Predicting Regime Shifts with Implied Principal Components Analysis

Jeffrey Wang
Harvard University

Abstract

Many participants in the financial system use interest rate swaps to exchange risk from cash flows with floating returns. Since they allow retail investors and institutions to hedge their future financial risk, over $100 trillion dollars of notional value flow through these instruments each year [4]. As a result, accurate pricing of swaps is crucial to allow market participants to price their risk. Many strategies to do so, such as performing Principal Components Analysis on the swaps curve, rely on past historical data to predict future rates. These methods, however, are prone to failure during periods of *regime shifts*, periods of fundamental changes in macroeconomic variables, policies or regulations. As such, identifying when regime shifts might occur is of critical importance. In this paper, we investigate Implied Principal Components Analysis (i-PCA), a forward-looking method that builds a covariance matrix using implied volatility information from options markets. We first describe the method and show how it is analogous to traditional PCA in many ways, but provides more robust, less noisy, real-time indicators of imminent regime shifts. We then introduce a method to systematically detect these regime shifts, with 5 years of experimental backtesting. Finally, we discuss how this might enable real-time detection of regime shifts.

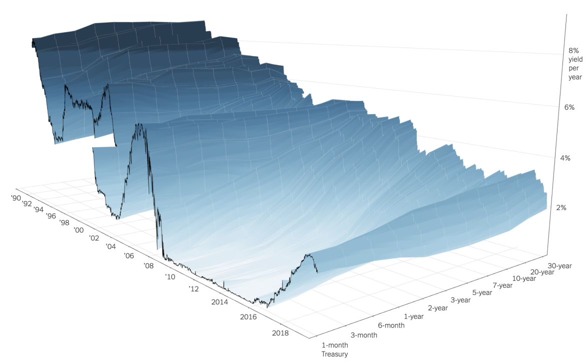

Figure 1: The swaps curve, from 1 month to 30 years, 1990 to 2018. An inverted curve, when rates are higher over shorter tenors than they are over the medium-to-long term, reflect the market's premonition that a recession may be incoming [3].

1 Introduction

The yield curve depicts the interest rates of bonds at differing maturities, reflecting investor expectations from short to long-term about macroeconomic conditions. Swaps, where one party exchanges the floating rate of a securitized cash flow for a fixed one, are an instrument that allows counterparties to exchange cashflow risk. The swaps curve, generally tightly bound to the yield curve except in periods of exceptionally low investor trust, illustrates fixed rates one can pay over different tenors (1 month to 30 years in the image below).

In a swap, the *payer* benefits when the interest rate is higher than expected, and the *receiver* benefits when the interest rate is lower than expected. While the rates market is affected by many factors, one long-standing method to decompose structural influences on rates is Principal Components Analysis [1]. By performing PCA on the rates curve, we obtain eigenvalues $e_1 \dots e_n$ and an orthogonal eigenbasis $p_1 \dots p_n$, where the variance of the data explained by the basis vector p_i is equal to $\frac{p_i}{\sum_j p_j}$.

Using our eigenbasis, we decompose our dataset into uncorrelated factors, which incorporate structural influences. By taking linear combinations of exclusively the first two principal components, the vast majority of variance (around 99%) in the data can be explained. PCA on the rates curve is useful because it is *interpretable*; the first two components broadly decompose the values into *level* and *slope* factors. The level factor describes the movement and shape of the entire curve, whereas the slope factor describes relative values between different tenors [1]; PC1 typically corresponds to the level factor and PC2 to the slope factor. Reconstructing our data using the first and second PC's (Figure 2), we see close alignment with the 10 year yield and 3s/30s spread.

Mechanically, the values of the first principal component

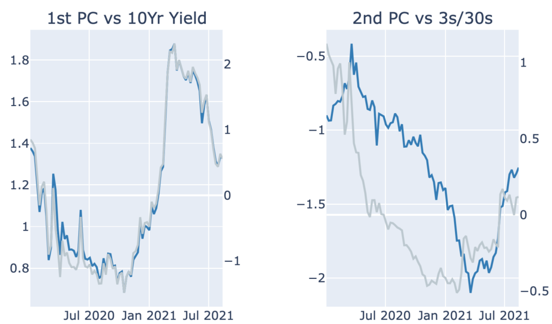

Figure 2: Reconstructing the swap curve using the first and second principal components reveals close alignment with the level of the curve (for PC1) and relative spreads (PC2).

$[v_1 \dots v_n]$ indicate that a 1-σ shift in the level factor will cause the i-th rate to shift by v_i.[1]

2 Implied PCA

To illustrate an example of a regime shift affecting a historical analysis, we consider the relative movement of different tenors (e.g. 2s and 10s). Investors looking for relative value opportunities and those seeking to hedge risk must do so by relying on estimates of betas between different tenors t_1 and t_2, calculated as $\frac{\mathrm{Cov}(t_1,t_2)}{\mathrm{Var}(t_2)}$. Historical data can be a poor indicator during regime shifts, however; consider three sets of three month average betas from October 2022 to June 2022:

Time Interval	2s β	10s β	2s/10s Ratio
Oct 21 - Dec 21	0.720	1.574	0.457
Jan 22 - Mar 22	1.608	2.006	0.767
Apr 22 - June 22	3.027	2.997	1.010

During a period of immense macroeconomic changes, from rate hikes to COVID whiplash to inflation, using the previous 3 month's beta at any point would be wildly off the *ex-post* realized beta. Here, we outline Implied PCA, a framework that uses forward-looking data to rectify this issue.[2]

2.1 Methodology

Rather than using past data of swap rates, we instead use implied volatility of options: on the rates themselves, and on the spread between rates (known as yield curve spread options, or YCSOs). Notably, since the implied volatility of an option is the value of volatility that infers the current market price of the option, if we assume the options markets are efficient, then we effectively are using investors' best, most up-to-date guess about forward-looking conditions for swap rates.[3]

Specifically, let's assume we are considering rates $r_1, \dots, r_n$. We can write that

$$\mathrm{Var}(r_i - r_j) = \mathrm{Var}(r_i) + \mathrm{Var}(r_j) - 2\mathrm{Cov}(r_i, r_j)$$

Rearranging, we have that

$$\mathrm{Cov}(r_i, r_j) = \frac{1}{2}(\mathrm{Var}(r_i) + \mathrm{Var}(r_j) - \mathrm{Var}(r_i - r_j))$$

As such, we can compute the implied covariance of r_i and r_j, since the Var terms are the square of implied volatility from the rates options market (for r_i and r_j), which we denote σ_{r_i}, and the yield curve spread options market (for $r_i - r_j$), which we denote $\sigma_{r_i - r_j}$. Notably, we can infer β in a similar manner as well. Given implied covariances for $r_1, \dots, r_n$, we can write an implied covariance matrix:

$$\begin{bmatrix} \sigma_{r_1}^2 & \cdots & \frac{1}{2}(\sigma_{r_1}^2 + \sigma {r_n}^2 - \sigma_{r_n - r_1}^2) \\ \vdots & \ddots & \vdots \\ \frac{1}{2}(\sigma_{r_1}^2 + \sigma {r_n}^2 - \sigma_{r_n - r_1}^2) & \cdots & \sigma_{r_n}^2 \end{bmatrix}$$

We can then perform an eigendecomposition on this matrix using our preferred scientific computing library to obtain eigenvectors and eigenvalues, written in matrix form as PDP^T, where P contains eigenvectors and D is a diagonal matrix of associated eigenvalues in descending order. Finally, we can compute principal components by taking $V = P\sqrt{D}$, which are scaled such that a 1-sigma move in the level factor will shift rates at different tenors by their corresponding values in the first principal component [5]. All of this data is available from market sources, such as J.P. Morgan DataQuery.

Since this process is mechanically the same as performing PCA on the rates curve[4], except uses forward-looking data, we expect the PC values to be similarly interpretable. To test this, we ran iPCA on a 4 x 4 covariance matrix containing data for 2s, 5s, 10s, and 30s. We then plotted PC1 values at six different dates. The resulting values were interpretable in context of wider macroeconomic undercurrents:

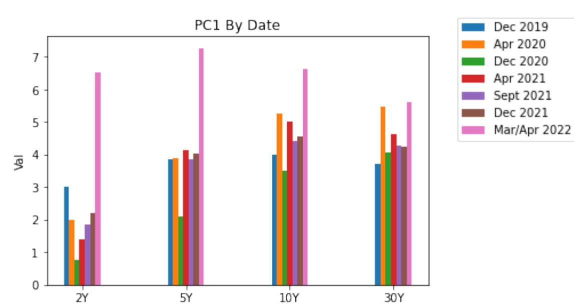

Figure 3: PC1 values at 2s, 5s, 10s, and 30s across six dates.

- **Dec 2019 - Apr 2020**: COVID. With the world shut down and many warning that the pandemic might last years, the two year PC1 value dropped, indicating the investor attitude that the short-term macroeconomy would be less volatile (likely depressed). The 5 year value remains nearly constant, and the 10/30 year values increase (as such an unprecedented event reflects greater long-term uncertainty about the macroeconomy).

- **Apr 2020 - Dec 2020**: Pandemic Normalization. As a slow economic recovery begins, markets are less volatile than expected. PC1 values across all tenors drop.
- **Apr 2021 - Dec 2021**: Relative Calm. PC1 values stay relatively constant in all tenors throughout 2021 as the macroeconomy continues its recovery.
- **Dec 2021 - Mar 2022**: Rate Hiking. With the sudden presence of inflation and massive rate hiking from the Fed, uncertainty (and expected future volatility) spikes in all tenors.

2.2 Detecting regime shifts with the percent variance explained by PC1

During periods of regime shifts, the structural components that compose the level factor change; as a result, in addition to PC1 values changing, we also expect the percent variance explained by PC1 (PVPC1) to fall as the market responds to the change. Plotting PVPC1 from iPCA from the past 5 years, we see that there is a strong qualitative alignment of drops in PVPC1 and regime shifts.

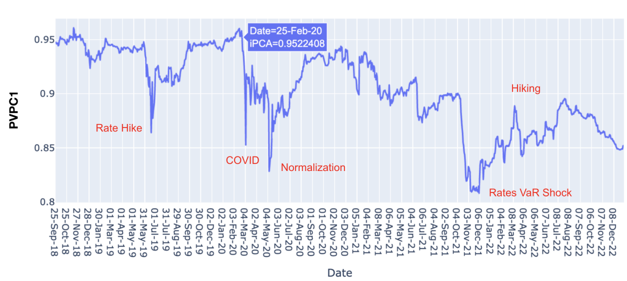

Figure 4: PVPC1 closely aligns with regime shifts over the past 5 years.

In Section 3, we further corroborate this hypothesis with data from the Fed, and propose a systematic way of detecting these shifts in real time.

2.3 iPCA is real-time and robust

Compared to traditional PCA on historical swaps data, iPCA offers a few advantages – primarily that it is real-time and robust. To test the former, we ran iPCA on each day's data for the past 5 years, and ran traditional PCA with a 60-day window of past data on each day for the past five years. We then graphed PVPC1 for iPCA and historical PCA; the results are in Figure 5. As seen in the graph, the drops in PVPC1 in historical PCA lag those in iPCA by approximately the window size. For instance, the deepest drop in both graphs, during COVID, begins in late April with historical PCA. Implied PCA offers an indicator that is much closer to real-time of a regime shift.

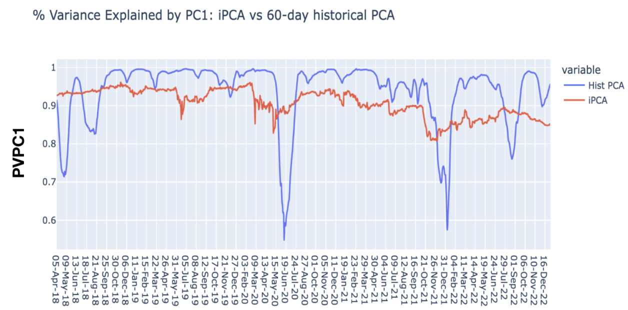

Figure 5: Comparison of PVPC1 between Historical PCA with a 60-day lookback window and iPCA.

Finally, we note that iPCA introduces a parameter that is not present in traditional PCA of the swap curve: that of options expiry dates. To test the effect of different expiries, we compared important measures of swap pricing with options 1 month and 1 year from expiry: see Figure 6 for β.

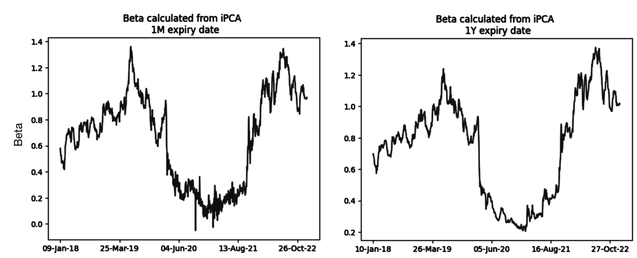

Figure 6: Implied beta from iPCA, using one month and one year expiry dates. The shapes are nearly identical; the one-month expiry horizon is slightly more noisy, as would be expected.

3 Systematic Detection of Regime Shifts

The "golden bullet" here is systematic detection of regime shifts, preferably as soon as they begin to happen. In each instance of a regime shift, PVPC1 undergoes a sudden change from relative stochasticity to a sudden drop. We can model this change with a simple autoregressive model using a moving threshold. For instance, consider a PVPC1 vector **p** and corresponding date vector **d**; we wish to know if there is a significant regime shift beginning on index d_i. We propose the following algorithm:

```
let i = 0
while i <= p.length {
  let bck = [p_{i-1}, ..., p_{i-c_2}]
  let f1 = [p_i, ..., p_{i+c_1}]
  let f2 = [p_{i+c_1+1}, ..., p_{i+2c_1+1}]
  let regimeshifts = []
  if max(f1) - min(f2) > k · stdev(bck) {
    regimeshifts.append(i).
    i = i + max(c_2, 2c_1)
  }
  else {
    i = i + 1.
  {
}
```

where c_1, c_2, and k are hyperparameters. Depending on the parameters used, we can get “loose" or “strict" detection of events (Figure 7). To train these parameters, we combed through 5 years of FOMC notes and compiled a list of important monetary regime shifts. We then split our data into four years of training data and one year of test data, gridsearched the hyperparameter space, and found “strict" (higher specificity; lower sensitivity) and “loose" (higher sensitivity; lower specificity) parameter sets for our method. Figure 7 illustrates these different parameter sets.[5]

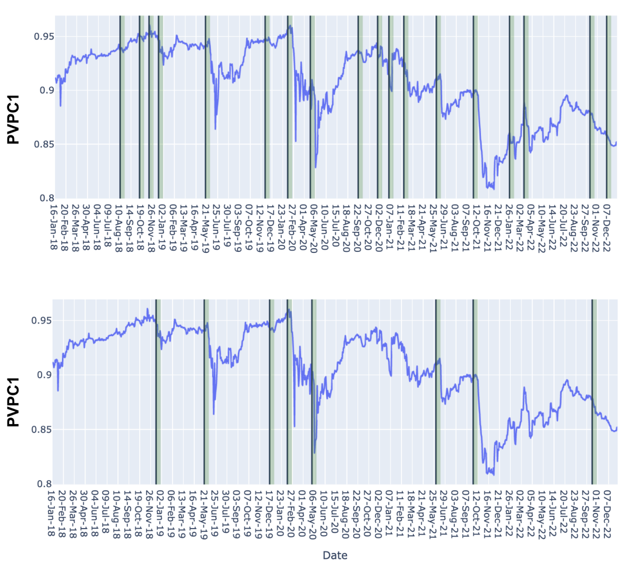

Figure 7: Significant regime shifts by PVPC1, as determined by “loose" and “strict" hyperparameters (top and bottom).

The method appears to capture significant regime shifts, like COVID, recalibration, and rate hiking well. However, there are some tradeoffs to it; the foremost of them is that there is a sensitivity-detection tradeoff in the magnitude of the different hyperparameters. This is most obvious with k, but also is especially present with c_2 (the size of the backward window). When c_2 is large, the algorithm is exceptionally good at finding COVID-esque regime shifts: when PVPC1 is relatively constant and then undergoes a sudden shift (most of the changes detected in the bottom of Figure 7 are these types of changes). However, in periods of exceptional volatility, when PVPC1 shifts a lot in the course of a few days (such as the “spike" moments in the graph), c_2 must be smaller to detect them. When c_2 is smaller, however, the automated detection regime may split a single large drop (like the COVID shift) into two smaller ones. It may also capture more fine graph noise unless k is raised to be extremely high (15+). As such, it may be useful to add a filtering step after this algorithm.

Despite these drawbacks, the combination of iPCA and the simple systematic detection algorithm allows for simple real-time detection of regime shifts, with relatively accurate leading predictions. The method is easy to implement *and* interpret; code is publicly available on our GitHub (see Section 5).

4 Discussion and Conclusion

In this paper, we introduce implied Principal Components Analysis, a method for using forward looking options data to perform PCA on the swaps curve. We demonstrate that iPCA is analogous to traditional PCA on the swaps curve, highly interpretable, robust, and allows us to identify regime shifts by analyzing the % variance explained by PC1; we also illustrate that it is more effective than historical PCA on this measure. Finally, we introduce a method to identify these systematically and backtest it on 5 years of data – showing high alignment with macroeconomic movements.

5 Data and Code Availability

Data and Jupyter Notebooks containing the code are available at this GitHub repository.

6 Acknowledgements

This work was done under the guidance of the swaps team at Bracebridge Capital. The original project was done alongside Alkiviades Boukas, Monie Choi, Sam Fuller, Joshua Salazar, and Jehan Boher. All work in this report is work of Jeffrey Wang, the author. The author thanks Arrunava Moondra, Ajay Balaji, and Vicky Zhao for providing mentorship over this project; the author also thanks Sue Mullaney, Monica Hatch, and Nicole Migliozzi for organizing the program.

References

[1] David Redfern, D. M. Principal component analysis for yield curve modelling. *Moody's Analytics: Enterprise Risk Solutions* (2014), 1–27.

[2] Ian T. Jolliffe, J. C. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society* (2016), 1–10.

[3] Irwin, N. What's the deal with that inverted yield curve? a sports analogy might help. *The New York Times* (2019).

[4] Lee Baker, Richard Haynes, J. R. R. S. B. T. Introducing enns: A measure of the size of interest rate swap markets. *Office of the Chief Economist, Consumer Financial Trade Commission* (2020), 1–2.

[5] Srini Ramaswamy, Ipek Ozil, P. M. M. F. Interest rate derivatives: Implied principal components analysis. *JP Morgan North America Fixed Income Strategy* (2022), 1–10.

Notes

[1] See [2] for more.

[2] Historical PCA's window lag isn't exactly its window size; enough changed data after the regime shift can cause the PVPC1 value to begin falling, as one might expect.

[3] Market efficiency can be a difficult assumption, especially during periods of little liquidity. We note that the method introduced in this paper may not be particularly useful for forecasting sudden market crunches with no warning (e.g. Russia's invasion of Ukraine) but may be particularly good at catching slow-rumbling crises (e.g. 2008 crisis, COVID) before they go "mainstream."

[4] There is one additional parameter here, the options expiry date. However, we show that iPCA is robust to different expiries.

[5] The parameter sets used were $k = 7$, $c_1 = 9$, $c_2 = 20$ and $k = 10$, $c_1 = 9$, $c_2 = 20$. Some discussion of c_1 and c_2 sizes is provided in the text.