4455K

MPhil in Economics

Dissertation Title: Lasso Regularisation of Parametric Portfolio Policies.

Deadline Date: 24th of July 2017.

I acknowledge that this is my own work, and I have read and understood WKH Section on plagiarism.

Actual word count: 8082.

Lasso Regularisation of Parametric Portfolio Policies

4455K
Faculty of Economics
University of Cambridge

July 24, 2017

Abstract

We apply lasso regularisation to Parametric Portfolio Policies (PPP) model, which reduces the risk of overfitting when estimated with a large number of characteristics. Our model selects the most relevant characteristics in every time period, allowing different characteristics to be relevant at different times. We presented a numerical method for estimation by imposing a constraint on the optimisation problem. We estimate the model for foreign exchange (FX) using economic fundamentals as characteristics. We find that our model outperforms the baseline PPP model, with a lower standard deviation, a higher average return and a higher CE return. The results of model's variable selection and estimation is mostly consistent with the literature and theoretical predictions about the relationships between economic fundamentals and currency movements, and hence provides support for macroeconomic theories of FX. Our findings suggest that there exists predictability for FX movements in macroeconomic fundamentals.

Keywords: Parametric Portfolio Policies, Lasso, Regularisation, Foreign Exchange

Contents

1 Introduction

Economic models of exchange rates have so far been unsuccessful in predicting foreign exchange (FX) movements. These are widely thought of as unconnected from economic fundamentals in what is known as the "exchange rate disconnect puzzle". Meese and Rogoff (1983) claim that models of exchange rates which condition on economic fundamentals do not outperform a naive random walk. This view excludes the possibility of constructing FX portfolios on the basis of economic fundamentals. In recent years, regularisation methods such as the Lasso (Tibshirani 1996) have dramatically improved the performance of "kitchen-sink" predictive regressions which incorporate a large number of factors (see Li, Tsiakas and Wang 2015). The Lasso estimator performs a variable selection and shrinks large coefficients to reduce the risk of overfitting, and hence automatically determines the most relevant economic fundamentals. In this paper, we will explore how the Lasso can be applied to portfolio optimisation.

Coincidently, we have also seen the introduction of Parametric Portfolio Policies (PPP) (Brandt et al. 2007), which model the weight of each asset as a linear function of its characteristics. An optimal portfolio of currencies would therefore produces weights as a function of economic fundamentals. The empirical literature on PPP has been focused on carefully selecting a small set of variables on the basis of economic theory. For instance, Brandt et al (2007) constructs an equity portfolio using the book-to-market and size as characteristics, to reflect the Fama and French (1993) 3-factor model. Although it is accepted that returns are related to many more characteristics, these are omitted to limit the risk overfitting. Indeed, including characteristics which are not relevant can lead the excessively complex models with poor out-of-sample performance. However, this practice comes at the cost of ignoring large numbers of characteristics which contain potentially predictive information. If PPP is to be extended to account for the wide range of variables which determine exchange rates, there is clearly a need for regularisation.

Can Parametric Portfolio Policies be estimated with a large number of characteristic whilst *minimising the risk of overfitting ?* In this paper, we extend the PPP methodology to estimation with a large number of coefficients with a Lasso regularisation. We are consciously adopting an agnostic view of economic theory to discover predictability by allowing different macroeconomic fundamentals to be relevant at different times, in a way that incorporates regime switching automatically. The Lasso performs a variable selection and shrinks regression coefficients, which reduces the risk of overfitting. We evaluate three types of regularisations: the Ridge, Lasso and Elastic Net. The Ridge regression shrinks coefficients estimates towards 0, thereby reducing the risk of overfitting. The Lasso performs both variable selection by setting some coefficient estimates to 0 as well as shrinking the remaining coefficients. The Elastic Net (Zou and Hastie 2005) combines the Ridge and Lasso and has interesting properties for highly correlated variables. We select the optimal level of regularisation through k-fold cross validation estimates of the out-of-sample Sharpe Ratio. We propose a numerical method for estimation by imposing a constraint on the optimisation using a differential evolution algorithm.

We evaluate the proposed model empirically with the three types of regularisation for a basket of 9 currencies, using 52 economic fundamentals as characteristics (many of which are likely to be unrelated to FX movement, as with a "kitchen-sink" regression). We also estimate the tangency portfolio, the equally weighted portfolio and the baseline PPP model (with no regularisation). To evaluate the out-of-sample performance, we perform quarterly rebalancing between 1993 and 2017 with an estimation window of 30 observations. We find that the regularisation of PPP leads to a large increase in average return for every type of regularisation relative to the baseline PPP model, as the regularisation of coefficients reduces overfitting. Whilst the ridge regularisation increases the standard deviation relative to the baseline PPP model, the lasso leads to a reduction in standard deviation, and even comes close the standard deviation of the EW portfolio. All-in-all, the lasso increases the Sharpe ratio to 0.4342 (from 0.2526), whilst decreasing turnover. We analyse the variables which are selected by the Lasso, as well as the size and sign of estimates, and find that they are mostly consistent with the literature and theoretical predictions about the relationships between economic fundamentals and currency movements. Our results suggest clear economic profitability and hence predictability using macroeconomic fundamentals once the correct empirical methods are used to discover it, so support in some way macroeconomic theories of FX rates.

The remainder of this paper is organised as follows. In the next section, we review the previous relevant literature. In section 3, we present our dataset. In section 4, we present the baseline PPP model, then outline our model of regularised PPP and finally discuss transaction costs and weight constraints. In section 5, we detail the numerical method used for estimation, the method for selecting regularisation parameters and present the results. Section 6 outlines the limitations of our approach and potential avenues for future work. We conclude in section 7.

2 Literature Review

Since traditional of exchange fail to predict price movements (see Engel, West and Mark 2007), it has been claimed by many that exchange rate movements are disconnected from economic fundamentals, in what is known as the "exchange rate disconnect" puzzle. This idea negates the possibility of achieving superior returns on currency portfolios by conditioning on economic statistics. Cheung, Chinn, and Pascual (2005) claim that fundamental-based models have no predictive ability, whilst Meese and Rogoff (1983) argue that models of exchange rates that condition on economic fundamentals cannot outperform a naive random walk. The random walk model assumes a constant equilibrium exchange rate, and with time-varying risk premia we would expect time varying predictability, which is not inconsistent with rational behaviour in the markets. We also know that returns have an ARCH structure so the random walk hypothesis is not relevant - however returns might follow a martingale difference.

The unpredictability of exchange rates paradoxically presents a challenge to the efficient market hypothesis (Fama 1970). Since currencies have different interest rates, investors can profit by borrowing at low-yielding currencies to invest in high-yielding ones, in what is known as the carry trade (Fama 1984). Burnside, Eichenbaum, and Rebelo (2008) find that a well-diversified portfolio of carry trades has double the Sharpe Ratio of the U.S. stock market. Abnormal returns should not persist in efficient markets.

An extensive literature has been devoted to providing an economic explanation to this market anomaly. Some have argued that the volatility of high-yielding currencies implies that the carry trade is highly risky. An alternative explanation is the "peso problem" of the strategy: Barro (2006) argues that potential financial collapse that justifies the high returns may have not yet occurred. The momentum strategy has also received considerable interest. Menkhoff, Sarno, Schmeling, and Schrimpf (2012) provide an empirical investigation of momentum strategies in FX markets. They find limits to arbitrage which prevents easily exploitable returns. Burnside, Eichenbaum, and Rebelo (2011) study the combination of momentum and carry trade strategies. Nevertheless, this literature is generally focused on single, isolated strategies, without considering optimal combinations of strategies. Momentum might just be reflecting economic fundamentals, which we use directly in this paper. Our results implicitly suggest predictability in economic terms once the correct empirical methods is used to discover it.

Our methodology is based on the parametric portfolio policies (PPP) approach to portfolio optimisation introduced by Brandt et al. (2007). They model weights in each asset as a linear function of asset characteristics, which are estimated by maximising with respect to an investor's utility over an estimation window. By incorporating the book-to-market, size and momentum factors suggested by Fama and French (1993) and Carhart (1997), they obtain an equity portfolio with an annualised certainty equivalent gain of 11% relative to the market portfolio. There has since appeared an extensive empirical literature, which seeks to adapt and extend the approach to other assets, with a focus on characteristic selection. In particular, Barroso and Santa-Clara (2015) form currency PPP using economics statistics as characteristics. They perform tests on a range of characteristics, and conclude that it is optimal to include the interest rate spread, momentum and value reversal. They obtain a 0.29 gain in Sharpe ratio over the equal weighted portfolio, which shows that PPP can be successfully applied to FX even with a small number of economic indicators.

In recent years, a number of studies have discovered predictability through sophisticated models, which uses high dimensional datasets. Li, Tsiakas and Wang (2015) show that a kitchen-sink predictive regression for exchange rates can lead to reliable forecasts, by conditioning on a wide range of fundamentals and using an elastic net estimator. The elastic net combines the ridge and lasso regularisations, which reduces the influence of less informative predictors. This approach is known as the "efficient kitchen- sink regression", which is shown to deliver superior out-of-sample performance. This illustrates that the key to revealing the predictive information in economic fundamentals is by using a large number of fundamentals, then applying lasso and ridge regularisations to the estimates. Our paper builds on this work by adapting it to the PPP framework, which so far has been estimated with only a small number of fundamentals.

3 Data

3.1 Asset and Characteristic Selection

We use quarterly FX series for G10 currencies (AUD, CAD, EUR, JPY, NZD, NOK, SEK, CHF and GBP), using the USD as the base currency. These are the most heavily traded and liquid currencies in the world, which therefore have the lowest transaction costs. The first period is 01/01/1986 and the last is 01/01/2017 (with the exception of the Euro which starts on 01/01/1999). We therefore have 125 observations. Table 1 presents summary statistics for the 9 currencies. We observe large differences in average returns between currencies.

Table 1: Summary Statistics for Currency Returns

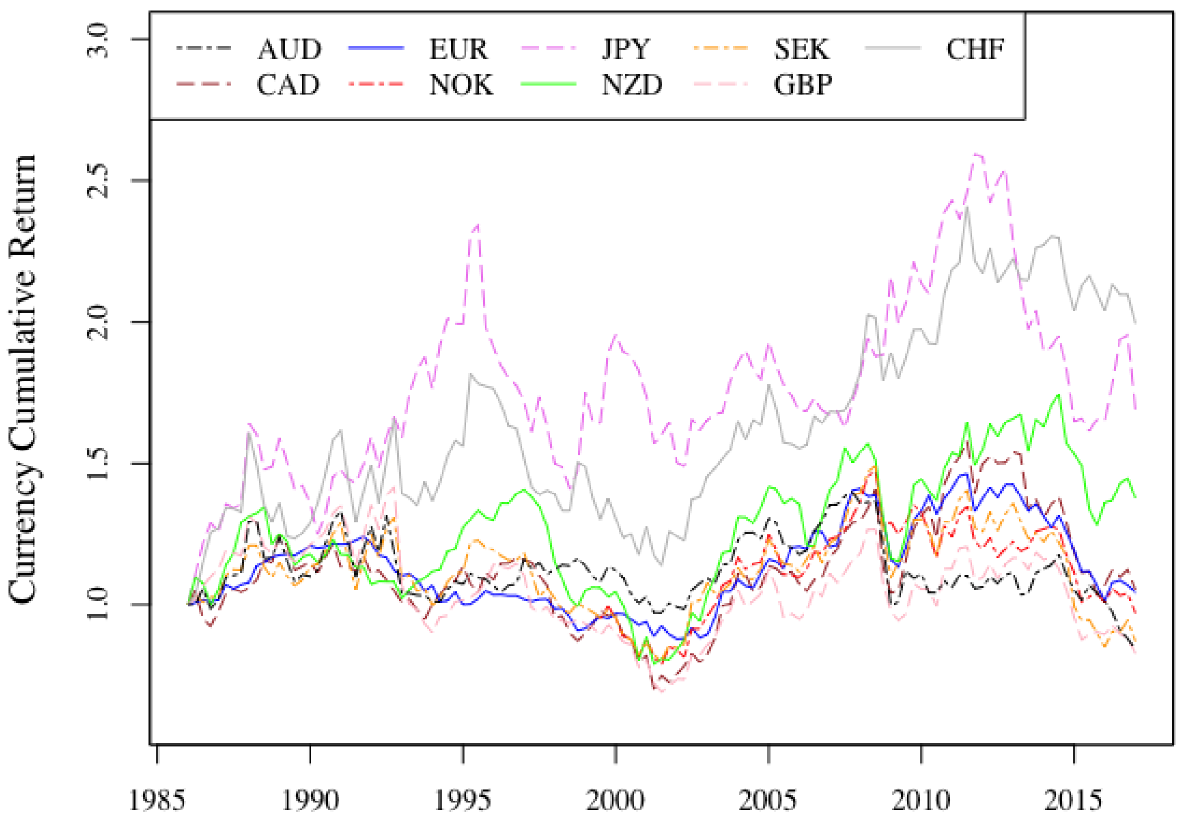
Currency	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev
AUD	-0.1970	-0.0316	0.0081	0.0020	0.0376	0.1550	0.0560
CAD	-0.1080	-0.0200	0.0020	0.0009	0.0192	0.0973	0.0348
EUR	-0.1320	-0.0337	0.0068	0.0009	0.0385	0.1210	0.0513
JPY	-0.1560	-0.0343	0.0009	0.0062	0.0468	0.2120	0.0636
NZD	-0.1370	-0.0256	0.0083	0.0041	0.0398	0.1420	0.0548
NOK	-0.1700	-0.0356	0.0027	0.0005	0.0357	0.1710	0.0558
SEK	-0.2460	-0.0337	0.0050	0.0001	0.0341	0.1340	0.0569
CHF	-0.1600	-0.0310	0.0000	0.0073	0.0472	0.2120	0.0604
GBP	-0.1750	-0.0255	0.0020	-0.0001	0.0245	0.1660	0.0502

Note: Statistics are provided are for 125 quarterly observations (with the exception of the Euro which starts in 1999 and hence has 71 observations).

The Swiss Franc has the highest, with an average of 0.0073 whilst the British Pounds has the lowest with an average of -0.0001. There is less dispersion in the standard deviations of returns, the lowest being the Canadian Dollar with 0.0348 and the highest being the Japanese Yen with 0.0636. The first and third quartiles inform us on the distribution without the influence of extreme observations. Figure 1 is the standardised cumulative return of the 9 currencies. We can see clearly that the majority of the currencies fluctuate around 1 over the sample period. Furthermore, it appears that fluctuations are highly correlated between the series. This is likely to reflect the movements of the USD, which is the base currency.

We also obtain characteristics of the corresponding countries in the form of economic fundamentals. These are obtained from the Thompson Reuters "International Comparable Economics" dataset, which provides series in the same units, scale and seasonality adjustments, therefore enabling accurate comparisons between countries. We select all 21 series available from this dataset, including many which are not obviously related to FX movements (such as the unemployment rate). These are GDP, Consumer Spending, Government Spending, Gross Fixed Capital Investment, Exports, Imports, M0, M1, M2, M3, Consumer Price Index (CPI), Core CPI, Purchasing Price Index, Export Prices, Import Prices, Terms of Trade, Unemployment Rate, Industrial Production, Government Deficits, Government Debt, Total External Debt, Current Account, Current Account (as % of GDP), Foreign Trade Balance,

Figure 1: Standardised Cumulative Returns of FX Series



Note: All series are standardised to 1 on 01/01/1986 (with the exception of the Euro which starts in 1999 and hence has 71 observations).

International Reserves and Interest Rates[1]. Both the level and the growth rate and included for each variable, totalling 52 characteristics. [2] We use quarters as the sampling frequency, since the overwhelming majority of these indicators are reported quarterly. Data for the Euro starts on 01/01/1999.

3.2 Data Adjustments

Many of the indicators sampled do not go back to 1986 in the Reuters dataset, and some do not exist for certain countries. For instance, Japanese GDP only goes back to 1994 and Norway does not publish M1 money supply data. If we have some observations, we forecast backwards to 1986 using an ARIMA model, which is specified using information criteria. For a number of characteristics, we do not find any series in the Thompson Reuters dataset. These are Australia M2, Canada Interest Rates, Eurozone M0, PPI, Export Prices, Import Prices, Terms of Trades, Japanese Government Deficit, New Zealand M2, Industrial Production, Government Deficit, Norway M0, M1, M3 and UK interest rate. These represent only 3.2% of variables. Since observations are normalised cross-sectionally and PPP only models deviations from some benchmark portfolio, we can address this issue by plugging in 0s as it should not introduce significant biases.

The final transformation that is applied to the data is a cross-sectional normalisation, which is required for PPP estimation. In every time period, we collect the currencies which are active (to exclude the Euro before 1999). Then for every characteristic, we compute the mean and standard deviation of the characteristics for the active currencies of the given characteristic. We then subtract the mean and divide by the standard deviation the characteristics, which we used to compute the sum and standard deviation. These now have a mean of 0 and a standard deviation of 1, whilst the characteristics of currencies that are not active remain unchanged.

[1] We use 3-month sovereign bond yields as the interest rate.

[2] Every characteristic is observed for every currency, so we have a total 469 variables (52*9).

4 Methodology

4.1 The Baseline Model Parametric Portfolio Policies

We first present the baseline PPP model of Brandt et al. (2009). We assume that in every period t, asset return $r_{i,t+1}$ from t to $t+1$ is related to a vector of characteristics of the asset $x_{i,t}$. In our case, currency returns are related to the economic characteristics (such as the inflation rate) of the currency's home country. The investor chooses portfolio weights $w_{i,t}$ for asset i in the investable universe, in order to maximise his utility conditional on the portfolio return:

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} E_t\Big(u(r_{p,t})\Big) = E_t\bigg(u\big(r_{p,t}\big)\bigg) \tag{1}$$

where portfolio weights are parametrised as a linear function of asset characteristics:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta^T \hat{x}_{i,t} \tag{2}$$

where $\bar{w}_{i,t}$ is weight of asset i at time t of a benchmark portfolio, θ is a vector of coefficients to be estimated (and is constant across assets) and $\hat{x}_{i,t}$ is a vector of characteristic for asset i at time t, which is standardised cross sectionally to have a mean of 0 and a standard deviation of 1 for every characteristic across assets, in every period (as described in the data adjustments subsection). The benchmark portfolio is chosen to be the naive diversification portfolio (also referred to as equal weighted (EW)), in which 1/N of wealth is allocated to each of the N_t assets available in every period t:

$$\bar{w}_{i,t} = \frac{1}{N_t} \tag{3}$$

DeMiguel, Garlappi and Uppal (2007) argue for its use as a benchmark to assess the performance of various portfolio rules proposed in the literature, as it does not require the estimation of moments. The intuition behind the PPP model is that the asset characteristics should fully contain the distribution of returns which is relevant for forming portfolios. The parameter θ is estimated by maximising the sample analog (average utility over the estimation period as a function of θ):

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \frac{1}{T} \sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} \left(\bar{w}_{i,t} + \frac{1}{N_t} \theta^T \hat{x}_{i,t} \right) r_{i,t+1} \right) \tag{4}$$

where T is the estimation period and N_t is the number of assets in period t.

4.2 Regularisation of Parametric Portfolio Policies

Brandt et al. (2007) and the subsequent empirical literature (such as Barroso and Santa-Clara 2015) have estimated the model with a small number of characteristics, which are predicted by economic theory to have relevance for the asset returns. For instance, Brandt et al. (2007) use three characteristics for equities: the market capitalisation, book to market ratio and past returns. This has the elegant theoretical property of nesting the long-short portfolios construction of Fama and French (1993). However, one can argue that there are more than three characteristics which are relevant to forming portfolios, and hence the model's performance can be improved by using a large number of asset characteristics. When doing so, we are aware of the risk of overfitting. If unrelated characteristics are included, the estimation procedure could still estimate a large coefficient (and hence pick up "noise"). This would lead worse out-of-sample performance, similarly to a kitchen-sink regression. To overcome this issue, we propose three regularisation strategies for penalising the size of coefficients and selecting the optimal combination of characteristics.

We introduce the baseline PPP model in (4) by adding a ridge constraint:

$$\hat{\theta}_R = \underset{\theta}{\operatorname{argmax}} \frac{1}{T} \sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} \left(\bar{w}_{i,t} + \frac{1}{N_t} \theta^T \hat{x}_{i,t} \right) r_{i,t+1} \right) \text{ s.t. } \sum_{j=1}^{p} \theta_j^2 \leq S_R \tag{5}$$

where S_R is a regularisation tuning parameter. This equation trades off two different criteria. As well as the baseline PPP model, it seeks parameters that maximise the average utility over the estimation period. In addition, the ridge constraint has the effect of shrinking the estimates of θ towards 0 as S_R increases. Hence, it reduces the flexibility and variance of the estimator, at the expense of an increase in bias. As the model produces different estimates for every value of the tuning parameter S_R, it's selection is crucial. We defer this issue to the next section, in which we discuss cross validation.

The ridge constraint will include all p characteristics, and hence does not address the fact that some characteristics may not be relevant to the joint distribution of the return process. The penalty will shrink all of the coefficients but will not set any of them at exactly 0. To address this issue, we introduce a lasso constraint:

$$\hat{\theta}_L = \underset{\theta}{\operatorname{argmax}} \frac{1}{T} \sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} \left(\bar{w}_{i,t} + \frac{1}{N_t} \theta^T \hat{x}_{i,t} \right) r_{i,t+1} \right) \text{ s.t. } \sum_{j=1}^{p} |\theta_j| \leq S_L \tag{6}$$

In addition to shrinking the coefficient estimates, the lasso forces some of the coefficient estimates to be close to 0 for a sufficiently large tuning parameter S_L, and hence performs a variable selection (of the asset characteristics). Hence, the lasso constraint produces "sparse" models with only the subset of characteristics that are relevant. As with the ridge constraint, we defer to the next section the discussion of the selection of S_L. The lasso has severe limitations for high dimensional data (large p) with few observations (small n), as it selects at most n variables before it saturates. Furthermore, if there are groups of strongly correlated variables, it has a tendency to select one and drop the others from these groups. As we are using quarterly data, the number of observations in the estimation period is small. Meanwhile, we have a high dimensional dataset (p=52) and economic indicators can be highly correlated, so we are likely to encounter these problems. In order to overcome these limitations, we present the Elastic Net (Zou and Hastie 2005), which combines both the

Ridge and Lasso constraint:

$$\hat{\theta}_{EN} = \underset{\theta}{\operatorname{argmax}} \frac{1}{T} \sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} \left(\bar{w}_{i,t} + \frac{1}{N_t} \theta^T \hat{x}_{i,t} \right) r_{i,t+1} \right) \text{ s.t. } \begin{cases} \sum_{j=1}^{p} \theta_j^2 \leq S_R \\ \sum_{k=1}^{p} |\theta_k| \leq S_L \end{cases} \tag{7}$$

The lasso constraint (L1 norm) selects characteristics whilst the ridge constraint (L2 norm) stabilises the coefficients and reduces variance. The elastic net can select more than n characteristics (with n being the number of observations in the estimation period).

4.3 Transaction Costs and Weight Constraints

To realistically evaluate the performance of our portfolio strategies, we account for transaction costs. We define the portfolio returns net of transaction costs as:

$$r_{p,t+1} = \sum_{i=1}^{N_t} w_i r_{i,t+1} - c \sum_{i=1}^{N_t} |w_{i,t} - w_{i,t-1}| \tag{8}$$

where c reflects transaction costs as a proportion of turnover, and is assumed to be constant through time. Hence, transaction costs are proportional to the amount of turnover. As we are working with FX, which is a highly liquid market, transaction costs are very low. We follow the previous literature by assuming that c = 3bps. In reality, transaction costs are decreasing over time and depend on the size of transactions. Since we are rebalancing every quarter, we do not expect these small transaction costs to significantly erode performance. We impose a non-negativity constraint on the portfolio weights, as follows:

$$w_{i,t}^{+} = \frac{\max\{0, w_{i,t}\}}{\sum_{j=1}^{N_t} \max\{0, w_{j,t}\}} \tag{9}$$

We set to 0 weights which are negative and scale the remaining weights by the sum of the positive ones. This formulation does not affect the parametrisation of the portfolio policy and increases stability by ensuring that all weights are between 0 and 1. We are implicitly allocating across the 9 currencies without ever going long on the base currency. As in Brandt et al. (2007) we assume CRRA utility over returns:

$$u(r_{p,t+1}) = \frac{(1 + r_{p,t+1})^{1-\gamma}}{1-\gamma} \tag{10}$$

with a risk aversion parameter γ of five.

5 Estimation and Results

5.1 Numerical Method for Estimation

The estimation of $\hat{\theta}$ is performed numerically through the differential evolution algorithm proposed by Brest et al. (2006). The algorithm takes as an input a function to minimise, which typically performs calculations on a dataset in the global environment (which depends on the input vector) and returns a value to the algorithm, which then proceeds to the next iteration until it has found an input vector which minimises the objective function. In this case the objective function is the average utility over the estimation periods which is computed by calculating the weights, returns and finally the utility in every period (which is then preceded by a minus sign since the algorithm performs minimisation whereas we want maximisation), whilst the input vector is the vector of coefficients. Unlike many other optimisers, it searches for a global maximum by maintaining a population of candidate solutions, which evolve randomly. As the evolutions are stochastic, convergence is stochastic and hence depends on the seed values. To ensure reproducibility, we set a seed of 1234 for all optimisations. It is implemented in R as part of the "DEoptimR" package. As the search space has 52 dimensions, convergence takes around two days with the default maximum number of iterations (52*200). For every model, we determine the number of iterations after testing the algorithm for a single observation, and observing the asymptotic behaviour of parameter estimates. As expected, we find that adding constraints significantly increases the number of iterations needed to ensure convergence. We set the number of iterations to 60 for the baseline PPP model and to 70 for the ridge regularisation. For the Lasso and Elastic Net, we set the number of iterations to 200. Indeed, adding the lasso constraint leads to an increase as the algorithm converges towards a corner of a hypercube in 52 dimensions. Since the parameters converge asymptotically to 0 for the characteristics that are dropped, estimates are infinitesimally close to 0. When plotting the kernel density of the parameters, a large spike is found around 0, indicating convergence. The search space is constrained to -1 and 1 for every element of θ, which is not binding since the parameters do not converge to the boundaries.

To evaluate the out-of-sample performance of our model, we estimate the parameters in every period using a rolling estimation window of past observations. Since we have 125 observations in total, we set to size of the estimation window to 30, to balance the tradeoff between the length of the estimation window and the number of out-of-sample returns, which we can use for inference. Since we are using quarterly data, there is a 3-month lag between the estimation window and returns. This ensures that all economic statistics used in our estimation would have been available at the time of the rebalancing decision. The full backtest requires that the model is estimated 95 times (at every rebalancing) in a "for loop". Since every iteration of the for loop is independent of each other (there is no need to communicate between two given iterations), they can be performed by a separate thread and at the same time on different CPU cores, using the "parallel" R package. We use the Amazon EC2 instance service, which consists in virtual servers in the cloud, which can be rented at a per-hour price. This allows us to use an Ubuntu Linux server with 64 CPU cores (the run time is roughly proportional to the number of CPU cores due to the use of parallelism).

5.2 K-Fold Cross Validation

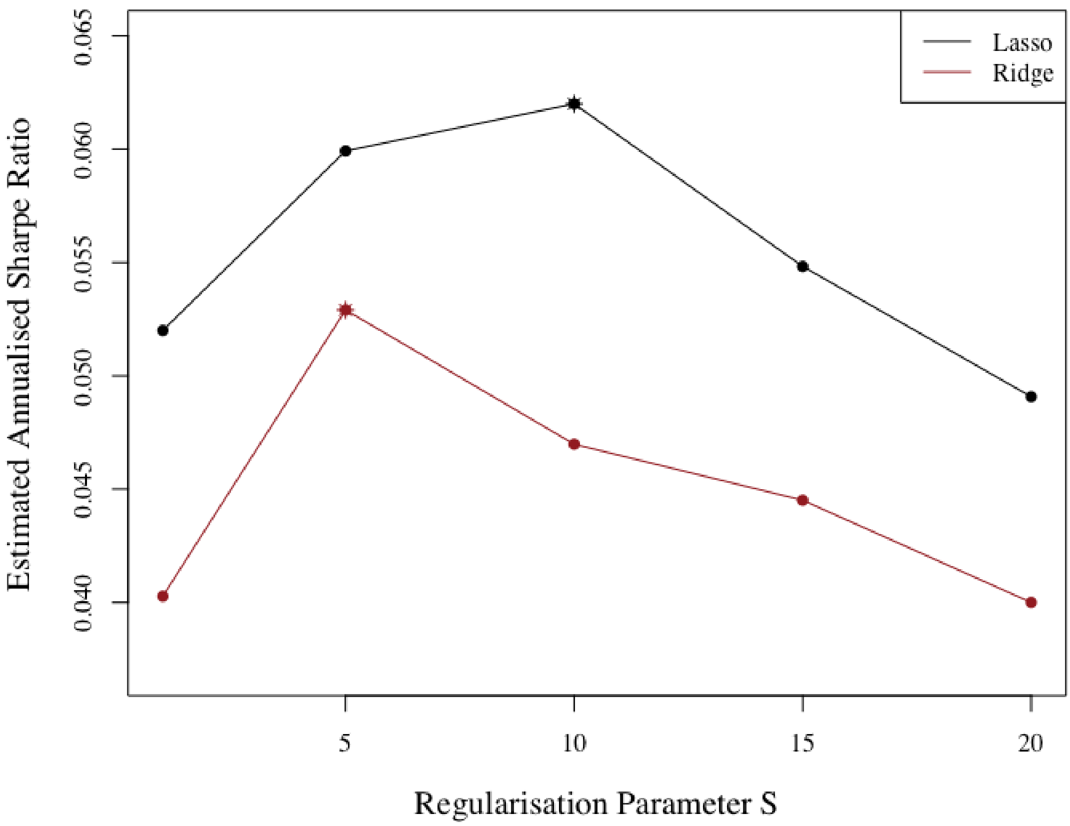
To select the regularisation parameters S_R and S_L, we estimate the out-of-sample Sharp ratio of our model through k-fold cross validation for a range of values. We randomly divide the observations into k groups of equal size, and form k estimation sets of k-1 groups and one validation group, which is held out of the estimation. For every value of S in which we are interested in, we estimate the model for the k groups and compute the mean Sharpe ratio over the validation groups. This forms our estimate of the out-of-sample Sharpe ratio of our model for the given regularisation parameter. We repeat this process for a range of plausible values using the same groups. Hence, for every value S, we estimate the model k times with $\frac{k-1}{k}$-times the entire dataset, which is computationally complex. We select k to optimise the tradeoff between computer run-time and estimator variance. Since we are using an 8-core computer, setting k=8 allows us to efficiently estimate the models in parallel. We display the results in Figure 2. Firstly, we can see that the estimated Sharpe ratio is much higher for the lasso than the ridge for every value of the regularisation parameter. This is a first indication that the lasso performs better. Secondly, the estimated Sharpe ratio is maximised at $S_R = 5$ and $S_L = 10$ for the Ridge and Lasso regularisations respectively, which we will therefore use in our estimation.

For the Elastic Net, we re-use the optimal parameters of the lasso and ridge cross-validation. As it combines the two regularisations, cross validation of the Elastic Net would require that we evaluate a matrix of potential parameters. In this case, we would need to run the estimation 16 times, which would be too computationally intensive.

5.3 Presentation of Results

We present the out-of-sample performance of our model with the different regularisations. We also estimate the equally weighted (EW) portfolio (sometimes referred to as “1/N”). Since our PPP models deviations from this portfolio, it is a natural benchmark (and does not require the estimation of the moments of the distribution of returns). In addition,

Figure 2: 8-fold Cross Validation Estimates of Sharpe Ratios



Note: 8-fold cross validation was run for 4 values of S_R and S_L, using the same groups every time. Since there are 125 observations in total, the 8 groups have an average size of 15,625, with estimation sets of average size 109,935.

we estimate the sample-based tangency portfolio (TP) with a rolling estimation window of identical size and a risk free rate of 0. We display these results in Table 2. We find that the Tangency Portfolio performs poorly, with a Sharpe Ratio of -0.1530 and a CE return of -0.4428 (i.e. an investor would be indifferent between holding it or losing 44% in every period with certainty). We can explain this result by the fact that the vector of expected returns and covariance matrix of expected returns have to be estimated. Merton (1980) argues that a very long estimation period is needed to estimate the latter precisely, whilst we have an estimation window of 30 observations. This results in extreme weights, with an average smallest weight of -286%. As in DeMiguel, Garlappi and Uppal (2007), the EW portfolio outperforms the TP since the losses due to estimation error outweigh the gains from optimal diversification. We will therefore use the EW portfolio as our benchmark to evaluate PPP models. The EW portfolio has an annualised average return of 0.0046, which captures the fact that, on average, our basket of currencies have appreciated by a very small amount against the USD over Q3 of 1993 to Q1 2017. The EW portfolio's Sharp Ratio is 0.0299, annualised standard deviation is 0.1539 and finally the CE is -0.0690. We proceed to compare our benchmark to that of the baseline PPP model, which is found to have a a much higher Sharp ratio of 0.2566 (due to an increase in average return), whilst the CE return increases slightly to -0.0628. This shows that PPP outperforms the EW portfolio when a large number of (potentially unrelated and noisy) characteristics are used for estimation. There is evidence of overfitting in the large amount of turnover and extreme weights, with a max w_ix100 of 41.1738. However, this does not erode performance as dramatically as with the Tangency Portfolio. This is a strong starting point to improve upon through regularisation. The introduction of the ridge constraint leads to an increase in Sharpe ratio to 0.3485, as the average return increases noticeably to 0.0601 and the standard error increases to 0.1725 whilst the CE return falls to -0.0304. This finding clearly demonstrates the benefit of regularisation of high-dimension parametric portfolio policies. Increasing the rigidity of the estimator by penalising extreme coefficients reduces the variance of the estimator. This, in

Table 2: Summary Statistics for Currency Returns

	TP	EW	PPP	Ridge	Lasso	ElasticNet
$\bar{r}$	-0.1439	0.0046	0.0428	0.0601	0.0687	0.0704
$\sigma(r)$	0.9404	0.1539	0.1666	0.1725	0.1566	0.1622
SR	-0.1530	0.0299	0.2566	0.3485	0.4387	0.4342
CE	-0.4428	-0.0690	-0.0628	-0.0304	0.0095	0.0071
$\sum \lvert w_{i,t} - w_{i,t-1} \rvert$	15.1111	0.0000	1.0459	0.9345	0.9751	0.9507
min $w_i \times 100$	-286.8808	11.1111	0.0000	0.0000	0.0000	0.0000
max $w_i \times 100$	369.3958	12.5000	41.1738	37.0147	38.0867	38.4416

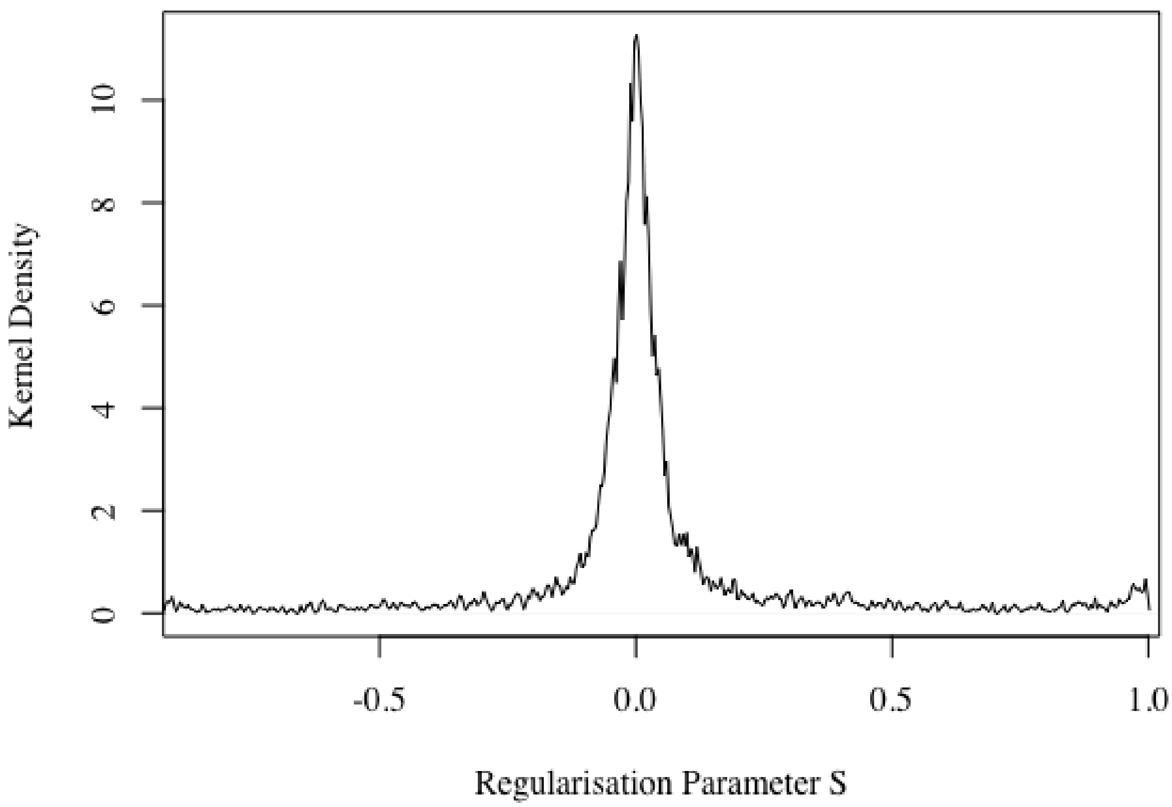
Note: The models considered are: TP the Tangency Portfolio, EW the equally weighted, PPP the baseline PPP model with no regularisation, Ridge the PPP model with a ridge constraint, Lasso the PPP model with lasso constraint and finally ElasticNet the PPP with Elastic net constraint. We present estimates of the annualised average return, standard deviation and Sharpe ratio, which is computed as the ratio of the annualised average return and standard deviation. The certainty equivalent return is computed numerically as the return which has a CRRA utility equal to the average utility of annualised returns of the investor over time. $\sum |w_{i,t} - w_{i,t-1}|$ is the average turnover. min $w_i \times 100$ and max $w_i \times 100$ indicate the average smallest and largest weight. Statistics are from a sample from Q2 1993 to Q1 2017, totalling 95 observations since the first 30 periods are used for estimation. The fact that max w_ix100 and min w_i x100 are different for EW is due to the increase in the number of currencies from 8 to 9 when the Euro is introduced in 1999.

turn, reduces the influence of individual factors, which can potentially be unrelated to the joint distribution of returns. This also leads to a 10% fall in portfolio turnover and an 11% fall in max w_ix100 as smaller coefficients lead to more stable weights. Whereas the adverse influence of unrelated factors is eroded, it is not completely eliminated. Hence, we turn our attention to the lasso regularisation, which performs a variable selection.

We find that the lasso regularisation performs significantly better than the ridge, with a Sharp Ratio of 0.3485. This is due to both an increase in average return to 0.0687 and a large fall in the standard deviation to 0.1566. This is a noticeable finding, since the standard deviation is smaller than for the baseline PPP and is almost as small as that of the EW portfolio (0.1539) but the return is almost 15X higher on average. The CE increases to 0.0095. We can explain this by the fact that the lasso constraint produces a "sparse" PPP model with only the subset of variables that are relevant. This eliminates completely the influence of unrelated variables, and increases the relative influence of relevant ones, which are no longer "crowded out" by noise. To verify this claim, we examine the parameters to find evidence of variable selection. Figure 3 is the kernel density of every coefficient estimate for every time period for the lasso model, which allows us to see the distribution of coefficients.

We find that the coefficients have a Laplace distribution centred around 0, which is what we would expect from traditional lasso estimates. This finding confirms that the adaptation of the lasso regularisation to PPP estimation has had the intended effect. One can see that many of the coefficients are not exactly equal to zero but rather are concentrated around zero. This can be explained by the numerical optimisation process, which converges towards a corner of a hypercube in 52 dimensions. The coefficients of the variables that are dropped converge asymptotically to 0, and would require a very large number of iterations to reach 0 up to 8 digits. The number of iterations controls the width of the spike, as we obtain an ever thinner spike as coefficients become ever closer with every iteration. Since we re-estimate the model for every rebalancing, we constrain the number of iterations such that

Figure 3: Kernal Density Estimation of Lasso Coefficients



Note: This is a kernel density estimate plot of every coefficient estimate of the lasso regularisation model. As the model is estimated 95 times (for every rebalancing) and there 52 parameters for every time period, we have a total of 4940 coefficients. The kernel bandwidth is 0.00344.

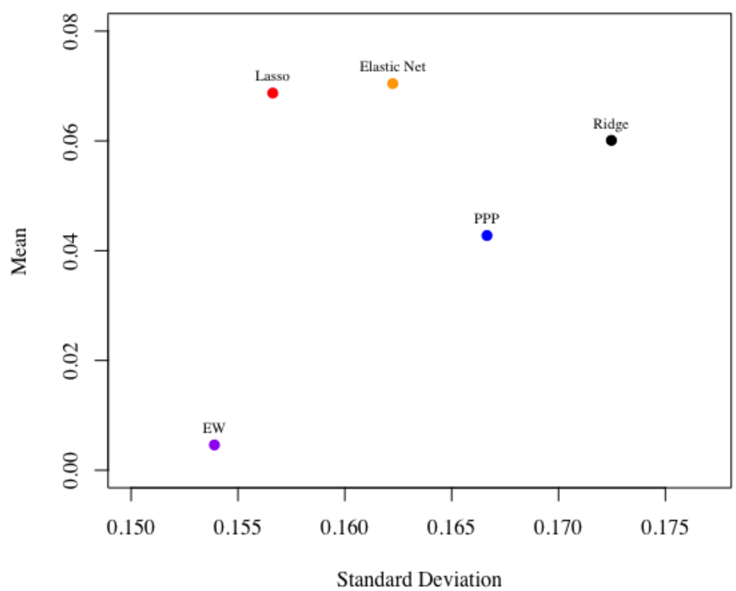
we can compute a backtest in a reasonable amount of time (as detailed in the numerical optimisation section). This explain why variables that are dropped have coefficients that are very close but not equal to zero. We can suppose that the vast majority of the coefficients which are within the interval $[-0.1, 0.1]$ are converging to zero. The fact that they do not equal zero should not have a strong effect on our results, since the size of the bias is small and exogenously determined by the optimisation algorithm.

Finally, we evaluate the performance of the elastic net regularisation. We find that relative to the lasso (our best performing model so far), the Elastic Net performs slightly worse due to an increase in standard deviation to 0.1622 (from 0.1566). The Sharpe Ratio falls slightly to 0.4342 (from 0.4387) and the CE return falls to 0.0071 (from 0.0095). Hence, the two models are practically indistinguishable. The lasso has the limitation that if there are groups of strongly correlated variables, it has a tendency to select one and drop the others from the groups. The Elastic Net overcomes this problems, and therefore should produce stronger returns by selecting variables despite the strong correlation between economic indicators.

To summarise our results, we visualise the estimated average mean and standard deviation of returns in the mean-variance space in Figure 4. We can see clearly that all of the PPP models have a higher mean return than the EW portfolio. They also all have a higher standard deviation. All regularised PPP model have a higher mean return than the baseline PPP model. Whilst the Ridge regularisation increases the standard deviation, the Lasso and Elastic Net lead to a reduction in standard deviation. The lasso even comes very close to the EW portfolio in terms of standard deviation, but with a significantly higher mean return. It is useful to bear in mind that PPP model deviation from a benchmark portfolio, with is the EW portfolio in this case. Hence, the lasso regularisation has led to a deviation from the EW portfolio, which maintains its low standard error whilst dramatically increasing its mean return.

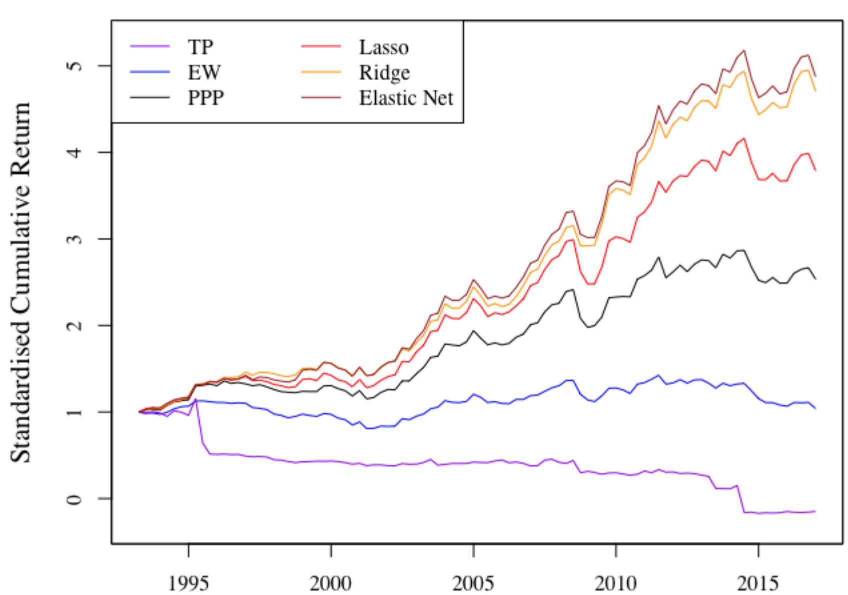
Finally, Figure 5 shows the standardised cumulative return of the 6 models that we studied. We learn that the TP is relatively stable but suffers from large negative losses, which leads

Figure 4: Return Statistics in the Mean Variance Space



Note: The data used in this plot corresponds to the first two lines of table X. Refer to the note of that table for a data description. We omit the tangency portfolio since it's standard deviation is so large (0.9404) and average return so small (-0.1439) that it not be possible to visualise on the same plot as the other 5 models whilst maintaining an acceptable level of clarity. As the return of the minimum variance portfolio is higher than that of the tangency portfolio, it is not possible to plot the mean-variance efficient frontier.

Figure 5: Cumulative Returns of FX series used



Note: All series are standardised to 1 on 01/01/1986.

the cumulative return to drop to close to 0. The baseline PPP model outperforms the EW model by a strong margin. The lasso and elastic net regularisation models are almost indistinguishable, and the ridge regularisation model lies midway between the baseline PPP model and the lasso.

5.4 Economic Interpretation of Coefficients

Our estimation has adopted an agnostic view to economic theory to discover empirical properties of the data, and is closely related to machine learning techniques. In this section, we discuss the economic interpretation of coefficient estimates. The Lasso regularisation has successfully selected economic indicators, which are then used to implicitly predict exchange rate movements through the selection of weights. Therefore, when a variable is selected, it can be interpreted as evidence of its predictive power over future exchange rate movements. Since the estimation of coefficients is repeated for every period, we have 95 sets of 52 coefficients. To facilitate interpretation, we average the coefficients over time and report a single coefficient per characteristic in Table 3 for the Lasso model. Since the PPP model imposes that all characteristics are standardised cross sectionally (as described in the data adjustment section), the coefficients are directly comparable. The normal econometric inference for model selection is replaced by the lasso selection process based on predictive value given the standardised data. Despite not having standard errors, we can accurately infer which variables were dropped and evaluate the sign and magnitude of coefficients in economic terms.

We observe that the lasso has performed a variable selection by the fact that many of the values are close to 0. Since we are reporting averages of estimates over time, it is possible that some variables were dropped in some time period but not in others. Hence, it is not surprising that the averages are not often exactly equal to 0. If a variable has an average coefficient of less than 0.1 in absolute value, we can be relatively certain that it has been dropped from the estimation over a large proportion of time periods. We find that this is the case for 34

Table 3: Average Coefficient Estimates

Variable	Level	Growth Rate
GDP	0.472	-0.009
Consumer Spending	0.178	-0.007
Government Spending	0.456	-0.008
Gross Fixed Capital Investment	0.062	-0.013
Exports	0.180	-0.019
Imports	-0.416	-0.010
M0	0.311	0.008
M1	0.009	0.008
M2	0.040	0.016
M3	0.124	-0.002
CPI	-0.034	-0.011
Core CPI	-0.315	-0.013
Purchassing Price Index	-0.040	0.003
Export Prices	-0.591	-0.345
Import Prices	-0.989	-0.264
Terms of Trade	-0.993	-0.099
Unemployment Rate	0.734	-0.048
Industrial Production Index	-0.338	-0.017
Government Deficit	-0.093	0.042
Government Debt	0.050	-0.015
Total External Debt	0.109	-0.031
Current Account	0.210	-0.028
Current Account as a % of GDP	-0.001	-0.009
Foreign Trade Balance	0.062	0.032
International Reserves	-0.197	-0.029
Interest Rate	-0.058	0.041

Note: We report the average coefficient over 95 time period for the 52 variables. These are are constructed from 26 economic indicators which are used both in level and growth rate (hence our table has 26 rows to increase clarity). Hence, the column "Growth Rate" reports the average coefficient estimates of the 26 variables in growth rate (i.e. these are separate characteristics).

out of the 52 variables. Hence, the Lasso has performed a stringent variable selection and produced sparse models, which dynamically selected the most relevant variables.

The variables that have an average coefficient of more than 0.1 in absolute value are GDP, Consumer Spending, Government Spending, Exports, Imports, M0, M3, Core CPI, Export Prices, Import Prices, Terms of Trade, the Unemployment Rate, the Industrial Production Index, Total External Debt, the Current Account, International Reserves, the growth rate of M2, the growth rate in the Export Price and the growth rate of Import Prices. Only 2 out of the 18 selected variables are growth rates. This indicates that the level of economic fundamentals have more predictive power then their growth rates. Many of the selected variables are predicted by economic theory to be strongly related to exchange rates movements, such as Imports, Exports, M0, M3, Core CPI, Terms of Trade, the Current Account and International Reserves. Furthermore, the sign of their average coefficients are consistent with theoretical models. For instance, we know that countries with a large trade deficit are more likely to have a depreciating currency. This is captured in our results by the fact that the average coefficient for Imports is -0.416 and is 0.180 for exports. Hence, currencies of countries with a large trade deficit will be underweighted. The average coefficient on Core CPI -0.315, which captures the notion that countries with high inflation will see their currencies depreciate due to lower real interest rates. The average coefficient on CPI is -0.034, indicating that CPI has been dropped by the Lasso estimation over many of the time periods. We know that if there are highly correlated variables, the Lasso has a tendency to select one of the variables and drop the others. In this case, the estimation has dropped CPI in favour of Core CPI, which reflects inflationary pressures better. Our model will underweight currencies with a high core inflation rate. The average coefficient on Terms of Trade is the largest out of all the coefficients in absolute terms, and hence Terms of Trade have the strongest influence over portfolio weights. Countries with a high relative price of imports in terms of exports will be strongly underweight. The unemployment rate is selected by the model, despite there not being any theoretical relationship with exchange rate movements. The average coefficient on the growth in export prices is -0.345 and -0.264 for import prices. Our model will underweight countries in which export prices are increasing faster than import prices, which is consistent with the economic idea that these are countries that are more like to devaluate their currencies. Overall our estimates are mostly consistent with the literature and theoretical predictions about the relationships between economic fundamentals and currency movements.

6 Limitations

The main limitation of our approach is the computational complexity of the estimation procedure, which is due to the large number of parameters, which need to be estimated at every rebalancing date. With a very small sample of 125 observations, it is feasible to estimate on a single powerful computer (assuming the code optimisations in section 5.1). However, if we choose a higher frequency we would need to distribute the computation over multiple computers, which is impractical. Whereas the baseline PPP with a small number of characteristics is easy to estimate, increasing the number characteristics and imposing constraints comes at a high cost in terms of estimation difficulty.

Another limitation to this work is the small number of observations (125) which is due to the quarterly sampling frequency of most economic fundamentals. This does not allow us to have a very long estimation window, which in turn leads to numerically unstable coefficient estimates, and therefore weights, as shown by the high turnover. Since we are using an asset class with low transaction costs, this does not dampen performance significantly. A possible extension would be to implement the gradual adjustment of transaction costs as in Brandt et al. (2007).

More work is needed to investigate why the Elastic Net performed worse than the lasso. It would be useful to perform the cross-validation of the Elastic Net separately, instead of using the regularisation parameters of the Ridge and Lasso.

7 Conclusion

In conclusion, we introduced the lasso regularised parametric portfolio policies model, which scales to a large number of characteristics without the risk of overfitting. Our model adopts an agnostic view of economic theory by selecting the most relevant characteristics in every time period. We presented a numerical method for estimation by imposing a constraint on the optimisation problem. We evaluated our model empirically using a basket of G10 FX series and 26 economic fundamentals as characteristics (which are present in both level and growth rate, totalling 52 characteristics). We performed a k-fold cross validation of the out-of-sample Sharpe Ratio to select the tuning parameter of the regularisation. We found that our model outperformed the baseline PPP model in this context, with a lower standard deviation of 0.1566, a higher average return of 0.0687 and a higher CE return of 0.0095. We analyse the variables which are selected by the Lasso, which documents the importance of Terms of Trade, Core CPI and the current account for explaining the optimal portfolio weights. Hence, our model's variable selection and estimation is mostly consistent with the literature and theoretical predictions about the relationships between economic fundamentals and currency movements, and hence provides support for macroeconomic theories of FX. Our findings suggest that there exists predictability for FX movements if we condition on a range of macroeconomic fundamentals and select variables dynamically.

8 Bibliography

Barro, Robert J. "Rare disasters and asset markets in the twentieth century." The Quarterly Journal of Economics 121.3 (2006): 823-866.

Barroso, Pedro, and Pedro Santa-Clara. "Beyond the carry trade: Optimal currency portfolios." Journal of Financial and Quantitative Analysis 50.5 (2015): 1037-1056.

Brest, Janez, et al. "Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems." IEEE transactions on evolutionary computation 10.6 (2006): 646-657.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov. "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns." The Review of Finan- cial Studies 22.9 (2009): 3411-3447.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. "Carry trade and momentum in currency markets." (2011).

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. "Carry trade: The gains of diversification." Journal of the European Economic Association 6.2?3 (2008): 581-588.

Carhart, Mark M. "On persistence in mutual fund performance." The Journal of finance 52.1 (1997): 57-82.

Cheung, Yin-Wong, Menzie D. Chinn, and Antonio Garcia Pascual. "Empirical exchange rate models of the nineties: Are any fit to survive?." Journal of international money and finance 24.7 (2005): 1150-1175.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?." The review of Financial studies 22.5 (2007): 1915-1953.

Engel, Charles, et al. "Exchange Rate Models Are Not as Bad as You Think [with Comments and Discussion]." NBER Macroeconomics annual 22 (2007): 381-473.

Fama, Eugene F. "Forward and spot exchange rates." Journal of monetary economics 14.3 (1984): 319-338.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." Journal of financial economics 33.1 (1993): 3-56.

Li, Jiahan, Ilias Tsiakas, and Wei Wang. "Predicting exchange rates out of sample: Can economic fundamentals beat the random walk?." Journal of Financial Econometrics 13.2 (2015): 293-341.

James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.

Malkiel, Burton G., and Eugene F. Fama. "Efficient capital markets: A review of the- ory and empirical work." The journal of Finance 25.2 (1970): 383-417.

Meese, Richard A., and Kenneth Rogoff. "Empirical exchange rate models of the seventies: Do they fit out of sample?." Journal of international economics 14.1-2 (1983): 3-24.

Menkhoff, Lukas, et al. "Carry trades and global foreign exchange volatility." The Jour- nal of Finance 67.2 (2012): 681-718.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society. Series B (Methodological) (1996): 267-288.

Zou, Hui, and Trevor Hastie. "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67.2 (2005): 301-320.