CQF Final Project - Jan 2021 Cohort

Gustavo Curi Amarante

August 23, 2021

In Part I, we explain why FX can be seen as the most liquid asset class, how we chose the currencies that will compose our portfolio and how these choices fit into the CQF final project requirements.

Part II, motivates the need for denoised/stable/robust covariance matrices and, as requested, explain in detail the Marchenko-Pastur covariance denoising method. A numerical example to illustrate intuition and test our code is also presented.

In part III, as requested, we present the full derivation of the Black-Litterman model and solve a hypothetical example, using the code we devolped, to show that our code is correct.

Part IV briefly outlines the optimization methdos that use: Markowitz/Max Sharpe (as requested) and Hierarchical Risk Partity (as our choice).

Parts V and VI respectively shows the output of a static optimization, to better visualize what the portfolio is doing in a single day, and the full backtest of the strategies, the evaluate their performance. These sections also show our conclusions.

Contents

I FX as an Asset Class **3**

1 Working with FX Data **3**

2 Choice of Assets **3**

3 CQF Requirements for Portfolio Choice and Data **8**

II Covariance Estimatimation **10**

4 Motivation **10**

5 Naive Covariance **11**

6 Robust Covariance **13**

III Expected Returns Based on Views **18**

7 Black-Litterman **18**

8 Using Signals to Generate Views **26**

IV Portfolio Construction **30**

9 Markowitz / Max Sharpe **30**

10 Detoned Hierarcical Risk Parity **30**

V The Static Portfolio **32**

VI Dynamic Portfolio (Backtesting) **35**

A Bloomberg Tickers **38**

Part I

FX as an Asset Class

1 Working with FX Data

There are some issues that we have to be very careful about when working with FX data. If we want our backtest to be truly representative of reality, it is not as simple as grabbing all of the spot quotes and using their percent changes as a measure of returns. This would be analogous to building a backtest based returns measured on stock prices changes, while ignoring dividend payments. In the case of currencies (or any other asset class), we have to include the true cash flows of the strategy and not just look at the asset price.

The first problem is that investor typically are not allowed to trade spot currencies, so they build their exposures to currency risk using forwards or futures. This type of contract has the advantage of the not requiring lots of cash (leverage), but it requires the backtesting to be more detailed. To go around this problem, we use data from currencies forward contracts, which are very liquid as we will see in section 2. By doing so, we have to be careful to roll positions properly when contracts are coming close to their maturity and take rolling and rebalancing costs into account when computing the profits or losses of the strategy. This means that all of our currency returns already takes into account funding and carry. The total return index that we use to measure returns is built by UBS, and are available in bloomberg terminals. Their tickers are presented in the appendix.

The second issue comes from the timestamp of the prices. Currencies are traded in different timezones and so spot rates and forward prices from different countries may reflect different information sets. One way that some people go around this problem is to look at weekly or monthly returns, in order to minimize the effect of marginal information on prices, but since we are interested in using daily returns, we were very careful to use forward prices always on the same time of day for all currencies, 4:00 PM London. The bloomberg tickers with this timestamp are presented in the appendix. These total return indexes are what we use to measure the return of a currency and is the data where all computations, like correlations, are based on. Figure 1 shows the total return indexes[1] for the selected currencies. The criteria for selection is going to be explained in section 2.

One last choice that we have to make is our base currency. As the academic literature usually does, we chose the USD as our home currency, so all currencies will be traded against the US dollar.

2 Choice of Assets

When dealing with systematic strategies, one issue that might arise is the lack of liquidity. Strategies that constantly roll and rebalance their positions to different signals need very liquid assets. Looking at data from the latest triennual central bank survey of the BIS [14] in figure 2, we can see that trading in FX markets reached an average of USD 6.6 trillion per day in April 2019, the last time the survey was conducted.

Figure 3 compares this number with the volume of other major global stock and bond markets and we can see that liquidity is probably not an issue for FX markets. Even when excluding the spot transactions, leaving only the financial instruments, the FX market is still bigger the other major stock and bond markets

[1] Since these return are based on currency forwrd contracts, they already include funding and carry in the price, so they can be interpreted as *excess* returns.

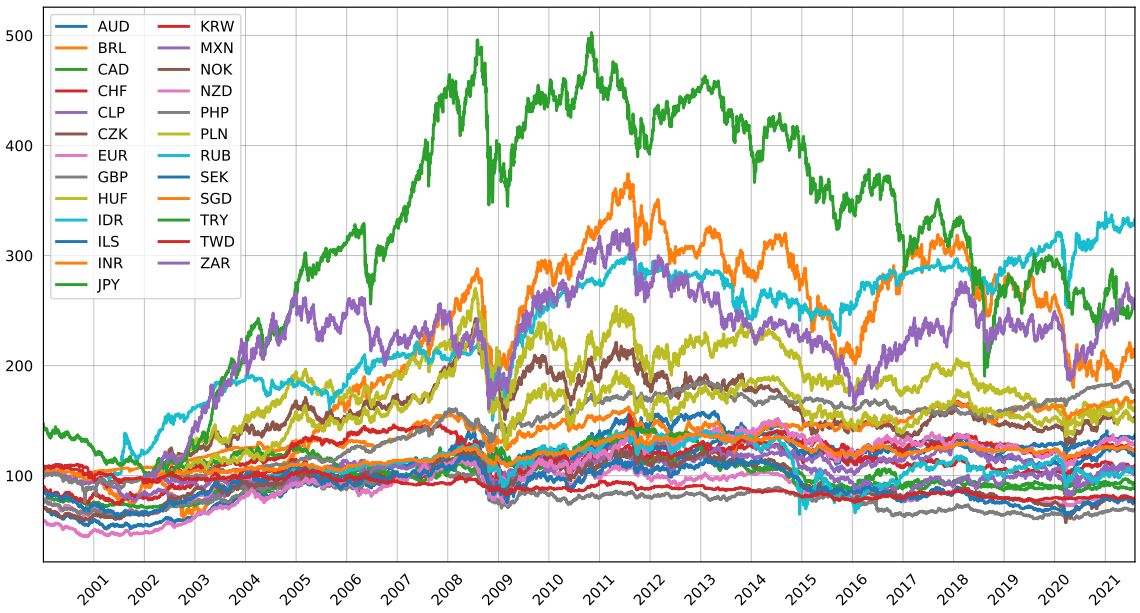

Figure 1: Excess Return Indexes for all selected currencies against the USD, computed by UBS

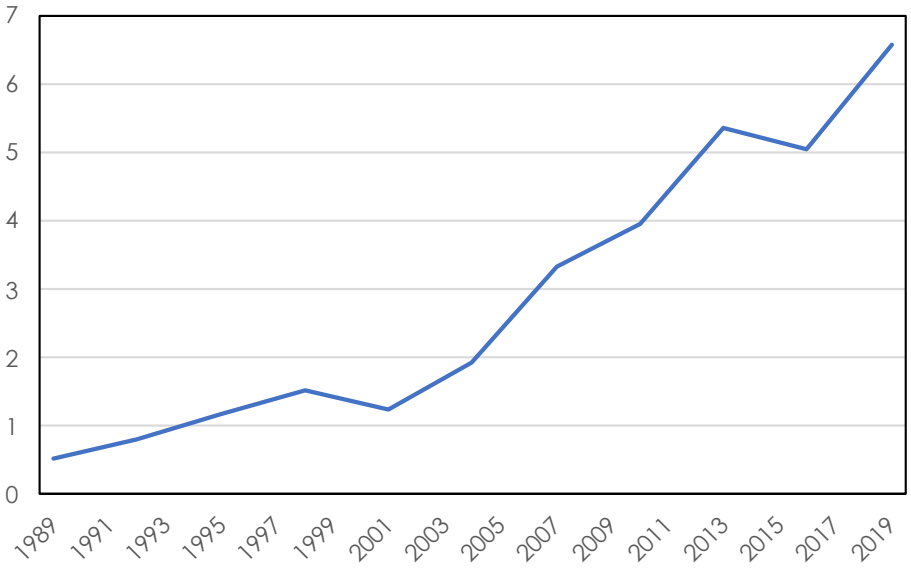

Figure 2: Average daily trading volumes (trillions of USD)

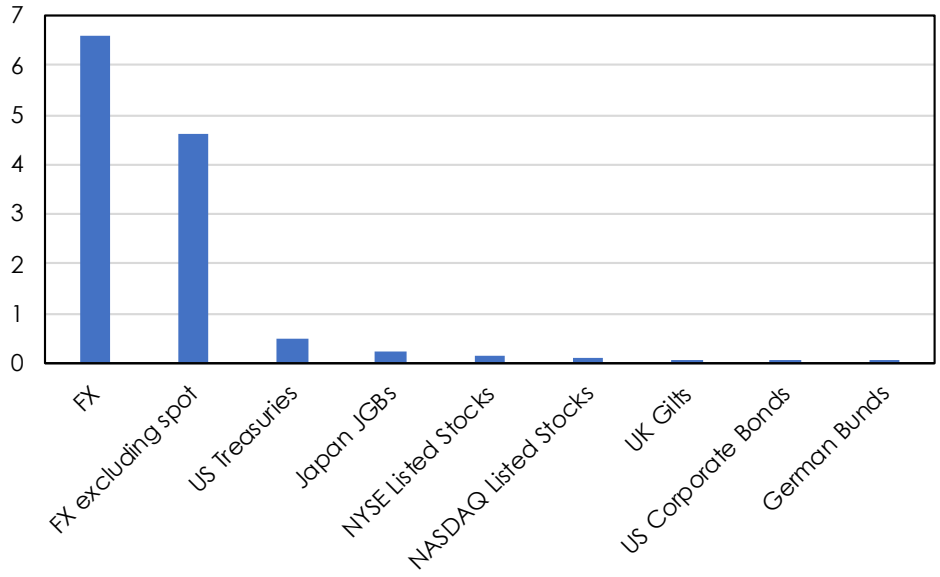

Figure 3: Average daily trading volumes (trillions of USD)

combined.

Having a closer look at the data from the BIS (ommited in this report), we can see that the growth of FX derivatives trading, especially in FX swaps, outpaced the growth of spot trading. This may be due to a growing popularity of FX as an asset class. Emerging market currencies reached 25% of overall global turnover, which makes them a viable asset class for strategies with higher trading frequencies.

Now we know that FX is a very liquid asset class. The second issue is that this liquidity is not evenly distributed on all currencies. According to the latest trianual central bank survey of the BIS [14] the 39 most traded currencies are shown in table 1. We want to select as many currencies as we can for our strategy since the covariance denoising method that we are going to use requires many observtations both in the timeseries dimension, which is the length of the series of returns, and cross-section dimension, which is the number of assets.

As explained in [13], we have to select which currencies will make into our strategy based on a few factors like liquidity, currency regime and other particularities. On the liquidity issue, we are going to eliminate currencies that correspondo to less than 0.1% of the average daily turnover (which means it has at least USD 6.6 billion average daily turnover). The currencies that have insufficient turnover to be part of our strategy are COP, SAR, MYR, RON, PEN, ARS, BHD and BGN. On the currency regime issue, we have to exclude currencies tha are pegged to another (usually the US dollar) which in our case are AED, BGN, BHD, DKK, HKD and SAR. Finally, the CNY, THB and MYR are also generally excluded when building strategies and backtesting because they present some particularities (short history, stressed events, past convertibility problems, etc).

This is why the market practice defines two groups of currencies: G10 and Emerging Markets (EM) and always trading them against the USD:

G10: AUD, CAD, CHF, EUR, GBP, JPY, NOK, NZD and SEK.

EM: BRL, CLP, CZK, HUF, IDR, ILS, INR, KRW, MXN, PHP, PLN, RUB, SGD, TRY, TWD and ZAR.

Table 1: List of Currencies

Currency	Name	Average Daily Turnover (USD Billions)	%	Reason for Exclusion
USD	United States Dollar	2912.00	44.15%	
EUR	Euro	1064.50	16.14%	
JPY	Japanese Yen	554.00	8.40%	
GBP	British Pound	422.00	6.40%	
AUD	Australian Dollar	223.50	3.39%	
CAD	Canadian Dollar	166.00	2.52%	
CHF	Swiss Franc	163.50	2.48%	
CNY	Chinese Yuan	142.50	2.16%	Particularities
HKD	Hong Kong Dollar	116.50	1.77%	Pegged
NZD	New Zealand Dollar	68.50	1.04%	
SEK	Swedish Krona	67.00	1.02%	
KRW	Korean Won	66.00	1.00%	
SGD	Singapore Dollar	59.50	0.90%	
NOK	Norwegian Krone	59.50	0.90%	
MXN	Mexican Peso	57.00	0.86%	
INR	Indian Rupee	57.00	0.86%	
RUB	Russian Ruble	36.00	0.55%	
ZAR	South African Rand	36.00	0.55%	
TRY	Turkish Lira	35.50	0.54%	
BRL	Brazilian Real	35.50	0.54%	
TWD	New Taiwan Dollar	30.00	0.45%	
DKK	Danish Krone	21.00	0.32%	Pegged
PLN	Polish Zotly	20.50	0.31%	
THB	Thai Bhat	16.00	0.24%	Particularities
IDR	Indonesian Rupiah	13.50	0.20%	
HUF	Hungarian Forint	13.50	0.20%	
CZK	Czech Koruna	13.00	0.20%	
ILS	Israeli New Shekel	10.00	0.15%	
CLP	Chilean Peso	9.50	0.14%	
PHP	Philippine Peso	9.50	0.14%	
AED	UAE Dirham	7.00	0.11%	Pegged
COP	Colombian Peso	6.00	0.09%	Liquidity
SAR	Saudi Riyal	6.00	0.09%	Liquidity, Pegged
MYR	Malasiyan Ringgit	4.50	0.07%	Liquidity, Particularities
RON	New Romanian Leu	3.00	0.05%	Liquidity
PEN	Peruvian New Sol	2.50	0.04%	Liquidity
ARS	Argentine Peso	2.00	0.03%	Liquidity
BHD	Bahraini Dinar	1.00	0.02%	Liquidity, Pegged
BGN	Bulgarian Lev	1.00	0.02%	Liquidity, Pegged
Others	-	64.50	0.98%	Liquidity
Total	-	6595	100%	

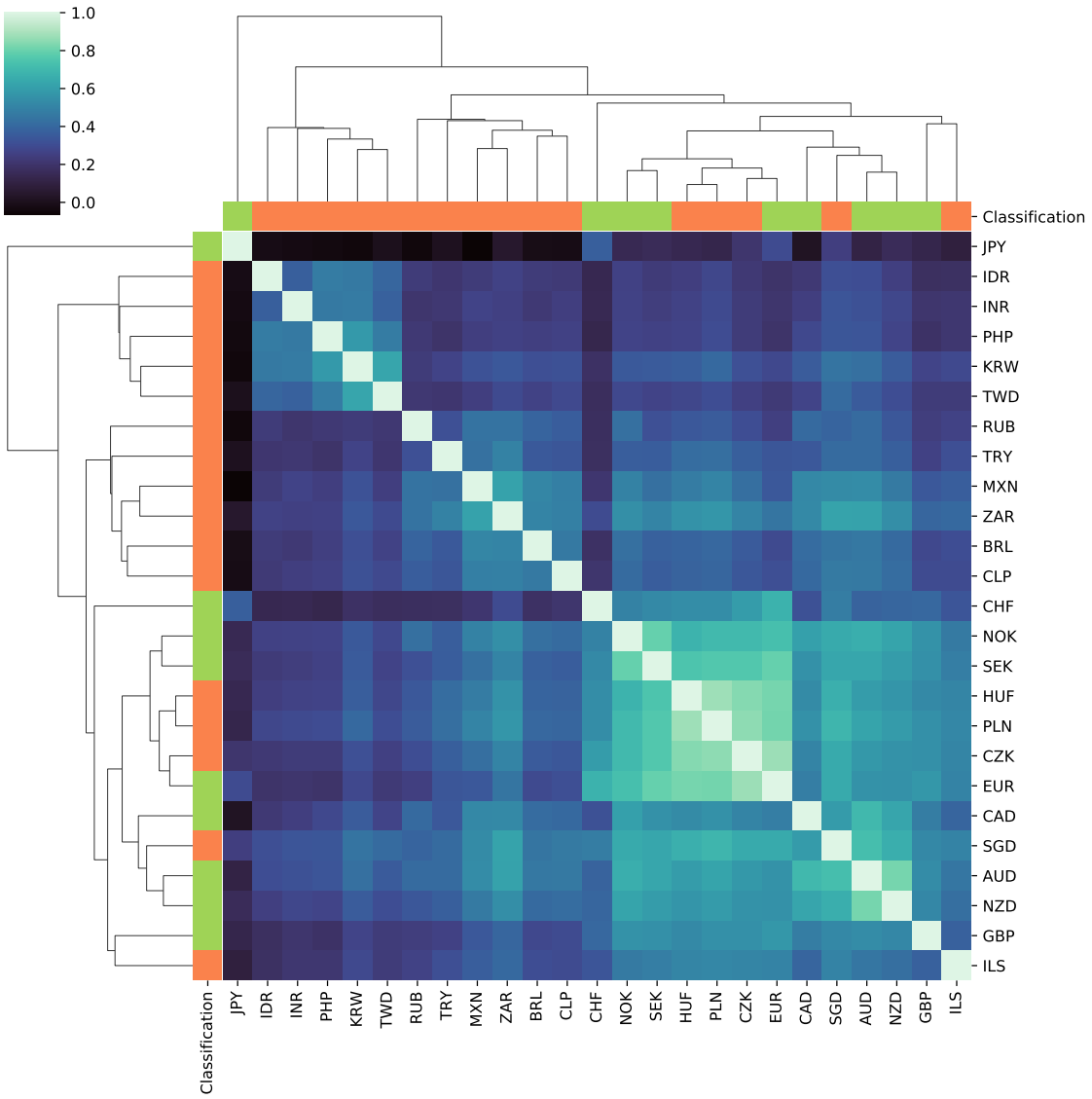

Figure 4: Correlation matrix and dendrogram of the total return of the currencies. Returns are mesured from forward contracts, so they can be interpreted as excess returns net of carry. The border of the correlation matrix (classification) shows if the currency comes from the G10 group (green) or EM group (red). Sample goes from 2009 (after the GFC) to 2019 (before covid).

The authors of [13] show that the resulting dynamics of rh G10 and EM groups are sometimes very different. They also point out that we can run a strategy separately on each group and then combine the strategies or use all currencies in a single strategy.

These different dynamics coming from G10 and EM play in our favor when it comes to diversification. This can be seen in figure 4, which shows the correlation matrix between our selected currencies and on its margins we have the dendrogram and the classification of that currency as either from G10 (green) or EM (red). The correlations were computed based on a sample from 2009 to 2019, so it leaves out the great financial crises of 2008 and the covid crises of 2020 and 2021. In figure 4 we can see the clusters of currencies that arise from their behavior. Close to the upper left corner we can see a group of 5 currencies (IDR, INR, PHP, KRW and TWD), which are all of the asian currencies in the EM group. The second most noticible block seems to be the high yielding EM currencies (RUB, TRY, BRL, MXN, CLP and ZAR) which are among the most volatile ones. The third most noticeble block is the one with the strongest correlation among its components (NOK, SEK, HUF, PLN, CZK and EUR) which are all of the european currencies, the nordics, the emerging eastern europeans and the euro itself. A fourth group is composed by the most developed economies (CAD, NZD, AUD, SGD), even though Singapore is not formally classified as a developed econmy, although it has many of the characteristics of one both in terms of economics and of asset behaviours. Outside of these 4 biggest groups, the remaing currencies seem to have low correlations with almost everyone. Thes are the last 4 currencies to join the groups in the dendrogram. The JPY is almost independent, with correlations close to zero with all other currencies and is most certainly a currency that brings a great deal of diversification to a portfolio. The CHF has some correlation with the european block, but it has very low correlations with the EM group, leaving GBP and ILS as the remaining diversifiers. Based on these correlations and measures of proximity we could classify our currencies based on their behaviour as:

EM Asia: IDR, INR, PHP, KRW and TWD

EM High Yielders: RUB, TRY, BRL, MXN, CLP and ZAR

Europeans: NOK, SEK, HUF, PLN, CZK and EUR

Most Developed: AUD, CAD, NZD and SGD

Diversifiers: JPY, CHF, GBP, ILS

3 CQF Requirements for Portfolio Choice and Data

- We have selected 25 viable currencies from a list of the 39 most traded, according to the BIS. Currencies were eliminated based on liquidity, currency regime and other particularities like short history, stressed events and past convertibility problems.

- Their correlations were estimated from 2009 to 2019 to exclude periods of major crises and regime changes. In this case, we had in mind the great financial crises of 2008 and covid crises of 2020 and 2021.

- We chose **not** narrow down the number of assets to less than 10 as this could compromise the diversification. Our programming tools allow us to handle as many assets as we need and we can later evaluate which currencies can be removed for performance issues. Also, the author of the Marchenko-Pastur

denoising method emphasizes that it requires a large number of observations on both dimensions (the length of the time-series of returns and the number assets) in order for the method to work properly.

- As we could see in the correlation matrix and dendrogram in figure 4, we have 4 diversifying currencies: JPY, CHF, GBP and ILS. These are the last 4 currencies that join the group based on the dendrogram. In other words, these are the currencies whose behaviour differ the most from others and will play the role of our "exogenous" assets.

- Although we are looking at the FX market and soon we will talk about carry, momentum and value of these currencies, we are **not** going to build factor portfolios. Our strategy consists of usign model-free characteristics of the currencies (carry, momentum and value) as judgements and estimates of expected returns to be used in the Black-Litterman model and not to go long/short these currencies in other to identify pricing factors.

- As stated in the final project Q&A, we were not limited to options A, B or C. We believe that our set of investible assets is closer to option A.

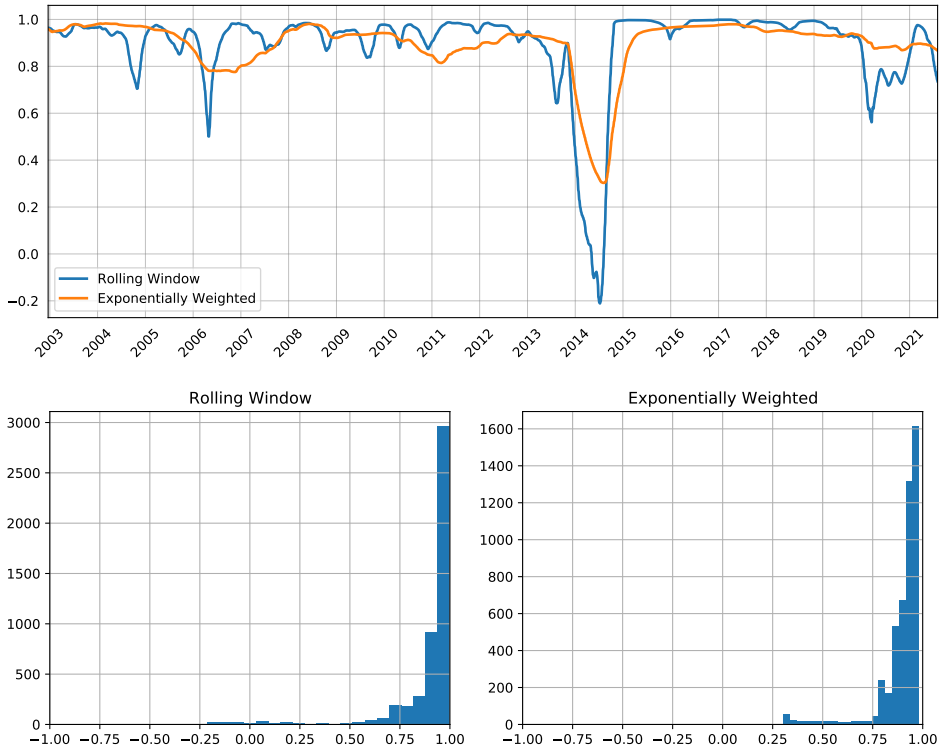

Figure 5: Correlations between the **EUR** and **CZK**. The size of the rolling window and center of mass of the exponentially weighted average is 1 year. Histograms show the frequency of each estimate.

Part II

Covariance Estimatimation

4 Motivation

To motivate the need of denoising and stabilizinhg methods let us look at the correlations between 2 pairs of currencies, EUR-CZK and JPY-BRL. In figure 5 we can see the time varying estimates of correlations between the EUR and the CZK. The size of the rolling window and center of mass of the exponential weights is 1 year. We can see that correlations stay close to 1, with the exception of 2014. In the histograms we can see that both methods generate consistent estimates that are close to one. This is one situation where stability does not seem to be an issue.

Figure 6 paints the opposite picture. There we see the correlations between the JPY and the BRL. There we see that correlations change a lot over time, but in magnitude and speed of change. The histograms show that both methods generate similar distributions but they span the complete parameter space from -1 to 1. In this case stability of estimates and noise are more likely to be an issue for portfolio constructions.

Empirical covariance matrices estimate the linear comovement of between random variables in a random vector. But since this is made on a noisy finite sample, the covariance estimates are also noisy. Unless this noise is treated, subsequent calculations like portfolio allocations will also be noisy.

The author of [15], Marcos Lopez de Prado, developed a way to reduce the noise in covariance matrices

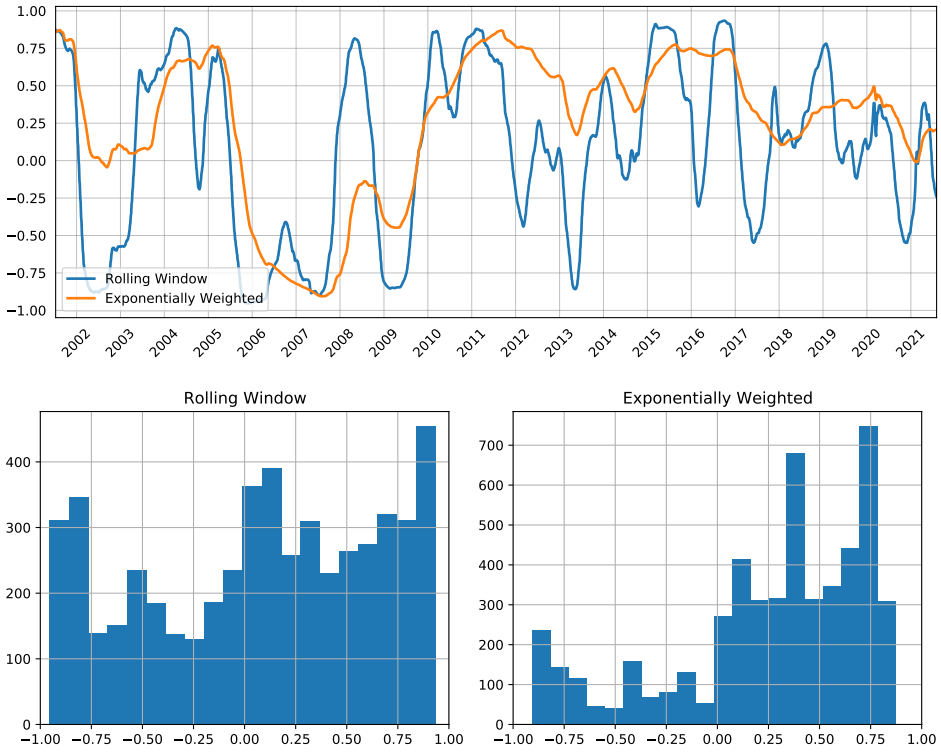

Figure 6: Correlations between the **JPY** and **BRL**. The size of the rolling window and center of mass of the exponentially weighted average is 1 year. Histograms show the frequency of each estimate.

with the help of the Marchenko-Pastur Theorem. This method is explained in section 6. The author emphasizes that his method is capable of reducing the noise without dilluting the signal, but that it is **not** capable of increasing stability of the estimates. That is why we also explain (and later apply) the shirinkage estimates of Ledoit-Wolfe [REFERENCIA AQUI], which are explicitly designed to improve stability and likely compromises the signal coming form the empirical covariance.

These two method do two different things and might even be combined. If we prioritize the correct allocation at all times, this will require enahncing the signal and probably bigger rebalances. On the other hand, if our rebalacing costs are too high, we might want to minimize them in order to save on costs, even if that makes the allocation not perfectly optimal at all times. This is analogous to the bias-variance tradeoff of statistical learning methods. The low bias/high variance situations rebalances the portfolio faster towards the optimum but incurs in more trading costs, while the high bias/low varaince situation saves on costs but the portfolio lags the optimal allocation. So we are going to tune the hyperparameters of the strategy (including the ones from the covariance stabilization methods) to find the best balance between a timely portfolio and trading costs.

5 Naive Covariance

When estimating empirical covariance matrices with financial data it is not hard to get in situations of ill-conditioned matrices. This usually happens when we are working with several different time-series of returns, each with different samples sizes. Computing the individual covariances between every possible pair of series

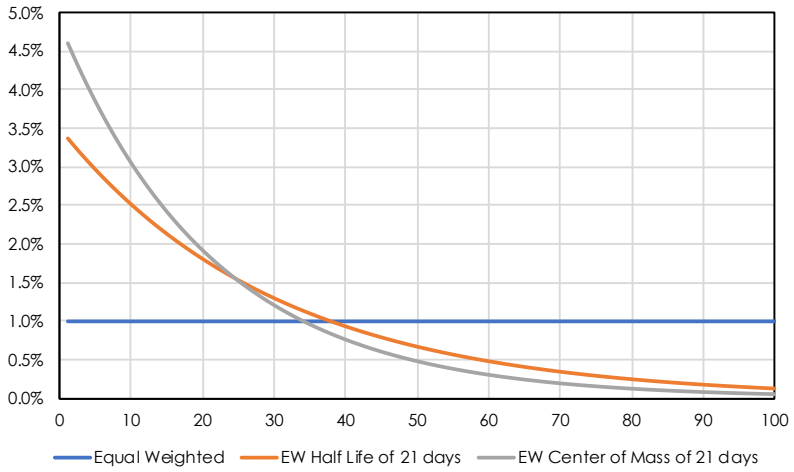

Figure 7: Comparison of equal weights and exponential weights with different values of α.

does not guarantee a covariance matrix that is positive-definite. To go around this problem we built long timeseries for total return indexes, in order to have a sufficiently big sample for all currencies.

Even if the genarated covariance is non-singular, it will typically have a very small value for the determinant which will propagate and amplify estimation errors when this matrix is inverted. This causes large estimation error, leading to misallocation of assets, and instability of the estimates, leading to increased transaction costs coming from unnecessary rebalances. This is why we are going to adjust our estimates of the empirical covariance matrix.

Even when the empirical covariance is adjusted we still have to think about how estimate it. The most commom ways are using a rolling window, where the number of observations turns into a hyperparameter of the strategy, and using an expanding window but with exponentially decaying weights, where the center of mass/half-life of the weights turns into a hyperparameter of the strategy that needs to be tuned.

The exponential weights (EW) for each observation on date t are given by $w_{t-i} = (1-\alpha)^i$, that the furthest in the past that an observation is, the smaller its weight will be. We can choose the α parameter directly, or based on the half-life[2] λ of the weights:

$$\alpha = 1 - e^{-\frac{\ln 2}{\lambda}}$$

or based on the center of mass (com) of the weights, which usually are interpreted as the "window" of the EW method:

$$\alpha = \frac{1}{1+com}$$

Figure 7 compares the observations weights between these two methods for two different values of α.

[2]number of periods until the weights decrease by half

6 Robust Covariance

6.1 The Marchanko-Pastur Theorem

Consider a matrix X of independent and indentically distributed random observations, of size $T \times N$, where the underlying process generating the observations has zero mean and variance σ^2. Define matrix C as:

$$C = \frac{1}{T} X'X$$

This matrix has eigenvalues $\lambda_1, \ldots, \lambda_N$. As T and N go to infinity, the distribution of the eigenvalues converge to the Marchenko-Pastur probability density function:

$$f(\lambda) = \begin{cases} \frac{T}{N} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda\sigma^2} & if\ \lambda \in [\lambda_-, \lambda_+] \\ 0 & otherwise \end{cases}$$

where the maximum expected eigenvalue is $\lambda_+ = \sigma^2 \left(1 + \sqrt{\frac{N}{T}}\right)^2$ and the minimum expected eigenvalue is $\lambda_+ = \sigma^2 \left(1 - \sqrt{\frac{N}{T}}\right)^2$. So, eigenvalues that lie in the $[\lambda_-, \lambda_+]$ interval are consistent with the random bahaviour.

The Marchenko-Pastur distribution is only valid when the whole dataset is randomly generated from a single distribution. In the financial world there are commom factors on these time-series, so not all the eigenvectors are random. Any eigenvalue found outside of this distribution can be thought as non-random, meaning they are the ones that carry the signal.

We can estimate σ^2 by finding the value that best fits the empirical distribution of our observed eigenvalues. This will give us the variance that is explained by the random eigenvectors of the correlation matrix. The approximation of the empirical distribution of eigenvalues to the Marchenko-Pastur distribution gets better as the number of observed eigenvalues increases, in other words, when the number of assets is big. This might be a bad approximation when working with a small number of assets. By finding the value of σ^2 that best fits the empirical distribution, we can find the λ_+ cutoff for the eigenvalues of the correlation matrix, anything above it can be interpreted as signal bearing eigenvalues.

The number of eigenvalues of the correlation matrix above this λ_+ cutoff is the estimate of the number of factors in the correlation matrix. The estimated value of σ^2 is the estimate of the percentage of noise in the correlation matrix.

Knowing this allows us to reconstruct the correlation matrix using only the eigenvalues that carry the non-random signals. The approach consists in setting a constant eigenvalue for all random eigenvectors, while keeping the trace of the correlation matrix. Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of the correlation matrix in descending order and let i be the index of the last eigenvalue outside the Marchenko-Pastur distribution, meaning that $\lambda_i > \lambda_+$ and $\lambda_{i+1} \leq \lambda_+$. Then, we set

$$\lambda_j = \frac{1}{(N-i)} \sum_{k=i+1}^{N} \lambda_k$$

for all $i < j \leq N$. Now, given the eigen decomposition of a correlation matrix $C = W\Lambda W'$ we re reconstruct

the denoised correlation matrix $\tilde{C}$ as

$$\tilde{C}_{aux} = W\tilde{\Lambda}W' \tag{1}$$

$$\tilde{C} = \tilde{C}_{aux}\left[\left(diag\left[\tilde{C}_{aux}\right]^{\frac{1}{2}}\right)\left(diag\left[\tilde{C}_{aux}\right]^{\frac{1}{2}}\right)'\right]^{-1} \tag{2}$$

where $\tilde{\Lambda}$ is a diagonal matrix holding the corrected eigenvalues and $diag\left[\bullet\right]$ zeros all non-diagonal elements of a square matrix. The $\tilde{C}_{aux}$ is just an intermediate step in computations, because the second computation makes sure that the diagonal of $\tilde{C}$ are all ones.

The author of the mehtod argues that this is better than shrinkage methods, because although shrinkage also reduces the condition number of the correlation matrix, it does not distinguish between signal and noise. If the signal is low on correlations matrices, shrinking it would only make things worse.

6.2 Targeted Shrinkage

As stated by Lopez de Prado (2019), the Marchenko-Pastur method is preferable to shrinkage because it removes the noise while preserving the signal. Alternatively, we could target the application of the shrinkage strictly to the random eigenvectors. Consider the correlation matrix given by:

$$\tilde{C}_{aux} = W_L\Lambda_L W_L' + \alpha W_R\Lambda_R W_R' + (1-\alpha)\, diag\left(W_R\Lambda_R W_R'\right)$$

where W_R and Λ_R are the eigenvectors and eigenvalues associated with noise ($\lambda_i \leq \lambda_+$), W_L and Λ_L are the eigenvectors and eigenvalues associated with signals ($\lambda_i > \lambda_+$) and α regulates the ammount of shrinkage. If $\alpha = 1$, there is no shrinkage and we end up in the Marchenko-Pastur case, and if $\alpha = 0$ we get total shrinkage on the noisy part of the correlation matrix.

Notice that after we compute $\tilde{C}_{aux}$ we still need to make the adjustment shown in equation (2) in order to make sure that the diagonal of $\tilde{C}$ are all ones.

6.3 Detoning

Financial assets are usually subject to a market factor, which would be characterized by the first eigenvector. In te context of clustering applications, it is useful to remove the market component, if it exists. The reason is, it is more difficult to cluster a correlation matrix with a strong market component, because the algorithm will struggle to find dissimilarities across clusters. By removing the market component, we allow greater portion of the correlation to be explained by components that affect specific subsets of the securities and not one big commom factor.

To get the detoned correlation matrix $\tilde{C}_D$, we can remove that market component from an already denoised correlation matrix $\tilde{C}$:

$$\tilde{C}_{aux} = \tilde{C} - W_M\Lambda_M W_M'$$

$$\tilde{C}_D = \tilde{C}_{aux}\left[\left(diag\left[\tilde{C}_{aux}\right]^{\frac{1}{2}}\right)\left(diag\left[\tilde{C}_{aux}\right]^{\frac{1}{2}}\right)'\right]^{-1}$$

where W_M and Λ_M are the eigenvectors and eigenvalues associated with the market component (usually just the first one). The $\tilde{C}_{aux}$ is just an intermediate step in computations, because the second computation makes

sure that the diagonal of $\tilde{C}_D$ are all ones.

With the removal of an eigenvector, the detoned correlation matrix is singular. This is not a problem for clustering applications, which usually do not require invertibility of the correlation matrix. Still, detoned correlation matrices cannot be used for mean-variance optimizations.

6.4 Numerical Example

In order to test if our code for the Marchenko-Pastur denoising is working properly we are going to simulate data with a similar size as our FX returns data. We are simulating 25 assets with 5000 observations each. For the correlation matrix we are generating a randomly generated correlation matrix with 3 underlying factors, meaning that 3 eigenvectors carry the signals and the other 22 eigenvectors are just noise. The details of how this correlation matrix can be simulated is detalied in [15] and the code is available in our files. Figure 8 shows the simulated empirical correlation.

The Marchenko-Pastur denoising detected exactly 3 eigenvectors containing signals, just as we simulated. Figure 9 shows the sorted eigenvalues on a logarithmic scale for the 3 correlation matrixes. We can see that the 3 biggest eigenvalues are significantly bigger then the remaining 22, as expected from the 3 uderlying factors in our simulation. From the 4th eigenvalue and on, they are all associated with noise and drecrease monotonically. The green line shows the sorted eigenvalues for the target-shrinkage correlation matrix with a shrinkage parameter of $\alpha = 1$, meaning that we have full shrinkage of the noise eigenvectors. Since these 22 noise eigenvectors correspond to the minority of the total variance of data, even with their total shrinking we still get eigenvalues that are not that different from the originals in the empirical correlation matrix. Finally, the orange line shows the sorted eigenvalues of the denoised correlation. We can see that all of the eigenvalues associated of noise eigenvectors have been normalized to the same value. Some of them are smaller, but the key here is that the very small eigenvalues are gone. These are the ones that would cause instability when inverting the covariance or correlation matrix.

Table 2 shows a few characteristics of each resulting correlation matrix of each method, which also brings some insights to the stability and signal issues. The maximium eigenvalues of each matrix are very similiar, meaning that the strength of the strongest signals are not compromised. The minimum eigenvalue of the empirical correlation matrix is 0.0293. This value increases 1.542 times in the targeted shrinkage matrix and 6.617 times in the denoised matrix. These are the extremes values of the eigenvalues, but the determinant is the characteristic of the matrix that generates the typical instability when applying methods that require inverting the correlation matrix. The stability gain in the denoising method is much bigger with the targeted shirinkage. The determinant of the denoised matrix is around 193 times bigger than the determinant of the empirical covariance. Although the determinant is still very small, it is many times bigger than the empirical determinant. In order to have a feel for how much the values of the correlations matrix change with these methods, the last two lines show norms of the difference matrix between each method and the empirical correlation. The biggest change in the targeted shrinkage is only 0.07. The change in the denoised matrix got up to 0.18 a much bigger change, but with the intent of reducing noise.

6.5 Application to FX Data

Now that we know that our code is works properly, with intuitive results when using simulated data, we can apply the same analysis to our selected assets. We have already seen the correlation matrix of the currencies in figure 4. For this case, the Marchenko-Pastur denoising also detected 3 eigenvectors containing signals.

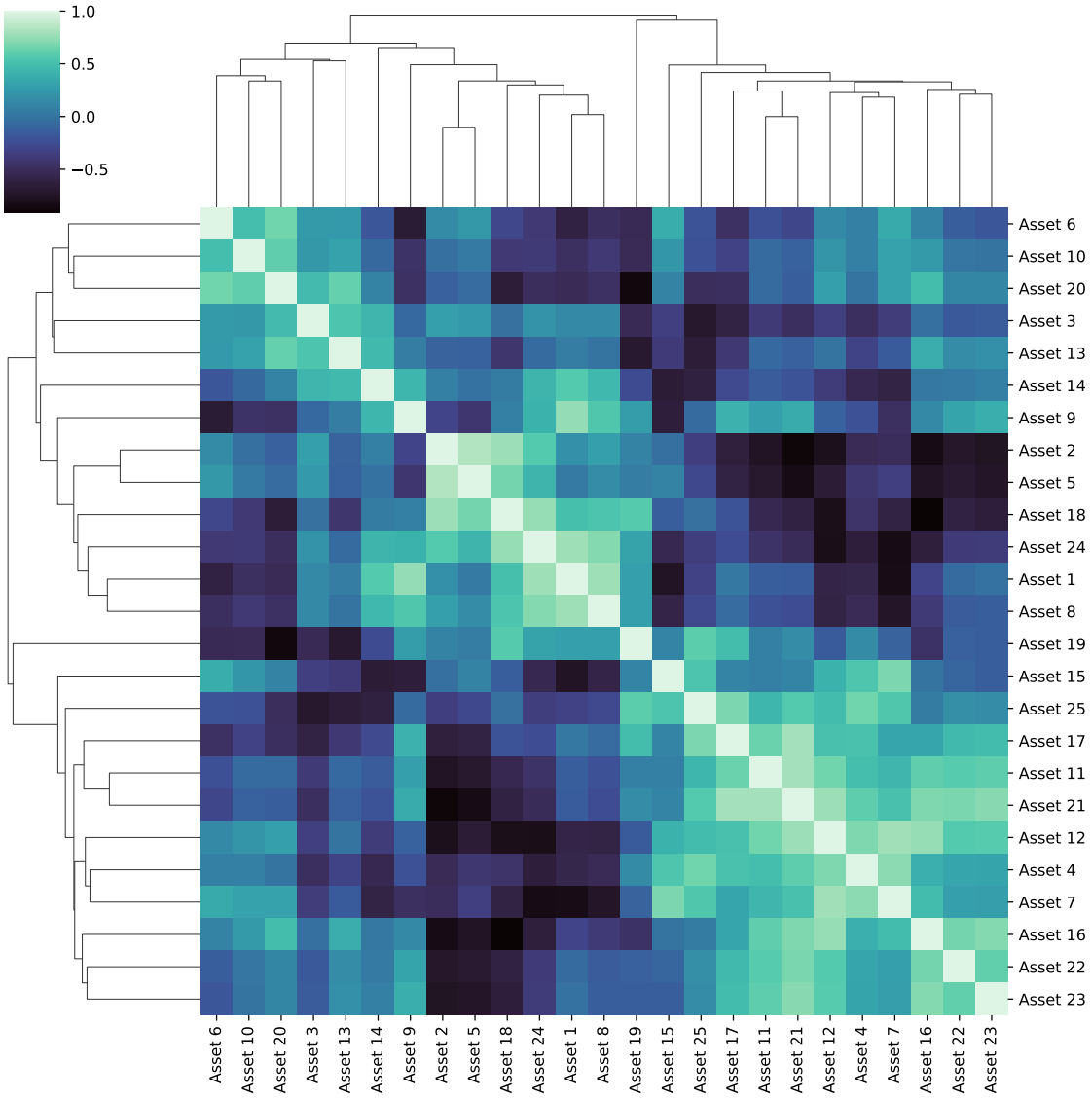

Figure 8: Correlation matrix of the 25 simulated asset returns with 3 underlying signal factors and 22 noise factors

Table 2: Characteristics of the correlation matrices of simulated returns

	Empirical	**Denoised**	**Targeted Shirinkage**
Maximum Eigenvalue	9.5140	9.2765	9.6769
Minimum Eigenvalue	0.0293	0.1939	0.0452
Determinant	7.2516e-15	1.3977e-12	2.2764e-15
Maximum Norm	0.0	0.1851	0.0704
Frobenius Norm	0.0	1.12490	0.4545

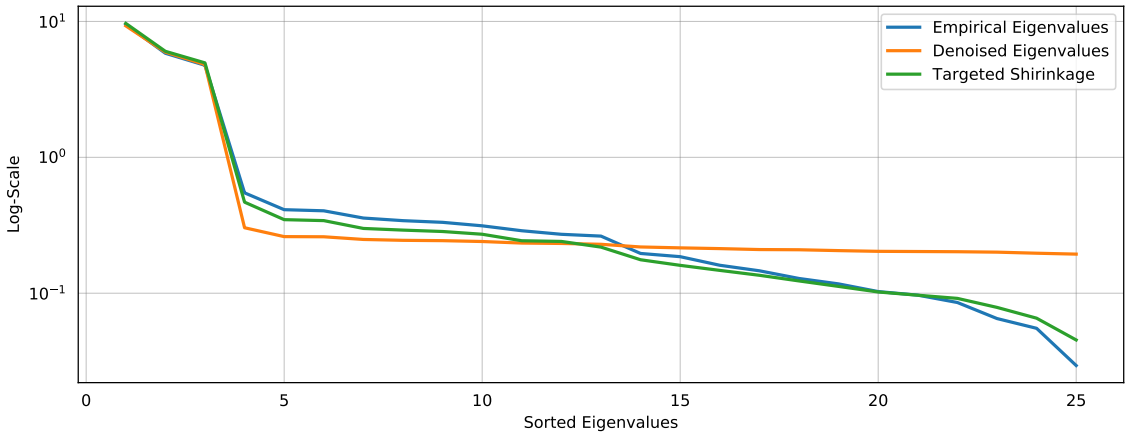

Figure 9: Sorted eigenvalues for each correlation matrix method

Figure 10 shows the sorted eigenvalues on a logarithmic scale for the 3 correlation matrixes. Just like in our simulated data, we can see that the 3 biggest eigenvalues are significantly bigger then the remaining 22. From the 4th eigenvalue and on, they are all associated with noise. Again, the green line shows the sorted eigenvalues for the target-shrinkage correlation matrix with a shrinkage parameter of $\alpha = 1$, meaning that we have full shrinkage of the noise eigenvectors. Since these 22 noise eigenvectors correspond to the minority of the total variance of data, even with their total shrinking we still get eigenvalues that are not that different from the originals in the empirical correlation matrix. If we set $\alpha = 0$ the eigenvalues would match the ones from the empircal correlation matrix. The orange line shows the sorted eigenvalues of the denoised correlation. We can see that all of the eigenvalues associated of noise eigenvectors have been normalized to similar values. Some of them are smaller, but the key here is that the very small eigenvalues are gone. These are the ones that would cause instability when inverting the covariance or correlation matrix. The big eigenvalues, which are the ones that carry the signals, that have high explanatory power, did not change that much.

Table 3 shows the same matrix characteristics as before for each method and the results are very in line with the ones from our simulated data. The maximium eigenvalues of each matrix are similiar, meaning that the strength of the strongest signals are not compromised. The minimum eigenvalue of the empirical correlation matrix is 0.1002. This value increases 1.397 times in the targeted shrinkage matrix and 3.647 times in the denoised matrix, which is not as much as in the simulated data, but it is still a significant increase. These are the extremes values of the eigenvalues, but the determinant is the characteristic of the matrix that generates the typical instability when applying methods that require inverting the correlation matrix. The stability gain in the denoising method is much bigger with the targeted shirinkage. The determinant of the denoised matrix is 48 times bigger than the determinant of the empirical covariance. In this case, the determinant is not as small as our simulated data, indicating the 22 noise eigenvalues carry more of the variability of the data than in our simulated case. The norms of the matrix with respect to the empirical correlation show bigger changes than in the simulated data. The biggest change, again, came in the denoise method with a maximum of 0.26 change on an estimate of correlation while the maximum change in the targeted shirinkage is only 0.19.

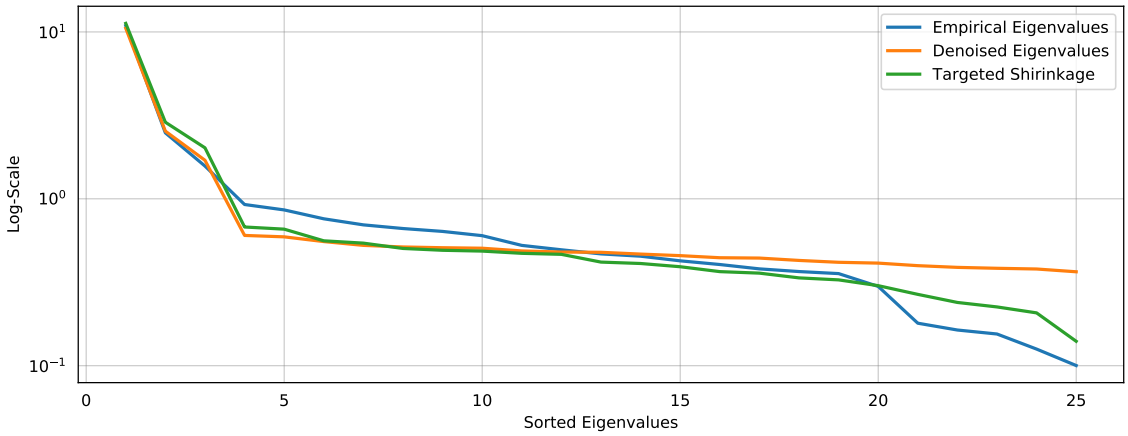

Figure 10: Sorted eigenvalues for each correlation matrix method

Table 3: Characteristics of the correlation matrices of FX returns

	Empirical	Denoised	Targeted Shirinkage
Maximum Eigenvalue	10.9011	10.5224	11.2545
Minimum Eigenvalue	0.1002	0.3655	0.1400
Determinant	3.6984e-08	1.7754e-06	2.8738e-08
Max Norm	0.0	0.2606	0.1940
Frobenius Norm	0.0	1.5243	1.0993

Part III

Expected Returns Based on Views

7 Black-Litterman

Traditional portfolio optimization usually takes as inputs the vector of expected returns and the covariance matrix of the assets. These look like few inputs for a task as important as finding an optimal asset allocation, especially for clients. The Black-Litterman (BL) model [1] looks at a larger set of inputs (views, confidence on views, expected returns, uncertainty of the reference model, covariances) in order to generate a portfolio that connects the views of the portfolio managers and market equilibrium. Here we are going to through the original model . Our references were [3, 8].

7.1 The Original Model

7.1.1 The Model for Returns

Consider a market of N securities or asset classes. Their returns are modelled by the following normal distribution:

$$X \sim N(\mu, \Sigma) \tag{3}$$

The covariance matrix Σ can be estimated using any robust method but it must be a positive definite matrix[3]. Since there is uncertainty around μ it cannot be treated as a given value, but modeled as a random variable. The original BL model states that μ is normally distributed:

$$\mu \sim (\pi, \tau\Sigma) \tag{4}$$

where π represents the best guess for μ and $\tau\Sigma$ is the uncertainty of the guess. To chose the value of π, BL use an equilibrium argument. Assuming there is no estimation error, the reference model (3) becomes:

$$X \sim N(\pi, \Sigma)$$

Investors maximize an unconstrained mean-variance trade-off

$$w_\lambda = \arg\max_w \{w'\pi - \lambda w'\Sigma w\}$$

The first order condition is given by the first derivative of the objective function with respect to w

$$\pi - 2\lambda\Sigma w = 0$$

If we have an average risk aversion $\bar{\lambda}$ and the equilibrium allocation $\tilde{w}$ we can estimate π as

$$\pi = 2\bar{\lambda}\Sigma\tilde{w}$$

The original BL paper uses $\bar{\lambda} = 1.2$, but we can change this parameter later. The initial portfolio $\tilde{w}$ can be based on a benchmark[4].

The mean value for the returns can be estimated using shrinkage:

$$\hat{\mu} = \sum_{t=1}^{T} X_t \qquad \pi^{(0)} = 2\bar{\lambda}\Sigma\tilde{w}$$

$$\mu^{(s)} = (1-s)\,\hat{\mu} + s\pi^{(0)} \tag{5}$$

The overall uncertainty parameter from τ can be set as

$$\tau = \frac{1}{T}$$

This allows for a simple measure of overall uncertainty. There are other ways to set this parameter, but they usually require more parameters and calibrations.

7.1.2 The Views

Next, we include views on the model. A view is a statement on the asset returns that can potentially disagree with the reference from the model (3). The BL model allows for linear views on μ. A set of K views is represented by a $K \times N$ "pick" matrix P, in which the k-th row determines the relative weight of each

[3]Methods like detoning, previously described in this report, generate covariance matrices that are not positive definite. This type of method cannot be used for this kind of application.

[4]As an example, in the case of single stocks, we can use the market-value weighted portfolio, and for asset classes or strategies we can use for example, a risk parity portfolio or a simple risk based allocation.

expected return in the respective view.

$$P\mu \sim N(v, \Omega) \tag{6}$$

where v and Ω quantify the views and their uncertainty. It is simple and convevient to set the uncertainty of the views as

$$\Omega = \frac{1}{c} P\Sigma P'$$

so the structure and correlations on uncertainties inherits the ones from the market and $c \in [0, \infty]$ represents an overall level of confidence relative to market uncertainty. If $c = 1$ the uncertainty of views is the same as the markets, if $c < 1$ the uncertainty in amplified and if $c > 1$ the uncertainty of the views are smaller than the market. We can also set different levels of condicence for each view with

$$\Omega = \frac{1}{c} diag(u) P\Sigma P' diag(u)$$

and $u \in [0, \infty]^K$.

The values for the views themselves can be set quantitatively in v, or we can make some adaptations and simplifications in order to set them qualitatively.

$$v_k = (P\pi)_k + \eta_k \sqrt{(P\Sigma P')_{k,k}}$$

Where $\eta_k \in [-\beta, -\alpha, 0, \alpha, \beta]$ with $\alpha < \beta$. This represents the qualitative level of confidence[5], in this case, they respectively mean "very bearish", "bearish", "neutral", "bullish" and "very bullish".

7.1.3 The Posterior Distribution

The last step of the model is to find the posterior distribution of the returns X. The first thing we need is the posterior $f_{\mu|v}(\mu)$ which gives us the best gues for μ and the uncertainty around it. After that we adjust it again to find the posterior $f_{X|v}(x)$.

To determine the posterior distribution of μ given V we apply the Bayes rule:

$$f_{\mu|v}(\mu) = \frac{f_{\mu,V}(\mu, V)}{f_V(v)} = \frac{f_{V|\mu}(v) f_\mu(\mu)}{\int f_{V|\mu}(v) f_\mu(\mu) d\mu} \tag{7}$$

The elements of the numerator are $f_\mu(\mu)$ and $f_{V|\mu}(v)$. The $f_\mu(\mu)$ p.d.f. for μ comes from (4) and is given by:

$$f_\mu(\mu) = \frac{|\tau\Sigma|^{-\frac{1}{2}}}{(2\bar{\pi})^{\frac{N}{2}}} e^{-\frac{1}{2}(\mu-\pi)'(\tau\Sigma)^{-1}(\mu-\pi)}$$

The conditional p.d.f. $f_{V|\mu}(v)$ is little more trouble to get to. We start by rewriting (6) as:

$$v = P\mu + Z \qquad Z \sim N(0, \Omega)$$

So we can model v as a random variable V whose distribution is conditioned on the realization of μ:

$$V|\mu \sim N(P\mu, \Omega)$$

[5] typical values for these parameters are $\alpha = 1$ and $\beta = 2$

The conditional p.d.f. of this variable is:

$$f_{V|\mu}(v) = \frac{|\Omega|^{-\frac{1}{2}}}{(2\bar{\pi})^{\frac{K}{2}}} e^{-\frac{1}{2}(v-P\mu)'\Omega^{-1}(v-P\mu)}$$

Now, we can write the joint p.d.f. $f_{\mu,V}(\mu, v)$ as:

$$\begin{aligned}
f_{\mu,V}(\mu,v) &= f_{V|\mu}(v) f_{\mu}(\mu) \\
&= \frac{|\Omega|^{-\frac{1}{2}}}{(2\bar{\pi})^{\frac{K}{2}}} e^{-\frac{1}{2}(v-P\mu)'\Omega^{-1}(v-P\mu)} \frac{|\tau\Sigma|^{-\frac{1}{2}}}{(2\bar{\pi})^{\frac{N}{2}}} e^{-\frac{1}{2}(\mu-\pi)'(\tau\Sigma)^{-1}(\mu-\pi)} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[(\mu-\pi)'(\tau\Sigma)^{-1}(\mu-\pi)+(v-P\mu)'\Omega^{-1}(v-P\mu)\right]} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[(\mu'-\pi')((\tau\Sigma)^{-1}\mu-(\tau\Sigma)^{-1}\pi)+(v'-\mu'P')(\Omega^{-1}v-\Omega^{-1}P\mu)\right]} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[\mu'(\tau\Sigma)^{-1}\mu-\mu'(\tau\Sigma)^{-1}\pi-\pi'(\tau\Sigma)^{-1}\mu+\pi'(\tau\Sigma)^{-1}\pi)+(v'\Omega^{-1}v-v'\Omega^{-1}P\mu-\mu'P'\Omega^{-1}v+\mu'P'\Omega^{-1}P\mu)\right]} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[\mu'(\tau\Sigma)^{-1}\mu-2\mu'(\tau\Sigma)^{-1}\pi+\pi'(\tau\Sigma)^{-1}\pi+v'\Omega^{-1}v-2\mu'P'\Omega^{-1}v+\mu'P'\Omega^{-1}P\mu\right]} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left\{\mu'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]\mu-2\mu'\left[(\tau\Sigma)^{-1}\pi+P\Omega^{-1}v\right]+\pi'(\tau\Sigma)^{-1}\pi+v'\Omega^{-1}v\right\}}
\end{aligned}$$

Now we define

$$\mu_{BL} = \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}v\right] \tag{8}$$

$$\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]\mu_{BL} = \left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}v\right]$$

So we can rewrite the joint p.d.f. as

$$\begin{aligned}
f_{\mu,V}(\mu,V) &= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left\{\mu'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]\mu-2\mu'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]\mu_{BL}(v)+\pi'(\tau\Sigma)^{-1}\pi+v'\Omega^{-1}v\right\}} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left\{\mu'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]\mu-2\mu'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right]\mu_{BL}+\pi'(\tau\Sigma)^{-1}\pi+v'\Omega^{-1}v+\mu'_{BL}\left((\tau\Sigma)^{-1}+P'\Omega^{-1}P\right)\mu_{BL}-\mu'_{BL}\left((\tau\Sigma)^{-1}+P'\right.\right.} \\
&= (2\bar{\pi})^{-\frac{K+N}{2}} |\tau\Sigma|^{-\frac{1}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}\left\{(\mu-\mu_{BL})'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right](\mu-\mu_{BL})+\pi'(\tau\Sigma)^{-1}\pi+v'\Omega^{-1}v-\mu'_{BL}\left((\tau\Sigma)^{-1}+P'\Omega^{-1}P\right)\mu_{BL}\right\}}
\end{aligned}$$

Let us simplify the last three terms inside the curly brackets

$$\pi'(\tau\Sigma)^{-1}\pi + v'\Omega^{-1}v - \mu'_{BL}\left((\tau\Sigma)^{-1} + P'\Omega^{-1}P\right)\mu_{BL}$$

$$\pi'(\tau\Sigma)^{-1}\pi + v'\Omega^{-1}v - \left[\pi'(\tau\Sigma)^{-1} + v'\Omega^{-1}P\right]\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}v\right]$$

$$\pi'(\tau\Sigma)^{-1}\pi + v'\Omega^{-1}v - \left[\pi'(\tau\Sigma)^{-1} + v'\Omega^{-1}P\right]\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}v\right]$$

$$\pi'(\tau\Sigma)^{-1}\pi + v'\Omega^{-1}v - \left\{\pi'(\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1} + v'\Omega^{-1}P\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\right\}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}v\right]$$

$$\pi'(\tau\Sigma)^{-1}\pi + v'\Omega^{-1}v - \left\{\pi'(\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}(\tau\Sigma)^{-1}\pi + 2\pi'(\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}P'\Omega^{-1}v + v'\Omega^{-1}P\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}P'\Omega^{-1}v\right\}$$

$$\pi'(\tau\Sigma)^{-1}\pi + v'\Omega^{-1}v - \pi'(\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}(\tau\Sigma)^{-1}\pi - 2\pi'(\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}P'\Omega^{-1}v - v'\Omega^{-1}P\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}P'\Omega^{-1}v$$

$$v'\left\{\Omega^{-1} - \Omega^{-1}P\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}P'\Omega^{-1}\right\}v + \pi'\left\{(\tau\Sigma)^{-1} - (\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}(\tau\Sigma)^{-1}\right\}\pi - 2\pi'(\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}P'\Omega^{-1}v$$

There is a matrix identity that says

$$\left(A - BD^{-1}C\right)^{-1} = A^{-1} - A^{-1}B\left(CA^{-1}B - D\right)CA^{-1}$$

So we can rewrite the first term in curly brackets

$$v'\left\{\Omega + P(\tau\Sigma)P'\right\}^{-1}v + \pi'\left\{(\tau\Sigma)^{-1} - (\tau\Sigma)^{-1}\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}(\tau\Sigma)^{-1}\right\}\pi - 2v'\Omega^{-1}P\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}(\tau\Sigma)^{-1}\pi$$

Now define

$$\tilde{v} = \left[\Omega + P\left(\tau\Sigma\right)P'\right]\Omega^{-1}P\left[\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\pi$$

$$\left[\Omega + P\left(\tau\Sigma\right)P'\right]^{-1}\tilde{v} = \Omega^{-1}P\left[\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\pi$$

so we get

$$v'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}v + \pi'\left\{\left(\tau\Sigma\right)^{-1} - \left(\tau\Sigma\right)^{-1}\left[\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\right\}\pi - 2v'\left[\Omega + P\left(\tau\Sigma\right)P'\right]^{-1}\tilde{v}$$

Add and subtract terms to make things neater

$$v'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}v+\pi'\left\{\left(\tau\Sigma\right)^{-1} - \left(\tau\Sigma\right)^{-1}\left[\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\right\}\pi-2v'\left[\Omega + P\left(\tau\Sigma\right)P'\right]^{-1}\tilde{v}+\tilde{v}'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}\tilde{v}-\tilde{v}'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}\tilde{v}$$

$$v'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}v-2v'\left[\Omega + P\left(\tau\Sigma\right)P'\right]^{-1}\tilde{v}+\tilde{v}'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}\tilde{v}+\pi'\left\{\left(\tau\Sigma\right)^{-1} - \left(\tau\Sigma\right)^{-1}\left[\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\right\}\pi-\tilde{v}'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}\tilde{v}$$

$$\left(v-\tilde{v}\right)'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}\left(v-\tilde{v}\right)+\pi'\left\{\left(\tau\Sigma\right)^{-1} - \left(\tau\Sigma\right)^{-1}\left[\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\right\}\pi-\tilde{v}'\left\{\Omega + P\left(\tau\Sigma\right)P'\right\}^{-1}\tilde{v}$$

We can now substitute all back to get

$$f_{\mu,V}\left(\mu,V\right) = \left|\tau\Sigma\right|^{-\frac{1}{2}}\left|\Omega\right|^{-\frac{1}{2}}e^{-\frac{1}{2}\left\{\left(\mu-\mu_{BL}\right)'\left[\left(\tau\Sigma\right)^{-1}+P'\Omega^{-1}P\right]\left(\mu-\mu_{BL}\right)+\left(v-\tilde{v}\right)'\left\{\Omega+P\left(\tau\Sigma\right)P'\right\}^{-1}\left(v-\tilde{v}\right)+\pi'\left\{\left(\tau\Sigma\right)^{-1}-\left(\tau\Sigma\right)^{-1}\left[\left(\tau\Sigma\right)^{-1}+P'\Omega^{-1}P\right]^{-1}\left(\tau\Sigma\right)^{-1}\right\}\pi-\tilde{v}'\left\{\Omega+P\left(\tau\Sigma\right)P'\right\}^{-1}\tilde{v}\right\}}$$

We have matrix identities that says $|AB| = |A||B|$ and $|I_j + AB| = |I_k + BA|$, so we can write:

$$\begin{aligned}
\left|\tau\Sigma\right|\left|\Omega\right|\left|\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right| &= \left|\Omega\right|\left|\tau\Sigma\left(\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right)\right| \\
&= \left|\Omega\right|\left|I + \tau\Sigma P'\Omega^{-1}P\right| \\
&= \left|\Omega\right|\left|I + \Omega^{-1}P\tau\Sigma P'\right| \\
&= \left|\Omega\left(I + \Omega^{-1}P\tau\Sigma P'\right)\right| \\
&= \left|\Omega + P\tau\Sigma P'\right|
\end{aligned}$$

So we have

$$\left|\tau\Sigma\right|\left|\Omega\right| = \frac{\left|\Omega + P\tau\Sigma P'\right|}{\left|\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right|}$$

$$\left|\tau\Sigma\right|^{-\frac{1}{2}}\left|\Omega\right|^{-\frac{1}{2}} = \frac{\left|\Omega + P\tau\Sigma P'\right|^{-\frac{1}{2}}}{\left|\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right|^{-\frac{1}{2}}}$$

$$\left|\tau\Sigma\right|^{-\frac{1}{2}}\left|\Omega\right|^{-\frac{1}{2}} = \left|\Omega + P\tau\Sigma P'\right|^{-\frac{1}{2}}\left|\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right|^{\frac{1}{2}}$$

Substituting back, we have

$$f_{\mu,V}\left(\mu,V\right) = \left(2\pi\right)^{-\frac{K+N}{2}}\left|\Omega + P\tau\Sigma P'\right|^{-\frac{1}{2}}\left|\left(\tau\Sigma\right)^{-1} + P'\Omega^{-1}P\right|^{\frac{1}{2}}e^{-\frac{1}{2}\left\{\left(\mu-\mu_{BL}\right)'\left[\left(\tau\Sigma\right)^{-1}+P'\Omega^{-1}P\right]\left(\mu-\mu_{BL}\right)+\left(v-\tilde{v}\right)'\left\{\Omega+P\left(\tau\Sigma\right)P'\right\}^{-1}\left(v-\tilde{v}\right)+\pi'\left\{\left(\tau\Sigma\right)^{-1}-\left(\tau\Sigma\right)^{-1}\left[\left(\tau\Sigma\right)^{-1}\right.\right.\right.}$$

We can simplify this expression to build some intuition. Since μ_{BL} and $\tilde{v}$ do not depend on either μ or v, we

can drop constants and values that have no μ or v:

$$f_{\mu,V}(\mu,V) \propto |\Omega + P\tau\Sigma P'|^{-\frac{1}{2}} \left|(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right|^{\frac{1}{2}} e^{-\frac{1}{2}\left\{(\mu-\mu_{BL})'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right](\mu-\mu_{BL})+(v-\tilde{v})'\left\{\Omega+P(\tau\Sigma)P'\right\}^{-1}(v-\tilde{v})\right\}}$$

$$f_{\mu,V}(\mu,V) \propto \left|(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right|^{\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_{BL})'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right](\mu-\mu_{BL})} \times |\Omega + P\tau\Sigma P'|^{-\frac{1}{2}} e^{-\frac{1}{2}(v-\tilde{v})'\left\{\Omega+P(\tau\Sigma)P'\right\}^{-1}(v-\tilde{v})}$$

$$f_{\mu,V}(\mu,V) \propto f_{\mu|V}(\mu)\, g(v)$$

We just showed that the joint distribution of μ and V is proportional to the conditional distribution times a function of v. So we now have

$$f_{\mu|V}(\mu) \propto \left|(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right|^{\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_{BL})'\left[(\tau\Sigma)^{-1}+P'\Omega^{-1}P\right](\mu-\mu_{BL})}$$

So we get

$$\mu|V \sim N(\mu_{BL}, \Sigma^{\mu}_{BL})$$

$$\mu_{BL} = \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\pi + P'\Omega^{-1}v\right]$$

$$\Sigma^{\mu}_{BL} = \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}$$

Now we have the distribution of μ given the views v. We are only missing the distribution of the returns X given the views. We can rewrite (3) as:

$$X = \mu + Z \qquad Z \sim N(0, \Sigma)$$

Therefore, since both terms are normally distributed, the posterior for the returns is

$$X|V, \Omega \sim N(\mu_{BL}, \Sigma_{BL})$$

where

$$\Sigma_{BL} = \Sigma + \Sigma^{\mu}_{BL}$$

7.2 Testing Our Code

To test if our code is working properly we generate a simple example. Table 4 show the characteristics of the assets. All these values were made up to build an intuitive example with a reasonable visualization of what the Black-Litterman class is doing. Assets A and B have a correlation of 0.5, while asset C has no correlation with any asset. For all the tests below, we are using a risk-free return of 0.75%, an average risk aversion of 1.2 and there is no shrinkage of the expected returns towards the historical. The value of τ is set to 0.002, which is equivalent to having 500 observations when estimating the covariance matrix.

The first view that we are going to test is simple, we just state that the expected return of Asset A is 10%. So our view matrices are given by

$$P = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \qquad v = \begin{pmatrix} 0.1 \end{pmatrix}$$

Table 4: Inputs to the test

(a) Charateristics of the Assets

Assets	**Volatility**	**Equilibrium Weights**	**Historical μ**
A	0.10	0.333	0
B	0.12	0.333	0
C	0.15	0.333	0

(b) Correlation Matrix

	Asset A	**Asset B**	**Asset C**
Asset A	1	0.5	0
Asset B	0.5	1	0
Asset C	0	0	1

Table 5: Comparison of Weights on the Risky Portfolio

	Without Views	$E(r_A) = 0.1$	$E(r_A) > E(r_B) > E(r_C)$
Asset A	0.1797	0.2816	0.3879
Asset B	0.4331	0.3793	0.3009
Asset C	03870	0.3390	0.3110

Figure 11 plots the results for the original portfolio without views (in red) and the view-adjusted portfolio (in green). Since we gave a higher expected return for asset A, it is natural that its expected return increased. But since asset B is correlated with asset A, asset B also inherits from the view and gets higher expected return as well. And since asset C is not correlated with any other asset, its expected return does not change. The actual weights of each asset in the portfolio are in table 5, in the second column. There we can seet that with this view, the wieght of asset increases from 18% to 28%, while the weight of the other assets decrease, which is also expected. As asset A's expected sharpe ratio increases more than of asset B's, the portfolio overweights asset A.

Now we test a relative view that is a bit more strict. Given our original inputs, the equilibrium expected returns of the assets as such that $E(r_A) < E(r_B) < E(r_C)$. Lets us test whats happens when we have views that go against this equilibrium. We will now input 3 views, $E(r_A) = E(r_B) + 0.01$, $E(r_B) = E(r_C) + 0.01$ and $E(r_A) = E(r_C) + 0.01$. These views say that each asset is expected to outperform other assets of higher risk by 1%. So our view matrices are given by

$$P = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix} \qquad v = \begin{pmatrix} 0.01 \\ 0.01 \\ 0.01 \end{pmatrix}$$

Figure 12 plots the results for the original portfolio without views (in red) and the view-adjusted portfolio (in green). Since we gave views that are trying to "invert" the order order of expected returns, we see that the expected return of asset A increases and the expected return from assets B and C decreases. The actual weights of each asset in the portfolio are in table 5, in the third column. There we can seet that with this view makes the portfolio more diversified, closer to the equal weighted portfolio, since the views made the assets have expected sharpe ratios that are closer to each other.

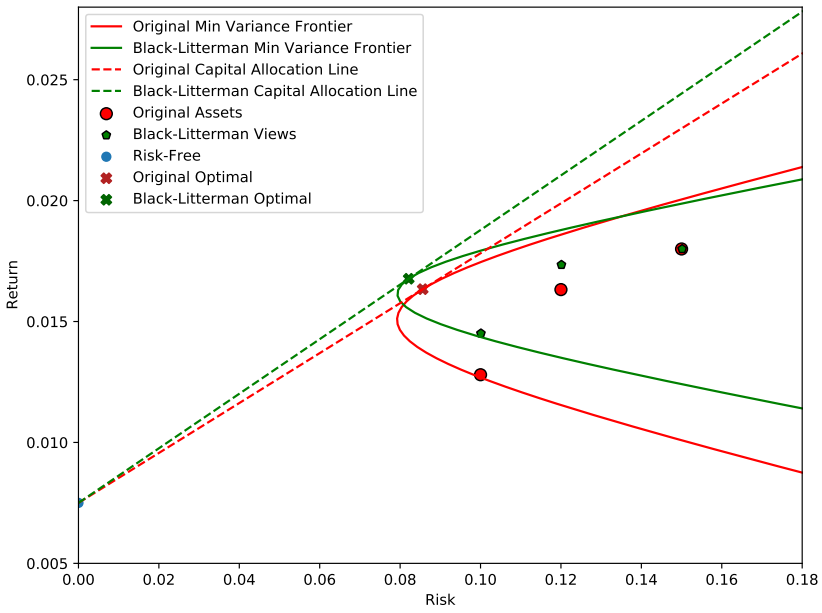

Figure 11: Minimal variance frontier, assets and optimal portfolios using the traditional model (red) and the Black-Litterman model (green) with the view the $E(r_A) = 0.1$. For this test, the value the view confidence is $c = 10$.

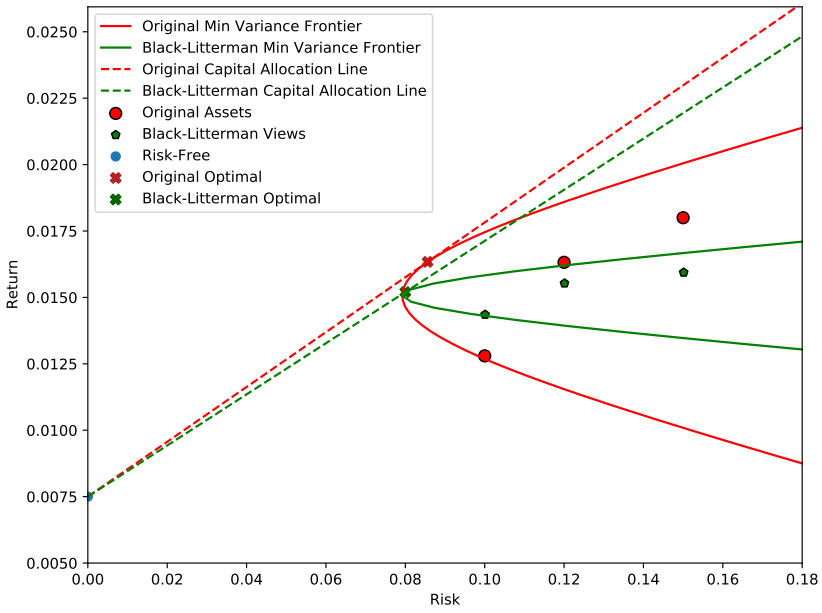

Figure 12: Minimal variance frontier, assets and optimal portfolios using the traditional model (red) and the Black-Litterman model (green) with the views $E(r_A) = E(r_B) + 0.01$, $E(r_B) = E(r_C) + 0.01$ and $E(r_A) = E(r_C) + 0.01$. For this test, the value the view confidence is $c = 100$, this is much higher just so that the chart makes it easier to visualize the effects of views.

8 Using Signals to Generate Views

If we want to run a backtest of a strategy, we need a systematic way of generating views. For this we will use the literature of factor investing. Each currency in our portfolio has carry, momentum and value measures. We will briefly explain how each of them are computed and a brief summary of the literature around predicting returns using these factors. This will give us an insight on how to convert carry, momentum and value into expected returns.

It is worht noting that we are **not** going to build a factor portfolio. We have a set of equilibrium expected returns that are going to be tilted based on the factor characteristics of each currency and not combine the currencies in a way to maximize exposure to a factor.

8.1 Carry

The seminal paper on carry is [9], where the authors define carry for all asset classes. There are also studies that focus specifically on carry for the FX markets like [13]. These authors definition of carry is the return that its future contract would get assuming that prices stay the same. In the case of currency futures, the no-arbitrage future price is given by:

$$F_{t,T} = S_t \frac{\left(1 + r_f\right)^T}{\left(1 + r_f^\star\right)^T}$$

where $F_{t,T}$ is the future price of the currency with maturity $T > t$, S_t is the spot excenhage rate at date t, r_f is the local risk-free rate and $r_f^\star$ is the foreign risk-free interest rate. The return from t to T of the of buying $F_{t,T}$ is

$$r_{t,T} = \frac{S_T}{F_{t,T}} - 1$$

Using the author's definition of carry, we assume that the current spot price S_0 does not change. So carry for a currency is defined as:

$$Carry_t = \frac{S_t}{F_{t,T}} - 1 \tag{9}$$

Notice that this measure is model-free and only depends on market prices that are currently observed.

Just to bring some intuition we can expand the formula above to get:

$$\begin{aligned} Carry_t &= \frac{S_t}{F_{t,T}} - 1 \\ &= \frac{S_t}{S_t \frac{\left(1+r_f\right)^T}{\left(1+r_f^\star\right)^T}} - 1 \\ &= \frac{\left(1 + r_f^\star\right)^T}{\left(1 + r_f\right)^T} - 1 \\ &\approx T\left(r_f^\star - r_f\right) \end{aligned}$$

So we can see that this measure of carry is the origin of the classical currency carry trade, which is based on the nominal interest rate differential.

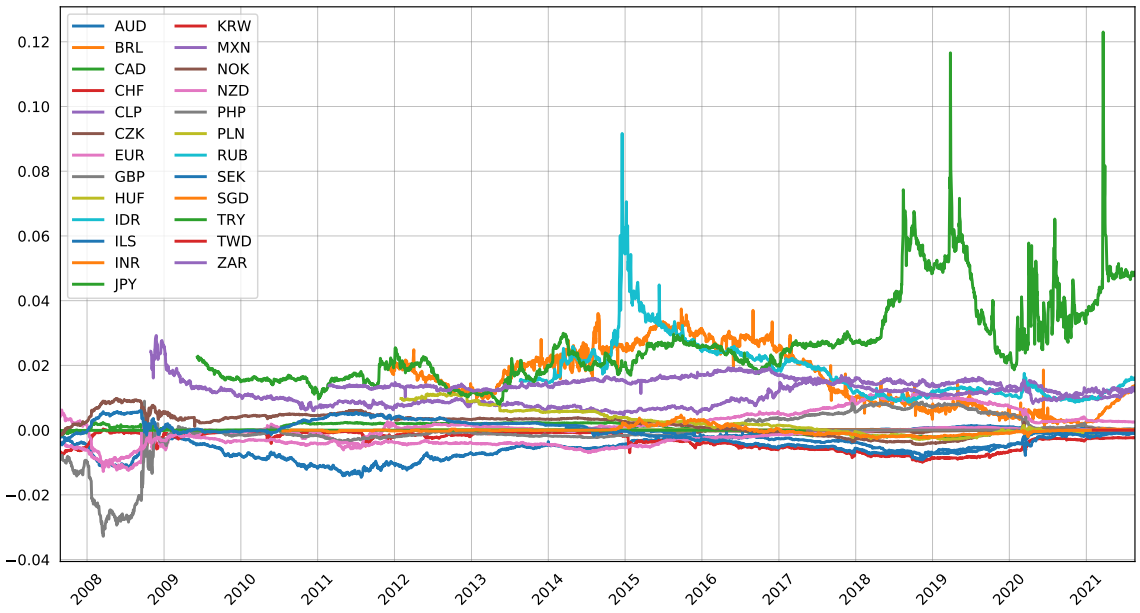

Figure 13: 3-month carry for all currencies

8.2 Value

The value measure is realted to a concept of fair price. The economic literature has several measures for the fair value of a currency. Fundametal equilibrium exchange rates (FEER), behavioral equilibrium exchange rates (BEER) and natural real exchange rate (NATREX) are commom models in the literature, but these rely on several hypothesis, non-observable variables, which leads to need for proxies, and their results are sensitive to sample size and methods of estimation. To remove this model uncertainty from our value measure, we decided to use a model-free measure: the **purchase power parity** (PPP). This is a well-known arbitrage condition which suggests that two currencies will be in equilibrium when similar goods are priced the same in the two countries, when converted to the same currency. If that won't be the case then we will observe a demand switch from the expensive good to the cheaper one. A corollary of this theory states that exchange rates adjust to reflect the difference between inflation rates among countries. For example, countries with higher inflation must depreciate their exchange rate. Empirical studies show that the PPP only holds on the long term which makes it a good "anchor" for market participants. OECD computes the PPP exchange rates with amazing detail, although only once per year. But since this is the most detailed and hypothesis-free computation, this will be our measure of choice. The details on how these PPP rates are computed (not estimated, because there are no models) are layed out in [5]. The authors [7] use a different model free measure for exchange rates, but it is inspired by the PPP and its computation does not go into as much detail and precision as the one from OECD.

So our chosen measure of value is the negative of the percent change from the spot rate towards the PPP_t :

$$\begin{aligned} Value_t &= -\left(\frac{PPP_t}{S_t} - 1\right) \\ &= 1 - \frac{PPP_t}{S_t} \end{aligned}$$

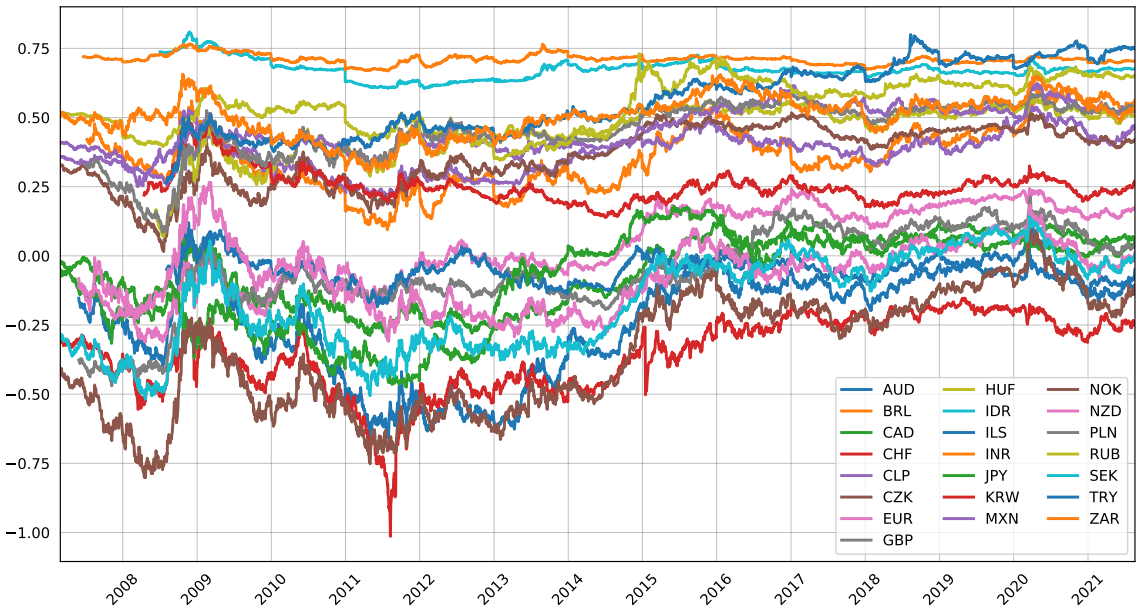

Figure 14: Value measure for all currencies based on the PPP

So a positive value measure means that the currency is undervalued and has room for appreciation. This value measure for each currency is shown in figure 14. This figure brings up another issue, which is the speed of reversion from the deviations from the PPP. Empirical studies like [11] have shown that the half-life of this reversion is typically between 3.2 to 4.4 years. So we cannot use the value measure directly as an expected return measure, we have to convert the undervaluation/overvaluation of the currency to expected return by adjusting for the size of the expected reversion. To keep horizons consistent, we will work with a 3-month expected return from the value measure. If the half-life (50% of the reversion) of PPP deviations has an average of 3.8 years (45.6 months), then the expected reversion in 3 months would be 4.45%. This means that we should multiply the value signal by 0.0445 in order to convert it to expected return over the next 3 months. This allows the magnitude of the expected returns generated by the value measure to be in line with those generated by the carry measure.

8.3 Momentum

For carry and value, we get measures the directly translate in expected returns, but converting momentum measures into expected retruns is not straightforward. For this measure, it is worth pointing out some issues with nomenclature from the literature. Seminal papers like [7] refer to "momentum" as the performance of an asset relative to other assets, which is sometimes called "cross-sectional momentum". Other important papers like [4] refer to "momentum" as the performance of an asset relative to itself, which is sometimes called "time-series momentum" or "trend following". In both cases, converting the momentum measure into expected return is not straightforward, but our interest here is to use the idea of "trend following". As [4] shows, there is momentum in performance in all asset classes. They find persistence in returns for one to 12-months that partially reverses over longer horizons, consistent with sentiment theories of initial under-reaction and delayed over-reaction. Specifically for currencies, their work points to an optimal lookback period of around 3 months. But that still does not solve the problem of how to convert the return over the

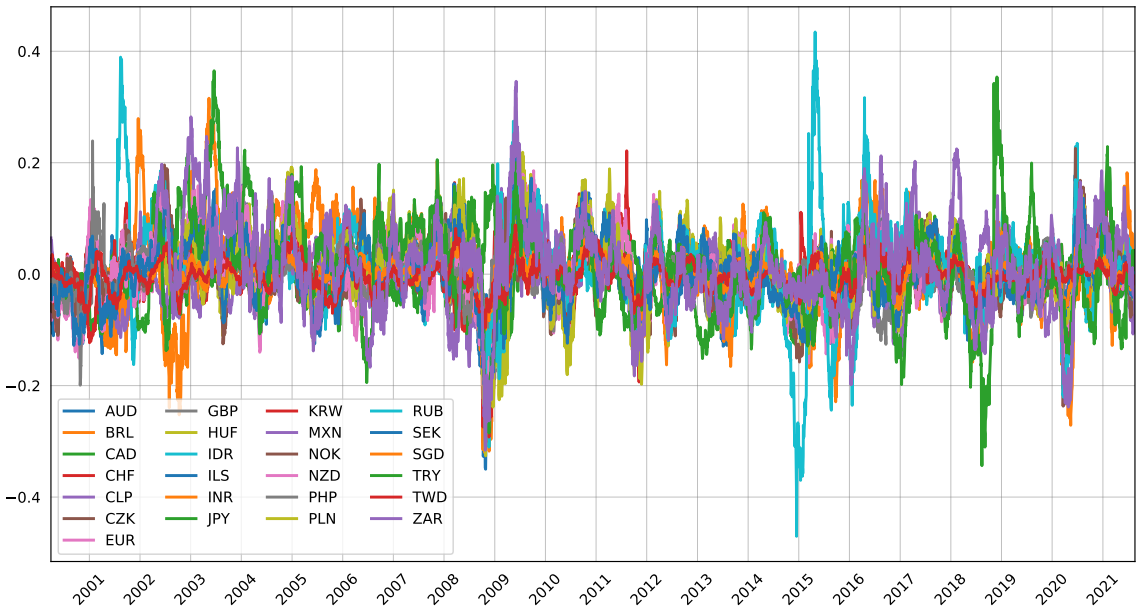

Figure 15: Momentum measure for all currencies based on the returns of the last 3 months.

last 3 months into a measure of expected returns.

The way we decided to include this signal in the portfolio construction process is to include the past 3-month return in the mistorical μ input of the Black-Litterman model. The shrinkage parameter of the optimization will be responsible for balancing the views on expected returns generated by carry and value with the momentum measure, which are the past returns. This shirinkage parameter will be set during the backtesting process. In other words, the historical returns will be in equation (5) as $\hat{\mu}$ and the hyperparameter s will be set on the backtesting stage.

And just for consistency, figure 15 shows the rolling 3-month returns for all currencies.

Part IV

Portfolio Construction

As required by the CQF, we implemented 2 portfolios optimizations. The first one is the traditional Markowitz / Max Sharpe portfolio, where the covariance matrix and expected returns used are the posteriors from the Black-Litterman model, and the second one is Lopez de Prado's Hierarchical Risk Parity (HRP) with a detoned covariance matrix where the first factor affecting all currencies (imagined to be the USD dollar) is removed before computing the weights. Both of them will be briefly explained.

9 Markowitz / Max Sharpe

Our objective is to make the best investment of our wealth for a period of T years. The decision of the final portfolio can be separated in steps, finding the optimall risky portfolio (Maximum sharpe index) and combine this risky portfolio with the risk-free asset, which depends on the risk aversion of the investor. Our investment universe is based on currency futures contracts so short-selling is allowed.

The first part of the problema is to find the optimal risky portfolio by solving the following problem:

$$\max_{w} \frac{Er\left(w\right) - r_f}{\sigma\left(w\right)}$$
$$w'\iota = 1$$

where

$$Er\left(w\right) = w'\mu \qquad \sigma\left(w\right) = \sqrt{w'\Sigma w}$$

This gives us the optimal weights $w^\star$ of the risky assets on the risky portfolio. In our case, μ and Σ will be the posteriors from the Black-Litterman model.

The investor's problem is to combine the optimal risky portfolio with the risk free asset in order to maximize its utility. So the problem is

$$\max_{w_I} Er_I\left(w_I\right) - \frac{\lambda}{2}\sigma_I^2\left(w_I\right)$$

where

$$Er_I\left(w_I\right) = w_I Er\left(w^\star\right) + \left(1 - w_I\right) r_f \qquad \sigma_I^2\left(w_I\right) = w_I^2\sigma^2\left(w^\star\right)$$

For the purpose of this project, we will use a risk-aversion coeficient of $\lambda = 1.2$ whenever needed, since we are more interested on how the risky portfolio changes.

We have generated a detailed code for a class on the markowitz optimization, which already has several functionalities (like making the classical charts) built into it. This class was already tested in the numerical example of section 7.2.

10 Detoned Hierarcical Risk Parity

The details of detoning were already explained in section 6.3 but it worth remembering that the covariance matrix that comes out of this method does not have full rank, so we cannot use in application that require

this matrix to be inverted and even numerical optimizations, because they are not guaranteed to converge. But the author of this detoning method points out that the matrix can still be used for clustering applications including his own Hierarchical Risk Parity (HRP). This method only takes the covariance matrix as an input, it does not rely on expected returns, but we are still going to use Black-Litterman's posterior covariance in this application. On the HRP side we still need to explain how the method works. The details of the method are outlined in [10] and coded in our project, but we will go through a brief explanation for the intuition.

It is well know that the Markowitz optimization has stability issues. One reason for the instability of quadratic optimization problems is that it assumes that all assets are possible substitutes for each other, or in a more mathematical way, the assets form a fully connected graph. Some investments are similar to each other, so it makes more sense to allow them to substitute each other. For investments that are less similar between themselves it makes sense to complement each other, not substitute. In other words, correlation matrices lack the notion of hierarchy. For these reasons, hierarchical structures are better designed to give not only stable but also intuitive results. Intuitively, what HRP does is a block-shrinkage of the covariance matrix.

This Hierarchical Risk Parity (HRP) method uses the information contained in the covariance matrix without requiring its inversion or positive-definitiveness. HRP can even compute a portfolio based on a singular covariance matrix, which is exactly what we are going to do after detoning it. The algorithm operates in three stages: tree clustering, quasi-diagonalization, and recursive bisection.

1. **Tree Clustering**: This stage uses the correlation matrix to combine the assets into a hierarchical structure of clusters, so that allocations can flow downstream through a tree graph. All of the computations in this stage do not require the correlation matrix to have full rank.

2. **Quasi-diagonalization**: This stage reorganizes the rows and columns of the correlation matrix, so that the largest values lie as close as possible along the diagonal, which renders a useful property that similar investments are placed together, and dissimilar investments are placed far apart. It even generates a nice visualization, which we already used in figure 4.

3. **Recursive Bisection**: Stage 2 has delivered a quasi-diagonal matrix and since the inverse-variance allocation is optimal for a diagonal covariance matrix, we can take advantage of these facts in two different ways: (a) bottom-up, to define the variance of a contiguous subset as the variance of an inverse-variance allocation; or (b) top-down, to split allocations between adjacent subsets in inverse proportion to their aggregated variances.

Part V

The Static Portfolio

In this section we are going to ilustrate how our portfolio works by building it for a **single date**. We chose August 13th 2021, as it is the most recent date that we have all of the data available. The code reads all the data, total return indexes, forwards prices, spot rates, PPP rates and the US 3-month libor. We compute the exponentially weighted covariance and correlation matrix using a center of mass of 3 months. Then the signals are computed, and their cross section for this date is show in figure 16.

We compute the weights for the portfolio using 5 different methods, 2 that are required for this project and 3 to use as benchmarks. We comment them below and plot them in figure 17.

1. **Equal Weighted**: Since we have 25 currencies in our investmet universe, their weights all equal $\frac{1}{25} = 4\%$.

2. **Inverse Volatility**: Very volatile currencies are underweighted. We can see that the BRL, TRY and ZAR are the currencies that get the smaller weights, while asian currencies like INR, SGD and TWD got the biggest share.

3. **Hierarchical Risk Parity**: As expected, this method generates allocations that are similar to the inverse volatility but with more intensity. Currencies with small weights in the inverse vol method end up with even smaller weights in HRP and the ones with bigger weights in inverse volatility get even bigger weights in the HRP method.

4. **Detoned Hierarchical Risk Parity**: An interesting thing came out of this allocation, since the correlation matrix is detoned (the first principal component is removed). We can interpret the first principal component as the main driver/factor behind the moves in all these currencies, and since they all trade against the USD, the first principal component is likely correlated with the strengh of the USD. When this factor is removed, the currencies whose allocation increased were the EUR, CZK, GBP, NOK and SEK. Although HUF's allocation decreased slightly, there seems to be concentration towards european currencies when this "dollar factor" is removed.

5. **Marchenko-pastur Denoising + Black-Litterman with signal views**: Even though currencies like BRL and MXN have postive positive carry, value and momentum, they have been allocated for a short position. Their high volatility likely plays a big role in this. And the currencies that ended up with the highest share of allocation, like SGD, JPY and TWD, are the ones with the lowest volatility, even though their signals are all close to zero or slightly negative. This behaviour is likely due to the markowitz optimization behind this method, which we know favors assets with low volatility and low correlations. This method does not seems to be too sensitive the signals that we input[6], as the weights are heavily dependent on the volatility and correlations of the currencies. This portfolio construction method, where signal are inputed in the Black-Litterman model as views on returns, does not seem to be best way to capture this type of premium. As the literature, factor portfolios might be better in isolating this type of premium. And after these factor portfolios are built, then the Black-litterman model can be used to better combining these individual strategies, instead of combing all the different signals into one strategy.

[6] We tested several levels of the confidence parameter c in the Black-Litterman model, but they all yield similar allocations.

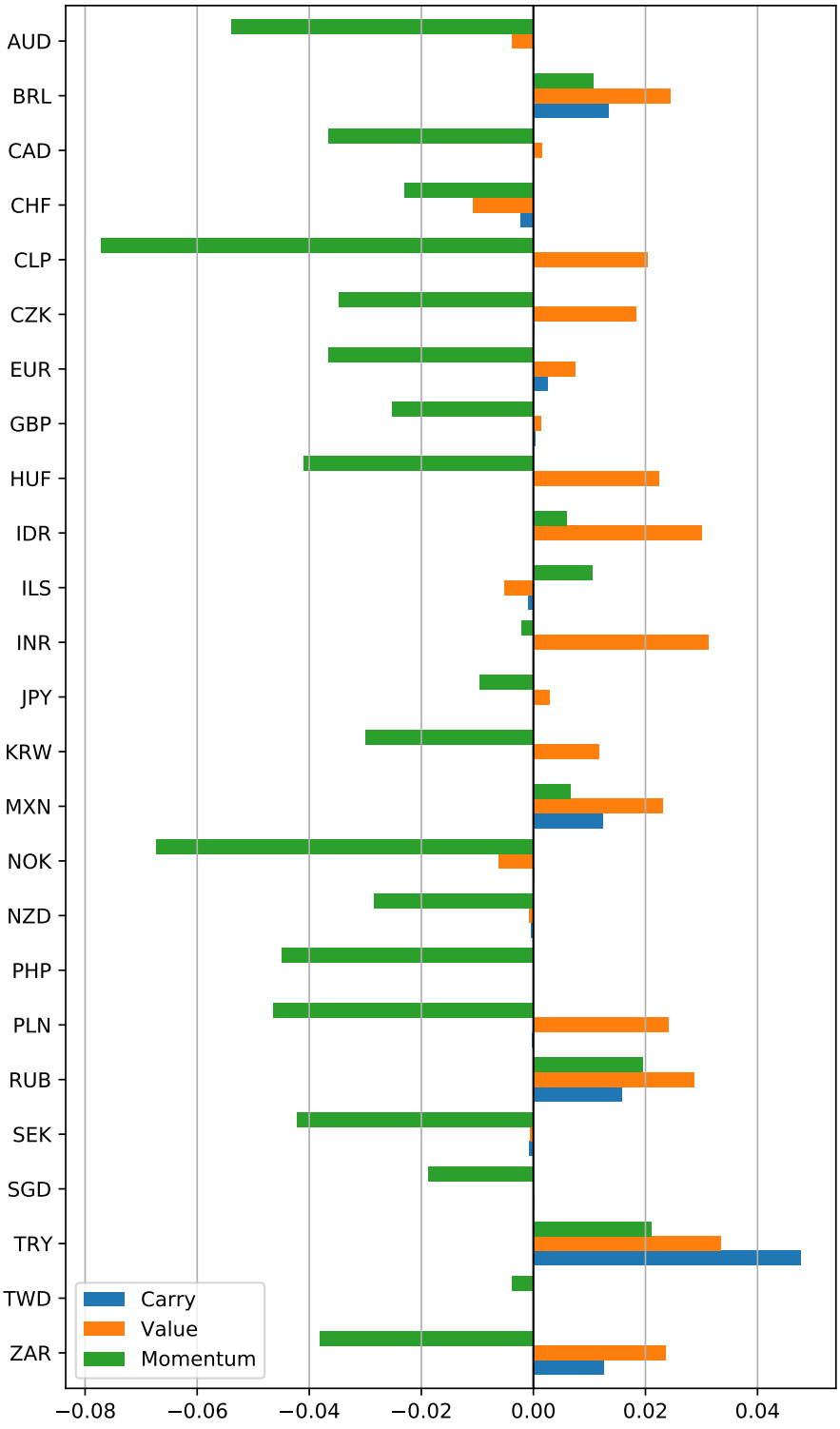

Figure 16: Cross-section of the carry, value and momentum signals for all curencies in August the 13th, 2021

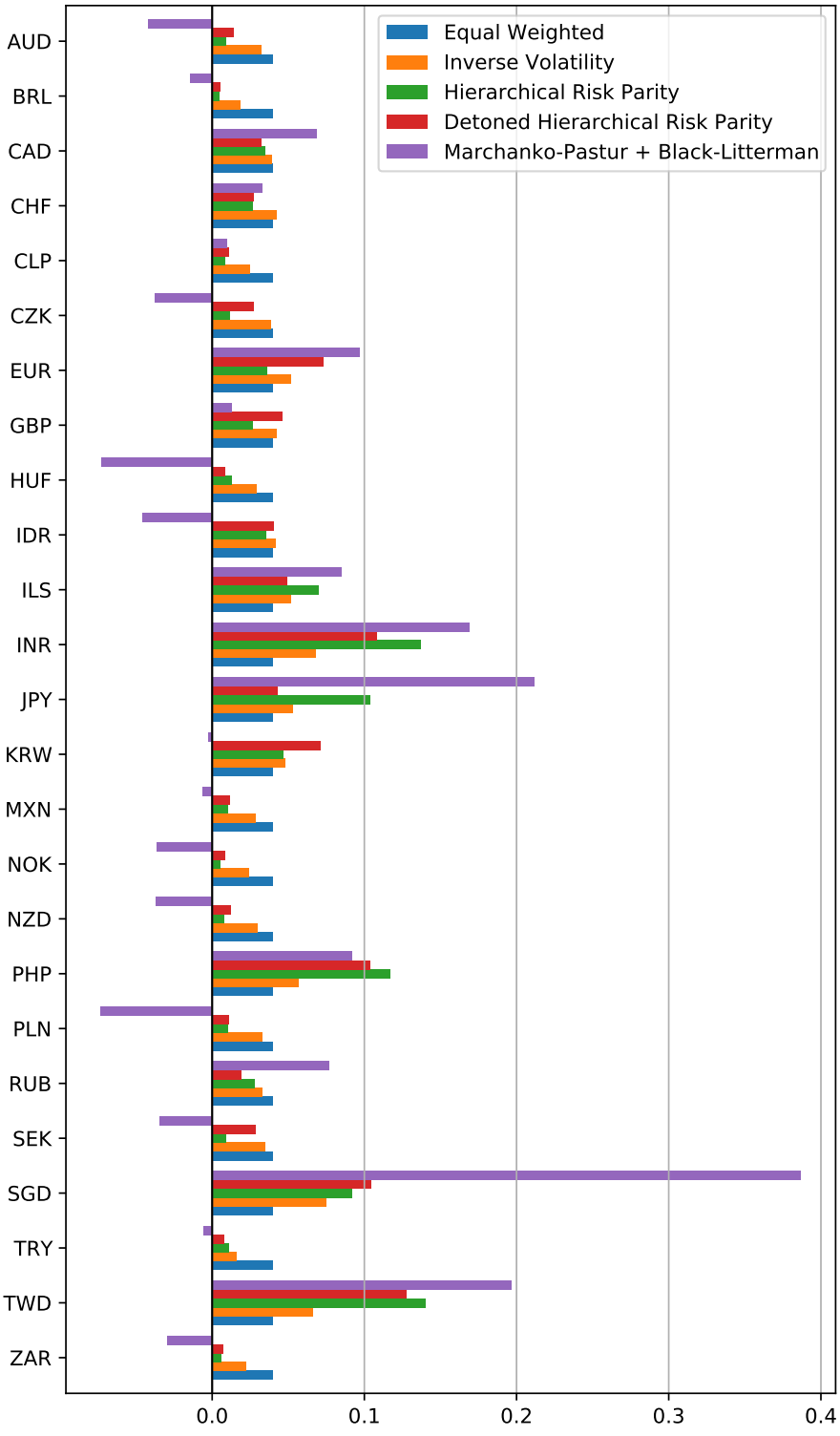

Figure 17: Portfolio weights using 5 different methods for August 13th, 2021

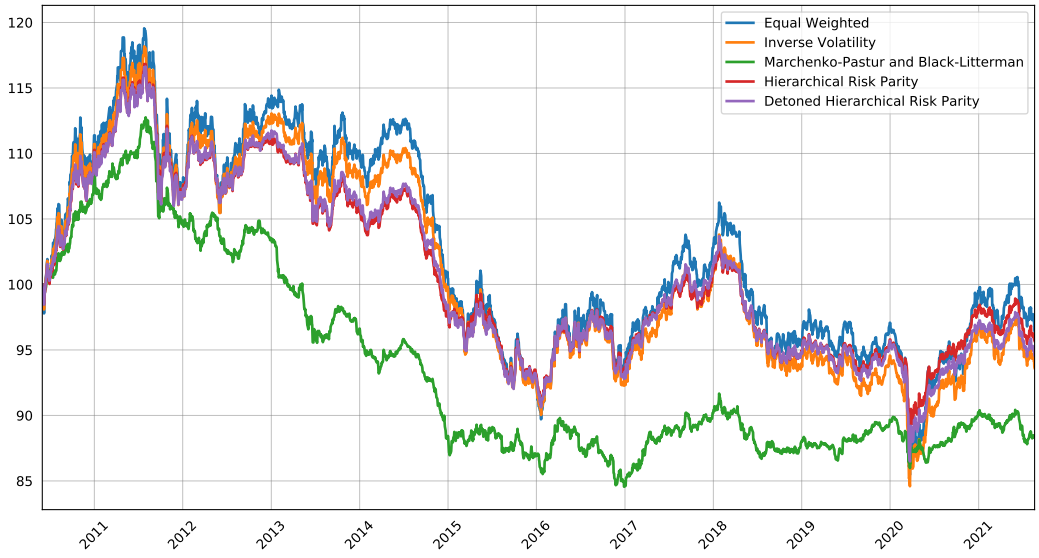

Figure 18: Total return indexes for the 5 portflio construction methods

Table 6: Performance measures for the 5 portfolio construction methods

	Equal Weighted	**Inverse Volatility**	**MP+BL**	**HRP**	**Detoned HRP**
Annualized Return	-0.31%	-0.57%	-1.06%	-0.38%	-0.48%
Annualized Volatility	6.33%	5.42%	3.08%	4.28%	4.43%
Sharpe Ratio	-0.05	-0.10	-0.34	-0.09	-0.11
Skewness	-0.20	-0.26	-0.17	-0.34	-0.37
Kurtosis	3.60	4.17	3.36	4.38	5.12

Part VI

Dynamic Portfolio (Backtesting)

The static portfolio does not seem to be the best way to evaluate a strategy. We need to implement it and backtest it in order to have an idea of their past performance. The backtest uses exactly the same procedure as the static portfolio. We chose to rebalance the portfolio once a month, meaning that every month the alloctions are re-computed and those weights are held for another month. We should not update the weights every day as this increases the trading costs of the strategies.

We can see in figure 18 that the overall performance (equal weighted) of currencies was negative in this period, relative to the dollar. All of the construction methods that rely only on the covariance matrix showed similar performance, while the target of this project, the Marchanko-Pastur denoised covariance coupled with views generated by the Black-Litterman model, had a worst performance, specially during 2011 to 2015. After that, the strategy has been as good as the other ones, but with a much lower volatility.

To evaluate this strategies in a more quantitative way we can compute performance measures. Table 6 shows annualized return, annualized volatility, Sharpe ratio, skewness and kurtosis of returns.

To visualize how the weights are changing, figure 19 plots the evolution of the weights for the BRL. Here it is easy to see that the weights for inverse volatility, HRP and detoned HRP are very similar, hence their similar performance. It is only the Marchenko-Pastur + Black-Litterman that has big changes in the weights,

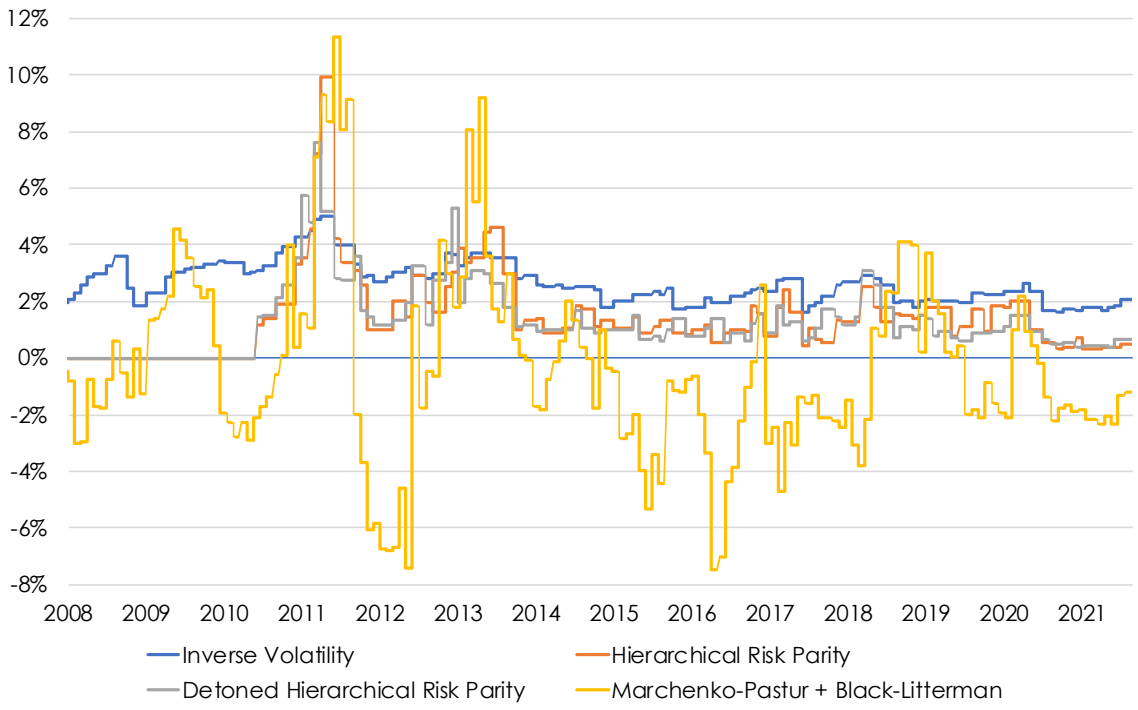

Figure 19: Evolution of the BRL weights for different portfolio construction methods

which is expected. But it also noting that the weights are not changing very fast, meaning that this strategy is probably not very costly in terms of rebalancing.

Of course, there are ways we could still make the backtest more detailed, but for the purpose of this project we covered enough to show that the Marchenko-Pastur + Black-Litterman model is probably not the best method to capture premiums like carry, momentum and value. We could include rebalance and rolling costs, but given how the weights change in the Marchenko-Pastur + Black-Litterman compared to the other ones, we know this is the strategy that will have the highest cost drag. We could also run some hyperparameter tuning in order to find better combinations of parameters like confidence of views and intensity of shrinkage[7]. Another thing that could help is run the strategies for EMs and DMs in parallel, and combine them later. This is typically done for off-the-shelf alternative risk premia (ARP) strategies which usually yields good results. One last thing, which in my opinion is likely to framatically improve the overall performance, is to build separate factor portfolios for each signal and for EMs and DMs. After these are build, then we can use Marchenko-Pastur + Black-Litterman model. This will reduce the number of assets in the optimization from 25 to 6, and each of them are already optimized strategies.

[7]We tested some combinations for these parameters and they all yield similar results. Again, Marchenko-Pastur + Black-Litterman is not the best way to capture carry, momentum and value signals.

References

[1] "*Asset Allocation: combining investros views with market equilibrium*", Goldman Sachs Fixed Income Research (1990)

[2] "*The Elements of Statistical Learning*", Hastie, Tibshirani & Friedman, 2nd Edition (2009)

[3] "The Black-Litterman Approach: Original Model and Extensions", Attilio Meucci, SSRN (2010)

[4] "*Time series momentum*", Moskowitz, Ooi & Pedersen, Journal of Financial Economics Volume 104, Issue 2 (2012)

[5] "Eurostat-OECD Methodological Manual on Purchasing Power Parities", Eurostat Methodologies and Working Papers (2012)

[6] "*An Introduction to Statistical Learning*", James, Hastie, Witten, Tibshirani (2013)

[7] "Value and Momentum Everywhere", The Journal of Finance, Asness, Moskowitz and Pedersen (2013)

[8] "*The Black-Litterman Model In Detail*", Jay Walters, SSRN (2014)

[9] "Carry", Koijen, Moskowitz, Pedersen and Vrugt (2016)

[10] "Advances in Financial Machine Learning", Wiley, Marcos Lopez de Prado (2018)

[11] "Speed of Reversion of Deviations of the Purchasing Power Parity for Brazilian Cities", Bastos, Ferreira and Arruda, Revista Brasileira de Economia (2018)

[12] "Machine Learning: An applied mathematics introduction", Wilmott (2019)

[13] "*Factor Investing in Currency Markets: Does it Make Sense?*" Baku et al, Amundi Asset Management Working Paper (2019)

[14] "*Triennual Central Bank Survey: Foreign exchange turnover in April 2019*". Monetary and Economic Department of the Bank for International Settlements (2019)

[15] "Machine Learning for Asset Managers", Marcos M. Lopez de Prado (2020)

A Bloomberg Tickers

Table 7 shows the bloomberg tickers with the data that we used for each currency.

Table 7: Bloomberg Tickers

	Classification	Total Return Index (UBS)	Forward 3m (in bps)	Spot	PPP
AUD	DM	UISFMAUE Index	AUD3M L160 Curncy	USDAUD L160 Curncy	PPP AS Index
BRL	EM	UISFMBRL Index	BCN3M L160 Curncy	USDBRL L160 Curncy	PPP BZ Index
CAD	DM	UISFMCUE Index	CAD3M L160 Curncy	USDCAD L160 Curncy	PPP CA Index
CHF	DM	UISFMFUE Index	CHF3M L160 Curncy	USDCAD L160 Curncy	PPP SZ Index
CLP	EM	UISFMCLP Index	CHN3M L160 Curncy	USDCLP L160 Curncy	PPP CL Index
CZK	EM	UISFMCZK Index	CZK3M L160 Curncy	USDCZK L160 Curncy	PPP CZ Index
EUR	DM	UISFMEUE Index	EUR3M L160 Curncy	USDCAD L160 Curncy	PPP EUAR Index
GBP	DM	UISFMGUE Index	GBP3M L160 Curncy	USDCAD L160 Curncy	PPP UK Index
HUF	EM	UISFMHUF Index	HUF3M L160 Curncy	USDHUF L160 Curncy	PPP HG Index
IDR	EM	UISFMIDR Index	IHN3M L1600 Curncy	USDIDR L160 Curncy	PPP ID Index
ILS	EM	UISFMILS Index	ILS3M L160 Curncy	USDILS L160 Curncy	PPP IL Index
INR	EM	UISFMINR Index	IRN3M L160 Curncy	USDINR L160 Curncy	PPP IN Index
JPY	DM	UISFMJUE Index	JPY3M L160 Curncy	USDCAD L160 Curncy	PPP JN Index
KRW	EM	UISFMKRW Index	KWN3M L160 Curncy	USDKRW L160 Curncy	PPP KO Index
MXN	EM	UISFMMXN Index	MXN3M L160 Curncy	USDMXN L160 Curncy	PPP MX Index
NOK	DM	UISFMNUE Index	NOK3M L160 Curncy	USDCAD L160 Curncy	PPP NO Index
NZD	DM	UISFMKUE Index	NZD3M L160 Curncy	USDCAD L160 Curncy	PPP NZ Index
PHP	EM	UISFMPHP Index	PPN3M L160 Curncy	USDPHP L160 Curncy	-
PLN	EM	UISFMPLN Index	PLN3M L160 Curncy	USDPLN L160 Curncy	PPP PO Index
RUB	EM	UISFMRUB Index	RUB3M L160 Curncy	USDRUB L160 Curncy	PPP RU Index
SEK	DM	UISFMSUE Index	SEK3M L160 Curncy	USDCAD L160 Curncy	PPP SW Index
SGD	EM	UISFMSGD Index	SGD3M L160 Curncy	USDSGD L160 Curncy	-
TRY	EM	UISFMTRY Index	TRY3M L160 Curncy	USDTRY L160 Curncy	PPP TK Index
TWD	EM	UISFMTWD Index	NTN3M L160 Curncy	USDTWD L160 Curncy	-
ZAR	EM	UISFMZAR Index	ZAR3M L160 Curncy	USDZAR L160 Curncy	PPP ZA Index