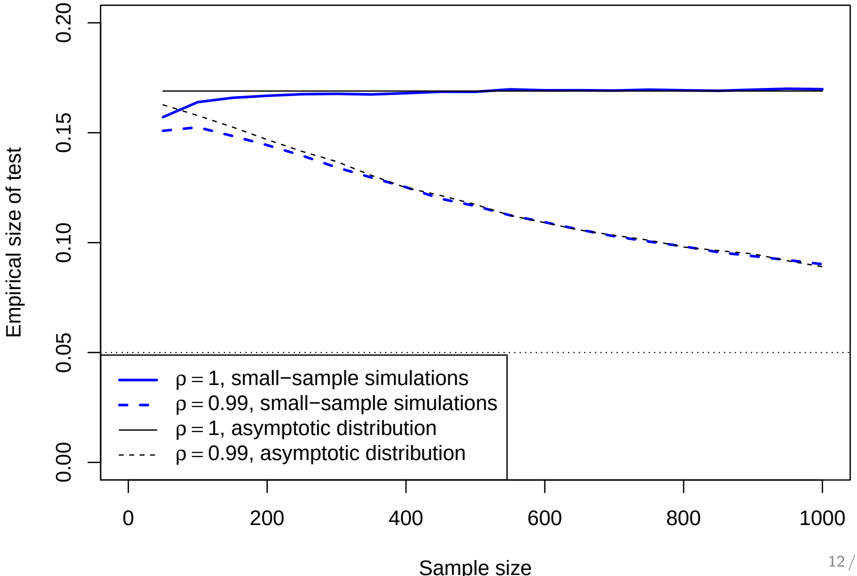
$$\tau = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} \xrightarrow{d} \delta Z_1 + \sqrt{1-\delta^2} Z_0$$

$$Z_0 \sim N(0,1), \quad \mathrm{E}(Z_1) = 0,\ \mathrm{Var}(Z_1) > 1, \quad \mathrm{Cov}(Z_0, Z_1) = 0$$

- t-test rejects too often when $\delta \neq 0$
- Problem would arise even if we knew the population value of the asymptotic variance that HAC methods try to estimate

Small-sample distribution vs. local-to-unity approximation

True size of t-test of $\beta_2 = 0$ with nominal size of 5%. DGP: $\delta = 1$



Warning flags

- Size distortions are large when
  - Correlation with lagged errors (δ) is strong
  - Persistence of x_{1t} and x_{2t} is high
  - Samples are small
- All these conditions arise in predictive regressions for yields or bond returns.

Recommendation: bootstrap procedure to gauge magnitude of potential size distortions

1. Extract three principal components of yields

$$x_{1t} = (PC1_t, PC2_t, PC3_t)'$$

$$i_{nt} = \hat{h}_n' x_{1t} + \hat{v}_{nt}$$

2. Estimate VAR for PCs

$$x_{1t} = \hat{\mu} + \hat{\phi} x_{1,t-1} + e_{1t}$$

3. Estimate VAR for proposed predictors

$$x_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1} + e_{2t}$$

4. Generate bootstrap sample $\{x_{1t}^*, x_{2t}^*\}_{t=1}^T$ from estimated VARs
   - Resample (e_{1t}^*, e_{2t}^*) jointly from VAR residuals (e_{1t}, e_{2t})
5. Generate artificial yield for security n from

$$i_{nt}^* = \hat{h}_n' x_{1t}^* + v_{nt}^* \quad v_{nt}^* \sim N(0, \sigma_v^2)$$

6. Calculate statistics of interest on the simulated data.
   - For example, regress excess bond return $rx_{n,t+h}^*$ on x_{1t}^* and x_{2t}^* and calculate Wald-test for $\beta_2 = 0$.

Features of our bootstrap procedure

- Delivers artificial data set with similar correlations and serial dependence as original but in which the spanning hypothesis holds by construction:

$$\mathsf{E}(y^*_{n,t+h}|x^*_{1t}, x^*_{2t}) = \mathsf{E}(y^*_{n,t+h}|x^*_{1t})$$

- Provides small-sample distribution of test statistics under H_0
- Designed to test *spanning* hypothesis
  - Previous studies used bootstrap to test *expectations* hypothesis

Alternative approach: Ibragimov and Müller (2010)

1. Divide original sample into say $q = 8$ subsamples
2. Estimate β_2 separately across each subsample
3. Calculate a t-test with q degrees of freedom from variation of b_{2i} across subsamples.

- Gets around “standard error bias”
- Simulation evidence shows excellent size and power properties
- Also shows whether results are robust across subsamples

Application 1: Joslin, Priebsch and Singleton (2014)

- Regressions of yields and returns on 3 yield PCs (x_{1t}) and measure of economic growth and inflation (x_{2t}).
- Found evidence for *unspanned macro risks*
- Warning flags
  - Autocorrelations are 0.91 for growth and 0.99 for inflation
  - 276 monthly observations (1985–2007)
  - Correlation between level and lagged forecast error is -0.37 (returns are low when level of yields is high)

JPS: predicting annual excess bond returns

		$\bar{R}_1^2$	$\bar{R}_2^2$	$\bar{R}_2^2 - \bar{R}_1^2$
Two-year	Data	0.14	0.49	0.35
bond	Simple bootstrap	0.30	0.36	0.06
		(0.06, 0.58)	(0.11, 0.63)	(-0.00, 0.22)
	BC bootstrap	0.38	0.44	0.06
		(0.07, 0.72)	(0.13, 0.75)	(-0.00, 0.23)
Ten-year	Data	0.20	0.37	0.17
bond	Simple bootstrap	0.26	0.32	0.07
		(0.07, 0.48)	(0.12, 0.54)	(-0.00, 0.23)
	BC bootstrap	0.27	0.34	0.08
		(0.06, 0.50)	(0.12, 0.57)	(-0.00, 0.27)
Average	Data	0.19	0.39	0.20
two- through	Simple bootstrap	0.28	0.35	0.07
ten-year		(0.08, 0.50)	(0.12, 0.56)	(-0.00, 0.23)
bonds	BC bootstrap	0.30	0.37	0.07
		(0.06, 0.55)	(0.13, 0.61)	(-0.00, 0.26)

JPS: predicting the level of the yield curve

	$PC1$	$PC2$	$PC3$	GRO	INF	Wald
Coefficient	0.928	-0.013	-0.097	0.092	0.118	
HAC statistic	40.965	1.201	0.576	2.376	2.357	14.873
HAC p-value	0.000	0.231	0.565	0.018	0.019	0.001
Simple bootstrap 5% c.v.				2.349	2.744	10.306
Simple bootstrap p-value				0.048	0.097	0.016
BC bootstrap 5% c.v.				2.448	2.985	12.042
BC bootstrap p-value				0.058	0.129	0.026
IM $q = 8$	0.000	0.864	0.436	0.339	0.456	
IM $q = 16$	0.000	0.709	0.752	0.153	0.554	
Estimated size of tests						
HAC				0.105	0.163	0.184
Simple bootstrap				0.047	0.066	0.057
IM $q = 8$				0.047	0.050	
IM $q = 16$				0.057	0.058	

JPS results when later data added

- JPS original sample: 1985-2008
- If we use instead 1985-2013:
  - Increases in $\bar{R}^2$ are smaller and squarely within bootstrap confidence intervals.
  - Coefficient on growth is not significant.
  - Coefficient on inflation has p-value of 0.042 using HAC standard errors but 0.125 using (simple) bootstrap.

Application 2: Ludvigson and Ng (2010)

- Studied predictive power of *macro factors* for bond returns
  - Macro factors are the first 8 PCs of 131 macro variables
- Selection of macro factors
  - They preselect factors and include squared and cubed terms.
  - We leave aside this specification search—use all 8 factors.
  - This simplifies things but results are similar in both cases.
- Controlling for information in the yield curve
  - They used Cochrane-Piazzesi factor.
  - We use level, slope and curvature instead.
- Original sample: 1964–2007

Ludvingson-Ng: predicting excess returns

	$PC1$	$PC2$	$PC3$	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$	$F8$
Coefficient	0.136	2.052	-5.014	0.742	0.146	-0.072	-0.528	-0.321	-0.576	-0.401	0.551
HAC statistic	1.552	2.595	2.724	1.855	0.379	0.608	1.912	1.307	2.220	2.361	3.036
HAC p-value	0.121	0.010	0.007	0.064	0.705	0.543	0.056	0.192	0.027	0.019	0.003
Bootstrap 5% c.v.				2.572	2.580	2.241	2.513	2.497	2.622	2.446	2.242
Bootstrap p-value				0.140	0.761	0.594	0.128	0.301	0.092	0.057	0.010
IM $q = 8$	0.001	0.001	0.225	0.098	0.558	0.579	0.088	0.703	0.496	0.085	0.324
IM $q = 16$	0.000	0.052	0.813	0.228	0.317	0.771	0.327	0.358	0.209	0.027	0.502
Estimated size of tests											
HAC				0.131	0.132	0.097	0.124	0.126	0.134	0.113	0.086
Bootstrap				0.058	0.055	0.053	0.061	0.055	0.053	0.049	0.046
IM $q = 8$				0.051	0.050	0.051	0.049	0.049	0.052	0.050	0.042
IM $q = 16$				0.051	0.048	0.051	0.050	0.051	0.045	0.055	0.046

- Wald-test of $\beta_2 = 0$
  - HAC p-value is 0.000, bootstrap p-value is 0.009
  - True size of 5% Wald test is 33.5%
- Regresion fit: $\bar{R}^2$
  - Increases from 0.25 to 0.35 when adding macro factors
  - But this increase is within bootstrap confidence interval

Return-forecasting factors

- Ludvigson and Ng also construct a “return-forecasting factor” from the original 8 macro factors to get an optimal predictor of interest rates.
- We use our bootstrap to examine the small-sample properties of this procedure.
- Here we do exactly what they did—same point estimates and HAC *p*-values.

Ludvigson-Ng return forecasting factor H8

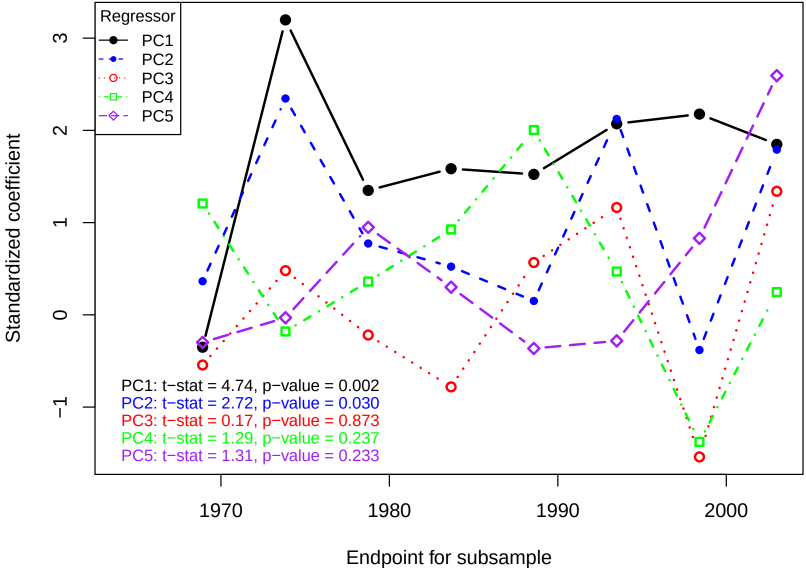
	Two years		Three years		Four years		Five years	
	CP	*H8*	*CP*	*H8*	*CP*	*H8*	*CP*	*H8*
Coefficient	0.335	0.331	0.645	0.588	0.955	0.776	1.115	0.937
HAC t-statistic	4.429	4.331	4.666	4.491	4.765	4.472	4.371	4.541
HAC p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bootstrap 5% c.v.		3.809		3.799		3.874		3.898
Bootstrap p-value		0.022		0.015		0.017		0.014
Estimated size of tests								
HAC		0.514		0.538		0.545		0.539
Bootstrap		0.047		0.055		0.057		0.050

- Increases in $\bar{R}^2$ are within bootstrap confidence intervals (except for the two-year bond)
- Results for later sample (1985–2007):
  Macro factors (and H8) have no significant predictive power

Application 3: Cochrane and Piazzesi (2005)

- Found that tent-shaped linear combination of forward rates—their "return-forecasting factor"—strongly predicts excess bond returns
- Also showed evidence that return-forecasting factor is *not* spanned by level, slope, and curvature
- We find:
  - Standard error bias cannot account for CP's findings.
  - But IM test fails to reject H_0
  - Reason: predictive power of PC4 and PC5 is highly sensitive to sample choice.

Standardized coefficients on principal components across 8 different subsamples for CP original data set



Other applications

Cooper and Priestley (2008)

Output gap appears to predict excess bond returns

- Did not accurately control for information in the yield curve (include orthogonalized CP factor)
- Apparently did not use appropriate HAC standard errors
- We find that the output gap has no incremental predictive power for bond returns.

Greenwood and Vayanos (2014)

Maturity composition of Treasury debt appears to predict return on long-term bond.

- But even using conventional HAC, p-value rises to 0.06 when level, slope and curvature added to regression.

Summary of contributions (econometrics)

- **We already knew:** if x_{1t} is highly persistent and not strictly exogenous, $\hat{\beta}_1$ is biased and hypothesis tests about β_1 are problematic (Mankiw and Shapiro, 1986; Stambaugh, 1999; Campbell and Yogo, 2006).
- **Our paper shows:** this is also a problem for inference about β_2 due to "standard error bias"
- **Warning flags:** lagged dependent variables, persistent regressors, small sample size—exactly the situation faced when predicting yields or bond returns.

Summary of contributions (finance)

- **We already knew:** expectations hypothesis is violated (Fama and Bliss, 1987; Campbell and Shiller, 1991).
- **Our paper confirms:** level and slope of yield curve are robust predictors of returns.
- **We thought we knew:** macro and other variables also help predict returns (Joslin, Priebsch, Singleton, 2014; Ludvigson and Ng, 2009, 2010; Cochrane and Piazzesi, 2005; Greenwood and Vayanos, 2014;, Cooper and Priestley, 2008).
- **Our paper concludes:** level and slope are all that is needed; there is no robust evidence against the spanning hypothesis.