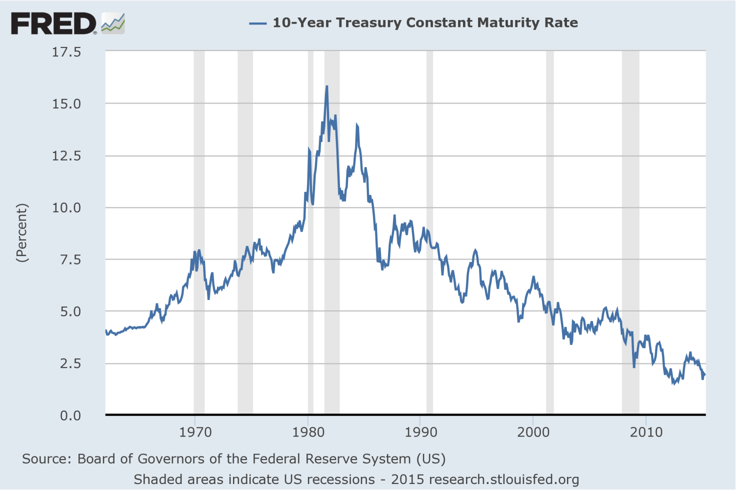
Robust Bond Risk Premia

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

[1]Federal Reserve Bank of San Francisco

[2]University of California, San Diego

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



long rate = expected future short rates + term premium

Could be low because short rates expected to stay low ...

- real rate near zero for a decade?
- Fed won't hit its 2% inflation target?

... or because term premium is low or negative

- flight to safety?
- effect of large-scale asset purchases by central banks?

Answer is critical for

- monetary policy
- investment decisions
- understanding lasting consequences of 2008 downturn

long rate $=$ expected future short rates $+$ term premium

In principle can measure if we have correct model of expected future short rates.

But what is information set on which these expectations are based?

Yield on any security at time t should be function of state vector z_t.

Under standard assumptions (e.g., Duffee, 2013) we would be able to back out z_t if we know the full yield curve at date t.

Three factors (level, slope, and curvature) summarize almost all information in yield curve.

"Spanning hypothesis": level, slope, and curvature are all that is needed to predict bond yields and excess returns.

This is weaker than expectations hypothesis.

Recent studies reporting evidence against spanning hypothesis

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

Evidence in all these studies comes from regressions of a 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

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

Our paper:

These studies did not adequately allow for small-sample implications of high persistence of x_{1t} and x_{2t}.

Once small-sample consequences of persistence are accounted for, evidence against spanning hypothesis is much less convincing.

Observation 1: serial correlation influences small-sample R^2

The more the serial correlation in x_{2t} and u_{t+h}, the more the R^2 will increase when x_{2t} is added to the regression, even if x_{2t} is completely independent of u_{t+h} at all leads and lags.

$$T(R_2^2 - R_1^2) \xrightarrow{d} r'Q^{-1}r/\gamma$$

$$\gamma = E[y_t - E(y_t)]^2$$

$$r \sim N(0, S)$$

$$Q = E(x_{2t}x_{2t}')$$

$$S = \sum_{v=-\infty}^{\infty} E(u_{t+h}u_{t+h-v}x_{2t}x_{2,t-v}')$$

$$E(u_{t+h}u_{t+h-v}x_{2t}x_{2,t-v}') = E(u_t u_{t-v})E(x_{2t}x_{2,t-v}') \neq 0$$

Overlapping returns induce serial correlation in u_t and give change in R^2 higher mean and variance

Observation 2: Lack of strict exogeneity of x_{1t} can affect small-sample properties of hypothesis tests

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

Conventional HAC is poor approximation to small-sample distribution when

- x_{1t} is correlated with lagged values of u_{t+h}
- x_{1t} and x_{2t} are highly persistent

This result does not depend on serial correlation of u_{t+h}

When H_0 is true, OLS coefficient can be written

$$b_2 = \left(\sum_{t=1}^{T} \tilde{x}_{2t} \tilde{x}_{2t}' \right)^{-1} \left(\sum_{t=1}^{T} \tilde{x}_{2t} u_{t+h} \right)$$

$\tilde{x}_{2t} =$ residuals from OLS regressions of x_{2t} on x_{1t}:

$$\tilde{x}_{2t} = x_{2t} - A_T x_{1t}$$

$$A_T = \left(\sum_{t=1}^{T} x_{2t} x_{1t}' \right) \left(\sum_{t=1}^{T} x_{1t} x_{1t}' \right)^{-1} .$$

Under conventional stationary asymptotics
$A_T \xrightarrow{p} 0$
b_2 has the same asymptotic distribution as

$$b_2^* = \left(\sum_{t=1}^{T} x_{2t} x_{2t}' \right)^{-1} \left(\sum_{t=1}^{T} x_{2t} u_{t+h} \right),$$

$$\sqrt{T} b_2 \xrightarrow{d} N(0, Q^{-1} S Q^{-1})$$

But when x_{1t} and x_{2t} are highly persistent, the regression of x_{2t} on x_{1t} behaves more like a spurious regression and A_T converges more slowly to zero.

When x_{1t} is correlated with lagged u_{t+h}, this effect is also correlated with $\sum \tilde{x}_{2t} u_{t+h}$ and causes variance of b_2 to be bigger than $Q^{-1} S Q^{-1}$.

Result: HAC correction based on conventional asymptotics understates the small-sample standard error of b_2.

Simple example

x_{1t} and x_{2t} scalars

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

$$x_{1,t+1} = \rho_1 x_{1t} + \varepsilon_{1,t+1}$$

$$x_{2,t+1} = \rho_2 x_{2t} + \varepsilon_{2,t+1}$$

ρ_1, ρ_2 near 1

$$\beta_2 = 0$$

$$x_{1t} = y_t$$

$\varepsilon_{1t}, \varepsilon_{2t}$ completely independent

Differs from Stambaugh bias

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

$$x_{1,t+1} = \rho_1 x_{1t} + \varepsilon_{1,t+1}$$

$$\rho_1 \text{ near } 1$$

$$x_{1t} \text{ correlated with } u_t$$

Stambaugh: weak exogeneity of x_{1t} makes b_1 biased confounding inference about β_1.

Our setting: x_{2t} is strictly exogenous but lack of strict exogeneity of x_{1t} confounds interence about β_2.

Can study effects of ρ_i near one using local-to-unity asymptotics

$$x_{i,t+1} = \rho_i x_{it} + \varepsilon_{i,t+1}$$

$$x_{i,t+1} = (1 + c_i/T)x_{it} + \varepsilon_{i,t+1}$$

E.g., approximate $\rho_i = 0.95$ and $T = 100$ with $c_i = -5$ and $T \to \infty$

$$\frac{b_2}{\hat{\sigma}_{b_2}} \xrightarrow{d}$$

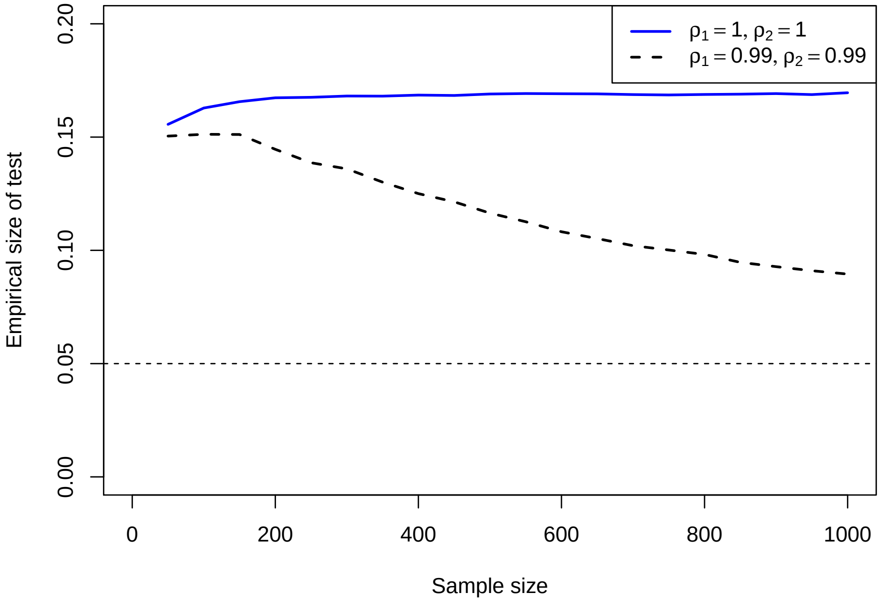
$$\frac{\int J_{c_2}(\lambda)dW_1(\lambda) - \left[\int J_{c_2}(\lambda)J_{c_1}(\lambda)d\lambda\right]\left[\int [J_{c_1}(\lambda)]^2 d\lambda\right]^{-1}\left[\int J_{c_1}(\lambda)dW_1(\lambda)\right]}{\left\{\int [J_{c_2}(\lambda)]^2 d\lambda - \left[\int J_{c_2}(\lambda)J_{c_1}(\lambda)d\lambda\right]^2\left[\int [J_{c_1}(\lambda)]^2 d\lambda\right]^{-1}\right\}^{1/2}}$$

Denominator strictly smaller than

$$\left\{\int [J_{c_2}(\lambda)]^2 d\lambda\right\}^{1/2}$$

t-statistic is larger than predicted by asymptotic distribution

Exact small-sample distributions based on simulation



Problem is *not*:

- biased estimation of b_2
- inconsistent estimation of $Q^{-1}SQ^{-1}$

Instead, problem is that true small-sample variance is bigger than $Q^{-1}SQ^{-1}$ (= "standard error bias")

A bootstrap procedure to calculate exact small-sample distribution

(1) Extract first three principal components of yields along with weight for yield n on components:

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

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

(2) Estimate VAR for components

$$x_{1t} = \hat{\mu} + \hat{\phi}_1 x_{1,t-1} + \hat{\phi}_2 x_{1,t-2} + \cdots + \hat{\phi}_{12} x_{1,t-12} + e_{1t}$$

(3) Generate artificial sample of $\{x_{11}^*, ..., x_{1T}^*\}$ from estimated VAR

(4) Generate artificial yield for security n from

$$y^*_{n\tau} = \hat{h}'_n x^*_{1\tau} + v^*_{n\tau} \quad v^*_{n\tau} \sim N(0, \sigma_v^2)$$

(5) Estimate VAR for proposed predictors

$$x_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1} + \hat{\alpha}_2 x_{2,t-2} + \cdots + \hat{\alpha}_p x_{2,t-p} + e_{2t}$$

(6) Generate artificial sample of $\{x^*_{21}, ..., x^*_{2T}\}$ from estimated VAR

- Note this is completely independent of $\{x^*_{11}, ..., x^*_{1T}\}$

(7) Calculate properties of any statistic of interest on the simulated data.

Advantage: Have artificial data set with same correlations and serial dependence as original but in which the spanning hypothesis holds by construction.

Since estimated VAR may underestimate true persistence, might get more accurate indication of size of problem by using bias-corrected VAR coefficients for the bootstrap simulation.

Robust HAC inference

Most HAC corrections do not address the issue.

Ibragimov and Müller (2010):

(1) Divide original sample into say $q = 8$ subsamples

(2) Estimate β_2 separately across each subsample

(3) If approximately independent and Normal, can calculate a t-test with q degrees of freedom from variation of b_{2i} across subsamples.

Application 1: Joslin, Priebsch and Singleton (2014)

Regressions of yields and returns on x_{1t} and measure of economic growth and inflation.

Estimated a structural no-arbitrage model based on these forecasting relations.

JPS: predicting 12-month holding yield on ten-year bond

	Ten-year bond		
	$\bar{R}_1^2$	$\bar{R}_2^2$	$\bar{R}_2^2 - \bar{R}_1^2$
Original sample: 1985–2008			
Data	0.20	0.37	0.17
Simple bootstrap	0.26	0.33	0.06
	(0.07, 0.48)	(0.12, 0.55)	(0.00, 0.25)
BC bootstrap	0.24	0.32	0.08
	(0.04, 0.50)	(0.09, 0.57)	(0.00, 0.29)

$\bar{R}_2^2$ decreases when more data are added

	Ten-year bond		
	$\bar{R}_1^2$	$\bar{R}_2^2$	$\bar{R}_2^2 - \bar{R}_1^2$
Original sample: 1985–2008			
Data	0.20	0.37	0.17
Simple bootstrap	0.26	0.33	0.06
	(0.07, 0.48)	(0.12, 0.55)	(0.00, 0.25)
BC bootstrap	0.24	0.32	0.08
	(0.04, 0.50)	(0.09, 0.57)	(0.00, 0.29)
Later sample 1985–2013			
Data	0.20	0.28	0.08
Simple bootstrap	0.22	0.28	0.06
	(0.03, 0.46)	(0.08, 0.51)	(0.00, 0.20)
BC bootstrap	0.24	0.30	0.06
	(0.03, 0.50)	(0.07, 0.54)	(0.00, 0.21)

In original sample $\bar{R}^2$ for predicting holding yield on two-year bond goes from 14% to 49% when x_{2t} is added.

Bias-corrected p-value for getting such a big increase is 0.034 (reject spanning).

But $\bar{R}_2^2$ drops to 28% when more data are included and p-value rises to 0.25.

JPS: predicting the level of the yield curve

	$PC1$	$PC2$	$PC3$	GRO	INF
Original sample: 1985–2008					
Coefficient	0.928	-0.013	-0.097	0.092	0.118
HAC statistic	41.205	1.312	0.508	2.214	2.400
HAC p-value	0.000	0.191	0.612	0.028	0.017
Simple bootstrap 5% c.v.'s				2.608	2.829
Simple bootstrap p-values				0.090	0.099
Simple bootstrap true size				0.120	0.171
BC bootstrap 5% c.v.'s				2.926	3.337
BC bootstrap p-values				**0.127**	**0.145**
BC bootstrap true size				0.159	0.224
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

In original sample, HAC Wald test of joint hypothesis that coefficients on GRO and INF are both zero has bootstrap bias-corrected *p*-value of 0.052.

In later sample, *p*-value is 0.265.

Application 2: Ludvigson and Ng (2010)

Looked at functions of up to 8 principal components in a data set of 131 macro variables for predicting excess bond returns.

They typically used Cochrane-Piazzesi factor to control for information in the yield curve.

We use level, slope and curvature instead.

Ludvingson Ng: predicting return on 2-year bond

	$PC1$	$PC2$	$PC3$	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$	$F8$
Original sample: 1964–2007											
Coefficient	-0.071	-0.973	2.825	0.471	-0.008	-0.085	-0.346	-0.083	-0.209	-0.133	0.254
HAC statistic	1.797	2.640	3.515	2.350	0.043	1.442	2.652	0.673	1.698	1.675	2.888
HAC p-value	0.073	0.009	0.000	0.019	0.966	0.150	0.008	0.501	0.090	0.095	0.004
Bootstrap 5% c.v.'s				2.703	2.632	2.256	2.736	2.790	2.859	2.494	2.430
Bootstrap p-values				0.092	0.974	0.218	0.060	0.607	0.221	0.194	0.026
Bootstrap true size				0.153	0.121	0.088	0.150	0.145	0.157	0.116	0.112
IM $q = 8$	0.002	0.007	0.356	0.052	0.404	0.217	0.007	0.526	0.545	0.177	0.241
IM $q = 16$	0.000	0.229	0.021	0.016	0.290	0.793	0.137	0.629	0.248	0.034	0.426

Ludvigson and Ng also construct a "return-forecasting factor" out of the original 8 macro factors to get an optimal predictor of interest rates.

We can perform identical calculations on our artificial samples to examine small-sample properties of this procedure.

Ludvigson-Ng return forecasting factor H8

	CP	*H8*	$\bar{R}_1^2$	$\bar{R}_2^2$	$\bar{R}_2^2 - \bar{R}_1^2$
Original sample: 1964–2007					
Data	0.335	0.331	0.31	0.42	0.11
HAC *t*-statistic	4.429	4.331			
HAC *p*-value	0.000	0.000			
Bootstrap 5% c.v./mean $\bar{R}^2$		4.044	0.27	0.31	0.03
Bootstrap *p*-value/95% CIs		0.029	(0.11, 0.45)	(0.14, 0.48)	(0.00, 0.11)
Bootstrap true size		**0.542**			
Five-year bond					
Data	1.115	0.937	0.33	0.42	0.09
HAC *t*-statistic	4.371	4.541			
HAC *p*-value	0.000	0.000			
Bootstrap 5% c.v./mean $\bar{R}^2$		4.031	0.27	0.31	0.03
Bootstrap *p*-value/95% CIs		0.018	(0.11, 0.47)	(0.15, 0.49)	(0.00, 0.11)
Bootstrap true size		**0.564**			

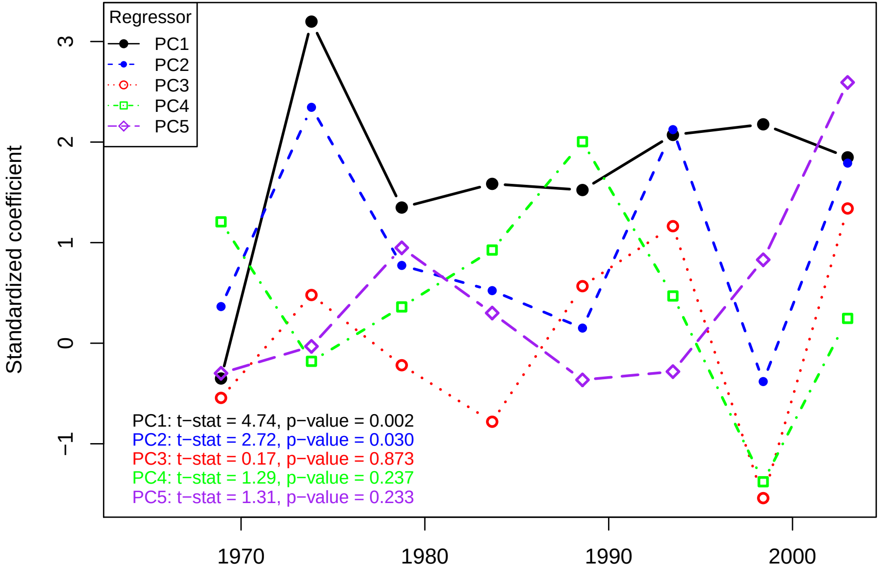
Application 3: Cochrane and Piazzesi (2005)

Cochrane and Piazzesi found that a function of current yields that is not spanned by level, slope, and curvature may be helpful for predicting yields and returns.

Cochrane-Piazzesi: predicting average holding returns

	$PC1$	$PC2$	$PC3$	$PC4$	$PC5$	Wald
Original sample: 1964–2003						
Data	0.127	-2.740	6.307	16.128	-2.038	
HAC statistic	1.724	5.205	2.950	5.626	0.748	31.919
HAC p-value	0.085	0.000	0.003	0.000	0.455	0.000
Bootstrap 5% c.v./mean $\bar{R}^2$				2.441	2.190	8.571
Bootstrap p-value/95% CIs				0.000	0.494	0.000
Bootstrap true size				0.097	0.078	0.116
IM $q = 8$	0.002	0.030	0.873	0.237	0.233	
IM $q = 16$	0.000	0.004	0.148	0.953	0.283	
Later sample: 1985–2013						
Data	0.104	-1.586	-3.962	-9.196	9.983	
HAC statistic	1.619	2.215	1.073	1.275	1.351	4.174
HAC p-value	0.106	0.027	0.284	0.203	0.178	0.124
Bootstrap 5% c.v./mean $\bar{R}^2$				2.656	2.367	11.321
Bootstrap p-value/95% CIs				0.317	0.283	0.289
Bootstrap true size				0.140	0.113	0.175
IM $q = 8$	0.011	0.079	0.044	0.803	0.435	
IM $q = 16$	0.001	0.031	0.215	0.190	0.949	

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



Application 4: Greenwood and Vayanos (2014))

Greenwood-Vayanos: measure of maturity composition of Treasury debt appears to predict return on long-term bond.

Even using conventional HAC, p-value drops to 0.06 when level, slope and curvature added to regression.

IM tests suggest p-value is above 0.8.

Summary of contributions (econometrics)

- **We already knew:** if x_{1t} is highly persistent and not strictly exogenous, b_1 is biased and hypothesis tests about β_1 are problematic (Mankiw and Shapiro, 1986; Stambaugh, 1999; Campbell and Yogo, 2006).
- **Our paper shows:** even if x_{2t} is strictly exogenous, high persistence of x_{1t} and x_{2t} along with lack of strict exogeneity of x_{1t} make hypothesis tests about β_2 problematic.

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; Ludbigson and Ng, 2009, 2010; Cochrane and Piazzesi, 2005; Greenwood and Vayasnos, 2014).
- **Our paper concludes:** level and slope are all that is needed; there is no robust evidence against spanning hypothesis.