Market Expectations in the Cross Section of Present Values

by Kelly, B., and S. Pruitt

Stefan Nagel

Stanford University, NBER, CEPR

January 2013

The problem tackled by this paper

- Market return
  \[ r_{mt+1} = \mu_t + \eta_{t+1} \]
- Present-value decomposition of market M/B ratio
  \[ x_{mt} \approx a - b\mu_t + b_gg_t \]
  with VAR(1) dynamics of book ROE, \( g_t \), and expected return, \( \mu_t \)
- Presence of \( g_t \) obscures “signal” \( \mu_t \)
- This paper: Cross-sectional information helps filter out \( \mu_t \) to improve predictive regressions
- Much of my discussion focuses on Kelly and Pruitt (2012, “3PRF”, WP), the paper that supplies the methodology
Figure 1: Out-of-Sample $R^2$ by Sample Split Date, One Year Returns

Notes: Out-of-sample $R^2$ across sample split dates. Forecasts are based on a single PLS factor from 100 book-to-market ratios of size and value-sorted portfolios, the aggregate book-to-market ratio, the cross-section premium of Polk et al. (2006), the consumption-wealth ratio of Lettau and Ludvigson (2001), and the first three principal components of the 100 book-to-market ratio cross section.

Figure 3: Out-of-Sample $R^2$ by Sample Split Date, One Month International Returns

Notes: Out-of-sample percentage $R^2$ across sample split dates for forecasts of one month international stock returns using a single PLS factor from 42 price-dividend ratios of high value and low value portfolios across 21 countries (Fama and French (1998)). See Section III.A.5 for list of countries.
How does it work? Example

- Define: $g_t \equiv \text{ROE component orthogonal to expected returns}$
- Assume: Three-asset cross-section where M/B ratios load on $\mu_t$ and $g_t$ with cross-sectionally **uncorrelated** loadings

  
  \[
  \begin{align*}
  x_{1t} &= 2\mu_t - 1.5g_t & \tilde{x}_{1t} &= \mu_t - 0.5g_t \\
  x_{2t} &= \mu_t & \tilde{x}_{2t} &= g_t \\
  x_{3t} &= -1.5g_t & \tilde{x}_{3t} &= -\mu_t - 0.5g_t
  \end{align*}
  \]

- First-stage t.s. regressions of $x_{it}$ on $r_{mt+1}$ (large $T$): slopes proportional to

  \[
  \begin{align*}
  \phi_1 &= 2 & \tilde{\phi}_1 &= 1 \\
  \phi_2 &= 1 & \tilde{\phi}_2 &= 0 \\
  \phi_3 &= 0 & \tilde{\phi}_3 &= -1
  \end{align*}
  \]

- Second-stage c.s. regressions of $\tilde{x}_{it}$ on $\tilde{\phi}_i$ each $t$: slopes

  \[
  F_t = \text{const.} \times \left[ (\mu_t - 0.5g_t) + 0 - (\mu_t - 0.5g_t) \right] = \text{const.} \times \mu_t
  \]

How does it work? Crucial assumption

- Assumption that loadings on $\mu_t$ and $g_t$ are c.s. **uncorrelated** is important: Second stage regression slopes $F_t$ are then...
  - “Long” in assets with positive M/B loadings on $\mu_t$
  - “Short” in assets with negative M/B loadings on $\mu_t$
  - “Long” and “short” in assets with M/B that load similarly on $g_t$: $g_t$ exposure cancels out
  - As $N \to \infty$, sample c.s. correlation closer to zero population c.s. correlation: $\mu_t$ consistently estimated

- But what if loadings on $\mu_t$ and $g_t$ are c.s. **correlated**?
  - Lucky: For stock market application in JF paper, correlation seems to be close to zero
  - But this may not be true in other applications
  - Remedy: Use proxies for $g_t$ in addition to $\mu_t$ proxy
  - But that means we have to take a stand on all of the systematic factors driving $M/B$ ratios
Concern: Large $N$, small $T$

- What are the properties of the estimator under the null hypothesis of no predictability?
- Simulation: No-predictability & pure-noise M/B null

\[
\begin{align*}
    r_{mt+1} &= \eta_{t+1} \\
    x_t &= \epsilon_t
\end{align*}
\]

where

\[
\begin{pmatrix}
    \eta_{t+1} \\
    \epsilon_t
\end{pmatrix} \sim \mathcal{N}(0, I_{N+1})
\]

Large $N$, small $T$: Spurious fit as $N$ grows

Mean third-stage R-squared under no-predictability null (100 simulations for each $(N, T)$ pair)
### Large $N$, small $T$: Example with $T = 3$ and $N = 9$

<table>
<thead>
<tr>
<th>$r_{mt+1}$</th>
<th>0.91</th>
<th>-1.49</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{it}$</td>
<td>-1.35</td>
<td>0.91</td>
</tr>
<tr>
<td>$\hat{\phi}_i$</td>
<td>-0.52</td>
<td>-1.26</td>
</tr>
<tr>
<td></td>
<td>-0.82</td>
<td>-0.44</td>
</tr>
<tr>
<td></td>
<td>-0.20</td>
<td>1.07</td>
</tr>
<tr>
<td>$\tilde{F}_t$</td>
<td>-0.93</td>
<td>-2.62</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>-0.06</td>
</tr>
<tr>
<td>$\pi_{t+1}$</td>
<td>1.94</td>
<td>-1.14</td>
</tr>
<tr>
<td>$\tilde{y}_t$</td>
<td>0.40</td>
<td>-0.22</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0044</td>
<td>0.0040</td>
</tr>
</tbody>
</table>

**Pure noise returns**

**First stage (sorted)**

**Pure noise M/B**

**Second stage**

**Third stage (fitted)**

### $N$ and $T$ grow simultaneously: Spurious fit with $N = T^2$

Mean third-stage R-squared under no-predictability null with $N = T^2$

(100 simulations for each $(N, T)$ pair)
Spurious predictability

- Thus, when $N$ is not small relative to $T$, there is a bias that overstates predictability
- It seems that this is not just a small-sample bias: also asymptotic bias if $N$ grows sufficiently fast relative to $T$
  - Theorem 1 in Kelly and Pruitt (2012, “3PRF”) seems to need additional assumption: $N$ cannot grow too fast relative to $T$
- Not a concern for the empirical results in the JF paper, as out-of-sample tests are not affected by this bias
- But concern underscores importance of out-of-sample testing
- Calls for further study of small-sample and asymptotic properties of this estimator

Summing up

- Nice idea
- Impressive empirical results: Strong out-of-sample predictability of stock market returns
- Some further work necessary on the properties of the 3PRF estimator
  - Correlated factor loadings
  - Small-$T$/large-$N$ behavior
  - Asymptotic behavior when $N$ grows fast relative to $T$
- These concerns do not affect out-of-sample tests: results in the JF paper are robust