Appendix to "Bond Risk Premia"

John H. Cochrane and Monika Piazzesi

December 9 2004. Revised Sept 15 2006; typos fixed in Equations (D.17)-(D.20)

A Additional Results

A.1 Unrestricted forecasts

Table A1 reports the point estimates of the unrestricted regression (1) together with standard errors and test statistics. The coefficient estimates in Table A1 are plotted in the top panel of Figure 1, and the R^2 and some of the statistics from Table A1 are reported in the right side of Table 1B. The table shows that the individual-bond regressions have the same high significance as the regression of average (across maturity) excess returns on forward rates documented in Table 1A.

A.2 Fama-Bliss

Table A2 presents a full set of estimates and statistics for the Fama-Bliss regressions summarized in Table 2. The main additions are the confidence intervals for R^2 and the small-sample distributions.

The R^2 confidence intervals show that the Fama-Bliss regressions do achieve an R^2 that is unlikely under the expectations hypothesis. However, the Fama-Bliss R^2 is just above that confidence interval where the 0.35 R^2 were much further above the confidence intervals in Table A1. Again, the multiple regression provides stronger evidence against the expectations hypothesis, accounting for the larger number of right hand variables, even accounting for small-sample distributions.

The small-sample standard errors are larger, and χ^2 statistics smaller, than the large-sample counterparts. The pattern is about the same as for the multiple regressions in Table 1. However, the small-sample statistics still reject. The rejection in Table 1 is stronger with small-sample statistics as well.

A.3 Comparison with Fama-Bliss

If the return-forecasting factor really is an improvement over other forecasts, it should drive out other variables, and the Fama-Bliss spread in particular. Table A3 presents multiple regressions to address this question. In the presence of the Fama-Bliss forward spread, the coefficients and significance of the return-forecasting factor from Table A2 are unchanged in

Table A3. The R^2 is also unaffected, meaning that the addition of the Fama-Bliss forward spread does not help to forecast bond returns. In the presence of the return-forecasting factor, however, the Fama-Bliss slope disappears. Clearly, the return-forecasting factor subsumes all the predictability of bond returns captured by the Fama-Bliss forward spread.

Table A1. Regressions of 1-year excess returns on all forward rates

Maturity n	const.	$y^{(1)}$	$f^{(2)}$	$f^{(3)}$	$f^{(4)}$	$f^{(5)}$	R^2	Level R^2	$\chi^2(5)$
2	-1.62	-0.98	0.59	1.21	0.29	-0.89	0.32	0.36	121.8
Large T	(0.69)	(0.17)	(0.40)	(0.29)	(0.22)	(0.17)			⟨0.00⟩
Small T	(0.86)	(0.30)	(0.50)	(0.40)	(0.30)	(0.28)	[0.19, 0.53]		32.9
EH							[0.00, 0.17]		⟨0.00⟩
3	-2.67	-1.78	0.53	3.07	0.38	-1.86	0.34	0.36	113.8
Large T	(1.27)	(0.30)	(0.67)	(0.47)	(0.41)	(0.30)			⟨0.00⟩
Small T	(1.53)	(0.53)	(0.88)	(0.71)	(0.53)	(0.50)	[0.21, 0.55]		38.6
EH							[0.00, 0.17]		⟨0.00⟩
4	-3.80	-2.57	0.87	3.61	1.28	-2.73	0.37	0.39	115.7
Large T	(1.73)	(0.44)	(0.87)	(0.59)	(0.55)	(0.40)			⟨0.00⟩
Small T	(2.03)	(0.71)	(1.18)	(0.94)	(0.71)	(0.68)	[0.24, 0.57]		46.0
EH							[0.00, 0.17]		⟨0.00⟩
5	-4.89	-3.21	1.24	4.11	1.25	-2.83	0.35	0.36	88.2
Large T	(2.16)	(0.55)	(1.03)	(0.67)	(0.65)	(0.49)			⟨0.00⟩
Small T	(2.49)	(0.88)	(1.46)	(1.16)	(0.88)	(0.85)	[0.21, 0.55]		39.2
EH							[0.00, 0.17]		⟨0.00⟩

Notes: The regression equation is

$$rx_{t+1}^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} y_t^{(1)} + \beta_2^{(n)} f_t^{(2)} + ... + \beta_5^{(n)} f_t^{(5)} + \varepsilon_{t+1}^{(n)}.$$

$\bar{R}^2$ reports adjusted R^2. "Level R^2" reports the R^2 from a regression using the level excess return on the left hand side, $e^{r_{t+1}^{(n)}} - e^{y_t^{(1)}}$. Standard errors are in parentheses "()." "Large T"standard errors use the 12 lag Hansen-Hodrick GMM correction for overlap and heteroskedasticity. "Small T" standard errors are based on 50,000 bootstrapped samples from an unconstrained 12 lag yield VAR. Square brackets "[]" are 95 percent bootstrap confidence intervals for R^2. "EH" imposes the expectations hypothesis on the bootstrap: We run a 12 lag autoregression for the 1-year rate and calculate other yields as expected values of the 1-year rate. "$\chi^2(5)$" is the Wald statistic that tests whether the slope coefficients are jointly zero. The 5 percent and 1 percent critical value for $\chi^2(5)$ are 11.1 and 15.1. All χ^2 statistics are computed with 18 Newey-West lags. "Small T" Wald statistics are computed from the covariance matrix of parameter estimates across the bootstrapped samples. Pointed brackets "<>" report probability values. Data source CRSP, sample 1964:1-2003:12.

Table A2. Fama-Bliss excess return regressions

Maturity n		α	β	R^2	$\chi^2(1)$	p-value
2		0.07	0.99	0.16		
	Large T	(0.30)	(0.26)		18.4	$\langle 0.00 \rangle$
	Small T	(0.16)	(0.33)	$[0.01, 0.33]$	9.1	$\langle 0.00 \rangle$
	EH			$[0.00, 0.12]$		$\langle 0.01 \rangle$
3		-0.13	1.35	0.17		
	Large T	(0.54)	(0.35)		19.2	$\langle 0.00 \rangle$
	Small T	(0.32)	(0.41)	$[0.01, 0.34]$	10.8	$\langle 0.00 \rangle$
	EH			$[0.00, 0.14]$		$\langle 0.01 \rangle$
4		-0.40	1.61	0.18		
	Large T	(0.75)	(0.45)		16.4	$\langle 0.00 \rangle$
	Small T	(0.48)	(0.48)	$[0.01, 0.34]$	11.2	$\langle 0.00 \rangle$
	EH			$[0.00, 0.14]$		$\langle 0.01 \rangle$
5		-0.09	1.27	0.09		
	Large T	(1.04)	(0.58)		5.7	$\langle 0.02 \rangle$
	Small T	(0.64)	(0.64)	$[0.00, 0.24]$	4.0	$\langle 0.04 \rangle$
	EH			$[0.00, 0.14]$		$\langle 0.13 \rangle$

Notes: The regressions are

$$rx_{t+1}^{(n)} = \alpha + \beta \left(f_t^{(n)} - y_t^{(1)} \right) + \varepsilon_{t+1}^{(n)}.$$

Standard errors are in parentheses, bootstrap 95 percent confidence intervals in square brackets "[]" and probability values in angled brackets "<>". The 5 percent and 1 percent critical values for a $\chi^2(1)$ are 3.8 and 6.6. See notes to Table A1 for details.

Table A3. Contest between $\gamma^\top f$ and Fama-Bliss

n	a_n	$\sigma(a_n)$	b_n	$\sigma(b_n)$	c_n	$\sigma(c_n)$	R^2
2	0.14	(0.25)	0.46	(0.04)	-0.05	(0.21)	0.31
3	0.13	(0.48)	0.87	(0.12)	-0.05	(0.41)	0.34
4	-0.03	(0.62)	1.22	(0.16)	0.05	(0.46)	0.37
5	-0.32	(0.71)	1.43	(0.15)	0.15	(0.35)	0.35

Notes: Multiple regression of excess holding period returns on the return-forecasting factor and Fama-Bliss slope. The regression is

$$rx_{t+1}^{(n)} = a_n + b_n \left(\gamma^\top f_t \right) + c_n \left(f_t^{(n)} - y_t^{(1)} \right) + \varepsilon_{t+1}^{(n)}.$$

Standard errors are in parentheses. See notes to Table A1 for details.

A.4 Forecasting the short rate

The return-forecasting factor also predicts changes in short-term interest rates. Short rate forecasts and excess return forecasts are mechanically linked, as emphasized by Fama and Bliss (1987), but seeing the same phenomenon as a short rate forecast provides a useful complementary intuition and suggests additional implications. Here, the expectations hypothesis predicts a coefficient of 1.0 – if the forward rate is one percentage point higher than the short rate, we should see the short rate rise one percentage point on average.

Table A4. Forecasting short rate changes

	const.	$f_t^{(2)} - y_t^{(1)}$	$y_t^{(1)}$	$f_t^{(2)}$	$f_t^{(3)}$	$f_t^{(4)}$	$f_t^{(5)}$	R^2	χ^2
				Fama-Bliss					
	-0.07	0.01						0.00	0.0
Large T	(0.30)	(0.26)							$\langle 0.98 \rangle$
Small T	(0.16)	(0.33)						$[0.00, 0.33]$	0.0
EH								$[0.05, 0.29]$	$\langle 1.00 \rangle$
				Unconstrained					
	1.62		-0.02	0.41	-1.21	-0.29	0.89	0.19	82.7
Large T	(0.69)		(0.17)	(0.40)	(0.29)	(0.22)	(0.17)		$\langle 0.00 \rangle$
Small T	(0.86)		(0.30)	(0.50)	(0.40)	(0.30)	(0.28)	$[0.15, 0.40]$	25.0
EH								$[0.07, 0.32]$	$\langle 0.01 \rangle$

Notes: The Fama-Bliss regression is

$$y_{t+1}^{(1)} - y_t^{(1)} = \beta_0 + \beta_1 \left(f_t^{(2)} - y_t^{(1)} \right) + \varepsilon_{t+1}.$$

The unconstrained regression equation is

$$y_{t+1}^{(1)} - y_t^{(1)} = \beta_0 + \beta_1 y_t^{(1)} + \beta_2 f_t^{(2)} + ... + \beta_5 f_t^{(5)} + \varepsilon_{t+1}.$$

χ^2 tests whether all slope coefficients are jointly zero (5 degrees of freedom unconstrained, one degree of freedom for Fama-Bliss). Standard errors are in parentheses, bootstrap 95 percent confidence intervals in square brackets "[]" and percent probability values in angled brackets "<>". See notes to Table A1 for details.

The Fama-Bliss regression in Table A4 shows instead that the two year forward spread has no power to forecast a one year change in the one year rate. Thus, the Fama and Bliss's (1987) return forecasts correspond to nearly random-walk behavior in yields. To find greater forecastability of bond excess returns, our return-forecasting factor must and does forecast changes in 1-year yields; a positive expected return forecasts implies that bond prices will rise. Indeed, in the "unconstrained" panel of Table A4, all forward rates together

have substantial power to predict one-year changes in the short rate. The R^2 for short rate changes rises to 19 percent, and the χ^2 test strongly rejects the null that the parameters are jointly zero.

To understand this phenomenon, note that we can always break the excess return into a one year yield change and a forward-spot spread,[7]

$$E_t\left(rx_{t+1}^{(2)}\right) = -E_t\left(y_{t+1}^{(1)} - y_t^{(1)}\right) + \left(f_t^{(2)} - y_t^{(1)}\right). \tag{A.1}$$

Intuitively, you make money either by capital gains, or by higher initial yields. Under the expectations hypothesis, expected excess returns are constant, so any movement in the forward spread must be matched by movements in the expected 1-year yield change. If the forward rate is higher than the spot rate, it must mean that investors must expect a rise in 1-year rates (a decline in long-term bond prices) to keep expected returns the same across maturities. In Fama and Bliss's regressions, the expected yield change term is constant, so changes in expected returns move one-for one with the forward spread. In our regressions, expected returns move more than changes in the forward spread. The only way to generate such changes is if the 1-year rate becomes forecastable as well, generating expected capital gains and losses for long-term bond holders.

Equation (A.1) also means that the regression coefficients which forecast the 1-year rate change in Table A4 are exactly equal to our return-forecasting factor $b_2\gamma^\top f_t$, which forecasts $E_t\left(rx_{t+1}^{(2)}\right)$, minus a coefficient of 1 on the 2-year forward spread. The factor that forecasts excess returns is also the state variable that forecasts the short rate.

A.5 Additional Lags

Table 5 reports our estimates of γ and α in the simple model for additional lags,

$$\overline{rx}_{t+1} = \gamma^\top\left[\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} ... + \alpha_k f_{t-\frac{k}{12}}\right] + \bar{\varepsilon}_{t+1}^{(n)}$$

Table A5 completes the model by showing how individual bond returns load on the common return-forecasting variable, i.e. estimates of b_n in

$$\overline{rx}_{t+1}^{(n)} = b_n\gamma^\top\left[\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} + ... + \alpha_k f_{t-\frac{k}{12}}\right] + \bar{\varepsilon}_{t+1}^{(n)}$$

In Panel A, the b_n rise with maturity. The b estimates using additional lags are almost exactly the same as those using only f_t, as claimed in the paper.

In Panel B, we see that the R^2 for individual regressions mirror the R^2 for the forecasts of bond average (across maturity) returns $\overline{rx}$. We also see that the R^2 from the restricted regressions are almost as high as those of the unrestricted regressions, indicating that the restrictions do little harm to the model's ability to fit the data. This finding is especially

cogent in this case, as the unrestricted regressions allow arbitrary coefficients across time (lags) as well as maturity. For example, with 3 additional lags, the unrestricted regressions use 4 bonds $\times$ 4 lags $\times$ (5 forward rates + 1 constant) = 96 parameters, while the restricted regressions use 4 b +6 γ + 4 α = 14 parameters.

In sum, Table A5 substantiates the claim in the paper that the single-factor model works just as well with additional lags as it does using only time t right hand variables.

Table A5. Return forecasts with additional lags

A. b_n estimates and standard errors

	Estimates				Standard errors			
	Lags				Lags			
	0	1	2	3	0	1	2	3
b_2	0.47	0.46	0.46	0.46	(0.06)	(0.04)	(0.04)	(0.04)
b_3	0.87	0.86	0.86	0.86	(0.11)	(0.09)	(0.08)	(0.08)
b_4	1.24	1.23	1.23	1.23	(0.17)	(0.13)	(0.12)	(0.11)
b_5	1.43	1.45	1.45	1.46	(0.21)	(0.16)	(0.15)	(0.14)

B. R^2

	Restricted				Unrestricted			
	Lags				Lags			
	0	1	2	3	0	1	2	3
$rx^{(2)}$	0.31	0.36	0.37	0.31	0.32	0.37	0.39	0.41
$rx^{(3)}$	0.34	0.39	0.41	0.41	0.34	0.40	0.42	0.44
$rx^{(4)}$	0.37	0.42	0.44	0.46	0.37	0.43	0.45	0.47
$rx^{(5)}$	0.34	0.41	0.43	0.44	0.35	0.41	0.44	0.46
$\overline{rx}$	0.35	0.41	0.43	0.44	0.35	0.41	0.43	0.46

Notes: Return forecasts with additional lags, using the restricted model

$$\overline{rx}_{t+1}^{(n)} = b_n \gamma^\top \left[\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} + \alpha_2 f_{t-\frac{2}{12}} + ... + \alpha_k f_{t-\frac{k}{12}} \right] + \bar{\varepsilon}_{t+1}^{(n)}$$

Estimates of γ and α are presented in Table 5.

A.6 Eigenvalue factor models for yields

We form the yield curve factors x_t used in Table 4 and Figure 2 from an eigenvalue decomposition of the covariance matrix of yields $var(y) = Q\Lambda Q^\top$ with $Q^\top Q = I$, Λ diagonal. Decompositions based on yield changes, returns, or excess returns are nearly identical. Then we can write yields in terms of factors as $y_t = Qx_t$; $cov(x_t, x_t^\top) = \Lambda$. Here, the columns of Q give how much each factor x_{it} moves all yields y_t. We can also write $x_t = Q^\top y_t$. Here, the columns of Q tell you how to recover factors from yields. The top right panel of Figure 2

plots the first three columns of Q. We label the factors "level," "slope," "curvature," "4-5" and "W" based on the shape of these loadings. We do not plot the last two small factors for clarity, and because being so small they are poorly identified separately. W is W shaped, loading most strongly on the 3 year yield. 4-5 is mostly a 4-5 year yield spread.

We compute the fraction of yield variance due to the kth factor as $\Lambda_{kk}/Q_k\Lambda_{kk}$. To calculate the fraction of yield variance due to the return-forecasting factor, we first run a regression $y_t^{(n)} = a + b\gamma^\top f_t + \varepsilon_t$, and then we calculate $\text{trace}\big[\text{cov}(b\gamma^\top f)\big]\,/\text{trace}[\text{cov}(y)]$.

A.7 Eigenvalue factor model for expected excess returns

We discuss in section B. an eigenvalue factor decomposition of the unconstrained expected excess return covariance matrix. We start with

$$Q\Lambda Q^\top = cov\left[E_t(rx_{t+1}), E_t(rx_{t+1})^\top\right] = \beta cov(f_t, f_t^\top)\beta^\top.$$

Now we can write the unconstrained regression

$$E_t(rx_{t+1}) = \beta f_t = Q\Gamma^\top f_t. \tag{A.2}$$

with $\Gamma^\top = Q^\top\beta$. Equivalently, we can find the factors Γ by regressing *portfolios* of expected returns on forward rates,

$$Q^\top rx_{t+1} = \Gamma^\top f_t + Q^\top\varepsilon_{t+1}.$$

Table A6 presents the results. The first column of Q in Panel A tells us how the first expected-return factor moves expected returns of bonds of different maturities. The coefficients rise smoothly from 0.21 to 0.68. This column is the equivalent of the b coefficients in our single-factor model $E_t(rx_{t+1}) = b\gamma^\top$. The corresponding first row of $\Gamma^\top$ in Panel B is very nearly our tent-shaped function of forward rates. Expressed as a function of yields it displays almost exactly the pattern of the top left panel in Figure 2: a rising function of yields with a strong 4-5 spread.

The remaining columns of Q and rows of $\Gamma^\top$ show the structure summarized by simple portfolio regressions in the text. When a portfolio loads strongly on one bond in Panel A, that bond's yield is important for forecasting that portfolio in Panel C. The remaining factors seem to be linear combinations (organized by variance) of the pattern shown in Table 7. Individual bond "pricing errors" in yields seem to be reversed.

The bottom two rows of Panel A give the variance decomposition. The first factor captures almost all of the variation in expected excess returns. Its standard deviation at 5.16 percentage points dominates the 0.26, 0.16 and 0.20 percentage point standard deviations of the other factors. Squared, to express the result as fractions of variance, the first factor accounts for 99.5 percent of the variance of expected returns.

Table A6. Factor decomposition for expected excess returns

A. Q matrix of loadings

	Factor			
Maturity	1	2	3	4
2	0.21	0.27	-0.69	0.64
3	0.40	0.84	0.34	-0.11
4	0.58	-0.16	-0.48	-0.64
5	0.68	-0.44	0.41	0.41
σ(factor)	5.16	0.26	0.16	0.20
Percent of var	99.51	0.25	0.09	0.15

B. $\Gamma^\top$ matrix; forecasting the portfolios

Factor	$y_t^{(1)}$	$f_t^{(2)}$	$f_t^{(3)}$	$f_t^{(4)}$	$f_t^{(5)}$
1	-4.58	1.68	6.36	1.81	-4.43
2	0.05	-0.07	0.54	-0.35	-0.13
3	-0.04	-0.12	0.20	-0.16	0.10
4	-0.10	0.27	-0.19	-0.17	0.23

C. $\Gamma^\top$ matrix with yields

Factor	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(3)}$	$y_t^{(4)}$	$y_t^{(5)}$
1	-6.27	-9.35	13.66	24.94	-22.14
2	0.12	-1.24	2.69	-0.89	-0.65
3	-0.08	-0.64	1.01	-1.07	0.51
4	0.37	0.93	-0.07	-1.58	1.13

Notes: We start with the unconstrained forecasting regressions

$$rx_{t+1}^{(n)} = \beta^{(n)} f_t + \varepsilon_{t+1}^{(n)}.$$

Then, we perform an eigenvalue decomposition of the covariance matrix of expected excess returns,

$$Q\Lambda Q^\top = cov\left[E_t(rx_{t+1}), E_t(rx_{t+1})^\top\right] = \beta cov(f_t, f_t^\top)\beta^\top.$$

Panel A gives the Q matrix. The last two rows of panel A give $\sqrt{\Lambda_i}$ and $\Lambda_i / \sum \Lambda_i$ respectively. Panels B and C give regression coefficients in forecasting regressions

$$Q^\top rx_{t+1} = \Gamma^\top f_t + Q^\top \varepsilon_{t+1}$$
$$Q^\top rx_{t+1} = \Gamma^\top y_t + Q^\top \varepsilon_{t+1}.$$

A.8 What measurement error can and cannot do

To understand the pattern of Figure 4, write the left hand variable as

$$\begin{aligned} rx_{t+1}^{(n)} &= p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)} \\ &= p_{t+1}^{(n-1)} + ny_t^{(n)} - y_t^{(1)}. \end{aligned}$$

Now, consider a regression of this return on to time-t variables. Clearly, measurement error in prices, forward rates or yields – anything that introduces spurious variation in time-t variables – will induce a coefficient of -1 on the one year yield and $+n$ on the $n-$year yield, as shown in the bottom panel of Figure 4. Similarly, if we write the left hand variable in terms of forward rates as

$$\begin{aligned} rx_{t+1}^{(n)} &= p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)} \\ &= p_{t+1}^{(n-1)} + \left[-p_t^{(n)} + p_t^{(n-1)}\right] + \left[-p_t^{(n-1)} + p_t^{(n-2)}\right] + \cdots + \left[-p_t^{(2)} + p_t^{(1)}\right] \cdots - p_t^{(1)} + p_t^{(1)} \\ &= p_{t+1}^{(n-1)} + 0 \times y_t^{(1)} + 1 \times f_t^{(1\to 2)} + 1 \times f_t^{(2\to 3)} + \cdots + 1 \times f_t^{(n-1\to n)} \end{aligned}$$

we see the step-function pattern shown in the top panel of Figure 4.

The crucial requirement for this pattern to emerge as a result of measurement error is that the measurement error at time t must be uncorrelated with in $p_{t+1}^{(n-1)}$ on the left hand side. If measurement error at time t is correlated with the measured variable at time $t+1$, then other time-t variables may seem to forecast returns, or the 1 and n year yield may forecast it with different patterns. Of course, the usual specification of i.i.d. measurement error is more than adequate for this conclusion. Also, measurement error must be uncorrelated with the true right hand variables, as we usually specify. Measurement errors correlated across *maturity* at a given time will not change this pattern. Multiple regressions orthogonalize right hand variables.

B Robustness checks

We investigate a number of desirable robustness checks. We show that the results obtain in the McCulloch-Kwan data set. We show that the results are stable across subsamples. In particular, the results are stronger in the low-inflation 1990s than they are in the high-inflation 1970s. This finding comfortingly suggests a premium for real rather than nominal interest rate risk. We show that the results obtain with real-time forecasts, rather than using the full sample to estimate regression coefficients. We construct "trading rule" profits, examine their behavior, and examine real-time trading rules as well. The trading rules improve substantially on those using Fama-Bliss slope forecasts.

B.1 Other data

The Fama-Bliss data are interpolated zero-coupon yields. In Table A7, we run the regressions with McCulloch-Kwon data, which use a different interpolation scheme to derive zero-coupon yields from Treasury bond data. Table A7 also compares the R^2 and γ estimates using McCulloch-Kwon and Fama-Bliss data over the McCulloch-Kwon sample (1964:1-1991:2). Clearly, the tent-shape of γ estimates and R^2 are very similar across the two datasets.

Table A7. Comparison with McCulloch-Kwon data

A. R^2

	All f_t		$\gamma^\top f_t$		$f_t^{(n)} - y_t^{(1)}$	
n	M-K	F-B	M-K	F-B	M-K	F-B
2	0.39	0.39	0.39	0.39	0.16	0.15
3	0.37	0.40	0.37	0.40	0.15	0.16
4	0.36	0.42	0.36	0.43	0.13	0.17
5	0.35	0.38	0.35	0.39	0.12	0.05

B. Coefficients

Dataset	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5
McCulloch-Kwon	−5.11	−2.52	1.78	3.19	1.94	−3.82
Fama-Bliss	−4.65	−1.91	1.11	2.98	0.51	−2.17

Notes: The data are McCulloch-Kwon and Fama-Bliss zero-coupon yields starting 1964:1 until the end of the McCulloch-Kwon dataset, 1991:12. The upper panel shows R^2 from the regressions corresponding to Table 1. The regressions run excess log returns $rx_{t+1}^{(n)}$ on the regressors indicated on top of the table: all forwards f_t, the return-forecasting factor $\gamma^\top f_t$, and the forward spread $f_t^{(n)} - y_t^{(1)}$. The lower panel shows the estimated γ coefficients in the regression of average returns on forward rates $\overline{rx}_{t+1} = \gamma^\top f_t + \overline{\varepsilon}_{t+1}$. McCulloch-Kwon data are downloaded from http://www.econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm.

B.2 Subsamples

Table A8 reports a breakdown by subsamples of the regression of average (across maturity) excess returns $\overline{rx}_{t+1}$ on forward rates. The first set of columns run the average return on the forward rates separately. The second set of columns runs the average return on the return-forecasting factor $\gamma^\top f$ where γ is estimated from the full sample. The latter regression moderates the tendency to find spurious forecastability with five right hand variables in short time periods.

The first row of Table A8 reminds us of the full sample result – the pretty tent-shaped coefficients and the 0.35 R^2. Of course, if you run a regression on its own fitted value you

get a coefficient of 1.0 and the same R^2, as shown in the two right hand columns of the first row.

The second set of rows examine the period before, during, and after the 1979:8-1982:10 episode, when the Fed changed operating procedures, interest rates were very volatile, and inflation declined and stabilized. The broad pattern of coefficients is the same before and after. The 0.78 R^2 looks dramatic during the experiment, but this period really only has three data points and 5 right hand variables. When we constrain the pattern of the coefficients in the right hand pair of columns, the R^2 is the same as the earlier period.

The last set of rows break down the regression by decades. Again, the pattern of coefficients is stable. The R^2 is worst in the 70s, a decade dominated by inflation, but the R^2 rises to a dramatic 0.71 in the 1990s, and still 0.43 when we constrain the coefficients γ to their full-sample values. This fact suggests that the forecast power derives from changes in the real rather than nominal term structure.

Table A8. Subsample analysis

	Regression on all forward rates							$\gamma^\top f$ only	
	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	$\gamma^\top f$	R^2
1964:01-2003:12	−3.2	−2.1	0.8	3.0	0.8	−2.1	0.35	1.00	0.35
1964:01-1979:08	−5.3	−1.3	1.4	2.5	−0.1	−1.7	0.31	0.73	0.26
1979:08-1982:10	−32.5	0.8	0.7	1.2	0.6	−0.8	0.78	0.77	0.24
1982:10-2003:12	−1.6	−1.8	1.0	2.0	1.2	−2.2	0.23	0.85	0.22
1964:01-1969:12	0.5	−1.4	0.3	2.1	0.4	−1.9	0.31	0.73	0.26
1970:01-1979:12	−9.6	−1.5	0.6	2.3	0.3	−0.6	0.22	0.65	0.16
1980:01-1989:12	−11.7	−2.4	1.8	2.6	1.0	−1.9	0.43	1.09	0.35
1990:01-1999:12	−13.9	−1.5	0.3	4.4	1.5	−2.5	0.71	1.57	0.43
2000:01-2003:12	−5.2	−1.1	0.2	2.2	0.3	0.02	0.65	0.60	0.34

Notes: Subsample analysis of average return-forecasting regressions. For each subsample, the first set of columns present the regression

$$\overline{rx}_{t+1} = \gamma^\top f_t + \overline{\varepsilon}_{t+1}.$$

The second set of columns report the coefficient estimate b and R^2 from

$$\overline{rx}_{t+1} = b\left(\gamma^\top f_t\right) + \varepsilon_{t+1}$$

using the γ parameter from the full sample regression. Sample: 1964-2003.

B.3 Real time forecasts and trading rule profits

How well can one forecast bond excess returns using real-time data? Of course, the conventional rational-expectations answer is that investors have historical information, and have

evolved rules of thumb that summarize far longer time series than our data set, so there is no error in using full-sample forecasts. Still, it is an interesting robustness exercise to see how well the forecast performs based only on data from 1964 up to the time the forecast must be made.

Figure 8 compares real-time and full-sample forecasts. They are quite similar. Even though the real-time regression starts the 1970s with only 6 years of data, it already captures the pattern of bond expected returns.

The forecasts are similar, but are they similarly successful? Figures 9 and 10 compare them with a simple calculation. We calculate "trading rule profits" as

$$\overline{rx}_{t+1} \times E_t(\overline{rx}_{t+1}) = \overline{rx}_{t+1} \times \left[\gamma^\top \left(\alpha_0 f_t + \alpha_1 f_{t-1} + a_2 f_{t-2}\right)\right].$$

This rule uses the forecast $E_t(\overline{rx}_{t+1})$ to recommend the size of a position which is subject to the ex-post return $\overline{rx}_{t+1}$. In Figure 10, we cumulate the profits so that the different cases can be more easily compared.[8] For the Fama-Bliss regressions, we calculate the expected excess return of each bond from its matched forward spread, and then we find the average expected excess return across maturities in order to compute $E_t(\overline{rx}_{t+1})$.

The full-sample vs. real-time trading rule profits in Figure 9 are quite similar. In both Figure 9 and the cumulated profits of Figure 10 all of the trading rules produce surprisingly few losses. The lines either rise or stay flat. The trading rules lie around waiting for occasional opportunities. Most of the time, the forward curve is not really rising a lot, nor tent shaped, so both rules see a small $E_t(\overline{rx}_{t+1})$. The trading rules thus recommend small positions, leading to small gains and losses. On infrequent occasions, the forward curve is either rising or tent-shaped, so the trading rules recommend large positions $E_t(\overline{rx}_{t+1})$, and these positions seem to work out.

The real-time forecast looks quite good in Figure 9, but the cumulative difference amounts to about half of the full-sample profits. However, the real-time trading rule does work, and it even works better than even the full sample Fama-Bliss forecast. We conclude that the overall pattern remains in real-time data. It does not seem to be the case that the forecast power, or the improvement over the Fama-Bliss forecasts and related slope forecasts, requires the use of ex-post data.

Of course, real "trading rules" should be based on arithmetic returns, and they should follow an explicit portfolio maximization problem. They also must incorporate estimates of the conditional *variance* of returns. Bond returns are heteroskedastic, so one needs to embark on a similar-sized project to understand conditional second moments and relate them to the conditional first moments we have investigated here.

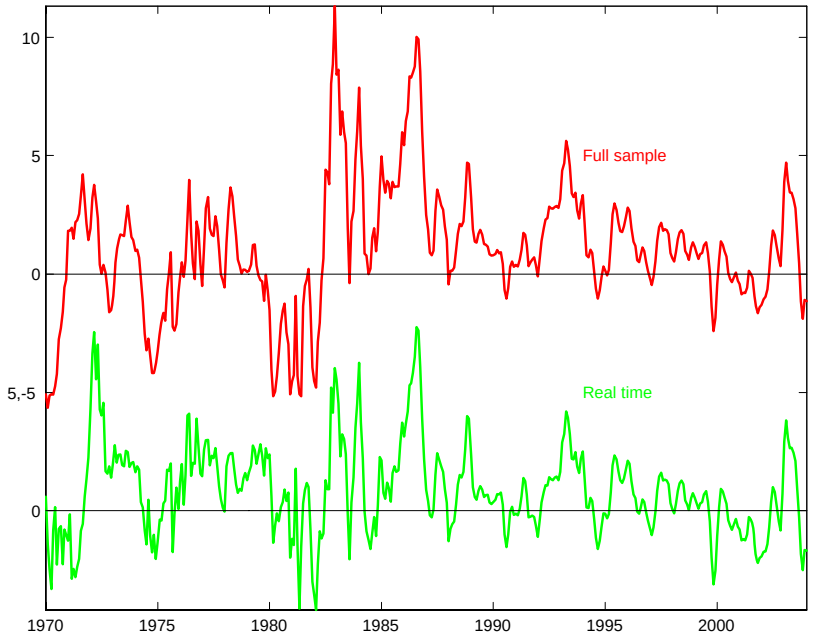

Figure 8: Full-sample and real-time forecasts of average (across maturity) excess bond returns. In both cases, the forecast is made with the regression $\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}$. The real-time graph re-estimates the regression at each t from the beginning of the sample to t.

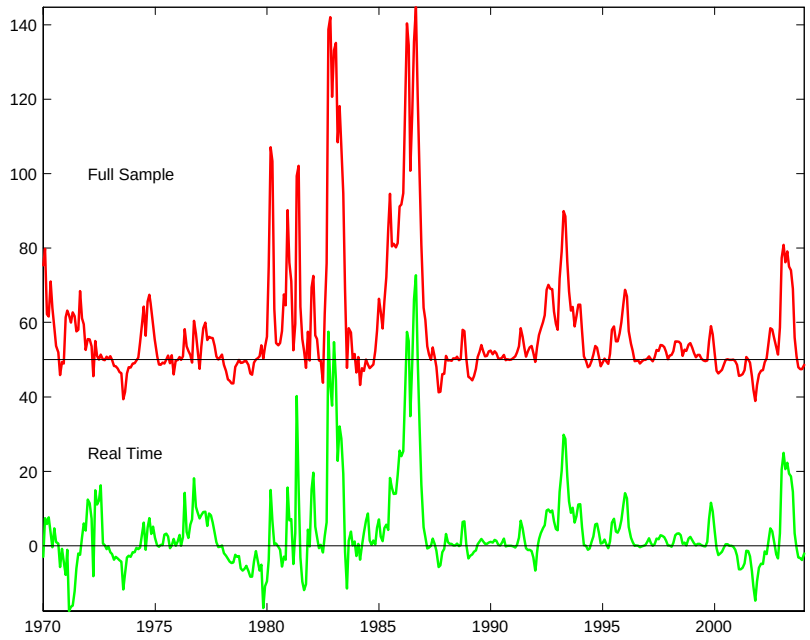

Figure 9: "Trading rule" profits, using full-sampe and real-time estimates of the return-forecasting factor.

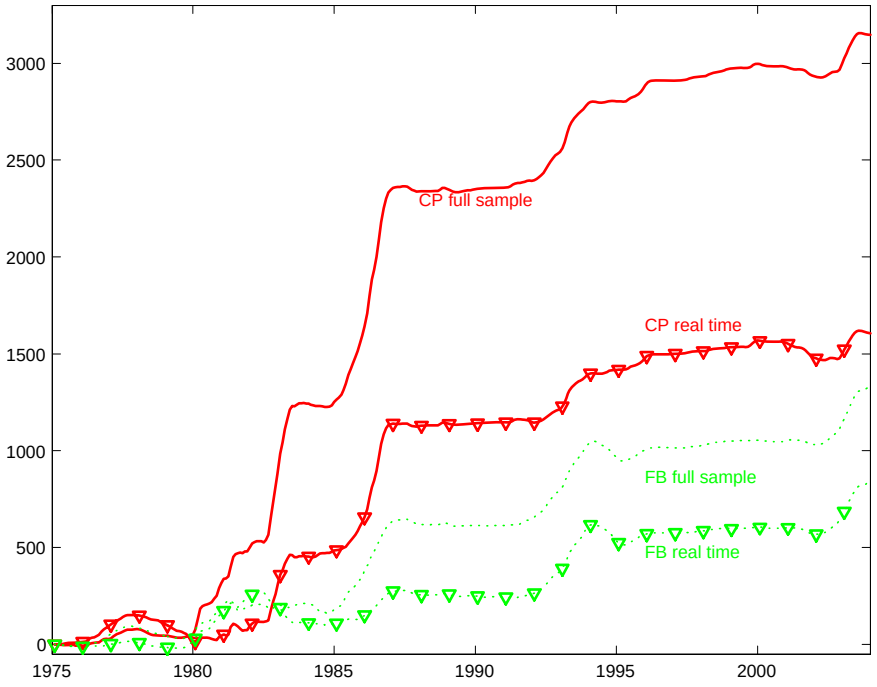

Figure 10: Cumulative profits from "trading rules" using full sample and real time information. Each line plots the cumulative value of $\overline{rx}_{t+1} \times E_t(\overline{rx}_{t+1})$. $E_t(\overline{rx}_{t+1})$ are formed from the full 1964-2003 sample or "real time" data from 1964-t as marked. The CP lines use the forecast $\overline{rx}_{t+1} = \gamma^\top(\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} + \alpha_2 f_{t-\frac{2}{12}})$. The FB (Fama-Bliss) lines forecast each excess return from the corresponding forward spread, and then average the forecasts across maturities.

C Calculations for the regressions

C.1 GMM estimates and tests

The unrestricted regression is

$$rx_{t+1} = \beta f_t + \varepsilon_{t+1},$$

The moment conditions of the unrestricted model are

$$g_T(\beta) = E(\varepsilon_{t+1} \otimes f_t) = 0. \tag{C.3}$$

The restricted model is $\beta = b\gamma^\top$, with the normalization $b^\top 1_4 = 4$.

We focus on a 2-step OLS estimate of the restricted model – first estimate average (across maturities) returns on f, then run each return on $\hat{\gamma}^\top f$:

$$\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}, \tag{C.4}$$

$$rx_{t+1} = b\left(\hat{\gamma}^\top f_t\right) + \varepsilon_{t+1}. \tag{C.5}$$

The estimates satisfy $1_4^\top b = 4$ automatically.

To provide standard errors for the two-step estimate in Table 1, we use the moments corresponding to the two OLS regressions (C.4) and (C.5),

$$\tilde{g}_T(b,\gamma) = \left[\begin{array}{c} E\left(\bar{\varepsilon}_{t+1}(b,\gamma) \times f_t\right) \\ E\left[\varepsilon_{t+1}(b,\gamma) \times \gamma^\top f_t\right] \end{array} \right] = 0.$$

Since the estimate is exactly identified from these moments ($a = I$) Hansen's (1982) Theorem 3.1 gives the standard error,

$$var\left(\begin{array}{c} \hat{\gamma} \\ \hat{b} \end{array} \right) = \frac{1}{T} \tilde{d}^{-1} \tilde{S} \tilde{d}^{-1\top}$$

where

$$\begin{aligned} \tilde{d} = \frac{\partial \tilde{g}_T}{\partial\left[\gamma^\top\ b^\top\right]} &= \frac{\partial}{\partial\left[\gamma^\top\ b^\top\right]} \left[\begin{array}{c} E\left[f_t\left(\overline{rx}_{t+1} - f_t^\top \gamma\right)\right] \\ E\left[\left(rx_{t+1} - b\gamma^\top f_t\right)\left(f_t^\top \gamma\right)\right] \end{array} \right] \\ &= \left[\begin{array}{cc} -E\left(f_t f_t^\top\right) & 0_{6\times 4} \\ E(rx_{t+1} f_t^\top) - 2b\gamma^\top E\left(f_t f_t^\top\right) & -\left[\gamma^\top E\left(f_t f_t^\top\right)\gamma\right] I_4 \end{array} \right]. \end{aligned}$$

Since the upper right block is zero, the upper left block of d^{-1} is $E\left(f_t f_t^\top\right)^{-1}$. Therefore, the variance of γ is not affected by the b estimate, and is equal to the usual GMM formula for a regression standard error, $var(\hat{\gamma}) = E(ff^\top)^{-1}\tilde{S}(1:6,1:6)E(ff^\top)^{-1}/T$. The variance of $\hat{b}$ is affected by the generated regressor γ, via the off diagonal term in $\tilde{d}^{-1}$.This is an interesting case in which the GMM standard errors that correct for the generated regressor are smaller than OLS standard errors that ignore the fact. OLS has no way of knowing that $\sum_n b_n = 1$, while the GMM standard errors know this fact. OLS standard errors thus find a common

component to the b standard errors, while GMM knows that common movement in the b is soaked up in the γ estimates.

We investigated a number of ways of computing S matrices. The regression is $\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}$. The Hansen-Hodrick ("HH") statistics are based on

$$cov(\hat{\gamma}) = E(f_t f_t^\top)^{-1} \left[\sum_{j=-k}^{k} E(f_t f_{t-j}^\top \bar{\varepsilon}_{t+1} \bar{\varepsilon}_{t+1-j}) \right] E(f_t f_t^\top)^{-1}.$$

The Newey-West ("NW") statistics use

$$cov(\hat{\gamma}) = E(f_t f_t^\top)^{-1} \left[\sum_{j=-k}^{k} \frac{k-|j|}{k} E(f_t f_{t-j}^\top \bar{\varepsilon}_{t+1} \bar{\varepsilon}_{t+1-j}) \right] E(f_t f_t^\top)^{-1}.$$

The "Simplified HH" statistics assume $E(f_t f_{t-j}^\top \bar{\varepsilon}_{t+1} \bar{\varepsilon}_{t+1-j}) = E(f_t f_{t-j}^\top) E(\bar{\varepsilon}_{t+1} \bar{\varepsilon}_{t+1-j})$ and $E(\bar{\varepsilon}_{t+1} \bar{\varepsilon}_{t+1-j}) = \frac{k-|j|}{k} E(\bar{\varepsilon}_{t+1}^2)$, hence

$$cov(\hat{\gamma}) = E(f_t f_t^\top)^{-1} \left[\sum_{j=-k}^{k} \frac{k-|j|}{k} E(f_t f_{t-j}^\top) \right] E(f_t f_t^\top)^{-1} E(\bar{\varepsilon}_{t+1}^2).$$

"No overlap" statistics use

$$cov(\hat{\gamma}) = E(f_t f_t^\top)^{-1} E(f_t f_t^\top \bar{\varepsilon}_{t+1}^2) E(f_t f_t^\top)^{-1}$$

averaged over 12 initial months.

To test the (inefficient) two step estimate in Table 6, we apply Hansen's Lemma 4.1 – the counterpart to the J_T test that handles inefficient as well as efficient estimates. To do this, we must first express the restricted estimate as a GMM estimate based on the unrestricted moment conditions (C.3). The two step OLS estimate of the restricted model sets to zero a linear combination of the unrestricted moments:

$$a_T g_T = 0, \tag{C.6}$$

where

$$a_T = \begin{bmatrix} I_6 & I_6 & I_6 & I_6 \\ \gamma^\top & 0_{1\times 6} & 0_{1\times 6} & 0_{1\times 6} \\ 0_{1\times 6} & \gamma^\top & 0_{1\times 6} & 0_{1\times 6} \\ 0_{1\times 6} & 0_{1\times 6} & \gamma^\top & 0_{1\times 6} \end{bmatrix} = \begin{bmatrix} 1_4^\top \otimes I_6 \\ I_3 \otimes \gamma^\top \quad 0_{3\times 6} \end{bmatrix}.$$

The first row of identity matrices in a_T sums across return maturities to do the regression of average returns on all forward rates. The last three rows sum across forward rates at a given return maturity to do the regression of each return on $\gamma^\top f$. An additional row of a_T of the form $\begin{bmatrix} 0_{1\times 6} & 0_{1\times 6} & 0_{1\times 6} & \gamma^\top \end{bmatrix}$ to estimate the last element of b would be redundant – the b_4 regression is implied by the first three regressions. The estimate is the same whether one

runs that regression or just estimates $b_4 = 1-b_1-b_2-b_3$. We follow the latter convention since the GMM distribution theory is written for full rank a matrices. It is initially troubling to see a parameter in the a matrix. Since we use the OLS γ estimate in the second stage regression, however, we can interpret γ in a_T as its OLS estimate, $\gamma = E_T(ff^\top)^{-1}E_T(\overline{rx}\ f)$. Then a_T is a random matrix that converges to a matrix a as it should in the GMM distribution theory. (I.e. we do not choose the γ in a_T to set $a_T(\gamma)g_T(\gamma, b) = 0$.)

We need the d matrix,

$$d \equiv \frac{\partial g_T}{\partial \begin{bmatrix} b^\top & \gamma^\top \end{bmatrix}}.$$

Recalling $b_4 = 4 - b_1 - b_2 - b_3$, the result is

$$d = \begin{bmatrix} \begin{bmatrix} -1 & & \\ & -1 & \\ & & -1 \\ 1 & 1 & 1 \end{bmatrix} \otimes E(ff^\top)\gamma & -b \otimes E(ff^\top) \end{bmatrix}.$$

Now, we can invoke Hansen's Lemma 4.1, and write the covariance matrix of the moments under the restricted estimate,

$$cov(g_T) = \frac{1}{T}(I - d(ad)^{-1}a)S(I - d(ad)^{-1}a)^\top.$$

The test statistic is

$$g_T^\top cov(g_T)^+ g_T \tilde{} \chi^2_{15}.$$

There are $4 \times 6 = 24$ moments and $6(\gamma) + 3(b)$ parameters, so there are 15 degrees of freedom. The $cov(g_T)$ matrix is singular, so the $+$ operator represents pseudoinversion. One can use an eigenvalue decomposition for $cov(g_T)$ and then retain only the largest 15 eigenvalues, i.e. write $cov(g_T) = Q\Lambda Q^\top$ where Q is orthogonal and Λ is diagonal and then $cov(g_T)^+ = Q\Lambda^+ Q^\top$ where Λ^+ inverts the 15 largest diagonal elements of Λ and sets the remainder to zero. We use the matlab pinv command.

To conduct the corresponding Wald test in Table 6, we first find the GMM covariance matrix of the unrestricted parameters. We form a vector of those parameters, $vec(\beta)$, and then

$$cov(vec(\beta)) = \frac{1}{T} d^{-1} S \left(d\right)^{-1\top},$$

$$d = \frac{\partial g_T}{\partial (vec(\beta)^\top)} = I_4 \otimes E(ff^\top).$$

To construct standard errors and test statistics for the restricted model with lags in Table 5, we proceed similarly. The estimated parameters are α and γ. The moments g_T set to zero

the two regressions on which we iterate,

$$\begin{aligned}
0 &= E(\gamma^\top f_t \times \varepsilon_{t+1}) \\
0 &= E(\gamma^\top f_{t-\frac{i}{12}} \times \varepsilon_{t+1}) \\
&... \\
0 &= E\left[(\alpha_0 f_t^{(1)} + \alpha_1 f_{t-1}^{(1)} + ..) \times \varepsilon_{t+1}\right] \\
0 &= E\left[(\alpha_0 f_t^{(2)} + \alpha_1 f_{t-1}^{(2)} + ..) \times \varepsilon_{t+1}\right] \\
&...
\end{aligned}$$

The standard error formula is as usual $cov(\left[\alpha^\top \gamma^\top\right]) = (d' S^{-1} d)^{-1}/T$. The d matrix is

$$d = \frac{dg_T}{d\left[\alpha^\top \gamma^\top\right]}.$$

Note that the first set of moments do not involve α and the second set of moments do not involve γ Thus, the d matrix is block-diagonal. This means that α standard errors are not affected by γ estimation and vice versa. Therefore, we can use regular GMM standard errors from each regression without modification.

C.2 Simulations for small-sample distributions

We simulate yields based on three different data-generating processes. Each of these processes is designed to address a specific concern about the statistical properties of our test statistics.

As the generic bootstrap for Wald statistics, we use a vector autoregression with 12 lags for the vector of yields

$$y_t = A_0 + A_1 y_{t-1/12} + \ldots A_{12} y_{t-1} + \varepsilon_t.$$

VARs based on fewer lags (such as one or two) are unable to replicate the one year horizon forecastability of bond returns or of the short rate documented in Table 5.

Second to address unit root fears, we use a VAR with 12 lags that imposes a common trend:

$$\Delta y_t = A_0 + B \begin{pmatrix} y_{t-1/12}^{(1)} - y_{t-1/12}^{(5)} \\ y_{t-1/12}^{(2)} - y_{t-1/12}^{(5)} \\ y_{t-1/12}^{(3)} - y_{t-1/12}^{(5)} \\ y_{t-1/12}^{(4)} - y_{t-1/12}^{(5)} \end{pmatrix} + \sum_{i=1}^{11} A_i \Delta y_{t-i/12} + \varepsilon_t,$$

where $\Delta y_t = y_t - y_{t-1/12}$.

Third, we impose the expectations hypothesis by starting with an AR(12) for the short

rate

$$y_t^{(1)} = a_0 + a_1 y_{t-1/12}^{(1)} + \ldots + a_{12} y_{t-1}^{(1)} + \varepsilon_t.$$

We then compute long yields as

$$y_t^{(n)} = \frac{1}{n} E_t \left(\sum_{i=1}^{n} y_{t+i-1}^{(n)} \right), n = 2, \ldots, 5.$$

To compute the expected value in this last expression, we expand the state space to rewrite the dynamics of the short rate as a vector autoregression with 1 lag. The 12-dimensional vector $x_t = \left[y_t^{(1)} \; y_{t-1/11}^{(1)} \; \ldots \; y_{t-11/12}^{(1)} \right]^\top$ follows

$$x_t = B_0 + B_1 x_{t-1/12} + \Sigma u_t$$

for

$$B_0 = \begin{pmatrix} a_0 \\ 0 \end{pmatrix}, B_1 = \begin{pmatrix} a_1 \cdots & a_{12} \\ I_{11\times 11} & 0_{11\times 1} \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0_{1\times 11} \\ 0_{11\times 1} & 0_{11\times 11} \end{pmatrix}.$$

The vector $e_1 = [1 \;\; 0_{1\times 11}]$ picks the first element in x_t, which gives $y_t^{(1)} = e_1 x_t$. Longer yields can then be easily computed recursively as

$$y_t^{(n)} = \frac{n-1}{n} y_t^{(n-1)} + \frac{1}{n} e_1 \left(\left(\sum_{i=0}^{n-1} B_1^i \right) B_0 + B_1^n x_t \right), n = 2, \ldots, 5,$$

with the understanding that B_1^i for $i = 0$ is the 12×12 identity matrix.

D Affine model

The main point of this section is to show that one can construct an affine term-structure model that captures our return-forecasting regressions exactly. To that end, we start by defining how an affine model works. Then we show how to construct an affine model – how to pick market prices of risk – so that a VAR of our five bond prices $p_{t+1} = \mu + \phi p_t + v_{t+1}$ *is* an affine model. The issue here is how to make the model "self-consistent," how to make sure that the prices that come *out* of the model are the same as the prices that go *in* to the model as "state variables."

Return regressions carry exactly the same information as a price VAR of course, so in a formal sense, finding market prices of risk to justify $p_{t+1} = \mu + \phi p_t + v_{t+1}$ as an affine model is enough to make our point. However, it is interesting to connect the affine model directly to return regressions, exhibiting the transformation from return regressions $rx_{t+1} = a + \beta f_t + \varepsilon_{t+1}$ to the price VAR and providing return-based intuition for the choice of the market prices of risk. We show the market prices of risk generate a self-consistent model if and only if they satisfy the one-period expected return equation (7). Our choice (8) is constructed to capture

the one-period expected return equation, so we now know that it will form a self-consistent affine model. There are many choices of market price of risk that satisfy the one-period expected return equation. We show that the choice (8) minimizes variance of the discount factor, and is thus analogous to the Hansen-Jagannathan (1991) discount factor and shares its many appealing properties.

D.1 Model setup

First, we set up the exponential-normal class of affine models that we specialize to account for return predictability. A vector of state variables follows

$$X_{t+1} = \mu + \phi X_t + v_{t+1};\ v_{t+1} \sim \mathcal{N}(0, V). \tag{D.7}$$

The discount factor is related to these state variables by

$$\begin{aligned} M_{t+1} &= \exp\left(-\delta_0 - \delta_1^\top X_t - \frac{1}{2}\lambda_t^\top V \lambda_t - \lambda_t^\top v_{t+1}\right) \\ \lambda_t &= \lambda_0 + \lambda_1 X_t. \end{aligned} \tag{D.8}$$

These two assumptions are no more than a specification of the time-series process for the nominal discount factor. We find log bond prices $p_t^{(n)}$ by solving the equation

$$p_t^{(n)} = \ln E_t(M_{t+1} \cdots M_{t+n}).$$

Proposition. *The log prices are affine functions of the state variables*

$$p_t^{(n)} = A_n + B_n^\top X_t. \tag{D.9}$$

The coefficients A_n and B_n can be computed recursively:

$$A_0 = 0;\ B_0 = 0$$

$$B_{n+1}^\top = -\delta_1^\top + B_n^\top \phi^* \tag{D.10}$$

$$A_{n+1} = -\delta_0 + A_n + B_n^\top \mu^* + \frac{1}{2} B_n^\top V B_n \tag{D.11}$$

where μ^* and ϕ^* are defined as

$$\phi^* \equiv \phi - V\lambda_1 \tag{D.12}$$

$$\mu^* \equiv \mu - V\lambda_0. \tag{D.13}$$

Proof. We guess the form (D.9) and show that the coefficients must obey (D.10)-(D.11). Of course, $p_t^{(0)} = 0$, so $A_0 = 0$ and $B_0 = 0$. For a one-period bond, we have

$$p_t^{(1)} = \ln E_t(M_{t+1}) = -\delta_0 - \delta_1^\top X_t,$$

which satisfies the first ($n = 1$) version of (D.10)-(D.11). We can therefore write the one-period yield as

$$y_t^{(1)} = \delta_0 + \delta_1^\top X_t.$$

The price at time t of a $n+1$ period maturity bond satisfies

$$P_t^{n+1} = E_t \left[M_{t+1} P_{t+1}^n \right].$$

Thus, we must have

$$\begin{aligned}
\exp\left(A_{n+1} + B_{n+1}^\top X_t\right) &= E_t \left[\exp\left(-\delta_0 - \delta_1^\top X_t - \frac{1}{2}\lambda_t^\top V \lambda_t - \lambda_t^\top v_{t+1} + A_n + B_n^\top X_{t+1} \right) \right] \\
&= \exp\left(-\delta_0 - \delta_1^\top X_t - \frac{1}{2}\lambda_t^\top V \lambda_t + A_n \right) E_t \left[\exp\left(-\lambda_t^\top v_{t+1} + B_n^\top X_{t+1} \right) \right]. \qquad \text{(D.14)}
\end{aligned}$$

We can simplify the second term in (D.14):

$$\begin{aligned}
&E_t \left[\exp\left(-\lambda_t^\top v_{t+1} + B_n^\top X_{t+1} \right) \right] \\
&= E_t \left[\exp\left(-\lambda_t^\top v_{t+1} + B_n^\top \mu + B_n^\top \phi X_t + B_n^\top v_{t+1} \right) \right] \\
&= E_t \left[\exp\left((-\lambda_t^\top + B_n^\top) v_{t+1} + B_n^\top \mu + B_n^\top \phi X_t \right) \right] \\
&= \exp\left(B_n^\top \mu + B_n^\top \phi X_t \right) \exp \frac{1}{2} (-\lambda_t^\top + B_n^\top) V (-\lambda_t + B_n) \\
&= \exp\left(B_n^\top \mu + B_n^\top \phi X_t \right) \exp \frac{1}{2} (\lambda_t^\top V \lambda_t - 2 B_n^\top V \lambda_t + B_n^\top V B_n).
\end{aligned}$$

Now, continuing from (D.14):

$$\begin{aligned}
A_{n+1} + B_{n+1}^\top X_t &= \left(-\delta_0 - \delta_1^\top X_t - \frac{1}{2}\lambda_t^\top V \lambda_t + A_n \right) + \left(B_n^\top \mu + B_n^\top \phi X_t \right) + \\
&\qquad + \left(\frac{1}{2}\lambda_t^\top V \lambda_t - B_n^\top V \lambda_t + \frac{1}{2} B_n^\top V B_n \right) \\
&= -\delta_0 - \delta_1^\top X_t + A_n + B_n^\top \mu + B_n^\top \phi X_t - B_n^\top V \lambda_t + \frac{1}{2} B_n^\top V B_n \\
&= -\delta_0 + A_n + B_n^\top \mu - B_n^\top V \lambda_0 + \frac{1}{2} B_n^\top V B_n - \delta_1^\top X_t + B_n^\top \phi X_t - B_n^\top V \lambda_1 X_t \\
&= \left(-\delta_0 + A_n + B_n^\top \mu - B_n^\top V \lambda_0 + \frac{1}{2} B_n^\top V B_n \right) + \left(-\delta_1^\top + B_n^\top \phi - B_n^\top V \lambda_1 \right) X_t.
\end{aligned}$$

Matching coefficients, we obtain

$$\begin{aligned}
B_{n+1}^\top &= -\delta_1^\top + B_n^\top \phi - B_n^\top V \lambda_1 \\
A_{n+1} &= -\delta_0 + A_n + B_n^\top \mu - B_n^\top V \lambda_0 + \frac{1}{2} B_n^\top V B_n.
\end{aligned}$$

Simplifying these expressions with (D.12) and (D.13), we obtain (D.10)-(D.11).

Comments

Iterating (D.10)-(D.11), we can also express the coefficients A_n, B_n in $p_t^{(n)} = A_n + B_n^\top X_t$ explicitly as

$$A_0 = 0;\ B_0 = 0$$

$$B_{n+1}^\top = -\delta_1^\top + B_n^\top \phi^* \tag{D.15}$$

$$A_{n+1} = -\delta_0 + A_n + B_n^\top \mu^* + \frac{1}{2} B_n^\top V B_n \tag{D.16}$$

$$B_n^\top = -\delta_1^\top \sum_{j=0}^{n-1} \phi^{*j} = -\delta_1^\top \left(I - \phi^{*n}\right)\left(I - \phi^*\right)^{-1} \tag{D.17}$$

$$A_n = -n\delta_0 + \sum_{j=0}^{n-1} \left(B_j^\top \mu^* + \frac{1}{2} B_j^\top V B_j \right). \tag{D.18}$$

Given prices, we can easily find formulas for yields, forward rates, etc. as linear functions of the state variable X_t. Yields are just

$$y_t^{(n)} = -\frac{A_n}{n} - \frac{B_n^\top}{n} X_t.$$

Forward rates are

$$\begin{aligned} f_t^{(n-1\to n)} &= p_t^{(n-1)} - p_t^{(n)} \\ &= (A_{n-1} - A_n) + (B_{n-1}^\top - B_n^\top) X_t \\ &= A_n^f + B_n^{f\top} X_t. \end{aligned}$$

We can find A^f and B^f from our previous formulas for A and B. From (D.17) and (D.18),

$$B_n^{f\top} = \delta_1^\top \phi^{*n-1} \tag{D.19}$$

$$A_n^f = \delta_0 - B_{n-1}^\top \mu^* - \frac{1}{2} B_{n-1}^\top V B_{n-1}. \tag{D.20}$$

In a risk neutral economy, $\lambda_0 = \lambda_1 = 0$. Thus, looking at (D.12) and (D.13), we would have the same bond pricing formulas (D.10)-(D.11) in a risk-neutral economy with $\phi^* = \phi$ and $\mu^* = \mu$. Equations (D.17) and (D.18) are recognizable as risk-neutral formulas with risk-neutral probabilities ϕ^* and μ^*. The forward rate formula (D.19) is even simpler. It says directly that the forward rate is equal to the expected value of the future spot rate and a Jensen's inequality term, under the risk-neutral measure (where ϕ^* is the autocorrelation matrix).

Note in (D.12) and (D.13) that λ_0 contributes only to the difference between μ and μ^*, and

thus contributes only to the *constant* term A_n in bond prices and yields. A homoskedastic discount factor can only give a constant risk premium. The matrix λ_1 contributes only to the difference between ϕ and ϕ^*, and only this parameter can affect the loading B_n of bond prices on state variables to give a *time-varying* risk premium. Equivalently, a time-varying risk premium needs conditional heteroskedasticity in the discount factor (D.8), and this is provided by the $\lambda_t v_{t+1}$ term of (D.8) and the variation in λ_t provided by $\lambda_1 \neq 0$.

D.2 One period returns

Here we derive the one-period expected return relation.

Proposition. *One-period returns in the affine model follow*

$$E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma_t^2(rx_{t+1}) = cov_t(rx_{t+1}, v_{t+1}^\top)\lambda_t. \tag{D.21}$$

This equation shows that the loadings λ_t which relate the discount factor to shocks in (D.8) are also the "market prices of risk" that relate expected excess returns to the covariance of returns with shocks. (This equation is similar to equation (7) in the paper. Here we use shock to prices, while the paper uses shocks to ex-post returns, but the two shocks are identical since $rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)}$.)

Proof. To show Equation (D.21), start with the pricing equation $1 = E_t\left[M_{t+1}R_{t+1}^{(n)}\right]$ which holds for the gross return $R_{t+1}^{(n)}$ on any n-period bond. Then, we can write

$$\begin{aligned}
1 &= E_t\left[M_{t+1}R_{t+1}^{(n)}\right] = E_t\left[e^{m_{t+1}+r_{t+1}^{(n)}}\right] \\
0 &= E_t\left[m_{t+1}\right] + E_t\left[r_{t+1}^{(n)}\right] + \frac{1}{2}\sigma_t^2(m_{t+1}) + \frac{1}{2}\sigma_t^2(r_{t+1}^{(n)}) + cov_t(r_{t+1}^{(n)}, m_{t+1})
\end{aligned}$$

with $m = \ln M$ and $r^{(n)} = \ln R^{(n)}$. Subtracting the same expression for the one-year yield,

$$0 = E_t\left[m_{t+1}\right] + y_t^{(1)} + \frac{1}{2}\sigma_t^2(m_{t+1}),$$

and with the 4×1 vector $rx_{t+1} = r_{t+1} - y_t^{(1)}$, we have

$$E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma_t^2(rx_{t+1}) = -cov_t(rx_{t+1}, m_{t+1}),$$

where $\sigma_t^2\left(rx_{t+1}\right)$ denotes the 4×1 vector of variances. Now, we substitute in for m_{t+1} from (D.8) to give (D.21).

D.3 A 'self-consistent' model

Equation (D.9) shows that log bond prices are linear functions of the state variables X_t. The next step is obviously to invert this relationship and infer the state variables from bond prices, so that bond prices (or yields, forward rates, etc.) themselves become the state variables. In this way, affine models end up expressing each bond price as a function of a few "factors" (e.g. level, slope and curvature) that are themselves linear combinations of bond prices.

Given this fact, it is tempting to start directly with bond prices as state variables, i.e. to write $X_t = p_t$, and specify the dynamics given by (9) as the dynamics of bond prices directly,

$$p_{t+1} = \mu + \phi p_t + v_{t+1}. \tag{D.22}$$

(It is more convenient here to keep the constants separate and define the vectors p, f to contain only the prices and forward rates, unlike the case in the paper in which p, f include a one as the first element.)

It is not obvious, however, that one can do this. The model with log prices as state variables may not be *self-consistent;* the prices that come *out* of the model may not be the same as the prices we started with; their status initially is only that of "state variables." The bond prices that come out of the model are, by (D.11) and (D.10), functions of the market prices of risk λ_0, λ_1 and risk-free rate specification δ_0 δ_1 as well as the dynamics μ and ϕ, so in fact the model will not be self-consistent for generic specifications of $\{\delta_0, \delta_1, \lambda_0, \lambda_1\}$. But there *are* choices of $\{\delta_0, \delta_1, \lambda_0, \lambda_1\}$ that ensure self-consistency. We show by construction that such choices exist; we characterize the choices, and along the way we show that any market prices of risk that satisfy the one-period return equation will generate a self-consistent model. In this sense, the market prices of risk λ which we *construct* to satisfy the return-forecasting regressions *do,* when inserted in this exponential-Gaussian model, produce an affine model consistent with the return regressions. Rather than just show that the choice (8) works, we characterize the set of market prices that work and how the choice (8) is a particular member of that set.

Definition *The affine model is* **self-consistent** *if the state variables are the prices, that is, if*

$$A_n = 0,\ B_n = e_n$$

where e_n is a vector with a one in the nth place and zeros elsewhere.

Proposition. *The affine model is self-consistent if and only if $\delta_0 = 0$, $\delta_1 = -e_1$ and the market prices of risk λ_0, λ_1 satisfy*

$$QV\lambda_1 = Q\phi - R \tag{D.23}$$

$$QV\lambda_0 = Q\mu + \frac{1}{2} Q \ \mathrm{diag}(V) \tag{D.24}$$

Here,

$$Q \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \quad R \equiv \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Q is a matrix that removes the last row, and R is a form of the first four rows of ϕ^* that generates the expectations hypothesis under the risk neutral measure, as we show below.

Proof. The proof is straightforward: we just look at the formulas for A_n and B_n in (D.10) and (D.11) and tabulate what it takes to satisfy $A_n = 0$, $B_n = e_n$. For the one-period yield we need $B_1^\top = e_1^\top$ and $A_1 = 0$. Looking at (D.10) and (D.11) we see that this requirement determines the choice of δ,

$$\delta_0 = 0; \ \delta_1 = -e_1.$$

This restriction just says to pick δ_0, δ_1 so that the one-year bond price is the first state variable. We can get there directly from

$$p_t^{(1)} = -\delta_0 - \delta_1^\top p_t.$$

For the $n = 2,\ 3,\ 4,\ 5$ period yield, the requirement $B_n^\top = e_n^\top$ in (D.10) and $A_n = 0$ in (D.11) become

$$\begin{aligned} e_n^\top &= e_1^\top + e_{n-1}^\top \phi^* \\ 0 &= e_n^\top \mu^* + \tfrac{1}{2} e_n^\top V e_n. \end{aligned} \tag{D.25}$$

In matrix notation, the self-consistency restriction $B_n = e_n$ is satisfied if ϕ^* has the form

$$\phi^* = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ a & b & c & d & e \end{bmatrix}. \tag{D.26}$$

The last row is arbitrary. We only use 5 prices as state variables, so there is no requirement that $e_6^\top = e_1^\top + e_5^\top \phi^*$.

The restriction $A_n = 0$ in equation (D.25) amounts to choosing constants in the market prices of risk to offset $1/2\ \sigma^2$ terms. The restriction is satisfied if

$$\mu_n^* = -\frac{1}{2} e_n^\top V e_n. \tag{D.27}$$

μ_5^* is similarly unrestricted.

Since only the first four rows of ϕ^* and μ^* are restricted, we can express the necessary restrictions on λ_0, λ_1 in matrix form by using our Q matrix that deletes

the fifth row,

$$Q\phi^* = R$$
$$Q\mu^* = -\frac{1}{2}Q \text{ diag}(V)$$

Finally, using the definitions of ϕ^* and μ^* in (D.12) and (D.13) we have

$$Q\left(\phi - V\lambda_1\right) = R$$
$$Q\left(\mu - V\lambda_0\right) = -\frac{1}{2}Q \text{ diag}(V)$$

and hence we have (D.23) and (D.24).

Comments

The form of (D.26) and (D.27) have a nice intuition. The one-period pricing equation is, from (D.21),

$$E_t\left(rx_{t+1}^{(n)}\right) + \frac{1}{2}\sigma_t^2\left(rx_{t+1}^{(n)}\right) = cov(rx_{t+1}^{(n)}, v_{t+1}^{\top})\lambda_t.$$

Under risk neutral dynamics, $\lambda_t = 0$. Now, from the definition

$$rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)},$$

so we should see under risk neutral dynamics

$$E_t^*\left(p_{t+1}^{(n-1)}\right) = p_t^{(n)} - p_t^{(1)} + \frac{1}{2}\sigma_t^2\left(p_{t+1}^{(n-1)}\right)$$
$$\sigma_t^2\left(rx_{t+1}^{(n)}\right) = \sigma_t^2\left(p_{t+1}^{(n-1)}\right) = \sigma^2(v_{n-1}) = V_{ii}.$$

The conditional mean in the first line is exactly the form of (D.26), and the constant in the first line is exactly the form of (D.27). In sum, the proposition just says "if you've picked market prices of risk so that the expectations hypothesis holds in the risk-neutral measure, you have a self-consistent affine model."

This logic does not constrain $E_t^*\left(p_{t+1}^{(5)}\right)$ since we do not observe $p_t^{(6)}$, and that is why the last row of ϕ^* is arbitrary. Intuitively, the 5 prices only define 4 excess returns and hence 4 market prices of risk.

From the definitions $\phi^* = \phi - V\lambda_1;\ \mu^* = \mu - V\lambda_0$ we can just pick any ϕ^* and μ^* with the required form (D.26) and (D.27), and then we can construct market prices of risk by

$$\lambda_1 = V^{-1}\left(\phi - \phi^*\right)$$
$$\lambda_0 = V^{-1}(\mu - \mu^*)$$

This is our first proof that it is *possible* to choose market prices of risk to incorporate the

return regressions, since return regressions amount to no more than a particular choice of values for ϕ, μ.

Since the last rows are not identified, many choices of ϕ^* and μ^* will generate an affine model, however. Equivalently, since we only observe four excess returns, we only can identify four market prices of risk. By changing the arbitrary fifth rows of ϕ^* and μ^*, we affect how market prices of risk spread over the 5 shocks, or, equivalently, the market price of risk of the fifth (orthogonalized) shock. The remaining question is how best to resolve the arbitrariness – how best to assign the fifth rows of ϕ^* and μ^*. At the same time, we want to produce a choice that ties the market prices of risk more closely to the actual return regressions than merely the statement that any return regression is equivalent to some ϕ, μ.

D.4 Connection to return regressions

Our next task is to connect the conditions (D.23) and (D.24) to return regressions rather than to the parameters of the (equivalent) price VAR. One reason we need to do this is to verify that the market prices of risk defined in terms of return regressions (8) satisfy conditions (D.23) and (D.24) defined in terms of the price VAR.

Denote the return regression

$$rx_{t+1} = \alpha + \beta f_t + \varepsilon_{t+1}; E(\varepsilon_{t+1}\varepsilon'_{t+1}) = \Sigma. \tag{D.28}$$

Proposition. *The parameters μ, ϕ, V of the affine model (D.22) and the parameters α, β, Σ of the return forecasting regressions (D.28) are connected by*

$$\begin{aligned} \alpha &= Q\mu \\ \beta &= (Q\phi - R)F^{-1} \\ \Sigma &= QVQ^\top. \end{aligned}$$

Here,

$$F \equiv \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

generates forward rates from prices, $f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)}$ so $f_t = Fp_t$.

Proof To connect return notation to the price VAR, we start with the definition that connects returns and prices,

$$rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)},$$

or

$$\begin{bmatrix} rx_{t+1}^{(2)} \\ rx_{t+1}^{(3)} \\ rx_{t+1}^{(4)} \\ rx_{t+1}^{(5)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_{t+1}^{(1)} \\ p_{t+1}^{(2)} \\ p_{t+1}^{(3)} \\ p_{t+1}^{(4)} \\ p_{t+1}^{(5)} \end{bmatrix} + \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} p_t^{(1)} \\ p_t^{(2)} \\ p_t^{(3)} \\ p_t^{(4)} \\ p_t^{(5)} \end{bmatrix},$$

or, in matrix notation,

$$rx_{t+1} = Qp_{t+1} - Rp_t.$$

Thus, if we have an affine model $p_t = \mu + \phi p_{t-1} + v_t$ it implies that we can forecast excess returns from prices p_t by

$$rx_{t+1} = Q\mu + (Q\phi - R)p_t + Qv_{t+1}. \tag{D.29}$$

Note that the return shocks ε_{t+1} are exactly the first four price shocks Qv_{t+1}.

The fitted value of this regression is of course exactly equivalent to a regression with forward rates f on the right hand side, and our next task is to make that transformation. From the definition $f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)}$, we can connect forward rates and prices with $f_t = Fp_t$, yielding

$$rx_{t+1} = Q\mu + (Q\phi - R)F^{-1}f_t + Qv_{t+1}.$$

Matching terms with (D.28) we obtain the representation of the proposition. We also have the covariance matrix of returns with price shocks,

$$cov(rx_{t+1}, v_{t+1}^\top) = Qcov(v_{t+1}, v_{t+1}^\top) = QV.$$

At this point, it's worth proving the logic stated in the text.

Proposition. *If market prices of risk satisfy the one period return equation, the affine model with log prices as state variables is self-consistent.* In equations, if the λ_t satisfy

$$E_t\left[rx_{t+1}^{(n)}\right] + \frac{1}{2}\sigma_t^2\left(rx_{t+1}^{(n)}\right) = cov(rx_{t+1}^{(n)}, v_{t+1}^\top)\lambda_t, \tag{D.30}$$

then they also satisfy (D.23) and (D.24).

Proof. Using the form of the return regression

$$rx_{t+1} = Q\mu + (Q\phi - R)p_t + Qv_{t+1},$$

the return pricing equation (D.30) is

$$Q\mu + (Q\phi - R)p_t + \frac{1}{2}\text{diag}(QVQ^\top) = QV\lambda_t = QV(\lambda_0 + \lambda_1 p_t).$$

Thus, matching the constant and the terms multiplying p_t, we obtain exactly the conditions (D.23) and (D.24) again

$$\begin{aligned} Q\mu + \frac{1}{2}\text{diag}(QVQ^\top) &= QV\lambda_0 \\ Q\phi - R &= QV\lambda_1. \end{aligned}$$

More intuitively, but less explicitly, we can always write a price as its payoff discounted by expected returns. For example,

$$\begin{aligned} p_t^{(3)} &= p_t^{(3)} - p_{t+1}^{(2)} + p_{t+1}^{(2)} - p_{t+2}^{(1)} + p_{t+2}^{(1)} \\ &= -\left[p_{t+1}^{(2)} - p_t^{(3)} - y_t^{(1)}\right] - y_t^{(1)} - \left[p_{t+2}^{(1)} - p_{t+1}^{(2)} - y_{t+1}^{(1)}\right] - y_{t+1}^{(1)} - y_{t+2}^{(1)} \\ &= -y_t^{(1)} - y_{t+1}^{(1)} - y_{t+2}^{(1)} - rx_{t+1}^{(3)} - rx_{t+2}^{(2)} \end{aligned}$$

so we can write

$$p_t^{(3)} = E_t\left[-y_t^{(1)} - y_{t+1}^{(1)} - y_{t+2}^{(1)} - rx_{t+1}^{(3)} - rx_{t+2}^{(2)}\right].$$

Obviously, if a model gets the right hand side correctly, it must get the left hand side correctly.

D.5 Minimum variance discount factor

Now, there are many choices of the market price of risk that satisfy our conditions; (D.23) and (D.24) only restrict the first four rows of $V\lambda_0$ and $V\lambda_1$. We find one particular choice appealing, and it is the one given in the text as Equation (8). The choice is

$$\begin{aligned} \lambda_{1:4,t} &= \Sigma^{-1}\left(E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma^2(rx_{t+1})\right) \\ &= \Sigma^{-1}\left(\alpha + \beta f_t + \frac{1}{2}diag(\Sigma)\right); \\ \lambda_{5,t} &= 0 \end{aligned} \tag{D.31}$$

(Equation (8) in the paper expresses a 4×1 vector market prices of risk, which multiply the four return shocks. Here, we want a 5×1 vector λ that can multiply the 5 price shocks in the self-consistent affine model $p_{t+1} = \mu + \phi p_t + v_{t+1}$. The return shocks are equal to the first four price shocks, so (D.31) and (8) are equivalent. Also, in this appendix we are using f_t to denote a vector that does not include a constant, so $\alpha + \beta f_t$ in (D.31) corresponds to βf_t in (8.))

Proposition. *The choice (D.31) is, in terms of the parameters $\{\mu, \phi, V\}$ of a price VAR,*

$$\lambda_0 = Q^\top \left(QVQ^\top\right)^{-1} \left(Q\mu + \frac{1}{2} Q\text{diag}\left(V\right)\right)$$
$$\lambda_1 = Q^\top \left(QVQ^\top\right)^{-1} \left(Q\phi - R\right).$$

The choices form a self-consistent affine model. (They are solutions to (D.23) and (D.24).)

Proof. As with Equation (8), Equation (D.31) satisfies the one period return equation (D.30) by construction, so we know it forms a self-consistent affine model. The rest is translating return notation to price notation. The matrix Q removes a row from a matrix with 5 rows, so the matrix

$$Q^\top \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

adds a row of zeros to a matrix with 4 rows. Therefore, we can write (D.31) as

$$\lambda_t = Q^\top \Sigma^{-1} \left(E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma^2(rx_{t+1})\right).$$

Substituting the regression in terms of the price-VAR, (D.29), and with $\Sigma = QVQ^\top$,

$$\lambda_t = Q^\top \left(QVQ^\top\right)^{-1} \left(Q\mu + (Q\phi - R)p_t + \frac{1}{2}\text{diag}\left(QVQ^\top\right)\right).$$

Matching constant and price terms, and with $\text{diag}\big(QVQ^\top\big) = Q\text{diag}(V)$

$$\lambda_0 = Q^\top \left(QVQ^\top\right)^{-1} \left(Q\mu + \frac{1}{2} Q\text{diag}\left(V\right)\right)$$
$$\lambda_1 = Q^\top \left(QVQ^\top\right)^{-1} \left(Q\phi - R\right).$$

These are easy solutions to (D.23) and (D.24); when we take $QV\lambda_i$, the terms in front of the final parentheses disappear.

This choice of market price of risk is particularly nice because it is the "minimal" choice. Justifying asset pricing phenomena with large market prices of risk is always suspicious, so why not pick the *smallest* market prices of risk that will do the job? We follow Hansen and Jagannathan (1991) and define "smallest" in terms of variance. In levels, these market prices of risk give the smallest Sharpe ratios. Our choice is the discount factor with *smallest conditional variance* that can do the job.

Proposition. *The market prices of risk (D.31) are the "smallest" market prices of risk possible, in the sense that they produce the discount factor with smallest variance.*

Proof. For any lognormal M, we have

$$\begin{aligned}\sigma_t^2(M) &= e^{2E_t(m)+2\sigma_t^2(m)} - e^{2E_t(m)+\sigma_t^2(m)} \\ &= e^{2E_t(m)+\sigma_t^2(m)}\left(e^{\sigma_t^2(m)} - 1\right).\end{aligned}$$

Thus, given the form (D.8) of the discount factor, we have

$$\sigma_t^2(M) = e^{-2\left(\delta_0+\delta_1^\top X_t\right)}\left(e^{\lambda_t^\top V\lambda_t} - 1\right).$$

Our objective is min $\sigma_t^2(m)$. We will take the one period pricing equation as the constraint, knowing that it implies a self-consistent model,

$$E_t\left[rx_{t+1}^{(n)}\right] + \frac{1}{2}\sigma_t^2\left(rx_{t+1}^{(n)}\right) = cov(rx_{t+1}^{(n)}, v_{t+1}^\top)\lambda_t = C\lambda_t,$$

where C is a 4×5 matrix. Thus, the problem is

$$\min_{\lambda_t}\frac{1}{2}\lambda_t^\top V\lambda_t \text{ s.t. } \alpha+\beta f_t+\frac{1}{2}Q\text{diag}(V) = C\lambda_t.$$

The first-order condition to this problem is

$$V\lambda_t = C^\top\xi_t,$$

where ξ_t is a 4×1 vector of Lagrange multipliers. We can now solve

$$\lambda_t = V^{-1}C^\top\xi_t.$$

Plugging this solution back into the constraint gives

$$\alpha+\beta f_t+\frac{1}{2}Q\text{diag}(V) = CV^{-1}C^\top\xi_t,$$

which we can solve for ξ_t, so that we get

$$\lambda_t = V^{-1}C^\top\left[CV^{-1}C^\top\right]^{-1}\left[\alpha+\beta f_t+\frac{1}{2}Q\text{diag}(V)\right].$$

With $C = QV$, we have the same market prices of risk we derived above,

$$\begin{aligned}\lambda_t &= V^{-1}VQ^\top\left[QVV^{-1}V^\top Q^\top\right]^{-1}\left[\alpha+\beta f_t+\frac{1}{2}Q\text{diag}(V)\right] \\ &= Q^\top\left[QVQ^\top\right]^{-1}\left[\alpha+\beta f_t+\frac{1}{2}Q\text{diag}(V)\right].\end{aligned}$$