Hedging and Replication

Replication and Basis Risk

In this week's homework, you're asked to consider replicating the HFRI index using a set of underlying factors (the Merrill Lynch factors). However, even if we have a high R^2 in our regression, this does not mean that our replication is perfect. In fact, there is *basis risk* involved in this strategy.

Let's consider the following replication:

$$r^{\text{HFRI}} = \alpha + \boldsymbol{\beta} \mathbf{X}^{\text{Merrill}} + \epsilon$$

Where r^{HFRI} is the return of the HFRI index (a $T \times 1$ vector), $\mathbf{X}^{\text{Merrill}}$ is the matrix of factor returns ($T \times K$), $\boldsymbol{\beta}$ is the vector of factor loadings ($K \times 1$), and ϵ is the error term.

If we consider the following hedged portfolio (re-arranging the above equation):

$$r^{\text{HFRI}} - \boldsymbol{\beta} \mathbf{X}^{\text{Merrill}} = \alpha + \epsilon$$

We see that there are *two* sources of our replicated return not being zero (i.e., if we can perfectly replicate the HFRI index, the left-hand side should be zero):

- α: The interpretation of this constant is that even if the Merrill factors have zero returns, the HFRI index will still generate some return. More specifically, it is the average return of the HFRI index that is not explained by the Merrill factors.
- ϵ: This is the error term of the regression, which captures the deviations of the HFRI index returns from the returns predicted by the Merrill factors. This term has an expected value of zero (by construction of OLS), but it can be positive or negative in any given period.

Let's explore these two components in more detail. When we consider hedge funds delivering "alpha", what we mean is that they are delivering returns that investors cannot easily replicate. That is, try as you

might to add Merrill factors to your portfolio, there is some portion of returns that hedge fund managers are able to deliver that cannot be explained.

There can be many reasons for this, including skill and access to unique assets (e.g. private equity, private credit, etc.). One of these unique assets that most people (including most hedge funds!) don't have access to is physical commodities. Many hedge funds invest in commodities, but only through futures and options on exchanges like the CME and ICE.

However, the actual physical commodities (e.g. oil barrels, bushels of wheat, natural gas in pipelines) are hard to access for most investors. Thus, if hedge funds are able to profitably trade these physical commodities, this can generate alpha. There is a *Financial Times* article **here** that discusses how some hedge funds are trying to access physical commodities to generate alpha. One can imagine that a physical commodities trading business (buying, storing, and delivering barrels of oil) is extremely hard to replicate using futures contracts alone.

Another thing to keep in mind in hedging and replication, is that the choice of whether or not to include a constant term in the regression matters. If we are trying to replicate the HFRI index, and *actually trade on it*, then we must exclude the constant term, because:

> *We can't invest in α!*

Intuitively, what we're saying if we include an α in the regression is that we are pulling returns out of thin air in our replication. This is fine if we are trying to *explain* returns, but if we are trying to bundle real assets into a tradable portfolio, we must exclude the constant term as we can't generate returns from nothing.

The second component ϵ is also important to consider. If we assume the constant term is zero (i.e. no alpha), then we still don't have a perfect replication! Namely, we're able to replicate it on average because:

$$\mathbb{E}[r^{\text{HFRI}}] = \mathbb{E}[\boldsymbol{\beta}\mathbf{X}^{\text{Merrill}}] + \mathbb{E}[\epsilon] = \mathbb{E}[\boldsymbol{\beta}\mathbf{X}^{\text{Merrill}}] + 0$$

Since $\mathbb{E}[\epsilon] = 0$ by construction. Therefore, our mean return will be correct (since in this class we take $\mathbb{E}[r] = \mu$, our sample average).

However, in any given period, the error term ϵ can be positive or negative, meaning that our replication can be off by a significant amount. This is known as basis risk, and it is the risk that the returns of our hedged portfolio deviate from zero due to the error term.

> *Even if we have the same expected return as the target index, our portfolio may still meaningfully deviate from the target due to ϵ.*

This risk of deviation is called *basis risk*, and can be extremely important to consider when replicating a portfolio. To see this, let's think about the variance of our replicated portfolio compared to the variance of the target index.

$$\begin{aligned}\text{Var}\left(r^{\text{HFRI}} - \boldsymbol{\beta}\mathbf{X}^{\text{Merrill}}\right) &= \text{Var}(\alpha + \epsilon) \\ &= \text{Var}(\epsilon) \\ &= \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\end{aligned}$$

So this is basically saying that the *difference* in variance between our replicated portfolio and the target index is equal to the variance of the error term ϵ.

A connection to R^2 in the regression can be made here. Recall that:

$$R^2 = 1 - \frac{\text{SS}_{\text{Res}}}{\text{SS}_{\text{Tot}}} = 1 - \frac{\sum_{t=1}^{T}\epsilon_t^2}{\sum_{t=1}^{T}\left(r_t^{\text{HFRI}} - \bar{r}^{\text{HFRI}}\right)^2}$$

So this is saying that the *higher* the R^2, the *lower* the variance of the error term ϵ, and thus the lower the basis risk of our replicated portfolio.

Some Python Advice

The `statsmodels` package in Python is my preferred way to run regressions. It also includes many powerful features in the regression output, here are some ways to access useful information:

```
import statsmodels.api as sm

# IMPORTANT: Statsmodels does not add a constant (intercept) by default!
# You must explicitly add it to your X matrix if you want an alpha.
X = sm.add_constant(X)

# Regular OLS regression
model = sm.OLS(y, X).fit()

# What happens if we want to drop missing values?
model = sm.OLS(y, X, missing='drop').fit()

# How can we access regression results?
print(model.summary())  # Full regression summary

# That looks pretty, but how do we get specific values?
# This assumes X has a constant at index 0 (which sm.add_constant does)
alpha = model.params[0]  # Intercept (alpha)

# And if I want the p-values?
p_values = model.pvalues

# What about the confidence intervals?
conf_intervals = model.conf_int()
```

Something else that we might want to do is if we believe our regressions have heteroskedasticity (non-constant variance of errors, seen in Data Analysis and Regression Review), we can use robust standard errors:

```
model = sm.OLS(y, X).fit(cov_type='HC3')  # Heteroskedasticity-Consistent SEs
```

We can also run confident intervals on our predictions

```
predictions = model.get_prediction(X_new)
pred_summary = predictions.summary_frame(alpha=0.05)  # 95% CI
```

If we want to detect outliers, that comes built-in as well:

```
influence = model.get_influence()
leverage = influence.hat_matrix_diag  # Leverage values
cooks_d, p_values = influence.cooks_distance  # Cook's Distance

# Summary of influence measures
influence_summary = influence.summary_frame()
```

Hedge funds pile into commodities in search of fresh source of returns

Funds such as Balyasny, Jain Global and Qube are expanding operations to allow them to trade underlying markets

---

Hedge funds and trading firms are piling into physical commodities markets in search of new sources of returns, despite lacking the decades of experience and information accumulated by established players such as Trafigura and Vitol.

Financial firms have a long history of trading contracts for power, natural gas and oil. But hedge funds such as Balyasny, Jain Global and Qube, as well as trading firm Jane Street, are expanding their operations to allow them to trade the underlying markets, deepening their exposure to global price swings.

This can involve buying the rights to transport natural gas over a pipeline, buying storage capacity for crude oil and storing electricity in advanced batteries before offloading it at peak demand times.

Trading these markets can confer an informational advantage to market participants. “It’s an information gold rush,” said Michael Alfaro, chief investment officer of hedge fund Gallo Partners, which is focused on energy and industrials. “When you’re trading physical commodities, you’re privy to a lot of information and you get a sense of what is actually happening from economic shifts before the actual data comes in.”

Multi-manager hedge fund Balyasny has expanded its power trading teams and researchers in Europe by hiring from utilities such as Centrica and Norlys, and has added natural gas traders. Multi-strategy hedge fund Jain Global this year bought Anahau Energy, which specialises in natural gas services, and is now actively trading the commodity.

Quantitative hedge fund Qube moved into European physical power in Europe via affiliate Volta, which also recently applied to join as a member of NEPOOL, an advisory group that helps draft rules for physical power markets across six US states. At least nine natural gas and power traders have joined Qube since 2024, according to an analysis of LinkedIn profiles.

Hedge funds have partly been inspired by the huge profits enjoyed by trading firms such as Trafigura and Vitol. Hedge fund Citadel also generated huge returns in 2022, as volatility in natural gas prices minted fortunes for top traders, particularly in Europe.

Citadel has long invested in energy assets to aid its trading activities, but this year the hedge fund has been particularly busy with deals. In March, it acquired Paloma Natural Gas, which was later renamed Apex Natural Gas, in Louisiana for $1.2bn. In October, Citadel bought German energy trader FlexPower, which has a subsidiary developing its own grid-scale battery projects.

More recently, Apex agreed in two different transactions to buy natural gas assets in Texas and Louisiana, according to a person familiar with the matter. A spokesperson for Citadel declined to comment.

This year has been more muted for hedge funds and trading firms compared with 2022, as commodities such as oil and gas have traded in tighter ranges. “It’s always been feast or famine in commodities,” said a hedge fund allocator at one of the world’s top asset managers.

For large hedge funds, going into physical commodities offers a separate return stream that theoretically makes a hedge fund more diversified. The potential upside from extremely volatile years such as 2022, when Russia invaded Ukraine, outweighs lower return periods.

Physical traders can benefit from surges in demand that hedge funds can anticipate through advanced weather reports and other data. An executive at a large hedge fund said physical power in particular was a sweet spot for hedge funds because of the ability to use analytical techniques to anticipate consumer demand for power across US states and European countries.

Hedge funds can also take delivery of commodities and store them for a period while prices dip and later sell them when those prices recover. Ilia Bouchouev, managing partner at New York hedge fund Pentathlon Investments, said financial players were more likely to lease batteries or take out contracts with options embedded rather than own the physical battery structures.

“This is similar to how the oil storage business has worked for decades and batteries are just a new form of storage,” he said.

But this also requires taking unfamiliar risks in areas outside their traditional expertise. Hedge fund firm Amaranth collapsed in spectacular fashion in the mid-2000s after a dalliance with commodities. The firm shifted away from investing in convertible bonds and lost 35 per cent of its $7.5bn of investor capital on disastrous natural gas wagers — although these trades were made using financial derivatives contracts rather than physical deals.

An executive at a top quantitative hedge fund that trades commodities but is not in the underlying physical markets questioned how hedge funds could hope to compete with trading firms such as Vitol, Trafigura and Gunvor as well as corporates including BP and Shell.

“The major commodity trading houses have huge balance sheets and they actually control the whole logistical supply chain in commodities, from transportation to refinement,” he said. “From this supply chain they can extract an enormous amount of valuable information.”