Discussion of
“Market Efficiency in the Age of Big Data”

Ian Martin, Stefan Nagel

Stéphane Bonhomme
University of Chicago

Conference “Expectations in Macroeconomic and Financial Models”
University of Chicago, June 26 2020
The actors

• **Nature** draws dividends according to the model

\[ \Delta y_1 = X'g + e_1, \quad \Delta y_2 = X'g + e_2, \]

where \((e_1, e_2) | X, g \sim \mathcal{N}(0, \Sigma_e)\), and \(g | X \sim \mathcal{N}(0, \Sigma_g)\).

• **Investors** know the model. They try to learn \(g\) using Bayesian updating. They know the fixed characteristics \(X\) and the (common) prior for \(g\). By contrast, \(e_2\) is unpredictable.

• **The econometrician** is interested in how prices are determined. She entertains the model

\[ p_1 = \mathbb{E}_1(y_2) = y_1 + X'\mathbb{E}_1(g). \]

• To assess whether this model is adequate, her strategy is to regress returns on characteristics and test for predictability.
Predicability (in-sample)

- However, in this setting the econometrician’s strategy is flawed, especially when the dimension of $X$ is large.

- To see this, note that returns are

  $$r_2 = y_2 - p_1 = \frac{X'(g - \mathbb{E}_1(g))}{\text{predictable?}} + e_2$$

- When there are few characteristics, $\mathbb{E}_1(g) \approx \hat{g}^{\text{OLS}}$ (by Bernstein von Mises), so $X'(g - \mathbb{E}_1(g)) \approx X'(\sum_i X_iX_i')^{-1}\sum_i X_ie_{1i}$, and it is easy to adjust the critical value of the test.

- However, in high dimensions (i.e., in a modern big data environment with many potential predictors) the equivalence breaks down and the authors show conventional tests become uninformative.
Many parameters... or just one?

- Although $g$ is high-dimensional, the model for dividends is in fact rather parsimonious.

- Consider the authors’ specification: $\Sigma_e = I$, $\Sigma_g = \theta J I$.

- This is a linear random coefficients regression model that depends on a single parameter: $\theta$.

... this is hardly a high-dimensional setting :)

- The problem the authors highlight comes from the fact that the econometrician does not exploit this simple structure and uses a conventional OLS-based test.
Can the econometrician do better? Case 1: when she knows the prior

- Assuming the econometrician bases her test of predictability on OLS residuals is well motivated given the asset pricing literature. However, this is clearly not the right approach here.

- Suppose first that the econometrician knows the model/prior.

- Then a simple approach to test for correct specification of the pricing model is to compare

  \[ E_{\text{data}}(r_2 \varphi(X)) \text{ and } E_{\text{model}}^{\theta}(r_2 \varphi(X)), \]

  where \( \varphi(X) \) is a low-dimensional vector of instruments, and \( E_{\text{model}}^{\theta} \) is computed based on the investors’ prior.

- How to choose \( \varphi(X) \) is interesting, but this is a question of efficiency. The test will be consistent under standard conditions.
Can the econometrician do better? Case 2: when she does not know the prior

- Suppose now, more realistically, that the econometrician does not know the prior, while keeping all the other assumptions of the model.

- Consider again the quantity $\mathbb{E}(r_2 \varphi(X))$, for some vector $\varphi(X)$.

- We have, under the null that the pricing model is correct,

$$
\mathbb{E}(r_2 \varphi(X)) = \mathbb{E}(X(g - \mathbb{E}_1(g)) \varphi(X)) \\
= \mathbb{E}(X(\mathbb{E}_1(g) - \mathbb{E}_1(g)) \varphi(X)) = 0,
$$

by the law of iterated expectations.

- Hence it is still easy to build a consistent test, irrespective of the prior, by choosing a low-dimensional $\varphi(X)$. 

A simple strategy

• Under the null that the pricing model is correct we have

\[ \tilde{h}_2 = \frac{1}{N} \sum_i \varphi(X_i)r_{2i} \mid X \sim \mathcal{N}(0, V(X)), \]

where \( V(X) \) depends on \( \Sigma_e \) and the posterior variance of \( g \).

• In the spirit of the paper, suppose the econometrician computes critical values under the (wrong) assumption that \( E_1(g) = g \). In this case

\[ \tilde{T}_{re} \equiv \tilde{h}'_2 \left[ \frac{1}{N^2} \sum_i \varphi(X_i)\varphi(X_i)' \right]^{-1} \tilde{h}_2 - m \approx \mathcal{N}(0, 1), \]

where \( m \) is the (small!) dimension of \( \varphi(X) \).

• To illustrate, I conduct a small simulation where \( \varphi(X) \) is the first component of \( X \).
Using a small subset of characteristics (here, only one) helps

Note: size of two tests, based on OLS as in the paper (in blue), and based on projected OLS (in green). The nominal size is shown in black. N=500, averages over 5000 simulations.
Revisiting the assumptions of the framework

- This simple strategy is possible here because: (1) investors’ prior coincides with the population distribution of $g$, and (2) the prior has a simple form.

- Let me first focus on (2).

- An important assumption is that $g$ is independent of $X$.

- How is this justified? $X$ are fixed over time. In general, in dynamic learning problems one expects priors to depend on initial conditions.

- Using a terminology from panel data econometrics, the authors’ setup assumes “independent random-effects”, which we know can be restrictive.
A correlated prior

• Suppose now that \( g \mid X \) follows a prior which is correlated with \( X \). This is still the distribution of \( g \mid X \) in the population.

• It turns out that the simple testing strategy based on \( \varphi(X) \) still works, even though \( g \) and \( X \) are correlated and the econometrician does not know the prior.

• Indeed, as before,

\[
\mathbb{E}(r_2 \varphi(X)) = \mathbb{E}(X(g - \mathbb{E}_1(g))\varphi(X)) = \mathbb{E}(X(\mathbb{E}_1(g) - \mathbb{E}_1(g))\varphi(X)) = 0.
\]

Critical values will depend on the (correlated) prior – maybe need \( X_t \) to vary over time?

• In this light a key assumption in the setup is (1), that investors’ prior is correct.
What about investors’ prior(s) being wrong?

• If (1) fails, the econometrician faces a twofold learning problem, since she does not know the distribution of $g$ in the population or the prior of investors.

• Then, in general

$$\mathbb{E}(r_2 \varphi(X)) = \mathbb{E} \left( X(g - \mathbb{E}_1 (g)) \varphi(X) \right) \neq 0.$$  

• Dividends may be used to learn about the distribution of $g$. Under independence: random coefficients models (Beran and Hall, 1992). Alternative: exploit the panel dimension (and $X_t$ varies over time?).

• Returns may be used to learn about the investors’ (common) prior...

... but (only?) under the assumption that they use Bayesian updating.
A (really!) great paper

- The main point that including many predictors can generate spurious predictability seems important. This may help explain some key empirical findings in the asset pricing literature.

- The framework is clean and parsimonious, and highlights many intuitions (e.g., out-of-sample versus in-sample tests).

- Although I see the historical motivation for focusing on OLS-based tests, I miss an exploration of other testing strategies that could be more successful in a big data environment.

- Can the analysis shed light on how to improve the practice of market efficiency testing? (perhaps in the next paper?)