Machine Learning in Asset Pricing Ch 5(5.3:end)

Nagel (2021) presented by Kevin Mei

April 2022

The University of Texas at Austin

Recall: where we left off

Ch 5.1: we model a world where dividend growth Δy_t is partially predictable based on firm characteristics X:

$$\Delta y_t = Xg + e_t$$

Ch 5.2: We modeled investors who try to learn about g. Suppose they use OLS and model:

$$\tilde{g}_{OLS,t} = (X'X)^{-1}X'\bar{\Delta y}_t$$

where $\bar{\Delta y}_t$ is the historical sample average cash flow growth. Then realized returns are:

$$r_{t+1} = y_{t+1} - p_t = X(g - \tilde{g}_{OLS,t}) + e_{t+1} = -X(X'X)^{-1}X'\bar{e}_t + e_{t+1}$$

Return Predictability

Suppose we start with the posterior g:

$$\tilde{g}_t = \Gamma_t (X'X)^{-1} X' \Delta \bar{y}_t$$

where Γ_t is the shrinkage matrix.

The expression for returns is now:

$$r_{t+1} = X(I_J - \Gamma_t)g - X\Gamma_t (X'X)^{-1} X' \bar{e}_t + e_{t+1}$$

With diffuse prior (i.e., OLS learning) then $\Gamma_t = I_J$ and the first term disappears.

With an informative prior, we can interpret this term as "underreaction" to the fundamental information in X, due to shrinkage

Return Predictability: In-Sample

The econometrician runs this regression, they get:

$$h_{t+1} = (X'X)^{-1}X'r_{t+1} = (I_J - \Gamma_t)g - \Gamma_t(X'X)^{-1}X'\bar{e}_t + (X'X)^{-1}X'e_{t+1}$$

Suppose we have an in-sample portfolio strategy with weights $w_t = \frac{1}{N}Xh_{t+1}$, then the portfolio returns $r_{IS,t+1} = w_t' r_{t+1} = h_{t+1}'(X'X)h_{t+1}$ would be:

$$\begin{aligned}
\mathbb{E}r_{IS,t+1} &= \frac{1}{N}\left(\mathbb{E}[g'(I_J - \Gamma_t)'X'X(I_J - \Gamma_t)g] + \mathbb{E}[\bar{e}'_t X(X'X)^{-1}\Gamma_t X'X\Gamma_t X'\bar{e}_t] + \mathbb{E}[\bar{e}'_{t+1} X(X'X)^{-1}X'e_{t+1}]\right) \\
&= \frac{1}{N}\sum_{j=1}^{J}\left(\frac{\lambda_j}{t\lambda_j + \frac{J}{N\theta}}\right)
\end{aligned}$$

When $\theta \to \infty$ this converges to what we saw in OLS last class. Interpretation?

Return Predictability: In-Sample

We can directly call $\mathbb{E}r_{IS,t+1}$ as expected total explained return variance, and we can construct an adjusted R^s measure:

$$R^2_{adj} = 1 - \left(1 - \frac{\mathbb{E}r_{IS,t+1}}{1 - J/N + \mathbb{E}r_{IS,t+1}}\right)\frac{N}{N-J}$$

In rational expectations, $\mathbb{E}r_{IS,t+1} = J/N$ and hence this adjusted $R^2 = 0$

With learning, there are in-sample predictable components in returns that raised this above zero

Return Predictability: In-Sample Adjusted R^2

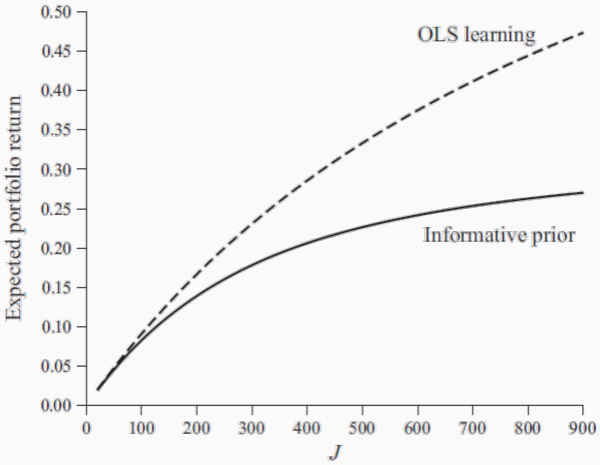

Figure 5.2. Adjusted R^2 in in-sample return prediction regression

Return Predictability: Out-Of-Sample

... or lack thereof.

Consider $r_{OOS,t+1} = w'_{OOS,t} r_{t+1}$ where $w'_{OOS,t} = \frac{1}{N} X h_t$.

Martin and Nagel (2019) show that $\mathbb{E} r_{OOS,t+1} = 0$, or that there is no return predictability. This is because investors are Bayesian and endowed with an objectively correct prior, so the econometrician on the same footing can no longer predict investors' forecast error and asset returns.

Takeaway: Researchers looking for risk premia or behavioral biases should focus on OOS, especially with high-dimensional data (large J), where there is a big wedge between in-sample and OOS predictability, which is driven by learning-induced investor forecast error.

Extensions: sparsity

Treat **sparsity** as a modification of prior beliefs. Suppose priors are drawn from a Laplace distribution:

$$f(g_j) = \frac{1}{2b} exp\left(-\frac{|g_j|}{b}\right)$$

where $2b^2 = \frac{\theta}{J}$ is the variance, and investors price assets based on the mode (rather than the mean) of the posterior distribution. This is similar to Lasso.

Expected OOS portfolio return is no longer zero, but simulations show that this is very similar to the shrinkage case.

Extensions: extra shrinkage or sparsity

In the Bayesian framework, shrinkage in cash flow forecasting was optimal, and founded as a consequence of informative prior beliefs.

Additional **shrinkage** and sparsity can be optimal due to boundedly rational limited attention or cost of data observation.

Suppose priors were close around zero than the objectively correct prior. Pick covariance matrix of the objective prior $\Sigma_g = \frac{\theta}{J} I$ with varying θ.

Extensions: extra shrinkage or sparsity, expected returns

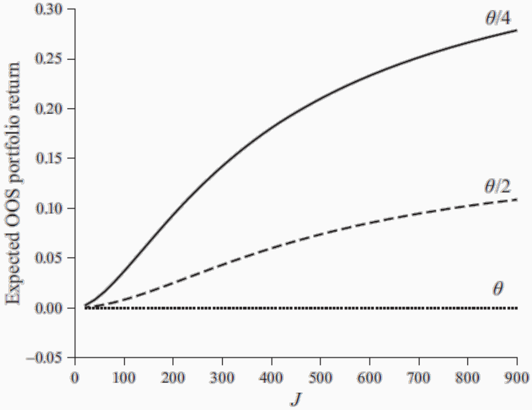

Figure 5.3. Expected OOS portfolio return when investors learn with excessive shrinkage

Summary

Suppose investors forecast cash flows on covariates.

- Investors make forecast errors that look predictable ex post with in-sample regressions
- Forecast errors contaminate results, making returns predictable
- Bayesian regression shrinkage reduces this predictability

The economic content of market efficiency that prices "fully reflect" all public information is not clear in a high-dimensional setting.

- Does this mean investors know the parameters g?
- Or that investors employ Bayesian updating

Argues that it isn't an interesting question to show in-sample deviations from if investors know g. Should focus on OOS or psuedo-OOS.