Bottom line

- Forecast 1 year treasury bond returns, over 1 year rate:
  \[ r x_{t+1}^{(n)} = a_n + b'_n f_t + \varepsilon_{t+1}^{(n)} \]

- \( R^2 \) up to 44%, up from Fama-Bliss / Campbell Shiller 15%

- A single factor \( \gamma' f \) forecasts bonds of all maturities. High expected returns in “bad times.”

- Tent-shaped factor is correlated with slope but is not slope. Improvement comes because it tells you when to bail out – when rates will rise in an upward-slope environment.
Background – Expectations and Fama-Bliss.

1. Expectations hypothesis. Expected returns are constant over time.

\[ r^{(n)}_{x,t+1} = a_n + 0 \times x_t + \varepsilon^{(n)}_{t+1} \]

2. Fama-Bliss.

(a) Expectations Hypothesis: \( \beta = 0 \). Instead, \( \beta \approx 1 \). If the \( n \) year forward is 1% higher than the spot, then the \( n \)-year bond will earn 1% more on average

(b) \( R^2 \approx 0.15 \); Held up well in 1990s

<table>
<thead>
<tr>
<th>Maturity ( n )</th>
<th>( \beta ) s.e.</th>
<th>( R^2 )</th>
<th>( \chi^2(1) )</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.99 (0.33)</td>
<td>0.16</td>
<td>18.4</td>
<td>\langle 0.00 \rangle</td>
</tr>
<tr>
<td>3</td>
<td>1.35 (0.41)</td>
<td>0.17</td>
<td>19.2</td>
<td>\langle 0.00 \rangle</td>
</tr>
<tr>
<td>4</td>
<td>1.61 (0.48)</td>
<td>0.18</td>
<td>16.4</td>
<td>\langle 0.00 \rangle</td>
</tr>
<tr>
<td>5</td>
<td>1.27 (0.64)</td>
<td>0.09</td>
<td>5.7</td>
<td>\langle 0.02 \rangle</td>
</tr>
</tbody>
</table>
Regressions of bond excess returns on all forward rates, not just matched $f - y$.

The same linear combination of forward rates forecasts all maturities’ returns.
A single factor for expected bond returns

\[ rx_{t+1}^{(n)} = b_n \left( \gamma_0 + \gamma_1 y_{t}^{(1)} + \gamma_2 f_{t}^{(1\rightarrow2)} + \ldots + \gamma_5 f_{t}^{(4\rightarrow5)} \right) + \varepsilon_{t+1}^{(n)}; \quad \frac{1}{4} \sum_{n=2}^{5} b_n = 1. \]

- Two step estimation; first \( \gamma \) then \( b \).

**Table 1 Estimates of the single-factor model**

**A. Estimates of the return-forecasting factor, \( r x_{t+1} = \gamma^\top f_t + \varepsilon_{t+1} \)**

<table>
<thead>
<tr>
<th>( \gamma_0 )</th>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_4 )</th>
<th>( \gamma_5 )</th>
<th>( R^2 )</th>
<th>( \chi^2(5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS estimates</td>
<td>(-3.24)</td>
<td>(-2.14)</td>
<td>(0.81)</td>
<td>(3.00)</td>
<td>(0.80)</td>
<td>(-2.08)</td>
<td>(0.35)</td>
</tr>
</tbody>
</table>

**B. Individual-bond regressions**

- Restricted

  \[ rx_{t+1}^{(n)} = b_n \left( \gamma^\top f_t \right) + \varepsilon_{t+1}^{(n)} \]

<table>
<thead>
<tr>
<th>( n )</th>
<th>( b_n )</th>
<th>( R^2 )</th>
<th>( \chi^2(5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.47</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.87</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.24</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.43</td>
<td>0.34</td>
<td></td>
</tr>
</tbody>
</table>

- Unrestricted

  \[ rx_{t+1}^{(n)} = \beta_n f_t + \varepsilon_{t+1}^{(n)} \]

<table>
<thead>
<tr>
<th>( n )</th>
<th>( R^2 )</th>
<th>( \chi^2(5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.32</td>
<td>121.8</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>113.8</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>115.7</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>88.2</td>
</tr>
</tbody>
</table>
• $\gamma$ capture tent shape.

• $b_n$ increase steadily with maturity.

• Restricted model $b_n \gamma$ almost perfectly matches unrestricted coefficients. (well below 1σ)

• $R^2 = 0.34 - 0.37$ up from $0.15 - 0.17$. And we’ll get to 0.44! Very significant rejection of $\gamma = 0$

• $R^2$ almost unaffected by the restriction. Restriction looks good in the graph.

• See paper version of table 1 for standard errors, joint tests including small sample, unit roots, etc. Bottom line: highly significant; EH is rejected, improvement on FB/3 factor models is significant.
More lags

\[ rx_{t+1}^{(n)} = a_n + b'_n f_{t-i} + \varepsilon_{t+1}^{(n)} \]
More lags are significant, same pattern. Suggests moving averages

\[ rx_{t+1}^{(n)} = a_n + b_n \gamma' (\alpha_0 f_t + \alpha_1 f_{t-1} + \ldots + \alpha_k f_{t-k}) + \varepsilon_{t+1}^{(n)} \]

\[ = a_n + b_n \left[ \alpha_0 (\gamma' f_t) + \alpha_1 (\gamma' f_{t-1}) + \ldots + \alpha_k (\gamma' f_{t-k}) \right] + \varepsilon_{t+1}^{(n)} \]

<table>
<thead>
<tr>
<th>( k )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 )</td>
<td>0.35</td>
<td>0.41</td>
<td>0.43</td>
<td>0.44</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Interpretation: Yields should be Markov, so a small transitory measurement error. \( f_{t-1/12} \) is informative about the true \( f_t \), so it enters with the same pattern.
**Stock Return Forecasts**

Table 3. Forecasts of excess stock returns (VWNYSE)

\[ r\bar{x}_{t+1} = a + bx_t + \varepsilon_{t+1} \]

<table>
<thead>
<tr>
<th>$\gamma^\top f$</th>
<th>(t)</th>
<th>$d/p$</th>
<th>(t)</th>
<th>$y^{(5)} - y^{(1)}$</th>
<th>(t)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.73 (2.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>3.56 (1.80)</td>
<td>3.29 (1.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td>1.87 (2.38)</td>
<td></td>
<td>-0.58</td>
<td>(−0.20)</td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>1.49 (2.17)</td>
<td>2.64 (1.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>MA $\gamma^\top f$</td>
<td>2.11 (3.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>MA $\gamma^\top f$</td>
<td>2.23 (3.86)</td>
<td>1.95 (1.02)</td>
<td>-1.41 (−0.63)</td>
<td></td>
<td>0.15</td>
<td></td>
</tr>
</tbody>
</table>

- 5 year bond had $b = 1.43$. Thus, $1.73 - 2.11$ is what you expect for a perpetuity.

- Does better than D/P and spread; Drives out spread; Survives with D/P

- A common term risk premium in stocks, bonds. Reassurance on fads & measurement errors
Interest Rate Forecasts

Table A4. Forecasting short rate changes $y_{t+1}^{(1)} - y_t^{(1)}$

<table>
<thead>
<tr>
<th>$f_t^{(2)} - y_t^{(1)}$</th>
<th>$y_t^{(1)}$</th>
<th>$f_t^{(2)}$</th>
<th>$f_t^{(3)}$</th>
<th>$f_t^{(4)}$</th>
<th>$f_t^{(5)}$</th>
<th>$R^2$</th>
<th>$\chi^2/p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB:</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>s.e.</td>
<td>(0.26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\langle 0.98 \rangle$</td>
</tr>
<tr>
<td>CP:</td>
<td>-0.02</td>
<td>0.41</td>
<td>-1.21</td>
<td>-0.29</td>
<td>0.89</td>
<td>0.19</td>
<td>82.7</td>
</tr>
<tr>
<td>s.e.</td>
<td>(0.17)</td>
<td>(0.40)</td>
<td>(0.29)</td>
<td>(0.22)</td>
<td>(0.17)</td>
<td></td>
<td>$\langle 0.00 \rangle$</td>
</tr>
</tbody>
</table>

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

- **Expectations:** $E_t \left( rx_{t+1}^{(2)} \right) = \text{constant}$
- **Fama-Bliss:** $E_t \left( y_{t+1}^{(1)} - y_t^{(1)} \right) = \text{constant}$
- **CP:** $E_t \left( rx_{t+1}^{(2)} \right)$ varies more than $\left( f_t^{(2)} - y_t^{(1)} \right)$: must be able to forecast short rates (capital gains on long bonds, not just “ride yields”)
- **Pattern of coefficients is exactly $\gamma$. $\gamma'f$ forecasts short rate changes.**
- $\gamma'f$ and slope are correlated. Both show a rising yield curve but no rate rise

- $\gamma'f$ improvement in many episodes. $\gamma'f$ says get out in 1984, 1987, 1994, 2004(?)
Regression forecasts $\hat{\gamma}^T f_t$. “Real-time” re-estimates the regression at each $t$ from 1965 to $t$. 
Trading rule

\[
x_{t+1} \times E_t(x_{t+1}) = x_{t+1} \times [\gamma^\top (\alpha_0 f_t + \alpha_1 f_{t-1} + \alpha_2 f_{t-2})].
\]

“Trading rule” profits, using full-sample and real-time estimates of the return-forecasting factor.
Cumulative trading rule profits; cumulative value of $\bar{x}_{t+1} \times E_t(\bar{x}_{t+1})$. 
\( \gamma' f \) is correlated with business cycles, and lower frequency. (Level, not growth.) “business cycle related risk premium.”
Relation to factor models (why is this news?)

A. Expected return factor $\gamma^*$

B. Yield factors

C. Return Predictions

D. Forward rate forecasts
Panel A: Yields are a linear combination of forwards. $\gamma' f = \gamma^* y$; which full set is a matter of taste.

$\gamma^* \approx$ Slope plus 4-5 spread.

Panel B: $\gamma' f$ has nothing to do with slope (symmetry: $\gamma' \ linear = 0$) and curvature (curved at the long, not short end).

Panel C: You can’t approximate $\gamma' f$ well with level, slope, and curvature factors. (See table below)

- Moral 1 Term structure models need L, S, C to get $\Delta y_{t+1}$ and $\gamma' f$ to get $E_{tr x_{t+1}}$
- Adding $\gamma' f$ will not help much to hit yields (pricing errors) but it will help to get transition dynamics right (i.e. expected returns, yield differences)

Moral 2. You can’t first reduce to L, S, C, then examine $E_{tr x_{t+1}} \rightarrow$ Reason #1 this was missed.

Panel D: Stable as we add forward rates.
• Is $\gamma'f$ forecast *significantly* better than forecasts using yield curve factors or simple spreads?

See Table 4 of paper,

regression of average returns on 3 forward factors

<table>
<thead>
<tr>
<th></th>
<th>const</th>
<th>curve</th>
<th>slope</th>
<th>level</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>-0.02</td>
<td></td>
<td>0.17</td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td>t</td>
<td>-0.98</td>
<td></td>
<td></td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>-0.04</td>
<td>1.39</td>
<td>0.17</td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td>t</td>
<td>-2.44</td>
<td>3.15</td>
<td>1.59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

regression of average returns on $x$

<table>
<thead>
<tr>
<th></th>
<th>const</th>
<th>x</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.00</td>
<td>0.47</td>
<td>0.36</td>
</tr>
<tr>
<td>t</td>
<td>0.08</td>
<td>7.99</td>
<td></td>
</tr>
</tbody>
</table>

regression of average returns on $x$ and 3 forward factors

<table>
<thead>
<tr>
<th></th>
<th>const</th>
<th>x</th>
<th>curve</th>
<th>slope</th>
<th>level</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>0.00</td>
<td>0.47</td>
<td>-0.06</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.36</td>
</tr>
<tr>
<td>t</td>
<td>0.02</td>
<td>4.95</td>
<td>-0.10</td>
<td>-0.10</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>
• How would you integrate this into an affine model?

Paper: shows you how to *construct* market prices of risk so that an affine model *exactly* matches this (any) return regression. (Also see “Decomposing the Yield Curve”)

Coefficients $rx_{t+1}^{(n)} = a_n + b_n f_t$ implied by $y_{t+1/12} = \phi y_t + \varepsilon_{t+12}$, then $\phi^{12}$. 
• Small $R^2$. No single-factor. Pattern looks like measurement error. Nothing there! Why? Lags matter – yields are not a VAR(1). (VAR(12) works here)

• Moral: Must look at annual returns directly (or fit ARMA(1,1), or deal with measurement error) to see annual horizon return forecasts.
Is this all measurement error?

Danger: if $p_t$ is measured too high, then $r_{t+1} = p_{t+1} - p_t$ will be too low, and a high $p_t$ will seem to forecast a low $r_{t+1}$. Is this all there is to our results?

1. No. Lags also forecast, with no common price.

2. No. $\gamma' f$ also forecasts stock returns with no common price.

3. Measurement error gives a pattern that the $n$ period yield at $t$ forecasts the $n$ period bond return. It does not give a common factor ($m$ yield helps to forecast $n$ bond return) Measurement error cannot produce our central finding, the single factor.
Testing the single factor model

- Paper Table 6: the single factor model is dramatically rejected! (Joint not individual coefficients)

- Paper Table 6: the single factor model is dramatically rejected! (Joint not individual coefficients)

\[
\begin{bmatrix}
rx_{t+1}^{(2)} \\
\vdots \\
rx_{t+1}^{(5)}
\end{bmatrix}
= (\alpha) + \begin{bmatrix}
\vdots & \vdots & \vdots \\
\beta & \vdots & \vdots
\end{bmatrix}
\begin{bmatrix}
y_{t}^{(1)} \\
f_{t}^{(2)} \\
\vdots \\
f_{t}^{(5)}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{t+1}^{(2)} \\
\varepsilon_{t+1}^{(3)} \\
\vdots \\
\varepsilon_{t+1}^{(5)}
\end{bmatrix}
\tag{1}
\]

vs

\[
\begin{bmatrix}
rx_{t+1}^{(2)} \\
\vdots \\
rx_{t+1}^{(5)}
\end{bmatrix}
= (\alpha) + \begin{bmatrix}
b_{2} \\
\vdots \\
b_{5}
\end{bmatrix}
\begin{bmatrix}
\gamma_{1} & \gamma_{2} & \cdots & \gamma_{5}
\end{bmatrix}
\begin{bmatrix}
y_{t}^{(1)} \\
f_{t}^{(2)} \\
\vdots \\
f_{t}^{(5)}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{t+1}^{(2)} \\
\varepsilon_{t+1}^{(3)} \\
\vdots \\
\varepsilon_{t+1}^{(5)}
\end{bmatrix}
\tag{2}
\]

- \[
rx_{t+1} = \beta f_t + \varepsilon_{t+1} \tag{1}
\]
- \[
rx_{t+1} = b \gamma' f_t + \varepsilon_{t+1} \tag{2}
\]
• GMM

\[ E_T(f_t \otimes \varepsilon_{t+1}) = 0 \]  \hspace{1cm} (1)

\[ E_T \left[ f_t \otimes (1^n_4 \varepsilon_{t+1}) \right] = 0 \rightarrow \gamma(2) \]

\[ E_T \left[ (\gamma' f_t) \otimes \varepsilon_{t+1} \right] = 0 \rightarrow b \]

• Results: If we do optimal GMM on (2), weird parameters. And huge rejections.

• Why??? Restriction:

\[ r x_{t+1}^{(n)} - b_n \gamma' f_t = \varepsilon_{t+1}^{(n)} \]

should not be predictable. \((E(\varepsilon \otimes f_t) = 0)\) Since

\[ \overline{r} x_{t+1} = \gamma' f_t + \overline{\varepsilon}_{t+1} \]

then (Portfolio interpretation)

\[ r x_{t+1}^{(n)} - b_n \times \overline{r} x_{t+1} = \overline{r}_n f_t + w_{t+1}^{(n)} \].
### Table 7. Forecasting the failures of the single-factor model

#### A. Coefficients and t-statistics

<table>
<thead>
<tr>
<th>Left hand var.</th>
<th>const.</th>
<th>$y^{(1)}_t$</th>
<th>$y^{(2)}_t$</th>
<th>$y^{(3)}_t$</th>
<th>$y^{(4)}_t$</th>
<th>$y^{(5)}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rx_{t+1} - b_2 \bar{rx}_{t+1}$</td>
<td>-0.11</td>
<td>-0.20</td>
<td><strong>0.80</strong></td>
<td>-0.30</td>
<td>-0.66</td>
<td>0.40</td>
</tr>
<tr>
<td>(t-stat)</td>
<td>(-0.75)</td>
<td>(-1.43)</td>
<td><strong>(2.19)</strong></td>
<td>(-0.90)</td>
<td>(-1.94)</td>
<td>(1.68)</td>
</tr>
<tr>
<td>$rx_{t+1} - b_3 \bar{rx}_{t+1}$</td>
<td>0.14</td>
<td>0.23</td>
<td>-1.28</td>
<td><strong>2.36</strong></td>
<td>-1.01</td>
<td>-0.30</td>
</tr>
<tr>
<td>(t-stat)</td>
<td>(1.62)</td>
<td>(2.22)</td>
<td>(-5.29)</td>
<td><strong>(11.24)</strong></td>
<td>(-4.97)</td>
<td>(-2.26)</td>
</tr>
<tr>
<td>$rx_{t+1} - b_4 \bar{rx}_{t+1}$</td>
<td>0.21</td>
<td>0.20</td>
<td>-0.06</td>
<td>-1.18</td>
<td><strong>1.84</strong></td>
<td>-0.82</td>
</tr>
<tr>
<td>(t-stat)</td>
<td>(2.33)</td>
<td>(2.39)</td>
<td>(-0.33)</td>
<td>(-8.45)</td>
<td><strong>(9.13)</strong></td>
<td>(-5.48)</td>
</tr>
<tr>
<td>$rx_{t+1} - b_5 \bar{rx}_{t+1}$</td>
<td>-0.24</td>
<td>-0.23</td>
<td>0.55</td>
<td>-0.88</td>
<td>-0.17</td>
<td><strong>0.72</strong></td>
</tr>
<tr>
<td>(t-stat)</td>
<td>(-1.14)</td>
<td>(-1.06)</td>
<td>(1.14)</td>
<td>(-2.01)</td>
<td>(-0.42)</td>
<td><strong>(2.61)</strong></td>
</tr>
</tbody>
</table>

#### B. Regression statistics

<table>
<thead>
<tr>
<th>Left hand var.</th>
<th>$R^2$</th>
<th>$\chi^2(5)$</th>
<th>$\sigma(\hat{\gamma}^\top y)$</th>
<th>$\sigma(\text{lhs})$</th>
<th>$\sigma(b(n)\gamma^\top y)$</th>
<th>$\sigma(rx^{(n)}_{t+1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rx_{t+1} - b_2 \bar{rx}_{t+1}$</td>
<td>0.15</td>
<td>41</td>
<td>0.17</td>
<td>0.43</td>
<td>1.12</td>
<td>1.93</td>
</tr>
<tr>
<td>$rx_{t+1} - b_3 \bar{rx}_{t+1}$</td>
<td>0.37</td>
<td>151</td>
<td>0.21</td>
<td>0.34</td>
<td>2.09</td>
<td>3.53</td>
</tr>
<tr>
<td>$rx_{t+1} - b_4 \bar{rx}_{t+1}$</td>
<td>0.33</td>
<td>193</td>
<td>0.18</td>
<td>0.30</td>
<td>2.98</td>
<td>4.90</td>
</tr>
<tr>
<td>$rx_{t+1} - b_5 \bar{rx}_{t+1}$</td>
<td>0.12</td>
<td>32</td>
<td>0.21</td>
<td>0.61</td>
<td>3.45</td>
<td>6.00</td>
</tr>
</tbody>
</table>
• These are predictable!

• With large $R^2$ and statistical significance!

• But they are tiny.

• Pattern: diagonal. $y^{(n)}$ out of line, it reverts back next period. No common factor.

• Tiny measurement errors or tiny (but profitable if you can leverage) “spread trades”

• The single-factor $\gamma'f$ accounts for all the economically important variation in expected returns.

• This is why we do a two-step OLS not efficient GMM estimation. Efficient GMM of a single factor model weights by $R^2$, not size.
More stuff in paper.

Q: What about...

• Subsamples? Yes.

• Other data? McCulloch-Kwan data, not just FB interpolation