Essays on Financial Forecasting

Kara Alice Ng

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Eric Zivot, Chair

Yu-Chin Chen

Chang Jin Kim

Program Authorized to Offer Degree:

Economics

University of Washington

Abstract

Essays on Financial Forecasting

Kara Alice Ng

Chair of the Supervisory Committee:
Professor Eric Zivot
Department of Economics

This dissertation investigates forecasting monetary policy, currency strategies, and equity returns. The three chapters are summarized below:

1. From December 2008 to November 2015, the Federal Reserve's fed funds target range was 0 to 25 basis points, at the effective zero lower bound for nominal interest rates. Although the fed funds target did not change for seven years, the underlying unobserved dynamics are not constant. Using a dynamic ordered probit model, I estimate the desired fed funds rate in the absence of a zero lower bound. Using real time data, the model has a good track record of predicting date of first rate hike. After the first rate hike, however, the model predicts a front-loaded normalization path. The actual pace of rate normalization was unusually slow in the first couple years of tightening, most likely

because the Fed felt asymmetric risk around the low interest rates. In contrast, the model deliberately allows the fed funds rate to move seamlessly in and out of the zero lower bound, neglecting this asymmetric risk.

2. Carry, momentum, and value are three established and profitable currency trading strategies. However, their performance varies through time such that a timing strategy could increase profits. I use a nonlinear model to combine these three strategies, where time-varying weights are allowed to vary depending on the degree of currency misevaluation. However, I find that even under a nonlinear model, currencies revert to fair value too slowly for a one-month investment horizon timing strategy. Because of this, a factor timing strategy using time-varying weights does not outperform equally weighting the three currency strategies over short horizons.
3. Buy-and-hold strategies are profitable because the equity market rises on average. However, equity timing could improve risk-adjusted portfolio profits. I generate out-of-sample return forecasts from a range of methods: multivariate regressions, univariate regressions, the average of univariate forecasts, and an expanding window historical average. Then I use these forecasts in a long-short equity strategy. Combined indicators result in 11 percent annualized returns, a 0.73 information ratio, and a 0.59 Sharpe ratio over a 22-year out-of-sample backtest, which outperforms a buy-and-hold strategy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Eric Zivot, for his advice, guidance and backing. I also express sincere gratitude to my committee members Yu-Chin Chen, Chang Jin Kim, and Phillip Thurtle.

This dissertation is only made possible because of the generous support, advocacy, and counsel from my Russell Investments friends and colleagues: Abraham Robison, Andrew Pease, Van Luu, Paul Eitelman, Amir Talebi, and Luke Lee. I am eternally grateful. Also, I appreciate the constant encouragement from Sangeetha Srinivasan, during our shared adventure of finishing a PhD while working a full-time job.

A special thank you to Michael Dueker for his mentorship, insight, and kindness.

DEDICATION

To Allen, who is also the best one.

Chapter 1. MODELING THE FED FUNDS TARGET RATE ABSENT THE ZERO LOWER BOUND USING A DYNAMIC ORDERED PROBIT MODEL

1.1 INTRODUCTION

On December 2008, the Federal Reserve lowered their fed funds target to a range between 0 and 25 basis points, hitting the zero lower bound for nominal interest rates. This paper models the unobserved desired fed funds target (also called the shadow rate or latent fed funds target) using a probit model, in a world where nominal interest rates could be negative. The unobserved desired rate is estimated from historical economic growth, inflation, labor market strength, observed rate changes, and estimated desired rates. In a world above the zero lower bound, if we see the Fed increase the fed funds target, it must mean the desired rate moved sufficiently above the previous target to warrant a change. When the observed rate hits and stays at the zero lower bound from December 2008 to November 2015, it must mean the desired rate is negative during that period. A probit model is a natural choice to estimate this shadow rate. Many researchers apply probit models to fed funds rates, but this paper is the first to apply a probit model to the unobserved fed funds target after the onset of the zero lower bound.

Since the onset of the zero lower bound, a large set of literature estimates the shadow short rate (i.e. shadow fed funds rate). The papers have some commonalities— everyone generally interprets shadow rates as the fed funds rate in the absence of the zero lower bound. Also, even though specific methodologies differ, everyone generally estimates the shadow rate using relationships between short rates and variables not constrained by the zero lower bound.

However, there are some nuances with the interpretations. Some are related to policy prescription. Some claim to be a proxy of the stance of monetary policy during the zero lower bound. Some just care about its use in a larger model. In the context of the shadow rate literature, my shadow rate is the unobserved desired fed funds rate, and can serve as a mix of policy prescription, estimate of monetary policy stance, and forecasting tool. I use my shadow rate to fill in zero values of the fed funds rate during the zero lower bound, forecast the path of fed funds rates, and conduct scenario analysis.

After estimating the desired target and its relationship with macroeconomic variables, one can forecast future monetary policy conditioned on scenario paths for those macroeconomic variables. A backtest with real time macroeconomic data from 2009 to 2015 shows that the vintage consensus forecasts for the date of the first rate-hike were consistently hawkish, while the probit model forecasts were more stable. If the model provides more accurate forecasts at dates farther out compared to consensus forecasts, it can provide more realistic expectations for policymakers and observers who rely on monetary policy.

After the fed funds target increased above the zero lower bound end of 2015, the model implied a more hawkish monetary path than the market was expecting. Furthermore, the path of monetary tightening was front-loaded— aggressive hikes at first, then gradual hikes later. This contrasted with Federal Reserve communication, where their economic projections showed a back-loaded monetary tightening path— gradual rate hikes through 2016 and more aggressive hikes 2017 to 2018. With the benefit of hindsight, the actual fed funds rate was raised more slowly than either the model or the Federal Reserve foresaw, but the pace was more back-loaded. The actual pace of rate normalization was unusually slow in the first year of tightening, most likely because the Fed felt asymmetric risk around the low interest rates. However, the model deliberately allows the fed funds rate to move seamlessly in and out of the zero lower bound, neglecting this asymmetric risk.

Section 1.2 is my literature review— I describe how researchers use the ordered probit model methodology to model discrete choice monetary policy decisions. I also explore how researchers estimate the fed funds rate during the zero lower bound period. Section 1.3 describes the model, data, and estimation. Section 1.4 shows the shadow rate time series and estimated parameters. I compare the model forecasts with historical consensus and Federal Reserve forecasts. Section 1.5 compares this model's latent fed funds rate with other researchers' shadow rates. Section 1.6 concludes.

1.2 LITERATURE REVIEW

1.2.1 *Probit models and discrete monetary policy*

The inspiration for this paper comes from the literature for discrete policy choice. Eichengreen, Watson, Grossman (1985) use a dynamic ordered probit equation to account for the discrete changes in Bank of England's target bank rate. The researchers specify that the desired bank rate reacts to variables such as the Bank of England's reserve position, exchange rates, interest rates, and economic activity. The Bank of England will make discrete bank rate changes of -50 basis points or +100 basis points whenever the desired (unobserved) bank rate strays too far from the last set (observed) bank rate. Eichengreen et al. estimate the model with maximum likelihood using weekly data from 1925 to 1931.

Dueker (1999) uses the dynamic ordered probit model of Eichengreen et al. on the US fed funds rate with weekly data from 1972 to 1997, estimated via Gibbs sampling rather than maximum likelihood. The desired (unobserved) fed funds rate reacts to lagged changes in fed funds rates and the commercial paper and treasury bill spread, which acts as a business cycle proxy. In this paper, the observed fed funds rate moves discretely in seven categories: < -0.50, -0.50, -0.25, 0.00, +0.25, +0.50, and > +0.50.

Hu and Phillips (2004) specify that the optimal fed funds target rate evolves as a function of money supply, initial unemployment claims, manufacturers' new orders. The data is monthly frequency from 1994 to 2001. The researcher's objective is to predict the timing and direction of changes; as such, they only have three ordered regimes: increase, no change, or decrease. They predict 78% meeting outcomes correctly over the sample period.

Van den Hauwe, Paap, and Dijk (2013) use a dynamic ordered probit methodology, estimated with real time data from 1990 to 2008, to predict fed funds target rate changes. Like Hu and Phillips, their order probit includes three regimes: target increase, no change, or decrease. However, van den Hauwe et al. include endogenous variable selection to test which macroeconomic and financial indicators have the best predictive ability in real time. They found economic activity indictors (e.g. industrial production, output gap, coincident index), term structure variables (e.g. 6-month treasury bill and effective fed funds rate spread), and forward expectation variables (e.g. professional forecasts, consumer confidence) have strong predictive power. They predict 82% of meeting outcomes correctly from 2001 to 2008 out-of-sample and 90% in-sample.

Ordered probit models seem like a natural choice for forecasting Federal Reserve policy decisions. However, analysis and use of this methodology stopped when the fed funds target hit the effective zero lower bound. This paper extends the ordered probit model through the world of near-zero short term interest rates.

1.2.2 *Modeling short term rates during the zero lower bound*

Other researchers do not approach modeling fed funds rate during the zero lower bound with a dynamic probit model. However, we broadly have something in common. Fed funds rate have relationships to other macroeconomic or financial variables not constrained by the zero lower bound, so we can reconstruct the fed funds rate using the estimated relationships and observed variables.

Taylor (1993) proposes a few iterations of simple monetary policy rules, but essentially says the federal funds rate should move as a function of inflation and output from their respective targets. The parameters of the rule are round values, specified based on intuition and prior research. With this policy prescription, the rule-based fed funds rate can be calculated in the period over the zero lower bound. Hakkio and Kahn (2014) estimate policy prescription rules using data from 1985 to 2001, a period of relative economic stability. In one specification, the effective fed funds rate is a function of lagged fed funds rate, core inflation gap, and unemployment gap. In a second specification, the effective fed funds rate is a function of the lagged fed funds rate, core inflation gap, and two labor market conditions indicators from the Kansas Federal Reserve. The two labor market indicators are the first two principal components of 24 labor market indicators. The relationships between these variables and the fed funds rate are significant at a 1% level during the researchers' sample period. Using the estimated coefficients, Hakkio and Kahn project the policy prescription from 2001 onward through the zero lower bound period. I show how the rule from Taylor, and rules from Hakkio and Kahn, compare to my estimated unobserved desired fed funds target later in this paper.

Black (1995) observes that the nominal short rate cannot be negative because there is always an option to hold cash at zero interest. He defines the observed interest rate as the maximum of zero interest (holding cash) or the shadow short rate (which can be negative). Using the term structure of interest rates and no-arbitrage conditions, one can price that shadow rate. The intuition is simple, but the mathematical solution is onerous. Many researchers, including Wu and Xia (2014) and Krippner (2012), build on Black's shadow rate term structure model (SRTSM) to calculate a shadow rate.

Wu and Xia propose an analytical approximation such that the SRTSM has a useable closed-form solution. Their paper data includes forward rates of 3 and 6 month, and 1, 2, 5, 7, and 10-year maturity from 1990 to 2013, and they extract three factors to capture information from the term structure. Wu and Xia calculate the shadow rate and confirm their shadow rate exhibits the same relationship with three main macroeconomic factors before December 2007 and after June 2009. I will use the shadow rate published on Wu and Xia's website, as of December 2015, for comparison purposes.

To insure the SRTSM has a useable closed-form solution, Krippner includes a function of maturity for the cash option. Krippner then estimates a shadow rate from two term structure factors. He argues the shadow rate is a good gauge on the stance of monetary policy, since the rate moves with quantitative easing events. The shadow series in Krippner's 2012 paper is very volatile, but the series he publishes on his website (which I will use for comparison) is smoother. Documentation on Krippner's website states that he used daily yield curve data with 0.25, 0.5, 1, 2, 3, 5, 10, and 30-year maturities. Krippner acknowledges that shadow rates estimated from SRTSMs are very sensitive to data (e.g. maturities), model specification, and estimation method, but asserts that Wu and Xia's three-factor shadow rate is less robust to specifications than his own two-factor shadow rate.

Some researchers, including Lombardi and Zhu (2014), Doh and Choi (2016), and Johannsen and Mertens (2016), avoid imposing Black's no-arbitrage method, and instead use time series models to estimate shadow rates. Time series for these three shadow rates are publicly available for comparison.

Lombardi and Zhu approach the shadow rate problem differently than Black. As opposed to a complicated structural model, they estimate an indicator to summarize the monetary policy stance that is intuitive and easy to compute in real time. First, they gather a pool of data that represents the Federal Reserve's monetary policy— interest rates (e.g. effective fed funds rate, treasury bills and bonds with 1-month to 30-year maturity, libor spread), monetary aggregates, and Federal Reserve balance sheets (assets and liabilities). Then, they estimate a dynamic factor model from 1970 to 2008 and choose their model specification (i.e. number of factors, number of lags). From December 2008 onward, when the fed funds rate hit the zero lower bound, Lombardi and Zhu treat maturities with sufficiently low rates as "missing" and let the dynamic factor model fill in the data using historical relationships between the rates and all other monetary policy variables. Then, the model-implied fed funds rate is Lombardi and Zhu's shadow rate. The shadow rate is not constrained to be equal to the effective fed funds rate above the zero lower bound and is not constrained to be negative during the zero lower bound, but generally tracks the observed fed funds rate before December 2008 and stays mostly negative during the zero lower bound period.

Doh and Choi observe that Wu and Xia (2014) and Lombardi and Zhu (2014) neglect private borrowing conditions, which would have been affected by the Federal Reserve's mortgage-backed security purchases. The methodology for Doh and Choi is straightforward. First, they take three principal components from interest rate levels (medium and long-term maturities only) and spreads (private borrowing conditions). Because the shortest maturity is 2 years, the data will not be censored during the zero lower bound period. Second, they regress the effective fed funds rate on these three components from 1976 to 2008. Third, they map the effective fed funds rate after the zero lower bound using the principal components.

Finally, Johannsen and Mertens use joint autoregressive process between inflation, unemployment gap, medium term rate (5-year yield), and shadow short rate (effective fed funds when it's above the zero lower bound), estimated with Bayesian techniques using quarterly data from 1960 to 2015. Each variable has a trend and cycle component, and the medium-term rate trend is equal to the short-term rate trend plus a constant term premium. During the zero lower bound period, the short rate is treated as missing data and projected with the estimated reaction function, only keeping draws from the posterior distribution if the shadow rate was below 25 basis points. Johannsen and

Mertens estimate the model with and without medium term yield. When the researchers exclude medium-term yields in their specification, they find the shadow rate's distribution is wider and lower. Also, they find the shadow short rate would rise above the zero lower bound sooner, when medium term rates are excluded.

From December 2008 to November 2015, the dynamics of the nominal fed funds rate are effectively censored by the zero lower bound. Many researchers interpolate the fed funds rate series by selecting macroeconomic and financial variables that are uncensored by the zero lower bound, and then mapping the theoretical and unbounded fed funds rate using historical relationships. However, no researchers use the same probit methodology as my model during the zero lower bound period.

1.3 METHODOLOGY

In the methodology section, I present the dynamic ordered probit model, define the data set, and show the estimation procedure.

1.3.1 *The model*

I use a dynamic ordered probit model to estimate the unobserved desired fed funds rate. First, I motivate and specify the probit equation. Then I show the probit equation in the context of the broader model.

The fed funds target is set discretely, such that one usually observes increases or decreases in 25 basis points increments (Table 1.1). However, I am interested in the unobserved latent variable that underlies the observed target, which can be interpreted as the desired fed funds target, and can smoothly evolve based on the strength of the economy or inflation outlook. If the desired fed funds target is sufficiently different from that last set (observable) fed funds target, then this should lead to an observable change in the fed funds target. One can infer the evolution of the latent fed funds target from the observable fed funds target changes using a probit framework.

Table 1.1: History of monthly fed fund target changes over January 1990 to March 2016.

Fed funds target changes		
Month change	**Count**	**Frequency**
[+0.5, +∞)	5	1.6%
(+0.25, +0.5)	0	0.0%
+0.25	27	8.6%
(0, +0.25)	0	0.0%
0	240	76.4%
(-0.25, 0)	0	0.0%
-0.25	25	8.0%
(-0.5, -0.25)	0	0.0%
(-∞, -0.5]	17	5.4%

Let X_t represent observable variables (i.e. indicators of economic health) and y_t^* represent the unobservable desired fed funds target. The latent fed funds target is a function of p lags of observable variables and itself.

$$y_t^* = f_1\left(y_{t-1}^*, \dots, y_{t-p}^*\right) + f_2\left(X_{t-i}, \dots, X_{t-p}\right) + \epsilon_t$$

Let y_t denote the observed fed funds target. Changes in the observable fed funds target indicate that the latent variable has moved sufficiently away from the last set target to warrant a rate change. I assume that the observed and unobserved rates evolve in the way shown in Table 1.2.

Table 1.2: Before the zero lower bound, actual changes are assumed to relate to desired rates in the following way.

If observe:	**Then implies:**	**Classify as category:**
$-\infty < (\boldsymbol{y_t - y_{t-1}}) \leq \boldsymbol{-0.5}$	$-\infty < (y_t^* - y_{t-1}) \leq -0.5$	Category 1 (Large Decrease)
$(\boldsymbol{y_t - y_{t-1}}) = \boldsymbol{-0.25}$	$-0.5 < (y_t^* - y_{t-1}) \leq -0.25$	Category 2 (Small Decrease)
$(\boldsymbol{y_t - y_{t-1}}) = \boldsymbol{0}$	$-0.25 < (y_t^* - y_{t-1}) < 0.25$	Category 3 (No Change)
$(\boldsymbol{y_t - y_{t-1}}) = \boldsymbol{0.25}$	$0.25 \leq (y_t^* - y_{t-1}) < 0.5$	Category 4 (Small Increase)
$\boldsymbol{0.5} \leq (\boldsymbol{y_t - y_{t-1}}) < \infty$	$0.5 \leq (y_t^* - y_{t-1}) < \infty$	Category 5 (Large Increase)

The mapping of observed behavior to unobserved behavior is standard in the ordered probit literature— just with different specification and number of categories. For example, Dueker (1999) assumes observed fed funds rates moved discretely in seven categories, but Hu and Phillips (2004) and van den Hauwe, Paap, and Dijk (2013) only use three ordered categories. Again, they use this methodology before December 2008, before the onset of the zero lower bound.

During the period of zero lower bound, one cannot observe further interest rate decreases. During this period, I assume observed and unobserved rates evolve in the way shown in Table 1.3.

Table 1.3: After the zero lower bound, actual (observed) changes are assumed to relate to desired rates in the following way.

If observe:	Then implies:	Classify as category:
$\boldsymbol{(y_t - y_{t-1}) = 0}$	$-\infty < (y_t^* - y_{t-1}) < 0.25$	Category 3 (No Change)
$\boldsymbol{(y_t - y_{t-1}) = 0.25}$	$0.25 \le (y_t^* - y_{t-1}) < 0.5$	Category 4 (Small Increase)
$\boldsymbol{0.5 \le (y_t - y_{t-1}) < \infty}$	$0.5 \le (y_t^* - y_{t-1}) < \infty$	Category 5 (Large Increase)

The behavior of the desired and observed fed funds rate evolves in a system with other macroeconomic variables. I use a vector autoregression (VAR) with *p=6* lags and *k* variables with normally distributed error.

$$\Phi(L)Y_t = \mu + \epsilon_t$$

$$\underbrace{Y_t = \mu + \sum_{i=1}^{p} \Phi_i Y_{t-i} + \epsilon_t}_{(k\ x\ 1) = (k\ x\ 1) \quad (k\ x\ k)(k\ x\ 1) \quad (k\ x\ 1)}$$

$$\underbrace{\epsilon_t \sim N(0, Q)}_{(k\ x\ 1)\ (k\ x\ 1),(k\ x\ k)}$$

Y_t is made of observable variables (X_t) and an unobservable variable (y_t^*), the desired fed funds target.

$$Y_t = \begin{bmatrix} X_t \\ y_t^* \end{bmatrix} = \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{k-1,t} \\ y_t^* \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{k-1} \\ \mu_k \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} \phi_{i,1,1} & \phi_{i,1,2} & \cdots & \phi_{i,1,k} \\ \phi_{i,2,1} & \ddots & \ddots & \phi_{i,2,k} \\ \vdots & \ddots & \ddots & \vdots \\ \phi_{i,k,1} & \phi_{i,k,2} & \cdots & \phi_{i,k,k} \end{bmatrix} \begin{bmatrix} X_{1,t-i} \\ \vdots \\ X_{k-1,t-i} \\ y_{t-i}^* \end{bmatrix} + \begin{bmatrix} \epsilon_{t,1} \\ \vdots \\ \epsilon_{t,k-1} \\ \epsilon_{t,k} \end{bmatrix}$$

Note that the last equation in the VAR is the probit equation:

$$y_t^* = \mu_1 + \sum_{i=1}^{p} \phi_{i,k,k}\, y_{t-i}^* + \sum_{i=1}^{p} [\phi_{i,k,1} \quad \cdots \quad \phi_{i,k,k-1}] X_{t-i} + \epsilon_{t,k}$$

Therefore, the desired fed funds rate is determined by past desired rates and observable lagged measures of economic growth, inflation, and labor market strength.

1.3.2 *Data*

My data set spans January 1984 to March 2016. When classifying ordered probit categories, I use the federal funds target rate from the Federal Reserve Bank of St. Louis Economic Data website (FRED).

I choose the observable variables with the goal of a parsimonious model that covers the Federal Reserve's mandates of economic health, maximum employment, and stable prices. I use two specifications. The first specification is for the base model.

- **Business cycle index (BCI):** The BCI captures economic growth and is estimated monthly in a separate probit model, using the qualitative VAR methodology from Dueker (2005), with real-time data from FRED. The index is restricted to be above zero during US economic expansion and below zero during economic recession, where recession dates are set by the National Bureau of Economic Research. The magnitude of the BCI is determined by macroeconomic variables such as nonfarm payroll employment growth, real personal consumption expenditure growth, core consumer price index (CPI) inflation, 10-year treasury yield, slope of the yield curve, libor and treasury spread, and corporate bond spreads from January 1967 onwards.
- **Unemployment**: Headline unemployment rate, from FRED, is a proxy for labor market strength.
- **Core inflation**: Monthly percent change in CPI, excluding food and energy, from FRED, is a proxy for price stability.

The second specification is for the scenario model.

- **Real gross domestic product (GDP)**: Quarterly frequency real GDP is from FRED, but I interpolate the data to make a monthly frequency series.
- **Unemployment**: Headline unemployment rate, from FRED, as in the first specification.
- **Core inflation**: Monthly percent change in personal consumption expenditure (PCE) price index, excluding food and energy, from FRED.

After estimating the fed funds target model using the two specifications above, I can produce conditional forecasts— forecasts of the desired fed funds rate, given assumed paths for economic health, employment, and inflation. For the first specification, I use real-time vintage forecasts of the BCI and core CPI inflation from the BCI model— the BCI model is a VAR and naturally generates forecast paths for the business cycle index and all other variables in the system. The unemployment rate path is from vintage issues of Blue Chip Economic Indicators, which surveys top economists monthly.

While it is straightforward to backtest the first specification using vintage as-was data, the second specification is more appropriate for scenario analysis— many economists produce GDP forecasts, but not many produce BCI forecasts. I choose real GDP, headline unemployment, and core PCE inflation for the scenario model because the Federal Reserve releases economic projections of these series four times a year when there is a post-meeting press conference. I can estimate the fed funds target model, produce forecasts conditional on the Federal Reserve's central, dovish, and hawkish projections, and check my model forecasts against the Fed's forward guidance.

1.3.3 *Estimation overview*

The VAR parameters (coefficients $\theta = \{\mu, \Phi_1, \dots, \Phi_p\}$), the error term's variance covariance matrix (Q), and the path of the latent variable ($y^* = \{y_t^* \; for \; t = 1, \dots, T\}$) are estimated via Gibbs sampling. My methodology is as follows:

1. Initial values of $y^*$are generated by adding random shocks to the observed fed funds target y_t.
2. Using the initial path of y^* in the VAR, I estimate the VAR coefficients θ and variance covariance matrix Q.
3. Using the VAR coefficients, I estimate the path of the latent variable y^* with a Kalman filter, where the latent variable is drawn from a truncated normal distribution. The bounds of the truncation are determined by the probit equation rules in Table 1.2 and Table 1.3. For example, if there is an observed 25 basis points increase in the target fed funds rate at time *t*, then the desired fed funds rate y_t^* drawn is such that: $0.25 \leq (y_t^* - y_{t-1}) < 0.5$. During the zero lower bound, if we observe no rate change, then the desired fed funds rate y_t^* drawn is such that: $-\infty < (y_t^* - y_{t-1}) < 0.25$.
4. I use the estimated path of the latent variable y^* to estimate the VAR parameters. After estimating VAR coefficients and covariance matrix via regression ($\widehat{\Phi}$, $\widehat{\mathrm{Q}}$), I randomize the parameters. The VAR coefficients (Φ^{new}) are drawn from a normal distribution and the covariance matrix ($\mathrm{Q}^{\mathrm{new}}$) is drawn from an inverted Wishart distribution.

$$\Phi^{\mathrm{new}} \sim \mathrm{N}\left(\widehat{\Phi}, \mathrm{var}\left(\widehat{\Phi} | \, Y\right)\right)$$

$$\mathrm{Q}^{\mathrm{new}} \sim W_{inverted}\left(T - k, \widehat{\mathrm{Q}}\right)$$

5. I repeat steps 3-4 for 3000 iterations. I throw out the first 1000 simulations of θ and y^* as the burn-in period. Then, I take the average of the last 2000 simulation of θ and y^* to get the final estimated parameters.
6. Forecasting:
    a. After the burn-in period, I forecast dynamically, one period ahead at a time until the desired forecast horizon. Because of the VAR structure, one can unconditionally forecast by generating shocks from the error term’s variance covariance matrix (Q). By taking the average of the last 2000 simulations of the desired fed funds forecast path, I obtain a final forecast.
    b. Another method is to conditionally forecast on the path of some explanatory variables. For example, one can condition the forecast of the desired fed funds rate

on expectations of unemployment, inflation, and business cycle by setting these future paths as given, and then generate shocks to the latent variable. This method allows the user to see the implied monetary policy for a given economic outlook. The conditional forecast methodology is from Dueker (2005).

1.3.4 *Estimation: Kalman filter for truncated normal variables*

After estimating coefficients and covariance matrix (Φ, Q) via regression, the latent time series is drawn from its conditional posterior using a Kalman filter for truncated normal variables, using the methodology from Dueker (2006). Suppressing the constants, the state space and measurement equations can be written as below.

State space equation:

$$S_t = F\,S_{t-1} + e_t$$

$$\underbrace{\begin{bmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{(kp\ x\ 1)\ =\ (kp\ x\ kp)\ (kp\ x\ 1)\ \ (kp\ x\ 1)}$$

Measurement equation:

$$X_t = H\,S_t$$

$$\underbrace{X_t = \begin{bmatrix} I_{k-1} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \end{bmatrix}}_{((k-1)\ x\ 1)\ =\ ((k-1)x\ kp)\ \ (kp\ x\ 1)}$$

Under the standard Kalman filter methodology, the prediction and update equations are as follows:

Standard prediction equations:

$$S_{t|t-1} = F\,S_{t-1|t-1}$$

$$P_{t|t-1} = F\,P_{t-1|t-1}F' + Q^S$$

$$Q^S = \begin{bmatrix} Q & \dots & 0 \\ \vdots & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Standard update equations:

$$K_t = P_{t|t-1}H'\left(H\, P_{t|t-1}H'\right)^{-1}$$

$$S_{t|t} = S_{t|t-1} + K_t\left(X_t - H\, S_{t|t-1}\right)$$

$$P_{t|t} = P_{t|t-1} - K_t\, H\, P_{t|t-1}$$

However, when I rewrite the Kalman filter for truncated normal variables, there are some modifications. Let the location of a and v correspond to the position of the latent variable (i.e. the kth row of the state space equation). Then, conditional on observable data at time t, the distribution of the error term is as follows:

$$\epsilon_t | Data_t \sim N\left(W \begin{bmatrix} 0 \\ \vdots \\ a \\ 0 \\ \vdots \\ 0 \end{bmatrix}, W \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & v & 0 \\ 0 & \cdots & 0 & I_{kp-k} \end{bmatrix} W' \right) \stackrel{\text{def}}{=} N(\widetilde{m}, \tilde{Q})$$

$$W \stackrel{\text{def}}{=} chol(Q^S)$$

If LB and UB denote the category rules for $y_t^* - y_{t-1}$(i.e. 25bp and 50bp), then I define a and v as:

$$a \stackrel{\text{def}}{=} -\frac{\left(PDF(\alpha_2) - PDF(\alpha_1)\right)}{\left(CDF(\alpha_2) - CDF(\alpha_1)\right)}$$

$$v \stackrel{\text{def}}{=} 1 - \frac{\left(\alpha_2 * PDF(\alpha_2) - \alpha_1 * PDF(\alpha_1)\right)}{\left(CDF(\alpha_2) - CDF(\alpha_1)\right)} - \frac{\left(PDF(\alpha_2) - PDF(\alpha_1)\right)^2}{\left(CDF(\alpha_2) - CDF(\alpha_1)\right)^2}$$

$$\alpha_1 \stackrel{\text{def}}{=} \left(LB - \left(S_{t|t,kk} - y_{t-1}\right)\right), \qquad \alpha_2 \stackrel{\text{def}}{=} \left(UB - \left(S_{t|t,kk} - y_{t-1}\right)\right)$$

Given these new definitions, the following are the modified Kalman filter equations:

Prediction equations:

$$S_{t|t-1} = F\, S_{t-1|t-1}$$

$$P_{t|t-1} = F\, P_{t-1|t-1}F' + \tilde{Q}$$

Update equations:

$$K_t = P_{t|t-1}H'\left(H\, P_{t|t-1}H'\right)^{-1}$$

$$S_{t|t} = S_{t|t-1} + K_t\left(X_t - H\, S_{t|t-1} - H\widetilde{m}\right)$$

$$P_{t|t} = P_{t|t-1} - K_t\, H\, P_{t|t-1}$$

1.4 RESULTS

1.4.1 *Base model output*

Table 1.4 shows the probit model parameters of the desired fed funds rate as a function of lagged BCI, core inflation, unemployment rate, and desired fed funds rate. On average, a higher BCI implies higher desired rates. Higher inflation implies higher desired rates, while higher unemployment implies lower desired rates. I also observe persistence between current and past desired rates.

Table 1.4: Parameter estimates for the desired fed funds rate equation, estimated as of March 2016, under the base model with BCI, CPI inflation, and headline unemployment.

Parameter		Mean	SD	5th percentile	95th percentile
t-1	BCI	0.26	0.19	-0.06	0.57
	inflation	0.03	0.03	-0.02	0.09
	unemployment	-0.45	0.34	-1.03	0.08
	Desired FFT	0.43	0.09	0.29	0.59
t-2	BCI	0.09	0.29	-0.38	0.60
	inflation	0.04	0.03	-0.01	0.09
	unemployment	-0.18	0.40	-0.82	0.51
	Desired FFT	0.23	0.11	0.04	0.41
t-3	BCI	-0.06	0.30	-0.53	0.45
	inflation	0.02	0.03	-0.04	0.07
	unemployment	0.14	0.41	-0.54	0.81
	Desired FFT	0.16	0.11	-0.01	0.35
t-4	BCI	-0.02	0.29	-0.51	0.45
	inflation	0.01	0.03	-0.04	0.06
	unemployment	0.33	0.40	-0.32	1.00
	Desired FFT	0.07	0.11	-0.11	0.25
t-5	BCI	0.11	0.30	-0.38	0.60
	inflation	0.04	0.03	-0.02	0.09
	unemployment	-0.13	0.39	-0.75	0.52
	Desired FFT	0.03	0.10	-0.14	0.20
t-6	BCI	-0.09	0.20	-0.41	0.26
	inflation	0.05	0.03	-0.01	0.10
	unemployment	0.12	0.32	-0.40	0.65
	Desired FFT	-0.02	0.10	-0.18	0.15
	Constant	0.48	0.20	0.15	0.81

Figure 1.1 shows the desired fed funds target estimated from the BCI, unemployment rate, and core inflation. Given the severity of the Great Recession, and if nominal rates could be negative, the desired monetary policy was to lower the fed funds target to almost -8%, before slowly

tightening conditions again over a period of a few years. The BCI fell to approximately 0.5 standard deviations below zero in the beginning of 2008, then plummeted to 3 standard deviations below zero the end of 2008, indicating a deep recession. The unemployment rate rose to 10% near the end of 2009 and remained elevated for many years. The steep drop in the desired fed funds rate was also driven by the steep drop in the observed fed funds rate before the onset of the zero lower bound, since the desired fed funds rate is a function of lagged desired rates.

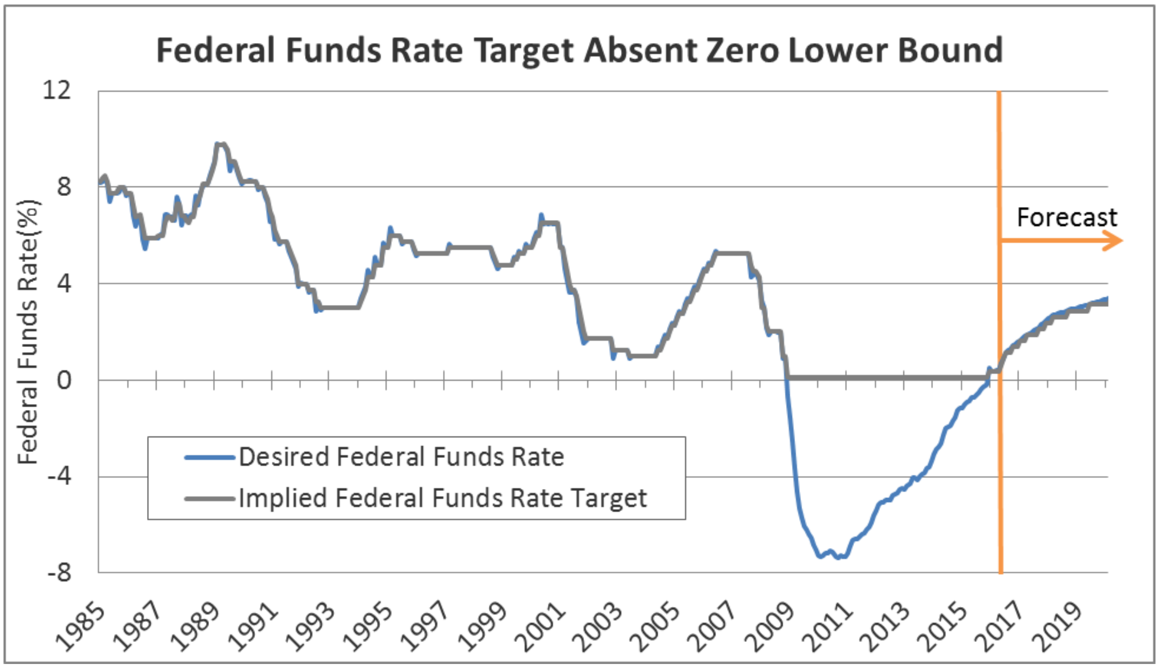

Figure 1.1: Estimated desired fed funds rate as of March 2016, under the base model with BCI, CPI inflation, and headline unemployment.

Figure 1.2 shows the model's predicted desired fed funds rate compared to Blue Chip's consensus forecast for the effective fed funds rate and the Federal Reserve's projections of the fed funds target. As of March 2016, the model's projected path of the desired fed funds target is quite aggressive at four 25 basis point rate hikes in 2016, four hikes in 2017, and two hikes in 2018. Notably, the tightening of monetary policy predicted by the fed fund target model is front loaded—the policy tightens aggressively in the first two years, then gradually thereafter. This contrasts with the Fed's March 2016 economic projections; their dot plot suggested the median FOMC member expects two hikes in 2016, four hikes in 2017, and almost 4.5 hikes in 2018.

Essentially, the Fed's projections show a back-loaded tightening path— monetary policy would tighten gradually in the beginning and aggressively later. However, my model suggests that given the economic, inflation, and labor backdrop, the Fed should want to raise rates more aggressively than the market is predicting or the Fed is communicating. However, with the benefit of hindsight and an extra two years of data, the realized path of the fed funds rate was slow and back-loaded (closer to the Fed's communication than my model). By the Fed's December 2017 meeting, the Fed had raised rates once in 2016 and three times in 2017, while predicting three hikes in 2018. A possible explanation is that the model neglects the asymmetric risk around the zero lower bound. The model's desired fed funds rate can move seamlessly below and above the zero lower bound, whereas the Fed's nominal rate cannot. By tightening slower than normal, the Fed is conservative about avoiding a policy mistake.

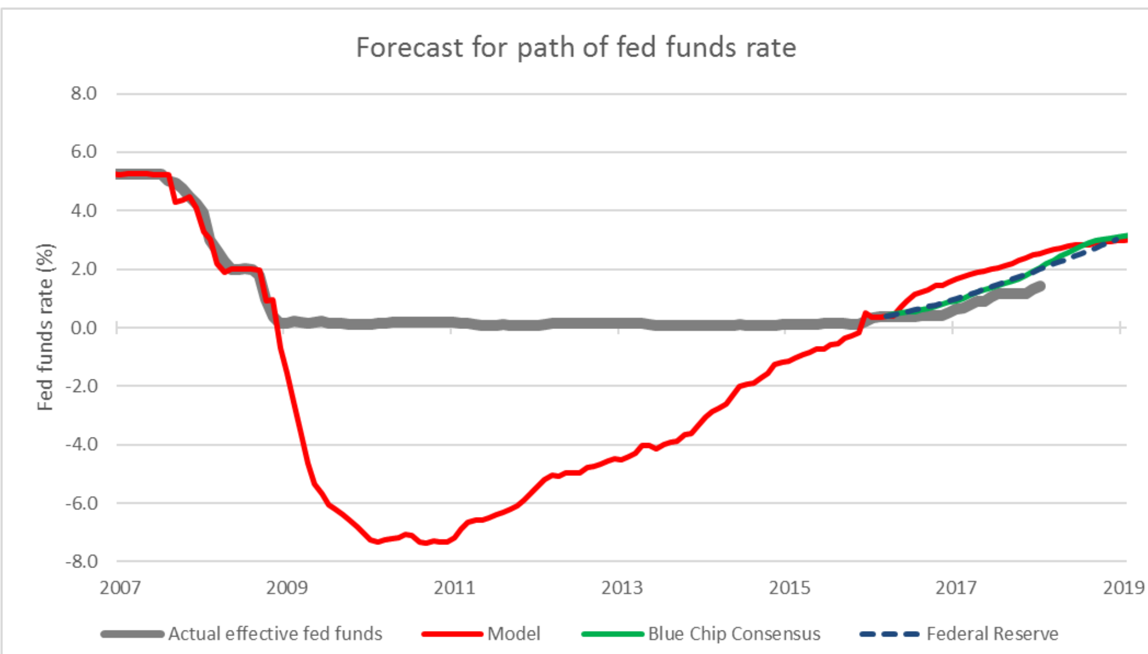

Figure 1.2: Fed funds forecast from the base model, Blue Chip consensus, and Federal Reserve, as of March 2016. These forecasts are plotted with the actual fed funds rate as of January 2018.

During the period when the desired rate was below the zero lower bound, the probit model forecasts were more stable compared against vintage backtests during the same period. Figure 1.3 shows how the forecasted first rate hike date has evolved across vintage backtests for the model versus consensus from 2009 to 2015. The horizontal axis shows the date of the real time backtest. The vertical axis shows the predicted date of the first fed funds rate hike, where the orange line is the model and the blue line are consensus forecasts from Blue Chip Economic Indicators. The Blue Chip predictions consistently underestimate how long the fed funds target would remain at the zero lower bound. For example, around the middle of 2010 economists expected a rate increase in the end of 2010, and in the middle of 2011 they expected an increase in the end of 2011. In contrast, the model forecasts were more stable.

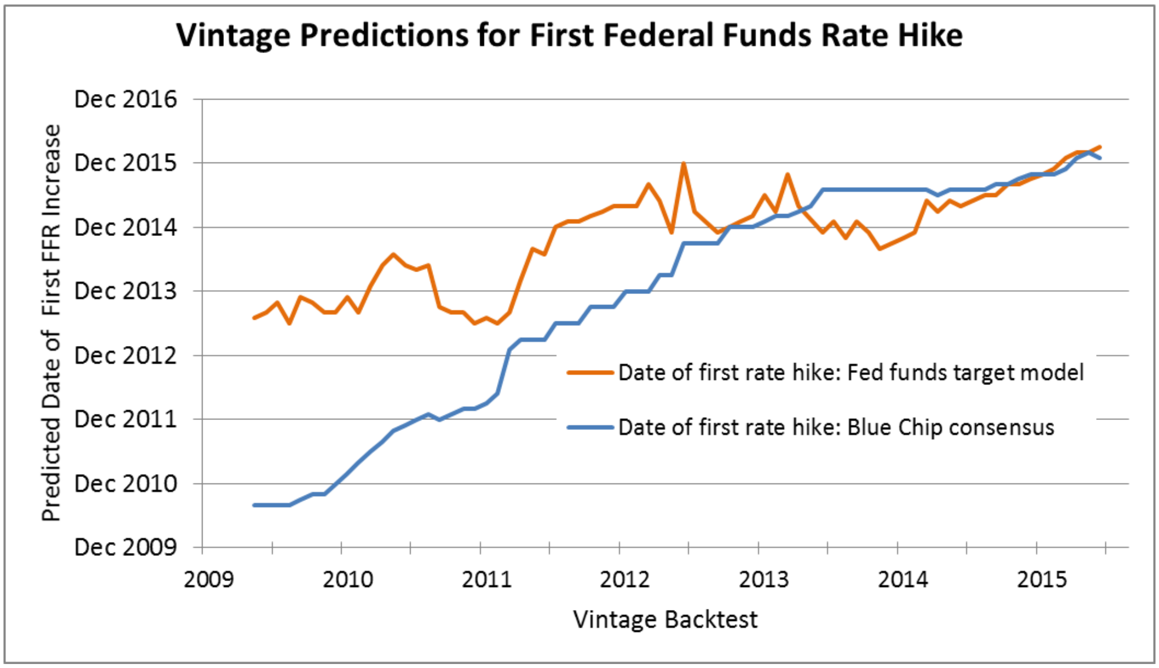

Figure 1.3: Real-time predictions of the base model versus Blue Chip consensus.

1.4.2 *Scenario model output*

Table 1.5 shows the probit model parameters of the desired fed funds rate as a function of lagged real GDP growth, core PCE inflation, unemployment rate, and desired fed funds rate. On average, a higher GDP implies higher desired rates, higher inflation implies higher desired rates, and higher unemployment implies lower desired rates. Like the base model, there is persistence between current and past desired rates.

Table 1.5: Parameter estimates for the desired fed funds rate equation, estimated as of March 2016, under the scenario model with real GDP, core PCE inflation, and headline unemployment.

Parameter		Mean	SD	5th percentile	95th percentile
t-1	GDP	0.15	0.08	0.02	0.27
	inflation	0.06	0.03	0.01	0.11
	unemployment	-0.55	0.35	-1.11	0.01
	Desired FFT	0.41	0.09	0.25	0.57
t-2	GDP	-0.14	0.15	-0.40	0.12
	inflation	0.03	0.03	-0.01	0.09
	unemployment	-0.15	0.43	-0.85	0.52
	Desired FFT	0.23	0.11	0.04	0.41
t-3	GDP	0.07	0.14	-0.17	0.30
	inflation	0.02	0.03	-0.02	0.07
	unemployment	0.04	0.43	-0.67	0.75
	Desired FFT	0.17	0.11	-0.01	0.35
t-4	GDP	0.11	0.13	-0.11	0.33
	inflation	0.03	0.03	-0.02	0.07
	unemployment	0.35	0.42	-0.34	1.06
	Desired FFT	0.09	0.11	-0.09	0.27
t-5	GDP	-0.15	0.15	-0.40	0.10
	inflation	0.04	0.03	-0.01	0.09
	unemployment	-0.24	0.41	-0.90	0.44
	Desired FFT	0.03	0.11	-0.14	0.21
t-6	GDP	0.04	0.08	-0.10	0.17
	inflation	0.03	0.03	-0.02	0.08
	unemployment	0.37	0.32	-0.16	0.90
	Desired FFT	-0.01	0.10	-0.19	0.15
	Constant	0.58	0.22	0.23	0.94

I use the model to conduct scenario analysis. First, I forecast the path of the desired fed funds rate using different assumed paths for real GDP, core PCE inflation, and headline unemployment. In the central case, I use the Fed's median economic projections. For the dovish scenario, I use the Fed's lowest GDP, lowest inflation, and highest unemployment projection. For the hawkish scenario, I use the Fed's highest GDP, highest inflation, and lowest unemployment projection.

Figure 1.4 shows the range of outcomes when I use the Federal Reserve's March 2016 economic projections. As summarized in Table 1.6, even when using the same assumptions for the future path of growth, inflation, and labor market, the Fed and the model arrive at slightly different conclusions. The model predicts four, four, then three hikes in 2016, 2017, and 2018. In contrast, the Fed projections show two, four, then four hikes in 2016, 2017, and 2018. The model front-loads the tightening process, whereas the Fed back-loads the tightening process. Again, the Fed's conservative policy normalization may reflect the asymmetric risk of policy mistakes when nominal interest rates are already low.

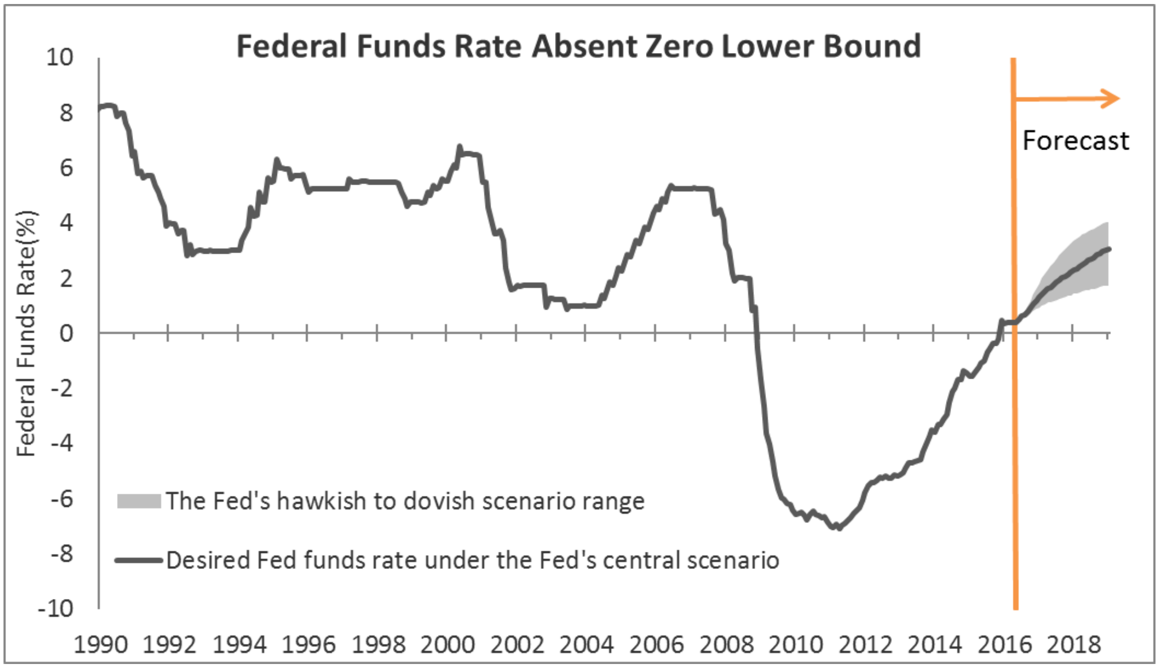

Figure 1.4: Estimated desired fed funds rate as of March 2016, under the scenario model with real GDP, core PCE inflation, and headline unemployment.

Table 1.6: Comparison of the number of hikes predicted by model and Fed as of March 2016.

	Fed		Model	
	Central	Range	Central	Range
2016	2	1 - 4	4	3 - 6
2017	4	4 -6	4	2 - 6
2018	4	2 -4	3	1 - 3

1.5 DISCUSSION: COMPARISON WITH OTHER METHODOLOGIES

1.5.1 *Interpretations of shadow rates*

How does one interpret these estimated shadow rates? After all, we can observe fed funds rates, but shadow rates are unobserved and hypothetical. Also, estimated shadow rates look different depending on the model.

Hakkio and Kahn interpret their rate as an estimated Taylor rule, or a policy prescription based on the Federal reserve’s historical reaction function. It is the recommended fed funds rate, if the rate could go negative.

Wu and Xia’s shadow rate and Krippner’s shadow rates are priced based on the rest of the yield curve under no arbitrage. Hypothetically, if a cash option did not exist, their shadow rate would represent the short-term yield. Their shadow rate is a theoretical construct, not a practical financial instrument. However, Krippner interprets SRTSM shadow rates as a measure of the stance of monetary policy.

Lombardi and Zhu's shadow rate and Doh and Choi's shadow rate are explicitly calculated to proxy the stance of monetary policy, by incorporating variables the Federal Reserve had control over. For example, the Federal Reserve deployed quantitative easing (QE) to lower interest rates and increase money supply— from late 2008 to late 2014, they bought long maturity and mortgage backed securities and they increased their balance sheet. Money supply, interest rates, balance sheets, borrowing conditions are indicators used when estimating these time series shadow rate. If the effective fed funds rate is an appropriate and complete measure of the easiness of monetary policy before the lower bound, then the shadow rate estimated by these researchers attempts to bridge the zero lower bound information gap. Even though the actual fed funds rate did not contribute to monetary policy during this time period, these researchers construct a hypothetical fed funds rate that reflects the overall stance of monetary policy.

Finally, the shadow rate estimated by Johannsen and Mertens is not an explicit proxy for the stance of monetary policy. They generically state that their shadow rate is the nominal interest rate under no zero lower bound. Instead, they concentrate on its use. By incorporating the lower bound, they intend to get better estimates of trend and cycle for the US economy. They also show that including a lower bound feature improves out-of-sample predictions for the short- and medium-term rate, both before and after the onset of the lower bound.

In the context of the shadow rate literature, my shadow rate is the unobserved desired fed funds rate. It is related to a policy prescription because it incorporates the Federal Reserve's historical reaction function. However, my shadow rate also captures some of the monetary policy stance since there exists feedback between the macroeconomy and my shadow rate. Therefore, I can use my shadow rate to fill out the policy rate during the zero lower bound, forecast the path of fed funds rates, and conduct scenario analysis.

1.5.2 *Comparison*

To compare the time series of my shadow rate with other methodologies, I categorize various publicly available shadow rate series into three groups: policy prescription, measurement of actual policy, and general tool. Within these three categories, I compare my estimated desired fed funds rate with these other shadow rates. I find that my shadow rate matches the shape of policy prescriptions since they have many of the same inputs (i.e. inflation, economic health).

However, the policy prescriptions tend to recommend rate hikes earlier than my fed funds model, and earlier than when the Fed raised rates in reality. Comparison of my shadow rate with other policy proxy shadow rates is not straight-forward, because the policy proxy shadow rates in the literature are not visually similar. However, in general the policy proxy shadow rates were less negative than the shadow rate estimated by my model. They were also generally the most accommodative in the last half of the zero lower bound, whereas my shadow rate bottomed in the first half of the zero lower bound. Visually, Johannsen and Mertens's shadow rate appear to be an equal blend of policy rules and shadow rates that proxy the stance of monetary policy. Conversely, my shadow rate is a blend of Johannsen and Mertens's shadow rate and the policy rules.

1.5.3 *Comparison: desired fed funds rate versus policy prescriptions*

The Hakkio-Kahn rule (HK) and the Taylor rule (TL) are interpreted as policy prescriptions.

There are a few iterations of Taylor's rule, but I use the following specification:

$$r^{TR} = (infl + 2) + 0.5\ (infl - 2) + 0.5\,\frac{\left(rGDP_{level} - rGDP_{level}^{potential}\right)}{rGDP_{level}^{potential}} * 100$$

$infl$: GDP implicit price deflator, in percentage change from a year ago

$rGDP_{level}$: real GDP, in chained dollars

$rGDP_{level}$: Congressional Budget Office's estimate of real potential GDP, in chained dollars

$(infl + 2)$: Assume 2% steady state real interest rate

Hakkio and Kahn have two specifications for their policy rules. One includes the unemployment gap and the second includes the Kansas Fed's labor market condition indices, which are the first two principal components of many labor market indicators. Inflation is calculated from the core PCE price index. I use the coefficients from Hakkio and Kahn's paper and data from FRED, as of February 2018, to replicate their shadow rates. Since Hakkio and Kahn's paper uses data as of April 2014, the replicated HK1 and HK2 series will differ slightly from the original because of data revisions.

$$r_t^{HK1} = \rho(r_{t-1}^{HK1} - 2) + (1 - \rho) * [c +\ \beta(infl_t - 2) + \gamma(ungap_t)]$$
$$r_t^{HK2} = \rho(r_{t-1}^{HK2} - 2) + (1 - \rho) * [c +\ \beta(infl_t - 2) + \gamma_1(LMCI1_t) +\ \gamma_2(LMCI2_t)]$$

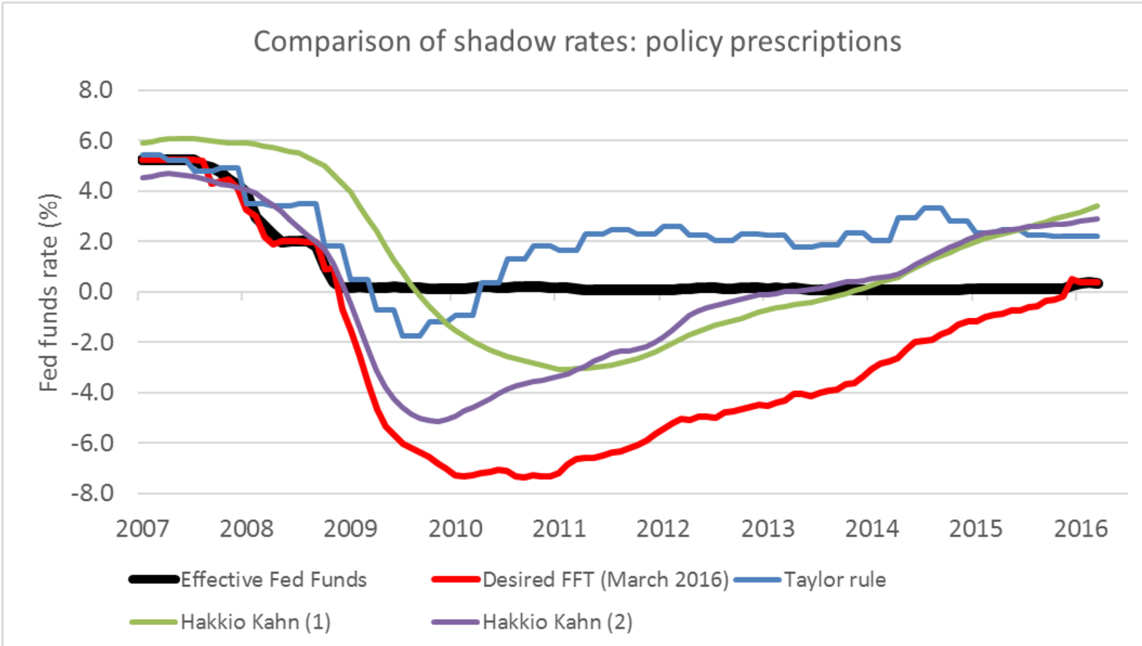

Figure 1.5: Comparison of estimated desired fed funds rate with various policy prescriptions.

Figure 1.5 compares the desired rate from my model and the policy prescriptions from the Taylor rule and Hakkio-Kahn rule. The model's desired fed funds rate has a similar shape as the policy rules from TL and HK. The model, TL, HK1, and HK2 all show the most negative rates in the first half the zero lower bound. Second, the model estimates much lower shadow rates than TL, HK1, or HK2. The minimum prescribed rates by TL, HK1, and HK2 are -1.7%, -3.1%, and 5.1%, respectively. In contrast, the model estimates that lowest desired rate is -7.4%.

The results show the policy rules bottom in September 2009, January 2011, and November 2009 for TL, HK1, and HK2, respectively. Conversely, the feds fund model bottoms in September 2010. The timing of the bottom can be explained by the choice of variable for economic health. For example, the TL uses real GDP output gap, which bottomed between Q2 to Q3 of 2009. This is roughly when the TL recommended the most accommodative policy. The fed funds model uses the BCI and unemployment rate, which bottomed end of 2008 and peaked in end of 2009; the shadow rate fell steeply during this macroeconomic decline but because the model contains inertia from lagged desired rates, the desired rate didn't bottom until months later. HK1 and HK2 also contain policy persistence, so their recommended policy is most accommodative months after their economic indicators bottom.

Finally, TL, HK1, HK2 all recommend the removal of accommodative policy very early. These rules would have recommended the Fed raise interest rates away from the zero lower bound on April 2010, February 2014, and October 2013 respectively. However, the Fed's actual rate hike was on December 2015. The fed funds model's prediction for the first rate hike is more stable (recall the real-time predictions of the model versus Blue Chip consensus in Figure 1.3).

1.5.4 *Comparison: desired fed funds rate versus shadow rates that proxy stance of monetary policy*

The Wu-Xia (WX) and Krippner (KP) shadow rate are calculated using the yield curve and no-arbitrage conditions. The Lombardi-Zhu (LZ) and Doh-Choi (DC) shadow rates are extracted from monetary policy measures that were not truncated during the zero lower bound. All these shadow rates can be interpreted as proxies for the actual stance of monetary policy.

Theoretically, WX and KP should look similar because they both price shadow rates from the shape of the yield curve. However, using slightly different yield curve maturities and model choices yield different looking series. Similar to WX and KP, the LZ shadow rate contains information about the yield curve; however, LZ also includes libor spreads, monetary aggregates, and Federal Reserve balance sheet to capture a wider view of Fed policy actions. The DC shadow rate incorporates the yield curve, but also includes indicators that reflect private-sector borrowing conditions. Differences between these shadow rates can be driven by methodology or variable inputs.

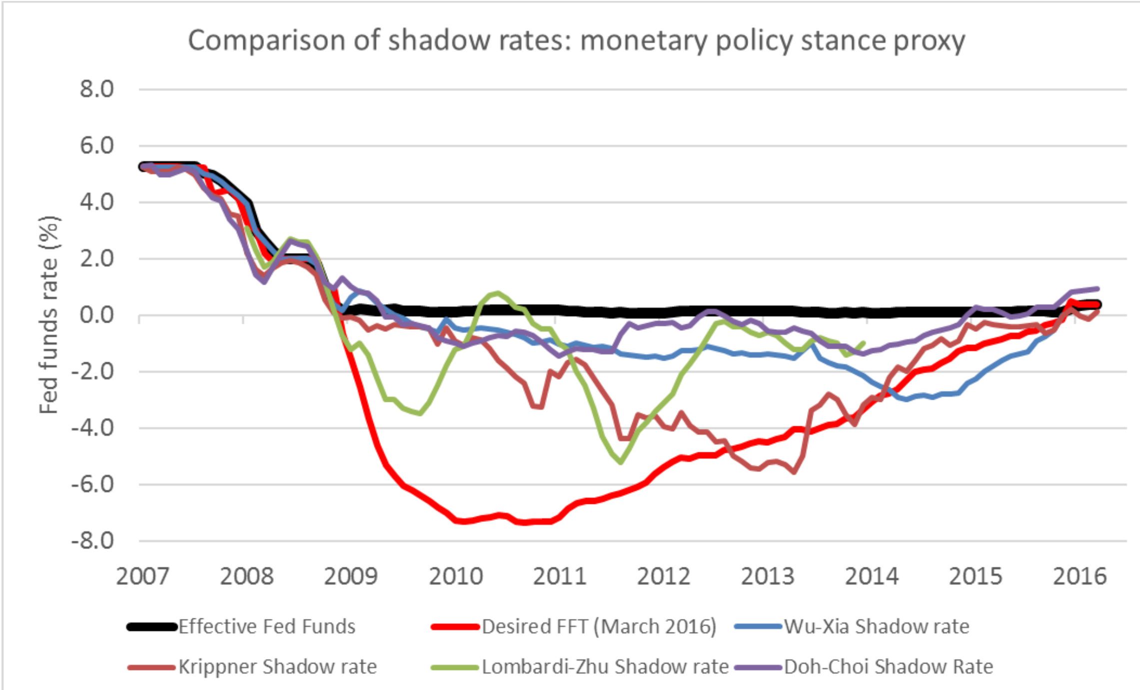

Figure 1.6: Comparison of estimated desired fed funds rate with other shadow rates. This set of alternative shadow rates can be interpreted as a proxy for stance of monetary policy.

Figure 1.6 compares the desired rate versus shadow rates that proxy monetary policy stances. The key takeaway is that even though WX, KP, LZ, and DC are interpreted as proxies for the actual stance of monetary policy, they vary drastically. The shadow rates estimated by WX and KP are both based on the intuition of no-arbitrage pricing of the yield curve, but produce different time series. The KP shadow rate bottomed in April 2013 at -5.6%. From 2013 to 2015, while KP estimated that monetary policy was becoming less accommodative, WX estimated that monetary policy was becoming more accommodative. The WX bottomed in May 2015, when the WX shadow rate was -3%.

DC critique the use of no-arbitrage conditions in pricing shadow rates. They recommend caution when distinguishing decline in longer term yield because of maturity specific factors versus common risk factors. They attribute some of the divergence between the DC and WX rates in mid-2011 to their model's different interpretation of fall in longer term yields.

I also observe that the LZ shadow rate is very volatile. After the onset of the zero lower bound, the LZ shadow rate fell until September 2009, most likely reflecting large scale asset purchases in

November 2008 and long-term treasury purchases in March 2009. The LZ shadow rate started rising, reflecting less accommodative policy from September 2009 to July 2010, when the shadow rate was positive. The shadow rate fell again thereafter, reflecting another round of large scale asset purchases in November 2010.

Furthermore, the LZ and KP shadow rates implied very different stances of monetary policy late-2012 to early-2013. LZ rate implied tighter policy around -0.6%, near a local peak, whereas the KP rate implied loose policy around -5.5%, near a local bottom.

Desired rates estimated from the fed funds target model are generally much lower than the policy proxy shadow rates. Though it is difficult to see a consensus stance of policy, the other shadow rates are the most accommodative at the middle to end of the zero lower bound sample. In contrast, the desired fed funds rate bottoms in the first half of the zero lower bound sample. The dynamics of the desired fed funds rate visually looks more similar to the policy prescriptions than the monetary stance proxies.

1.5.5 *Comparison: desired fed funds rate versus Johannsen and Mertens's shadow rate*

Finally, Johannsen and Mertens's shadow rate (JM) has no special interpretation other than the hypothetical federal funds rate in the absence of the zero lower bound. They model short rates in the context of a larger joint system. Visually, an average of the all the policy rules and policy proxy shadow rates produces a series that looks like the JM shadow rate. Therefore, it is reasonable to suggest that the JM shadow rate is a blend of policy prescription and policy measurement.

The desired fed funds rate is much lower than the JM shadow rate (Figure 1.7)— the desired rate hit a bottom of -7.4% in September 2010, but the JM shadow rate stayed around -2% through most of the lower bound sample, without much movement. However, the shape of the desired rate and JM shadow rate are similar— both show negative and falling shadow rates between end-2008 to 2010, and gradually rising shadow rates after 2013. The sole difference between the desired fed funds rate and JM is that the desired fed funds rate has a rounded dip, as opposed to the JM shadow rate's plateau.

Both models have a censoring methodology where in-sample draws of the shadow rate are rejected if it rises above the zero lower bound before observed fed funds rate rise. Both models include measures of unemployment and inflation, but the JM model also includes medium term interest rates. JM also produce a shadow rate without medium term interest rates and note the resulting shadow rates became much more negative with a steeper rise— this seems consistent with the behavior of the desired fed funds rate. Introducing medium term interest rates into the fed funds model may produce an estimated desired rate that is less negative and rises more gradually.

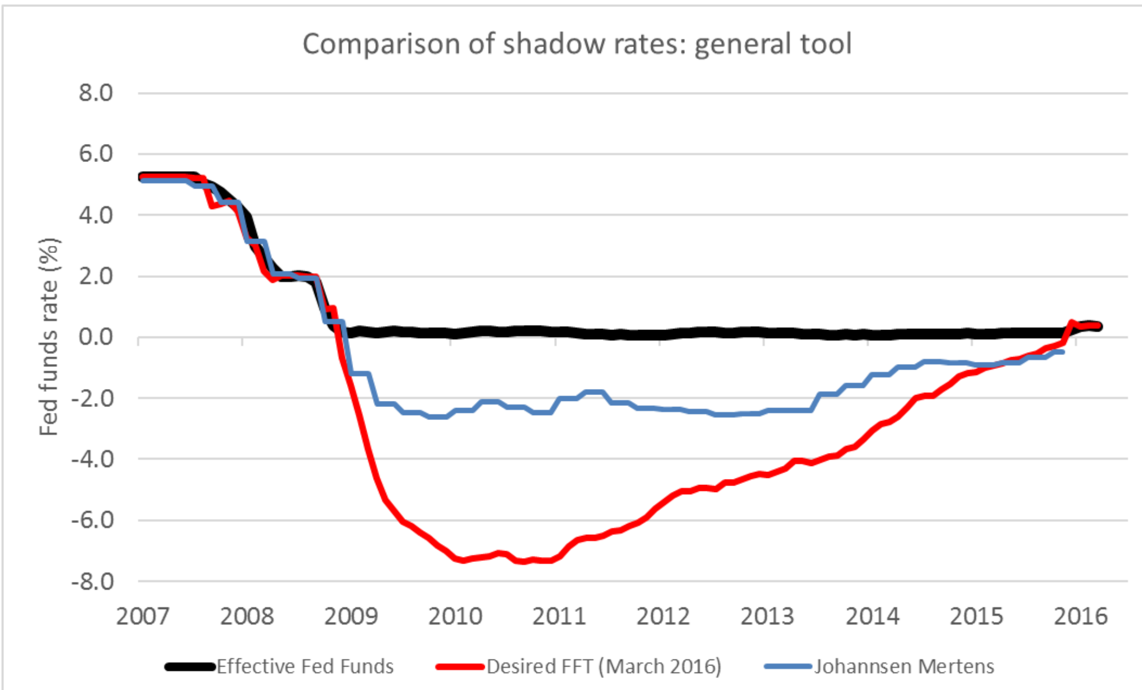

Figure 1.7: Comparison of estimated desired fed funds rate with Johannsen and Mertens's shadow rate.

1.6 CONCLUSION

This paper estimates the desired fed funds rate in the absence of a zero lower bound by using an ordered probit model.

Estimation of the in-sample shadow fed funds rate is important for filling in the policy rate time series during the zero lower bound period, so that it can be used with other macroeconomic models. I compare the ordered probit model's shadow rate with other shadow rates in the existing literature. Policy prescription rules from Taylor (1993) and Hakkio and Kahn (2014) have similar dynamics as my shadow rate, but rise from the zero lower bound much too early. Shadow rates that proxy the stance of monetary policy, such as shadow rates from Wu and Xia (2014), Krippner (2012), Lombardi and Zhu (2014), Doh and Choi (2016), and Johannsen and Mertens (2016) look different from each other because of differences in inputs or methodology. Because the shadow rate is unobservable, one cannot state definitively which shadow rate is the truth.

The model I present here has a good track record predicting the date of the first Federal Reserve rate hike. While analysts surveyed by Blue Chip Economic Indicators consistently underestimated how long the fed funds rate would remain below the zero lower bound, my model forecasts were more stable. Using policy prescriptions like Taylor rules would have underestimated how long the fed funds rate would remain below the zero lower bound as well.

However, after the fed funds rate rose above 25 basis points, the model forecasts were too aggressive. As of March 2016, the model estimated that the Fed, if they behaved similarly as the past, would have front-loaded their policy normalization— that is, the Fed would raise interest rates aggressively first, then gradually later. However, the Fed back-loaded their normalization with gradual rates hikes first, then more aggressive rate hikes later.

The Fed may have back-loaded their normalization because of asymmetric risk around the zero lower bound— if the Fed makes a policy mistake by tightening too quickly, there is less leeway for cutting rates to stimulate the economy when rates are already near zero. In contrast, the probit model assumes no rate restriction at zero, which is a deliberate feature of the model. Under normal circumstances, the Fed may have front-loaded their policy. In this paper, I show the fed funds target model outlined in this paper can be a valuable tool for forecasting monetary policy, and compare the model's estimated desired fed funds rates against other shadow rates in the literature.

Chapter 2. DYNAMIC WEIGHTING OF MOMENTUM, CARRY, AND VALUE CURRENCY STRATEGIES

2.1 Introduction

Carry, momentum, and value are common and profitable currency trading strategies. The carry trade buys currencies with high interest rates and sells currencies with low interest rates, the momentum trade buys currencies with recent strong returns and sells currencies with recent weak returns, and the value trade buys cheap currencies and sells expensive currencies.

"Real world" applications of these strategies include the Russell Conscious Currency Index (RCCI) and the Deutsche Bank Currency Returns (DBCR) index. Although their methodology for calculating momentum, carry and value differ slightly, both firms sort these G10 currencies[1] by a factor, then systematically buy the three top-ranked currencies and sell the three bottom-ranked currencies. Both firms also produce an aggregate index that equally weights their momentum index, carry index, and value index for currencies. Most individuals do not have the relationship with banks and enough pricing power to implement these currency factor positions in a cost-effective way, but a firm with proper infrastructure could. Using these indices, a firm could replicate the long and short currency positions as an overlay on top of a portfolio. The overlay could be an additional source of return or a risk-mitigating currency hedging strategy.

[1] The G10 currency universe includes: Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), euro (EUR), British pound (GBP), Japanese yen (JPY), Norwegian krone (NOK), New Zealand dollar (NZD), Swedish krona (SEK), and US dollar (USD).

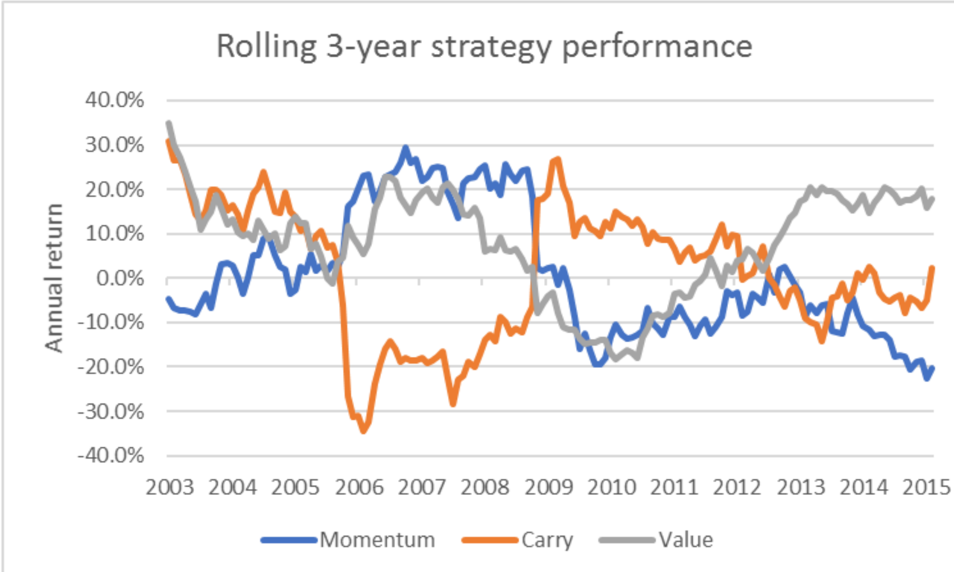

Figure 2.1**:** Rolling forward 3-year annualized returns from a momentum, carry, and value strategy in a G10 currency portfolio.

The historical relative performance of these three strategies varied through time, such that a constant equal weighting could be sub-optimal (Figure 2.1). For example, the carry strategy had strong three-year returns from 2003 to 2005, before suffering heavy losses. After the carry crash at the end of 2008, the carry strategy enjoyed positive returns until 2012. Conversely, the value strategy had positive returns between 2005 to 2008 (roughly the period when carry performed poorly) and negative returns 2008 to 2012 (roughly the period when carry performed well). A model that dynamically weights the factor strategies could increase profitability. To illustrate how potentially lucrative a factor timing strategy could be— if one could switch between the momentum, carry, or value factor strategy with perfect foresight every month from 2003 to 2017, then one could produce annualized returns of 70% and Sharpe ratio of 4.5. Although factor timing is a difficult problem, I propose a factor timing model that combines these three strategies in an intuitive way.

Intuitively, trading on misvaluation should likely be more profitable if the misvaluation is extreme. Also, trading on other factors should likely be more dangerous if a currency is prone to correction from extreme misvaluation. As such, I use a nonlinear model to combine these three strategies, where the time-varying weights are influenced by the degree of currency misvaluation.

Specifically, I use a variation of an exponential smooth transition autoregressive (ESTAR) model where the exponential transition is a function of currency misvaluation.

Although the proposed method sounds sensible, a naïve random walk benchmark has better forecasting accuracy and an equal weight benchmark has better portfolio performance. A possible explanation for my model's mediocre performance is that currency mean reversion to fair value happens slowly, such that value may not have strong explanatory power over short horizons. When fitting the model over a one-month investment horizon, my results show that value behavior is muted, such that the strategy is swamped by momentum and carry. Specifically, the return stream from my nonlinear model is highly correlated with the return stream from a pure momentum and pure carry strategy, but has low correlation with a pure value strategy. The main premise of the model is that the value factor should be powerful under periods of extreme misvaluation; however, the coefficient estimates from my proposed nonlinear model suggests that the premise is not empirically correct for a one-month investment horizon. Future research may entail estimating the model for a longer one-year investment horizon.

In my literature review (Section 2.2), I justify the use of momentum, carry, and valuation strategies. I also note that the literature suggests nonlinear models capture the dynamics of exchange rate reversion to fair value better than linear models. Section 2.3 explains my methodology. Section 2.4 shows the model's forecasting and investment performance. Section 2.5 discusses why the proposed model does not outperform the equal weight benchmark. Section 2.6 concludes.

2.2 LITERATURE REVIEW

In the existing literature, there is evidence that carry, momentum, and value can be profitable strategies. Numerous studies show the strategies' attractive Sharpe ratios over various sample periods and currency universes. Carry may be profitable on average as partial compensation for systematic risk. Momentum may be profitable on average because portfolio managers (or their clients) cannot survive the strategy's stretches of unprofitability. Exchange rates revert to fair value very slowly, but the literature suggests that nonlinear models capture the dynamics of reversion better than linear models.

2.2.1 *The carry strategy*

The carry trade involves borrowing low-yielding currencies and investing in high-yielding currencies, without hedging for exchange rate changes. According to the economic theory of uncovered interest parity (UIP), the high-yielding currency should depreciate or the low-yielding currency appreciate. This should offset any gains from the interest differential between the two regions. However, UIP fails empirically (see Engel 1996 or Hodrick 1987 for literature surveys), such that the carry trade strategy is profitable.

For example, Burnside et al. show superior carry trade performance. In Burnside, Eichenbaum, Kleshchelski, and Rebelo (2006), a carry strategy yielded positive Sharpe ratios over 1976 to 2005 for ten developed market currencies. The authors sold a currency when the forward rate was higher than the spot exchange rate, and bought when the forward rate was lower than the spot exchange rate— essentially borrowing the lower yielding currency and lending the higher yielding currency. When implementing this strategy for Belgium, Canada, France, Germany, Italy, Japan, Netherlands, Switzerland, USA, and Euro-area currencies individually against the UK pound, the individual Sharpe ratios ranged from 0.061-0.201 without transaction costs and 0.028-0.161 with bid-ask spread transaction costs. An equally weighted portfolio of these ten currency pairs yielded positive and statistically significant Sharpe ratios without transaction costs. They also find that the carry returns are uncorrelated with US stock market returns.

Burnside, Eichenbaum, and Rebelo (2007) extended their analysis to include emerging market currencies. By including both emerging and developed market currencies in a carry trade portfolio from 1997 to 2006, they increase the Sharpe ratio versus a carry portfolio with developed market currencies only. Because carry trade returns were uncorrelated with US stock market returns over this new sample period, they argue that it’s unlikely that carry payoffs were compensation for bearing risk.

Finally, Burnside, Eichenbaum, and Rebelo (2008) implement a diversified equally weighted carry trade strategy across 23 currencies from 1976 to 2007, and get a high Sharpe ratio of 0.828, which is statistically different from zero. They also build a high-low portfolio that trades on the most extreme interest differentials (similar to the DBCR and RCCI methodologies) and get a high Sharpe ratio of 0.538, which is statistically different from zero. A common theme from Burnside et al. is that the carry strategy is profitable.

Although there seems to be consensus that the carry strategy is profitable, it is still undetermined in the literature if the carry premium is compensation for risk. For example, Christiansen, Ranaldo, and Soderllind (2011) use an asset pricing model with regime-dependent factor loadings to partially explain carry trade profitability. They find that the carry trade is exposed to a time-varying systematic risk that increases in volatile markets, for example, when there is high exchange rate volatility or funding illiquidity. In addition, Menkhoff, Sarno, Schmeling, and Schrimpf (2012a) find carry trades fail when global exchange rate volatility is high, while Brunnermeier, Nagel, and Pedersen (2008) explore the skewness of exchange rates movements and find that currencies with high interest differentials are prone to crash risk.

Another justification of the carry premium is the “peso problem,” or that the carry trade is exposed to a not yet occurred rare disaster (Farhi and Gabaix 2016). To refute, Burnside, Eichenbaum, Kleshchelski, and Rebelo (2011b) use options to hedge against rare and extreme currency changes and find that the peso problem can’t explain all of carry trade profitability. In summary, the carry trade is empirically profitable, but its profitability may be partial compensation for risk.

2.2.2 *The momentum strategy*

The momentum strategy involves buying assets with recent strong returns and selling assets with recent weak returns. Menkhoff, Sarno, Schmeling, and Schrimpf (2012) test a cross sectional momentum strategy for 48 currencies over 1976 to 2010. They calculate lagged 1, 3, 6, 9, 12-month returns for all the currencies, then take a long position in the one-sixth of currencies with the highest lagged returns and take a short position in the one-sixth of currencies with the lowest lagged returns. They find that shorter lookback periods and shorter holding periods produce better Sharpe ratios, generally.

By implementing a high-low momentum strategy using one-month lagged returns and a one-month holding period, Menkhoff et al. generate a portfolio with a 9.46% annual return and a 0.95 Sharpe ratio. This profitability is driven by spot exchange rate changes, not carry trade interest rate differentials. In fact, the authors find that momentum and carry excess returns are uncorrelated, such that combining the two strategies can have diversification properties.

To check that momentum returns are not a manifestation of some risk factor, the authors regress momentum returns on risk factors in literature— macroeconomic variables like consumption growth, employment growth, inflation, and risk factors such as TED spread, term spread, carry trade long-short portfolio returns, and global exchange rate volatility. They also check momentum returns against three Fama-French factors and US stock return momentum factor. They find that these variables do not explain the momentum strategy returns.

Instead, Menkhoff et al. propose that the momentum strategy may be profitable because there are limits to arbitrage. First, bid-ask spreads cut into the profitability, reducing annual returns from 9.46% without transaction costs to 3.92% with transaction costs. The size of the currency cross section matters— in a simulation, decreasing the pool of investable currencies from 48 to 6 reduces the annual return from 9.46% to a little above 4%. Short investment horizon constraints could prevent arbitrage— the rolling three-year returns of the momentum strategy varies across time, such that myopic investors may not want to participate.

In another example, Burnside, Eichenbaum, and Rebelo (2011) explore an equally weighted momentum strategy for 20 currencies over 1976 to 2010. Unlike the top-sixth minus bottom-sixth momentum strategy of Menkhoff et al., Burnside et al. takes long position for all currencies with positive one-month lagged returns versus the US dollar base currency, then normalize the position to one US dollar. Although this implementation is different from Menkhoff's long-short strategy, the authors still find profitability— 4.5% annualized return and 0.62 Sharpe ratio for the equally weighted portfolio, and 4.9% average annualized return and 0.43 average Sharpe amongst the 20 individual currency strategies. The authors find that factors like three Fama-French, dollar risk factor, return differential factor, and global currency volatility cannot explain the momentum returns. Despite various implementations, a common theme in the existing literature is the profitability of trading strategies based on momentum.

2.2.3 *The value strategy*

The value trade is straight forward— buy cheap assets and sell expensive assets; the trick is correctly determining fair value. Purchasing price parity (PPP) is an intuitive way to price exchange rates. The PPP model says that the exchange rate between two countries is proportional to the price levels between the two countries. When the actual exchange rate is not at equilibrium, consumers can arbitrage goods until the PPP relationship holds. For example, suppose that a bundle of consumer goods is more expensive in the US than converting US dollars (USD) to Canadian dollars (CAD) and buying the goods in Canada; US consumers would then go to Canada to buy goods, increasing demand for CAD, until CAD/USD appreciates and the goods cost the same across the two regions. Realistically, transaction costs like transportation or taxes prevent arbitrage.

Given frictions, first I verify in the literature that reversion to PPP holds empirically. Second, I confirm that trading on variants of PPP can be a profitable strategy.

2.2.4 *The value strategy: exchange rate reversion to PPP*

The existing literature attempts to answer if PPP empirically holds— that is, does the real exchange rate (i.e. exchange rate adjusted by relative price levels) mean-revert? Taylor and Taylor (2004) nicely summarize the evolution of the PPP debate. Research from the 1980s had trouble rejecting the null hypothesis that real exchange rates do not revert because of low powered tests. These studies used test statistics where the time series was too short and the number of currency regions too few to distinguish low reversion speeds from unit roots. Taylor (2002), using a high-powered panel of twenty countries from 1850 to 1996, finds support for PPP.

However, the rate of reversion in linear models seemed very slow. In a linear model specification, the speed of reversion is constant regardless of the degree of misvaluation; instead, the speed of reversion should intuitively accelerate if misvaluation is large and slow if misvaluation is small. If currency mispricing is small, it's not worth incurring transportation costs to arbitrage; with a larger mispricing, arbitrage is more likely profitable, pushing the exchange rate towards PPP. Many researchers apply a class of threshold autoregressive models to study nonlinear reversion in real exchange rates, with promising results.

Taylor, Peel, and Sarno (2001) apply an exponential smooth transition autoregressive (ESTAR) methodology to UK sterling, German mark, French franc and Japanese yen versus the US dollar, over 1973 to 1996. They could not find statistically significant linear mean reversion, but do find statistically significant nonlinear mean reversion for each of the four real exchange rates. They also shed insight on the rate of reversion. Their impulse response shows that small shocks to real exchange rates that are near equilibrium are very persistent, just like shocks in linear real exchange rate models are very persistent. However, large shocks cause faster mean reversion in nonlinear models. Overall, these studies suggest that nonlinear models are better than linear models in capturing exchange rate reversion to fair value.

2.2.5 *The value strategy: profitability of trading on PPP*

Della Corte, Ramadorai, and Sarno (2016) apply a PPP long-short strategy for ten exchange rates from 1998 to 2013. They calculate real exchange rates using PPP, then take a long position in the one-fifth most undervalued currencies (lowest real exchange rates) and short position in the one-fifth most overvalued currencies (highest real exchange rates). This strategy generates a portfolio with 3.66% annual return and 0.41 Sharpe ratio, adjusted for transaction costs.

Because some currencies may face persistent misvaluation, a five-year relative PPP (RPPP) measure is an alternative and less strict way of capturing expensiveness. Asness, Moskowitz, Pedersen (2013) calculate five-year change in PPP to capture misvaluation for ten currencies over 1979 to 2011. They take long positions in the top-third and short positions in the bottom-third currencies ranked by RPPP and generate 3.3% annual return and 0.34 Sharpe ratio. Kroenke, Schindler, and Schrimpf (2014) also use this RPPP measure to value ten developed currencies. Over 1985 to 2009, their long-short RPPP strategy portfolio produces 4.8% annualized return and 0.45 Sharpe ratio without transaction costs, and 4.4% annualized return and 0.42 Sharpe ratio with transaction costs.

Clearly, there is evidence that trading on absolute PPP and relative PPP can both be profitable. As a result, I try both measures of valuation in my nonlinear model.

2.3 METHODOLOGY

2.3.1 *Data and definitions*

The data set spans from January 1999 to December 2017 at a monthly frequency for ten regions: United States, Australia, Canada, Switzerland, Euro zone, United Kingdom, Japan, Norway, New Zealand, and Sweden. The data set includes spot exchange rates and one-month forward exchange rates, interest rates (one-month, three-months, two-year, ten-year), consumer price indices, and PPP for each region. The US dollar is the base currency, such that the nine other currency regions will be compared to US.

- **Currency return**: The currency return is the return from holding a foreign currency for one period, and harvesting gains from spot exchange appreciation and the one-month interest differential between the region and US. Essentially, the currency return is the difference between the logged spot exchange rate and logged one-month forward exchange rate.

$$r_t = (s_t - f_{t-1})$$

- **Momentum:** I use the three-month spot return as the measure for momentum. This is the difference between the logged exchange rate today and the logged exchange rate from three months ago.

$$M_t = (s_t - s_{t-3})$$

- **Carry**: Carry is measured as the interest differential between the foreign region and the US. I use an average of the three-months, two-year, and ten-year interest rates to form the carry measure.

$$C_t = \frac{1}{3}\left[(i_{foreign,3m} - i_{US,3m}) + (i_{foreign,2y} - i_{US,2y}) + (i_{foreign,10y} - i_{US,10y})\right]$$

- **PPP valuation**: Measures the spot exchange rate deviation from fair value, as estimated by OECD's measure of PPP. I take the difference between the logged spot exchange rate and logged PPP measure. If the deviation is positive, then the spot exchange rate is higher than fair, or the currency is overvalued/expensive. If the deviation is negative, the currency is undervalued/cheap.

$$V_t = (s_t - ppp_t)$$

- **RPPP valuation**: Relative PPP is an alternative and less strict measure of value than absolute PPP. I define RPPP valuation as the spot exchange rate deviation from fair value, relative to five years ago. Fair value is derived from consumer price index levels.

$$V_t = \left(1 + R_{t,4.5y:5.5y}\right) * \frac{1 + \text{infl}_{foreign,t,4.5y:5.5y}}{1 + \text{infl}_{US,t,4.5y:5.5y}}$$

$$R_{t,4.5y:5.5y} = \frac{S_t}{\frac{1}{12}\sum_{i=4.5y}^{5.5y} S_{t-i}}$$

$$\text{infl}_{t,4.5y:5.5y} = \frac{CPI_t}{\frac{1}{12}\sum_{i=4.5y}^{5.5y} CPI_{t-i}}$$

2.3.2 *The model*

In my proposed model, currency returns (r) are a function of momentum (M), carry (C), and value (V); however, the coefficients on momentum, carry, and value are allowed to vary depending on the degree of currency misvaluation.

Currency relationship:

$$r_t = \pi_t * [\,\beta_M^{mis}\, M_{t-1} + \beta_C^{mis} C_{t-1} + \beta_V^{mis} V_{t-1}] + (1 - \pi_t) * [\,\beta_M^{fair} M_{t-1} + \beta_C^{fair} C_{t-1} + \beta_V^{fair} V_{t-1}] + e_t\,, e_t \sim iid(0, \sigma^2)$$

Smooth transition equation:

$$\pi_t = 1 - exp\left(-\frac{exp(\gamma)}{var(V_{t-1})} * (V_{t-1})^2 \right)$$

I use an exponential transition function similar to Taylor, Peel, and Sarno (2001). The transition function ranges between 0 and 1. When $\pi = 1$, currency returns are in the misvalued regime and when $\pi = 0$, currency returns are in the fair value regime. The parameter γ is estimated and determines the speed of transition between fair and misvalued regimes. The transition is symmetric, such that only the magnitude of misvaluation matters, and not if a currency is under- or over-valued.

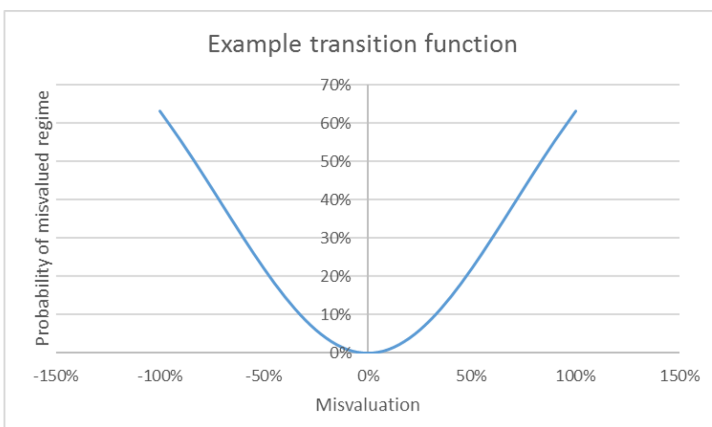

Figure 2.2: Example of a bell-shaped symmetric exponential transition function.

Let β^{mis} denote the coefficient for momentum, carry, and value under the misvaluation regime and β^{fair} denote the coefficient for momentum, carry, and value under the fair value regime. Intuitively, the coefficients should have certain signs. For momentum, past positive returns should lead to future positive returns, such that $\beta_M^{fair} \geq 0$ and $\beta_M^{mis} \geq 0$. Momentum should matter less during times of misvaluation, such that $\left|\beta_M^{mis}\right| \leq \left|\beta_M^{fair}\right|$. For carry, positive interest differential should lead to future positive returns, such that $\beta_C^{fair} \geq 0$ and $\beta_C^{mis} \geq 0$. Carry should matter less during times of misvaluation, such that $\left|\beta_M^{mis}\right| \leq \left|\beta_M^{fair}\right|$. For value, overvaluation should lead to future negative returns, such that $\beta_V^{fair} \leq 0$ and $\beta_V^{mis} \leq 0$. Valuation matters more during times of high misvaluation, such that $\left|\beta_V^{mis}\right| \geq \left|\beta_V^{fair}\right|$.

To better understand the currency relationship, I can reparametrize the equation as the sum of a fair value baseline and misvaluation regime, such that:

$$r_t = \pi_t * \left[\, \beta_M^{mis} M_{t-1} + \beta_C^{mis} C_{t-1} + \beta_V^{mis} V_{t-1}\right] + (1-\pi_t) * \left[\, \beta_M^{fair} M_{t-1} + \beta_C^{fair} C_{t-1} + \beta_V^{fair} V_{t-1}\right] + e_t$$

$$= \left[\beta_M^{fair} M_{t-1} + \beta_C^{fair} C_{t-1} + \beta_V^{fair} V_{t-1}\right] + \pi_t * \left[\,\left(\beta_M^{mis} - \beta_M^{fair}\right) M_{t-1} + \left(\beta_C^{mis} - \beta_C^{fair}\right) C_{t-1} + \left(\beta_V^{mis} - \beta_V^{fair}\right) V_{t-1}\right] + e_t$$

$$= \left[\beta_M^{fair} M_{t-1} + \beta_C^{fair} C_{t-1} + \beta_V^{fair} V_{t-1}\right] + \pi_t * \left[\,\theta_M M_{t-1} + \theta_C C_{t-1} + \theta_V V_{t-1}\right] + e_t$$

The interpretation after reparameterization is that currency returns evolve from a fair value baseline (the left half of the equation), and are adjusted by θ depending on the degree of the misvaluation regime π (the right half of equation). Under the adjustment framework, these are the intuitive signs:

- Momentum: $\beta_M^{fair} \geq 0, \theta_M \leq 0$
- Carry: $\beta_C^{fair} \geq 0, \theta_C \leq 0$a
- Value: $\beta_V^{fair} \leq 0, \theta_V \leq 0$

Parameters of the currency relationship $\left\{\left(\hat{\beta}_M^{fair}, \hat{\beta}_C^{fair}, \hat{\beta}_V^{fair}\right), \left(\hat{\theta}_M, \hat{\theta}_C, \hat{\theta}_V\right),\ \hat{\gamma},\ \hat{\sigma}\right\}$ are estimated with maximum likelihood (Appendix A) using the balanced panel of currencies. I recursively fit the model and generate a one-month ahead out-of-sample return forecast for the nine currency pairs, from 2003 to 2017.

Because the value portfolio literature uses both absolute and relative PPP, I try both specifications separately as the measure of value V_t. I also run a restricted version of the model, where the signs of β and θ hold as above. In total, there are four specifications of the nonlinear model: absolute PPP with unrestricted estimation, absolute PPP with restricted estimation, relative PPP with unrestricted estimation, and relative PPP with restricted estimation.

2.4 RESULTS

Currency forecasting and factor timing are notoriously difficult problems. I evaluate the performance of my model based on forecasting accuracy and portfolio profitability over the full sample from 2003 to 2017, in addition to three subsample periods. I find the nonlinear model has trouble outperforming a random walk benchmark on forecasting accuracy, and the equally weighted factor strategy on portfolio profitability. Interestingly, the one-month estimation horizon plays a key role in my nonlinear model's backtest performance (Section 5).

2.4.1 *Forecasting accuracy*

For forecasting accuracy, I compare the model's forecast of currency returns with the realized currency return, for currency *i* = AUD, CAD, CHF, EUR, GBP, JPY, NOK, NZD, SEK. I use ordinary least squares, random walk, and PPP as three distinct types of benchmarks for forecasting accuracy.

- **Actual realized currency return**:

$$r_{i,t+1}^{Actual} = \left(s_{i,t+1} - f_{i,t}\right)$$

- **Nonlinear model forecast**:

$$r_{i,t+1}^{model} = \left[\hat{\beta}_M^{fair} M_{i,t} + \hat{\beta}_C^{fair} C_{i,t} + \hat{\beta}_V^{fair} V_{i,t}\right] + \left[1 - exp\left(-\frac{exp(\hat{\gamma})}{var\left(V_{i,t}\right)} * \left(V_{i,t}\right)^2\right)\right] * \left[\hat{\theta}_M M_{i,t} + \hat{\theta}_C C_{i,t} + \hat{\theta}_V V_{i,t}\right]$$

There are four versions of the nonlinear model. I estimate the model with absolute PPP and relative PPP as the value measure V. I also estimate the model with unrestricted and restricted coefficients. Under the restricted version, the following hold:

- Momentum: $\hat{\beta}_M^{fair} \geq 0, \hat{\theta}_M \leq 0$
- Carry: $\hat{\beta}_C^{fair} \geq 0, \hat{\theta}_C \leq 0$
- Value: $\hat{\beta}_V^{fair} \leq 0, \hat{\theta}_V \leq 0$

- **Linear model benchmark**:

$$r_{i,t+1}^{OLS} = \hat{\beta}_M M_{i,t} + \hat{\beta}_C C_{i,t} + \hat{\beta}_V V_{i,t}$$

There are two versions of the linear model. I estimate the model with absolute PPP and relative PPP as the value measure V.

- **Random walk benchmark**:

$$r_{i,t+1}^{RW} = (s_{i,t} - f_{i,t})$$

- **PPP benchmark**:

$$r_{i,t+1}^{PPP} = (ppp_{i,t} - f_{i,t})$$

As in Meese and Rogoff (1983), my model does not outperform the random walk benchmark. The random walk benchmark had the lowest root-mean-squared error (RMSE) in aggregate, and the lowest RMSE for five of the nine currency pairs over the full sample period (Table 2.1). However, except for the large forecasting errors in the PPP benchmark, the errors were generally close for all the models. Restricting the signs of the coefficients β and θ to match intuition marginally decreases RMSE overall, and across most currencies. Using a nonlinear model, versus using a linear model with the same momentum, carry, and value variables does not seem to impact forecasting accuracy across the full sample.

In addition, these results are generally consistent in the subsamples (Table 2.2 - Table 2.4). The random walk benchmark has the lowest RMSE in aggregate for two of three subsample periods.

The random walk benchmark also has the lowest RMSE for five of the nine currency pairs over the 2003 to 2007 subsample period (Table 2.2), six of the nine currency pairs over the 2008 to 2012 subsample period (Table 2.3), but only three of the nine currency pairs over the 2013 to 2017 subsample period (Table 2.4). In contrast, a nonlinear model has the lowest RMSE for two of the nine currency pairs over the 2003 to 2007 subsample period (Table 2.2), three of the nine currency pairs over the 2008 to 2012 subsample period (Table 2.3), but five of the nine currency pairs over the 2013 to 2017 subsample period (Table 2.4). However, except for the PPP benchmark, all the other models (nonlinear, linear, and random walk) have similar RMSEs.

Although I cannot necessarily conclude that a random walk model forecasts significantly better than all other models, my results do show that the other more sophisticated models cannot significantly outperform a random walk, based on forecasting accuracy.

Table 2.1: Forecasting error over January 2003 to December 2017.

	Out-of-sample forecast RMSE									
Model	**All**	**AUD**	**CAD**	**CHF**	**EUR**	**GBP**	**JPY**	**NOK**	**NZD**	**SEK**
ESTAR with PPP, unrestricted	3.20	3.75	2.83	3.08	2.91	2.63	2.75	3.37	3.92	3.32
ESTAR with PPP, restricted	3.19	3.72	2.82	3.09	2.90	2.59	2.76	3.35	3.89	3.31
ESTAR with RPPP, unrestricted	3.20	3.71	2.83	3.14	2.94	2.58	2.79	3.40	3.90	3.30
ESTAR with RPPP, restricted	3.19	3.69	2.81	3.11	2.94	2.58	2.78	3.38	3.89	3.31
Benchmark OLS with PPP	3.19	3.73	2.81	3.09	2.90	2.60	2.78	3.36	3.91	3.31
Benchmark OLS with RPPP	3.19	3.72	2.82	3.08	2.90	2.60	2.78	3.36	3.90	3.31
Benchmark PPP	22.53	24.74	13.26	31.48	11.39	18.36	17.95	36.41	12.65	22.72
Benchmark Random Walk	3.17	3.70	2.78	3.02	2.86	2.62	2.74	3.33	3.91	3.30

Table 2.2: Forecasting error over subsample period January 2003 to December 2007.

Period: 1/2003 - 12/2007	Out-of-sample forecast RMSE									
Model	**All**	**AUD**	**CAD**	**CHF**	**EUR**	**GBP**	**JPY**	**NOK**	**NZD**	**SEK**
ESTAR with PPP, unrestricted	2.68	2.89	2.46	2.51	2.31	2.25	2.32	3.08	3.16	2.98
ESTAR with PPP, restricted	2.67	2.85	2.45	2.52	2.32	2.23	2.32	3.06	3.10	2.98
ESTAR with RPPP, unrestricted	2.72	2.86	2.48	2.62	2.36	2.24	2.37	3.16	3.17	3.02
ESTAR with RPPP, restricted	2.69	2.83	2.45	2.56	2.34	2.27	2.36	3.14	3.12	3.01
Benchmark OLS with PPP	2.68	2.87	2.47	2.52	2.32	2.24	2.35	3.06	3.13	2.98
Benchmark OLS with RPPP	2.68	2.86	2.47	2.55	2.33	2.23	2.33	3.07	3.09	2.99
Benchmark PPP	20.58	11.44	9.92	27.81	9.76	25.73	15.84	33.91	9.75	23.60
Benchmark Random Walk	2.66	2.93	2.48	2.44	2.31	2.23	2.19	2.99	3.17	2.98

Table 2.3: Forecasting error over subsample period January 2008 to December 2012.

Period: 1/2008 - 12/2012	Out-of-sample forecast RMSE									
Model	All	AUD	CAD	CHF	EUR	GBP	JPY	NOK	NZD	SEK
ESTAR with PPP, unrestricted	4.03	5.02	3.35	4.16	3.90	3.03	3.16	4.03	4.88	4.26
ESTAR with PPP, restricted	4.01	4.98	3.32	4.15	3.86	2.97	3.17	4.01	4.85	4.24
ESTAR with RPPP, unrestricted	4.01	4.97	3.36	4.20	3.93	2.91	3.21	4.00	4.84	4.20
ESTAR with RPPP, restricted	4.00	4.93	3.35	4.17	3.93	2.91	3.19	3.99	4.83	4.22
Benchmark OLS with PPP	4.02	4.97	3.33	4.16	3.85	2.99	3.21	4.01	4.90	4.25
Benchmark OLS with RPPP	4.00	4.94	3.32	4.12	3.85	2.99	3.19	3.96	4.90	4.23
Benchmark PPP	27.70	35.52	18.30	37.26	12.42	15.97	24.39	44.34	15.73	26.50
Benchmark Random Walk	3.97	4.88	3.27	4.07	3.77	3.05	3.15	3.95	4.91	4.21

Table 2.4: Forecasting error over subsample period January 2013 to December 2017.

Period: 1/2013 - 12/2017	Out-of-sample forecast RMSE									
Model	All	AUD	CAD	CHF	EUR	GBP	JPY	NOK	NZD	SEK
ESTAR with PPP, unrestricted	2.71	2.96	2.61	2.22	2.21	2.53	2.72	2.88	3.51	2.48
ESTAR with PPP, restricted	2.70	2.96	2.60	2.22	2.22	2.53	2.72	2.89	3.50	2.47
ESTAR with RPPP, unrestricted	2.71	2.93	2.56	2.24	2.23	2.54	2.74	2.95	3.50	2.44
ESTAR with RPPP, restricted	2.70	2.93	2.55	2.24	2.22	2.53	2.72	2.92	3.51	2.43
Benchmark OLS with PPP	2.69	2.95	2.56	2.23	2.23	2.51	2.72	2.89	3.47	2.44
Benchmark OLS with RPPP	2.71	2.98	2.58	2.23	2.24	2.52	2.74	2.95	3.48	2.47
Benchmark PPP	18.23	21.07	9.71	28.48	11.82	9.73	10.98	29.34	11.71	17.01
Benchmark Random Walk	2.70	2.95	2.53	2.20	2.24	2.51	2.79	2.96	3.41	2.48

2.4.2 *Portfolio profitability*

For portfolio profitability, I use a similar top-bottom strategy as the "real world" indices DBCR and RCCI. The nonlinear and linear models generate one-month ahead out-of-sample return forecast for the nine currency pairs, from 2003 to 2017. I rank the return forecasts and take an equal long position in the top three currencies and an equal short position in the bottom three currencies. For the momentum, carry, and PPP valuation benchmarks, I rank the nine currency pairs according to M, C, and V, respectively (as defined in the methodology section), and take an equal long position in the top three currencies and an equal short position in the bottom three currencies.

The equal weight benchmark is defined as the average of the returns generated by the benchmark momentum, benchmark carry, and benchmark value strategies. The nonlinear, linear, and pure factor strategies must have three short, three long, and three neutral currencies, but the equal weight portfolio is allowed different compositions as long as the currency weights sum to zero. Finally, the perfect factor switching strategy assumes perfect foresight— from benchmark momentum, benchmark carry, and benchmark value, the perfect factor switching strategy chooses the factor with the highest return over the month, each month. The perfect factor switching strategy is not realistic and not a benchmark, but illustrates how lucrative factor timing could be if done perfectly.

Table 2.5: Portfolio metrics from various currency long-short strategies, from January 2003 to December 2017.

	Strategy performance			
Model	**Mean**	**SD**	**Sharpe**	**Hit rate**
ESTAR with PPP, unrestricted	2.1%	21.6%	0.10	54%
ESTAR with PPP, restricted	1.4%	21.3%	0.07	54%
ESTAR with RPPP, unrestricted	3.5%	20.7%	0.17	51%
ESTAR with RPPP, restricted	5.4%	21.8%	0.25	52%
Benchmark OLS with PPP	2.1%	21.7%	0.09	54%
Benchmark OLS with RPPP	-1.1%	24.5%	-0.04	54%
Benchmark Momentum	-1.7%	23.7%	-0.07	53%
Benchmark Carry	4.7%	28.8%	0.16	56%
Benchmark Value, PPP	11.6%	20.6%	0.56	53%
Benchmark equal weight	7.1%	11.9%	0.60	57%
Perfect factor switching	69.8%	15.6%	4.48	91%

Using a nonlinear model to dynamically weight momentum, carry, and value based on the degree of currency misvaluation does not improve portfolio performance over an equally weighted portfolio of momentum, carry and value. Table 2.5 shows that over the full sample period, the nonlinear model strategies produce annualized returns between 1.4% and 5.4%, versus 7.1% from the equal weight benchmark. Adjusting for strategy volatility, the equal weight benchmark is still the most attractive strategy. The nonlinear models generate Sharpe ratios between 0.07 and 0.25, versus 0.60 from the equal weight benchmark. Hit rate is an imperfect summary statistic, since it calculates the percentage of months with positive return but neglects the magnitude of those returns, but the equal weight benchmark still beats the nonlinear models.

Table 2.6: Sharpe ratios from various currency long-short strategies, over various subsample periods.

	Subsample Sharpe ratios			
Model	**2003-2017**	**2003-2007**	**2008-2012**	**2013-2017**
ESTAR with PPP, unrestricted	0.10	0.76	-0.30	-0.12
ESTAR with PPP, restricted	0.07	0.89	-0.52	-0.12
ESTAR with RPPP, unrestricted	0.17	0.71	-0.01	-0.25
ESTAR with RPPP, restricted	0.25	0.85	-0.04	0.03
Benchmark OLS with PPP	0.09	0.92	-0.55	0.06
Benchmark OLS with RPPP	-0.04	0.83	-0.58	-0.14
Benchmark Momentum	-0.07	-0.21	0.30	-0.55
Benchmark Carry	0.16	0.92	-0.10	-0.07
Benchmark Value, PPP	0.56	1.01	-0.01	0.89
Benchmark equal weight	0.60	0.90	0.54	0.24
Perfect factor switching	4.48	4.26	4.75	5.09

There is ample evidence in existing literature that momentum, carry, and value performance can vary through time— momentum could be unprofitable for stretches, the carry strategy can be prone to crashes, and currencies can take a long time to revert to fair value. Also, Figure 2.1 shows that the factor with the highest three-year return switches between momentum, carry, and value over time. As such, I calculate performance metrics for five-year subsamples. Table 2.6 shows Sharpe ratios for the nonlinear models and benchmark strategies.

From January 2003 to December 2007, all strategies except for pure momentum had high Sharpe ratios, and the nonlinear strategies have comparable performance to the equal weight benchmark. From January 2008 to December 2012, all strategies except for pure momentum and the equal weight benchmark performed very poorly. From January 2013 to December 2017, the equal weight benchmark beat all strategies except for pure value.

Appendix Figure B.1-B.6 show the cumulative and month-over-month returns for select strategies. Pure momentum and pure carry suffer from periods of drawdowns, but the equal weight strategy has a smoother cumulative return stream than any of the individual momentum, carry, and value strategies. The cumulative return for the nonlinear models, however, still show periods of drawdown.

These results show that though the underlying pure momentum, carry, or value strategies have negative performance for extended periods, averaging the strategies (i.e. the equal weight benchmark) has a diversifying effect that improves overall performance. However, the nonlinear and linear regression weighting did not give this diversifying benefit. The next section explores why the backtest performance for the nonlinear model is weak.

2.5 DISCUSSION: WHY ISN'T PERFORMANCE BETTER?

The proposed nonlinear model does not forecast more accurately than a random walk benchmark and is not more profitable than an equally weighted portfolio of momentum, carry, and value. When Meese and Rogoff (1983) found that structural models had poor out-of-sample fit, they suggested misspecification or time-varying parameters as potential culprits. This section investigates why performance was poor and proposes a potential future improvement.

2.5.1 *Are relationships intuitive and significant?*

The intuition of this nonlinear strategy is that one-month ahead currency returns are predicted by momentum, carry, and valuation, and the importance of the three variables vary depending the degree of currency misvaluation. Table 7 shows the model coefficients and t-statistics. The key takeaway is that valuation as an explanatory variable for one-month currency returns is weaker than desired. Momentum and carry are stronger explanatory variables overall. In Table 2.7, the weights on the momentum and carry factor under fair value (beta_M and beta_C) have the intuitively correct sign and high t-stats. The weight on the value factor under fair value is close to zero, as expected; however, the adjustment to the value factor under extreme misvaluation, or theta_V, is also close to zero. The main premise of the model is that the value factor should be powerful under periods of extreme misvaluation; the estimation shows the premise is not correct. In addition, the benefit of weighting momentum and carry differently in times of currency misvaluation is unclear. In Table 2.7, the adjustment to the carry factor under extreme misvaluation, or theta_C, has an intuitively incorrect sign and low t-stat under the nonlinear model with relative PPP. The adjustment to the momentum factor under extreme misvaluation, or theta_M, has low t-stats under the nonlinear model with absolute PPP.

Table 2.7: Model coefficients and t-statistics, as estimated on December 2017 using the full data set.

Period: 1/1999 - 12/2017	**Model coefficients and (t-stats)**					
Model	**beta_M**	**beta_C**	**beta_V**	**theta_M**	**theta_C**	**theta_V**
ESTAR with PPP, unrestricted	0.05	0.17	0.00	-0.02	-0.13	-0.01
	(2.54)	(1.21)	(0.22)	(0.59)	(1.41)	(0.56)
ESTAR with PPP, restricted	0.05	0.17	0.00	-0.02	-0.13	-0.01
	(2.54)	(1.46)	(0.22)	(0.59)	(1.41)	(0.56)
ESTAR with RPPP, unrestricted	0.10	0.13	-0.01	-0.21	0.02	0.01
	(4.89)	(1.88)	(1.05)	(2.58)	(0.11)	(1.03)
ESTAR with RPPP, restricted	0.09	0.13	0.00	-0.21	0.00	0.00
	(4.58)	(1.96)	(0.18)	(2.38)	0.00	0.00
Benchmark OLS with PPP	0.04	0.11	-0.01			
	(3.38)	(2.61)	(3.04)			
Benchmark OLS with RPPP	0.04	0.12	-0.01			
	(3.27)	(2.89)	(1.64)			

Additionally, Table 2.7 shows that:

- When fitting one-month ahead currency returns for the nonlinear and linear models, the momentum coefficient (*beta_M*) is the only coefficient that is significant at the 5% level across the board. It also has the intuitively correct (positive) sign.
- The nonlinear models have an adjustment on momentum (*theta_M*) during the currency misvaluation regime. It has the intuitively correct (negative) sign across the board, but is only significant for the nonlinear model that use relative PPP.
- The carry coefficient has the intuitively correct (positive) sign for both the nonlinear and linear models. The coefficient is statistically significant for the linear models and vary for the nonlinear models.
- The nonlinear model's adjustment on carry (*theta_C*) during the currency misvaluation regime has the correct sign for the PPP nonlinear model, but not for the relative PPP nonlinear model.
- The value coefficient is the correct sign for the linear models. PPP value is statistically significant but relative PPP is not for the linear models.

- For the nonlinear models, the coefficient on value under no misvaluation should be close to zero, and they are. However, the adjustment on value (*theta_V*) during the currency misvaluation regime should be negative and significant. It is negative but not significant for the nonlinear PPP model, and positive (the wrong sign) and not significant for the nonlinear relative PPP model.

Overall, I can conclude that valuation is a weak explanatory variable for one-month currency returns, but momentum and carry are stronger. Also, the estimated coefficients show that it is better to use absolute PPP, rather than relative PPP, to adjust momentum and carry during periods of misvaluation.

2.5.2 *Are relationships stable through time?*

A strong relationship between value and one-month currency returns is core to the proposed nonlinear models. I estimate the linear and nonlinear models over a four-year rolling window from December 2002 to December 2017 to check if coefficients are stable through time. I hypothesize either model misspecification or that the underlying parameters shift over time.

I should expect that the value coefficient under the fair value is close to zero (i.e. when currencies are close to fair value, reversion towards fair value doesn't matter). However, a plot of the absolute PPP coefficients (Figure 2.3) shows that the value coefficient under the fair value travels above and below zero in the rolling estimation, and is surprisingly volatile. The value coefficient under the misvaluation regime is more stable and generally negative (the correct sign). However, the adjustment to valuation is positive (the wrong sign) in a few periods.

Figure 2.4 is the plot of the rolling value coefficients for the relative PPP nonlinear model. It shows that the value coefficient under the misvaluation regime is positive (the incorrect sign) for many periods. Furthermore, the adjustment to valuation is positive (the incorrect sign) frequently. In summary, the relationship between currency returns and value are not stable over time.

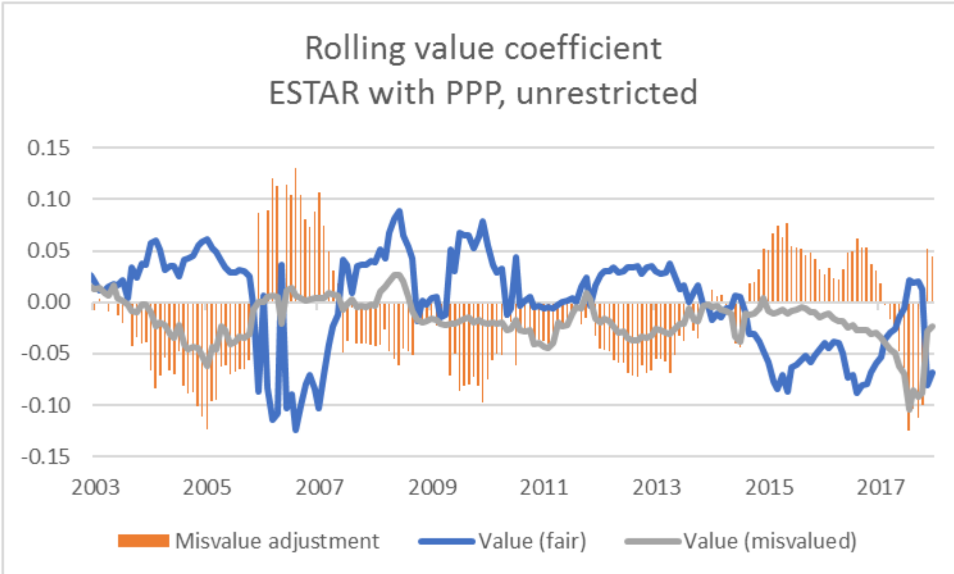

Figure 2.3: Value coefficient estimated over a rolling window, for the PPP nonlinear model.

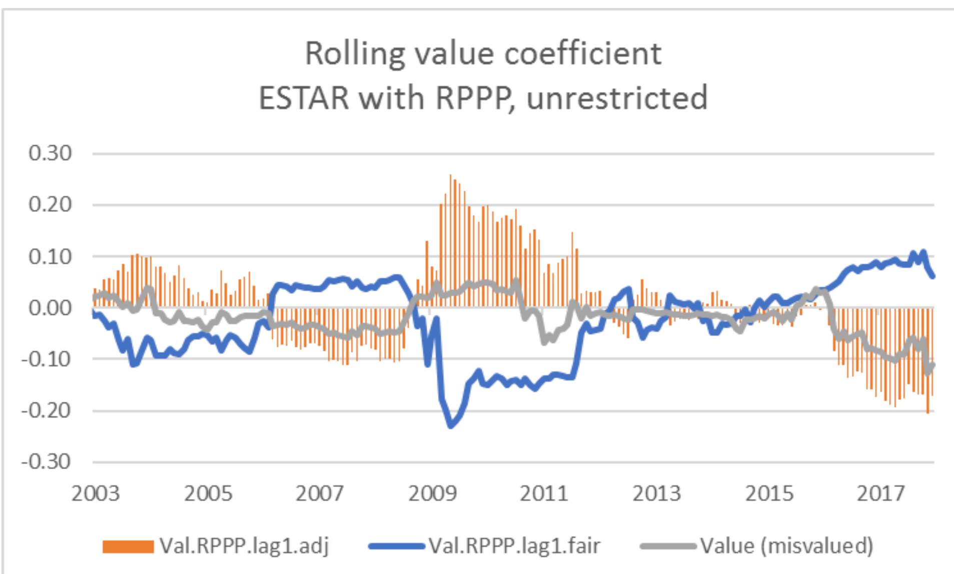

Figure 2.4: Value coefficient estimated over a rolling window, for the relative PPP nonlinear model.

See Appendix Figure B.7-B.8 for more rolling estimation graphics.

2.5.3 *Is valuation used appropriately?*

Perhaps not. The valuation trading strategy is clearly profitable (see literature review, Table 2.5, Table 2.6). It is possible that the model's transmission mechanism is the problem— valuation must enter through one-month ahead currency forecasts, then return forecasts are ranked across currencies, and then implement the long-short portfolio. Reversion to fair value takes time, even under severe misvaluation, such that a one-month time horizon is too short.

To get a clearer sense of strategy drivers and return attribution, I calculate return correlations (Table 2.8). The nonlinear models' returns are most correlated with the benchmark momentum and benchmark carry strategy returns; correlation with the benchmark value strategy return is very low. The nonlinear models are underutilizing value. Separately, the attractiveness of the equal weight benchmark is apparent— the momentum, carry, and value strategies are uncorrelated or negatively correlated with each other, making strong diversifiers.

To fix the transmission mechanism, it would be worthwhile to try a different data frequency. Value may transmit through annual currency returns better than one-month currency returns. The reduced number of data observations would make estimation harder, which would require a larger currency universe or longer time series.

Table 2.8: Return correlations between various currency long-short strategies, from January 2003 to December 2017.

	Monthly return correlation with benchmark			
Model	**Momentum**	**Carry**	**Value**	**Equal Weight**
ESTAR with PPP, unrestricted	22%	55%	8%	57%
ESTAR with PPP, restricted	28%	49%	16%	64%
ESTAR with RPPP, unrestricted	42%	32%	2%	57%
ESTAR with RPPP, restricted	40%	36%	4%	60%
Benchmark Momentum	100%	-23%	3%	52%
Benchmark Carry		100%	-20%	47%
Benchmark Value, PPP			100%	47%
Benchmark equal weight				100%

2.6 CONCLUSION

Carry, momentum, and value are promising strategies studied in literature and implemented in "real life" investible currency indices. In the existing literature, there are promising applications of nonlinear models to explain currency mean reversion. Together combined, I propose a nonlinear model that dynamically weighs carry, momentum, and value based on the degree of currency misvaluation. Although the idea is intuitively attractive, the backtest results are disappointing. The nonlinear model has trouble beating the forecasting ability of a random walk model, and is less profitable than an equally weighted strategy of carry, momentum, and value.

The inferior performance may be from a horizon mismatch— I provide evidence that value has little power over a one-month investment horizon. This hypothesis is consistent with previous studies. For example, Kilian and Taylor (2003) apply ESTAR to seven exchange rates and found in-sample return predictability at the two and three-year forecast horizon, but less predictability at the one quarter and one-year forecast horizon.

Unfortunately, a two- or three-year forecast may be too long of a horizon for a for tactical asset allocation currency product. A logical extension of this paper would be to run the analysis for a larger currency universe and longer time series, but at an annual sampling frequency. Value may have more explanatory power for a one-year investment horizon. However, these results suggest the combination of carry, momentum, and value strategies in a nonlinear valuation model does not work with a short-term forecast horizon.

Chapter 3. COMBINED INDICATORS AS AN EQUITY TIMING STRATEGY

3.1 INTRODUCTION

A large literature exists pertaining to whether equity returns are forecastable. My interest in equity return regressions is for use in US equity long-short trading strategies. I investigate methods of combining return forecasts, comparing their efficacy to the univariate forecasts which many studies focus on. These methods result in annualized returns of 11% over a 22-year sample period, which outperforms a baseline buy-and-hold equity strategy.

Despite the ever-expanding literature on new predictors of return predictability, the feasibility of market-timing strategies is far from certain. The seminal argument that returns are not predictable can be found in Welch and Goyal (2008), who show that most models have poor out-of-sample performance, and argue that univariate forecasts (as well as some methods that rely on multiple variables) cannot help investors time the market. However, Cochrane (2008) shows that poor out-of-sample R-squared statistics should be expected, even if all dividend-price variation were due to time-varying returns (by construction in a simulated data-generating process). This is due to the persistence of the dividend yield in relatively short samples. A proposed statistical test using long-horizon regression coefficients provides evidence for the predictability of returns.

I contribute to this broader literature by examining how multivariate forecasts or combinations of univariate forecasts can be leveraged to improve risk-adjusted portfolio profits. Specifically, my results show that the forecasts generated by my models can be used to profitably exit or short equities during economic downturns.

There are several studies that use multiple predictors. Notably, Rapach, Strauss, and Zhou (2010) illustrate that the inclusion of multiple predictors can outperform the historical average, while Panopoulou and Vrontos (2015) use principle components analysis and ridge regressions to show that integrating multiple sources of information can improve the predictability of hedge fund returns. In this paper I compare univariate and multivariate regressions, the average of univariate regressions, and a naïve benchmark of average historical returns. By examining long-short portfolio performance it is evident that significant gains can be obtained during economic contractions by using multivariate regressions.

These results mirror those of Hull and Qiao (2017), who incorporate a broad set of variables in an OLS regression while systematically excluding predictors across time based on a method of correlation screening (Hero and Rajaratnam 2011). Their strategy also generates a Sharpe ratio that beats a buy-and-hold strategy during their 2001-2015 sample period, with the bulk of their gains during market downturns in 2002 and 2008. Hull and Qiao fix the combination method and vary the choice of indicators. On the other hand, I vary the combination method and fix the choice of indicators.

Section 3.2 describes my data and model methodology. Section 3.3 shows my out-of-sample forecasts and portfolio backtest results. Section 3.4 concludes.

3.2 METHODOLOGY

3.2.1 *Data and definitions*

My set of covariates includes a reasonable but noncomprehensive set of economic and financial indicators that could have predictive power, and I note below the choices of variables that are supported by existing literature. The data set, available as of mid-February 2018, spans from January 1986 to January 2018 at monthly frequency. Table 3.1 shows summary statistics of the data series used in this paper.

- **US equity total returns index**: DataStream's US market index, which contains the top 1000 market-cap stocks. I use the total returns index which includes dividends.
- **Dividend yield (DY)**: DataStream's US market index dividend yield. A high DY is associated with higher future returns (Campbell and Shiller 1988).
- **Price to earnings (PE)**: DataStream's US market index price to earnings. A high PE is associated with lower future returns (Campbell and Shiller 1988). Intuitively, a high PE means a stock is expensive relative to its earnings.
- **Price to book (PB)**: DataStream's US market index price to book. A high PB is associated with lower future returns (Pontiff and Schall 1998).
- **Cyclically adjusted price to earnings (CAPE)**: After deflating DataStream's US market index price and earnings by core consumer price index, CAPE is the real price divided by average real earnings over the last seven years. This methodology is like the ten-year Shiller (2000) PE ratio. Intuitively, an asset with a high CAPE is considered expensive, and is associated with lower future returns.
- **Composite value**: Since DY, PE, PB, and CAPE are very correlated with each other, I standardize the four indicators with z-scores then take the first principal component. The composite value indicator is a single series that contains most of the variation of the four valuation indicators.
- **US Business cycle index (BCI):** The BCI uses the qualitative VAR methodology from Dueker (2005) to capture economic growth. The index is above zero during US economic expansion and below zero during economic recession. The BCI is estimated using

macroeconomic data such as nonfarm payroll employment growth, real personal consumption expenditure growth, core consumer price index inflation, bond yields, slope of the yield curve, TED spread, and corporate bond spreads. Intuitively, stronger economic growth should be associated with higher future returns.

- **US leading index (LEI):** The Federal Reserve Bank of Philadelphia produces a monthly leading index that predicts six-month economic growth. It is estimated using variables like nonfarm payroll employment growth, unemployment rate, average hours, average wages, housing permits, initial claims, Institute for Supply Management surveys, and slope of the yield curve. Like the BCI, a higher LEI should be associated with higher future returns.
- **Composite cycle**: The composite cycle is an average of the BCI and LEI.
- **Labor market conditions index (LMCI)**: The Federal Reserve Bank of Kansas City produces an index that is positive when labor market conditions are above average and negative when below average. They track 24 labor indicators including unemployment, job flows, quits, initial claims, and many labor surveys. Kansas LMCI series starts in 1992. To create a longer series, I splice the Kansas LMCI series with the Board of Governors' discontinued labor market conditions index, which uses 19 labor market indicators and goes back to 1976. The LMCI may act as a proxy for capacity constraints for labor where an overheating labor market lowers expected returns.
- **Consumption-wealth ratio (CAY):** The consumption-wealth ratio (CAY) is the gap between actual consumption (C) and consumption as predicted by wealth assets (A) and income (Y). Lettau and Ludvigson (2001) find that movements in the CAY strongly predicted stock returns. Intuitively, the CAY is a proxy for market expectations of future returns. In the long run, the amount people spend should be equal to the amount of wealth and income they have (i.e. consumption is cointegrated with wealth and income); in the short run, actual and predicted consumption can vary. Under models where consumers optimize their consumption over time, when consumers spend more than their wealth predicts today, it means they expect higher portfolio performance in the future. If consumers expect portfolio returns to fall in the future, they will lower their spending today to smooth consumption. The original CAY series was quarterly frequency, but this paper constructs a monthly frequency version using monthly versions of the underlying data.

- **Variance risk premium (VRP):** Bollerslev, Tauchen, and Zhou (2009) find that the VRP, or the difference between implied and realized variance, helps predict stock returns. The VRP series used in this paper is the difference between VIX and realized volatility. Intuitively, a high VRP is a contrarian buy signal— when options-implied volatility is much higher than actual volatility, then it means the market is overly fearful.

Table 3.1: Summary statistics of data

Period: 1/1986 - 1/2018	**Data summary statistics**								
Indicator	**Mean**	**SD**	**Skew**	**Ex. Kurtosis**	**Min**	**Q1**	**Q2**	**Q3**	**Max**
Composite value	0.57	1.49	0.70	0.37	-2.10	-0.41	0.52	1.08	4.79
DY	2.20	0.74	0.69	-0.39	0.95	1.70	2.00	2.70	4.10
PE	19.21	4.43	0.62	0.14	11.10	16.20	19.00	21.40	31.40
PB	2.71	0.76	1.10	0.90	1.40	2.10	2.65	2.88	5.12
CAPE	21.72	6.39	0.97	0.68	11.21	16.93	21.31	24.28	40.74
Composite cycle	1.07	0.72	-2.49	8.09	-2.76	0.89	1.28	1.47	1.92
BCI	0.90	0.69	-2.93	11.55	-3.53	0.78	1.06	1.28	1.75
LEI	1.25	0.79	-2.03	5.77	-2.79	0.95	1.46	1.75	2.39
LMCI	0.07	0.96	-0.71	-0.27	-2.21	-0.46	0.26	0.79	1.72
CAY	0.01	0.04	0.48	-1.05	-0.06	-0.03	-0.01	0.04	0.10
VRP	16.23	19.22	-3.34	52.92	-210.97	8.51	12.91	21.80	115.85
Fwd 1m return	10.69	52.64	-1.07	3.46	-279.07	-19.10	16.30	44.73	150.74
Fwd 3m return	10.41	31.31	-1.31	3.86	-147.20	-1.06	13.75	27.99	91.01
Fwd 6m return	10.24	22.61	-1.41	4.24	-108.37	0.19	13.31	23.51	67.33
Fwd 12m return	10.18	16.48	-1.28	2.17	-56.30	4.08	13.74	20.23	43.39

Table 3.2 shows the correlation of the indicators and forward returns with each other. I see that the valuation indicators (DY, PE, PB, CAPE) are very correlated with each other. This justifies the use of principal components to combine the information into a single composite value indicator. In addition, BCI and LEI are also very correlated with each other. Their information can be combined with the composite cycle indicator. Finally, the LMCI is moderately correlated with composite value. Just like expensive valuations may signal the age of the bull market is old, an overheated labor market may signal there's not much further for equities to rise given an old economic cycle.

In Table 3.2, the correlation between indicators and forward returns match intuition over the full sample period. Expensive valuations are associated with lower future returns. A strong economic cycle is associated with higher future returns. An overheated labor market is a headwind for future returns. High CAY and high VRP are associated with higher future returns.

Table 3.2: Correlation matrix for data over the full sample

Period: 1/1986 - 1/2017						Indicator Correlation					
Indicator	**Comp. value**	**DY**	**PE**	**PB**	**CAPE**	**Comp. cycle**	**BCI**	**LEI**	**LMCI**	**CAY**	**VRP**
Composite value	100%										
DY	-86%	100%									
PE	95%	-78%	100%								
PB	97%	-73%	90%	100%							
CAPE	97%	-80%	85%	96%	100%						
Composite cycle	18%	-4%	17%	18%	23%	100%					
BCI	21%	-11%	23%	19%	25%	97%	100%				
LEI	14%	2%	12%	15%	20%	98%	89%	100%			
LMCI	45%	-11%	34%	58%	55%	23%	19%	25%	100%		
CAY	-41%	62%	-35%	-29%	-40%	15%	13%	16%	-2%	100%	
VRP	8%	-3%	11%	9%	7%	-4%	1%	-9%	3%	8%	100%
Fwd 1m return	-11%	11%	-12%	-11%	-8%	11%	13%	9%	-5%	12%	21%
Fwd 3m return	-18%	17%	-20%	-18%	-13%	16%	18%	13%	-8%	19%	30%
Fwd 6m return	-24%	24%	-27%	-24%	-18%	16%	19%	12%	-12%	25%	24%
Fwd 12m return	-33%	32%	-33%	-33%	-28%	18%	18%	16%	-18%	37%	16%

Table 3.3, Table 3.4, and Table 3.5 show the correlation matrix calculated across three subsample periods. The relationships between the indicators and forward returns are mostly consistent with intuition during each subsample. Unfortunately, for the subsample period between 1986 and 1995, composite cycle has a negative relationship with forward returns. The BCI and LEI reflected the August 1990 to March 1991 recession by turning negative, but the bulk of the poor equity performance was leading up to the recession— if investors had bought equity during the short eight-month recession, they would have experienced huge gains.

Table 3.3: Correlation matrix for data over subsample 1986 to 1995

Period: 1/1986 - 12/1995						Indicator Correlation					
Indicator	**Comp. value**	**DY**	**PE**	**PB**	**CAPE**	**Comp. cycle**	**BCI**	**LEI**	**LMCI**	**CAY**	**VRP**
Composite value	100%										
DY	-97%	100%									
PE	93%	-83%	100%								
PB	99%	-94%	91%	100%							
CAPE	94%	-97%	75%	93%	100%						
Composite cycle	18%	-25%	16%	10%	20%	100%					
BCI	29%	-36%	26%	22%	30%	96%	100%				
LEI	8%	-15%	8%	0%	10%	97%	87%	100%			
LMCI	-54%	41%	-72%	-57%	-29%	27%	17%	33%	100%		
CAY	74%	-72%	67%	73%	71%	3%	16%	-7%	-51%	100%	
VRP	-41%	47%	-32%	-40%	-42%	-43%	-51%	-34%	4%	-18%	100%
Fwd 1m return	-13%	14%	-15%	-12%	-11%	-5%	-5%	-6%	-3%	5%	0%
Fwd 3m return	-21%	20%	-25%	-18%	-16%	-11%	-7%	-13%	-3%	10%	17%
Fwd 6m return	-21%	19%	-28%	-16%	-13%	-17%	-11%	-20%	-1%	9%	15%
Fwd 12m return	-22%	18%	-31%	-18%	-12%	-14%	-11%	-16%	8%	11%	21%

Table 3.4: Correlation matrix for data over subsample 1996 to 2005

Period: 1/1996 - 12/2005					**Indicator Correlation**						
Indicator	**Comp. value**	**DY**	**PE**	**PB**	**CAPE**	**Comp. cycle**	**BCI**	**LEI**	**LMCI**	**CAY**	**VRP**
Composite value	100%										
DY	-87%	100%									
PE	91%	-88%	100%								
PB	99%	-82%	85%	100%							
CAPE	95%	-72%	74%	96%	100%						
Composite cycle	13%	21%	-12%	17%	36%	100%					
BCI	5%	25%	-15%	8%	26%	95%	100%				
LEI	18%	16%	-9%	22%	41%	97%	85%	100%			
LMCI	84%	-65%	58%	89%	90%	22%	10%	30%	100%		
CAY	-22%	60%	-39%	-12%	-6%	30%	27%	31%	5%	100%	
VRP	16%	-6%	14%	19%	15%	-5%	-5%	-5%	17%	14%	100%
Fwd 1m return	-15%	24%	-24%	-12%	-6%	15%	12%	16%	-5%	19%	18%
Fwd 3m return	-26%	41%	-39%	-22%	-10%	32%	29%	33%	-8%	32%	33%
Fwd 6m return	-33%	55%	-50%	-28%	-13%	46%	45%	43%	-13%	47%	31%
Fwd 12m return	-46%	73%	-60%	-41%	-26%	53%	52%	50%	-27%	61%	19%

Table 3.5: Correlation matrix for data over subsample 2006 to 2017

Period: 1/2006 - 1/2017					**Indicator Correlation**						
Indicator	**Comp. value**	**DY**	**PE**	**PB**	**CAPE**	**Comp. cycle**	**BCI**	**LEI**	**LMCI**	**CAY**	**VRP**
Composite value	100%										
DY	-69%	100%									
PE	82%	-50%	100%								
PB	94%	-51%	65%	100%							
CAPE	95%	-64%	63%	95%	100%						
Composite cycle	54%	-59%	57%	35%	48%	100%					
BCI	53%	-63%	58%	30%	45%	97%	100%				
LEI	53%	-52%	53%	38%	48%	98%	91%	100%			
LMCI	74%	-33%	35%	86%	86%	8%	5%	10%	100%		
CAY	-69%	25%	-40%	-77%	-76%	-21%	-17%	-23%	-81%	100%	
VRP	-18%	3%	-5%	-26%	-22%	0%	9%	-9%	-27%	22%	100%
Fwd 1m return	-10%	3%	0%	-17%	-10%	17%	22%	12%	-13%	11%	34%
Fwd 3m return	-20%	13%	-5%	-29%	-19%	20%	25%	14%	-21%	12%	36%
Fwd 6m return	-34%	31%	-15%	-43%	-31%	14%	19%	8%	-29%	14%	25%
Fwd 12m return	-44%	47%	-13%	-53%	-45%	9%	12%	7%	-41%	29%	14%

3.2.2 *The models*

To generate a time series of out-of-sample forecasts, I estimate the models iteratively. That is, the earliest forecast uses ten years of data to fit the model, and then forms a forecast. The next out-of-sample forecast uses ten years and one month of data to fit the model, and then forms a forecast. This expanding window estimation happens every period until the end of the data sample.

- **Univariate OLS regression**: I regress forward 12-month returns on each of the 11 indicators individually.

$$r_{12m,t} = p_{t+12} - p_t = \hat{c} + \hat{\beta} X_{t-1}$$

Note the conservative time lag on the indicators. An indicator for month *t* will form a forecast for buying equity at time *t+1* and holding for 12 months. The extra lag helps account for slow data releases (i.e. macroeconomic data comes out with some slight delay) and delays in portfolio implementation. Having a less conservative time lag improves backtest performance for all estimated models, but would decrease practical relevance.

- **Multivariate OLS regression**: I regress forward 12-month returns on composite value, composite cycle, LMCI, CAY, and VRP. The OLS regression estimates the coefficients by minimizing the sum of squared errors.

$$r_{12m,t} = p_{t+12} - p_t = \hat{c} + \widehat{\boldsymbol{\beta_{ols}}} \begin{bmatrix} Value_{t-1} \\ Cycle_{t-1} \\ LMCI_{t-1} \\ CAY_{t-1} \\ VRP_{t-1} \end{bmatrix}$$

- **Multivariate ridge regression**: A ridge regression minimizes the sum of squared errors subject to a coefficient penalty. Large coefficients are penalized and the penalization parameter λ is estimated with a cross-validation algorithm. The ridge regression is not unbiased, but generates lower prediction errors than OLS when using correlated explanatory variables.

$$\min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \beta' \beta$$

Because ridge penalizes large coefficients, the indicators need to be standardized to the same scale. I use a y-aware methodology where the time series for each indicator is mapped to returns space using a univariate regression, then centered to have zero mean.

$$r_{12m,t} = p_{t+12} - p_t = \hat{c} + \widehat{\boldsymbol{\beta_{ridge}}} \begin{bmatrix} \widehat{Value}_{t-1} \\ \widehat{Cycle}_{t-1} \\ \widehat{LMCI}_{t-1} \\ \widehat{CAY}_{t-1} \\ \widehat{VRP}_{t-1} \end{bmatrix}$$

- **Multivariate principal components regression (PCR)**: I take the first two principal components of y-aware transformed composite value, composite cycle, LMCI, CAY, and VRP. Then I regress forward 12-month returns on the first two principal components.

$$r_{12m,t} = p_{t+12} - p_t = \hat{c} + \widehat{\boldsymbol{\beta_{PCR}}} \begin{bmatrix} PC_{1.t-1} \\ PC_{2.t-1} \end{bmatrix}$$

- **Average of univariate OLS regressions**: I take the forecasts from the five univariate regressions using composite value, composite cycle, LMCI, CAY, and VRP, then take a simple average.

- **Naïve benchmark**: An expanding window of the historical average return.

3.3 RESULTS

3.3.1 *Forecast behavior*

Figure 3.1 shows the out-of-sample forecasts. The dynamics of OLS, ridge, PCR and the average of univariate OLS regressions forecasts are very similar. Table 3.7 shows the correlation of the out-of-sample forecasts with each other, and the correlation between these four methods are very high, between 77% to 93% correlation. However, the average of univariate forecasts is less volatile. Table 5 shows the out-of-sample forecast summary statistics. Over the 1996 to 2018 forecast period, the OLS, ridge and PCR forecasts have relatively high standard deviations (9.3% to 11.8%) versus the average univariate standard deviation (3.8%). As expected, the expanding historical average (the benchmark) evolves very smoothly, with only a 2.4% standard deviation over the sample period.

The realized 12-month returns are mostly positive, but dip negative leading up to the dot-com bust and global financial crisis. The dot-com bubble burst early in the year 2000, but OLS, ridge, and PCR do not forecast negative returns until 2001 to 2002. The univariate forecasts fall during this period, but the average univariate is never negative. The US equity market peaks around mid-2007 leading up to the US recession— conversely, the OLS forecast turns negative slightly early in mid-2006 and the ridge forecast turns negative slightly late in mid-2008. The average of univariate forecasts does not turn negative until late-2008.

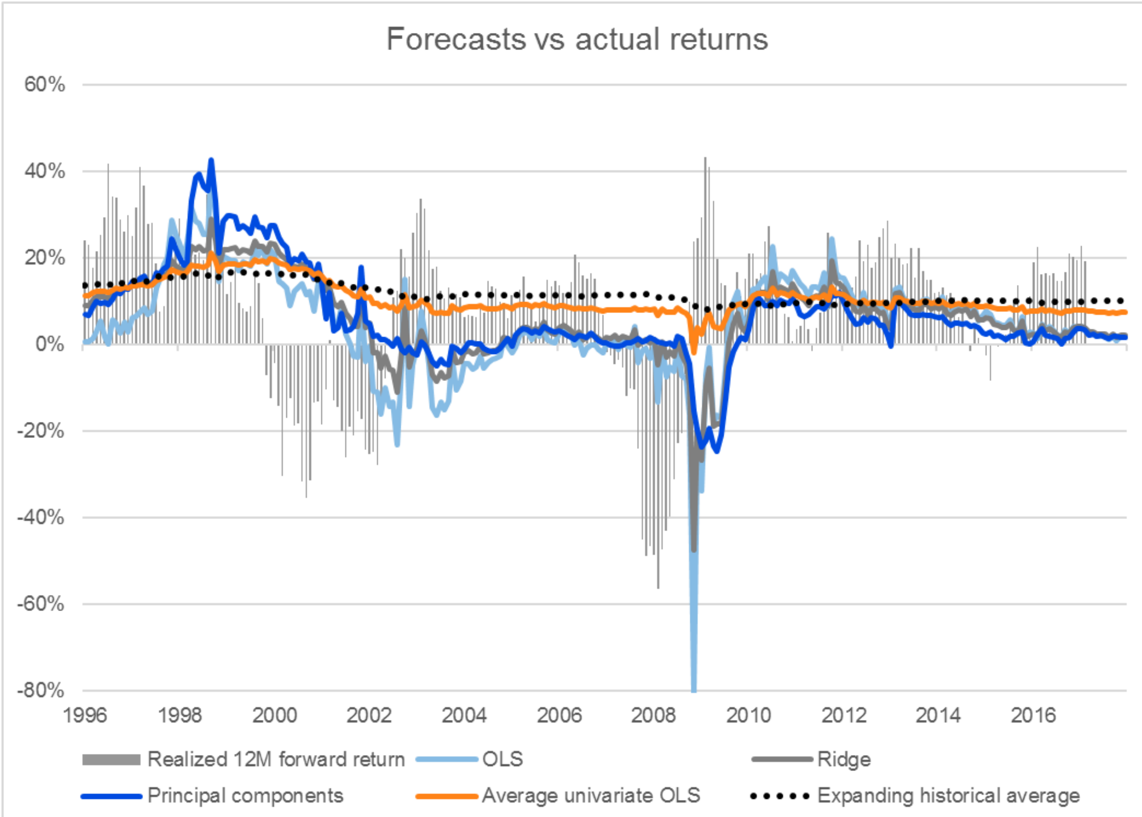

Figure 3.1: Out-of-sample forecasts compared with realized forward 12-month returns

Table 3.6: Summary statistics

Period: 1/1996 - 1/2018	**Out-of-sample forecast summary statistics**								
Model: 12m returns	**Mean**	**SD**	**Skew**	**Ex. Kurtosis**	**Min**	**Q1**	**Q2**	**Q3**	**Max**
Univariate: Composite value	12.4%	4.6%	1.32	1.01	5.1%	9.4%	11.1%	13.2%	24.9%
Univariate: DY	10.4%	5.0%	1.51	0.89	4.6%	7.6%	8.4%	10.3%	22.9%
Univariate: PE	11.5%	3.8%	-0.07	-0.40	1.2%	8.8%	11.6%	14.2%	20.2%
Univariate: PB	13.6%	4.7%	1.16	0.97	4.9%	10.7%	12.4%	14.9%	27.6%
Univariate: CAPE	13.3%	6.0%	1.58	1.84	2.7%	9.9%	11.7%	13.6%	33.5%
Univariate: Composite cycle	9.9%	8.0%	-3.58	14.62	-33.9%	9.9%	11.2%	13.4%	16.6%
Univariate: BCI	10.2%	8.1%	-3.79	16.13	-38.5%	10.0%	11.4%	13.8%	16.9%
Univariate: LEI	9.9%	7.2%	-3.27	13.08	-32.1%	9.8%	11.3%	13.3%	16.5%
Univariate: LMCI	14.3%	3.5%	0.39	-0.87	8.6%	11.5%	13.7%	17.4%	21.9%
Univariate: CAY	5.3%	5.6%	0.47	-0.85	-5.2%	1.2%	3.8%	9.1%	16.8%
Univariate: VRP	11.8%	5.0%	-0.63	17.85	-28.0%	9.2%	10.0%	14.2%	39.3%
OLS	5.1%	11.6%	-1.54	10.89	-80.6%	-0.1%	4.6%	12.3%	40.8%
Ridge	6.5%	9.3%	-0.85	4.19	-47.4%	1.7%	5.2%	12.2%	29.0%
Principal components	6.5%	10.5%	0.65	1.99	-24.7%	1.1%	3.6%	10.0%	42.8%
Average univariate OLS	10.7%	3.8%	0.85	0.51	-1.8%	8.2%	9.4%	12.1%	21.3%
Expanding historical average	11.8%	2.4%	0.84	-0.58	8.0%	10.0%	11.3%	13.7%	16.9%

Table 3.7 shows the correlation between model out-of-sample forecasts and forward returns. This tells us if a model's past out-of-sample signal history can tell us anything useful about forward returns— i.e. if a lower than normal forecast is a harbinger of lower than normal future returns.

Because the benchmark expanding average is relatively stable, it is uncorrelated with future returns and not useful as a timing tool. In contrast, my results show that OLS and ridge historical forecasts are positively correlated with future returns, and is thus a potentially useful timing tool.

Table 3.7: Correlation matrix

Period: 1/1996 - 1/2017	Out-of-sample forecast correlation matrix															
Model: 12m returns	**U1**	**U1a**	**U1b**	**U1c**	**U1d**	**U2**	**U2a**	**U2b**	**U3**	**U4**	**U5**	**OLS**	**Ridge**	**PCR**	**Avg U**	**Bench**
Univariate: Composite value (U1)	100%															
Univariate: DY (U1a)	90%	100%														
Univariate: PE (U1b)	72%	56%	100%													
Univariate: PB (U1c)	98%	85%	67%	100%												
Univariate: CAPE (U1d)	95%	91%	51%	95%	100%											
Univariate: Composite cycle (U2)	6%	21%	-20%	7%	15%	100%										
Univariate: BCI (U2a)	5%	18%	-24%	7%	15%	98%	100%									
Univariate: LEI (U2b)	10%	27%	-14%	9%	18%	98%	93%	100%								
Univariate: LMCI (U3)	76%	64%	44%	80%	77%	3%	4%	4%	100%							
Univariate: CAY (U4)	62%	72%	46%	59%	62%	21%	18%	25%	72%	100%						
Univariate: VRP (U5)	53%	61%	17%	51%	61%	24%	28%	23%	48%	52%	100%					
OLS	47%	47%	22%	51%	50%	54%	55%	52%	48%	57%	64%	100%				
Ridge	52%	59%	25%	53%	56%	68%	67%	67%	55%	73%	64%	93%	100%			
Principal components (PCR)	64%	72%	27%	64%	71%	68%	65%	69%	58%	73%	60%	79%	91%	100%		
Average univariate OLS (Av.U.)	73%	80%	34%	72%	78%	59%	58%	60%	72%	80%	74%	77%	91%	93%	100%	
Expanding historical average (Bench)	63%	80%	32%	56%	69%	45%	43%	49%	46%	64%	64%	32%	54%	70%	79%	100%
Fwd ret 1M	3%	0%	2%	4%	3%	3%	6%	0%	2%	5%	12%	16%	12%	7%	6%	-2%
Fwd ret 3M	4%	-1%	4%	8%	4%	-2%	1%	-5%	3%	6%	21%	22%	14%	5%	6%	-6%
Fwd ret 6M	5%	-1%	4%	10%	6%	-7%	-4%	-9%	6%	6%	11%	17%	9%	4%	3%	-10%
Fwd ret 9M	2%	-4%	-2%	8%	5%	-9%	-7%	-11%	6%	7%	4%	13%	6%	2%	0%	-13%
Fwd ret 12M	-3%	-9%	-11%	2%	1%	-10%	-9%	-12%	4%	6%	1%	11%	3%	1%	-3%	-17%

3.3.2 *Portfolio profitability*

From the model forecasts, I investigate a trading strategy that takes a full long position in US equities if the model forecast is positive, and a full short position if the forecast is negative. Although the model forecasts are for a 12-month investment horizon, I rebalance every month. Table 3.8 shows the portfolio performance of these strategies. Because the expanding historical average is always positive, it is equivalent to a buy-and-hold strategy.

Table 3.8 shows that over the sample period, a buy-and-hold strategy has a 9.0% average annualized return, a 0.45 Sharpe ratio, and a 0.59 information ratio. The ridge strategy beats buy-and-hold with an 11.0% average annualized return, a 0.59 Sharpe ratio, and a 0.73 information ratio.

In addition, I examine how the hit rate, which captures the percentage of months with positive returns, varies across strategies. Overall, the hit rate for the benchmark buy-and-hold strategy is higher than the hit rate for ridge, but only because monthly US equity returns tend to be positive. Conditional on a negative monthly return, the conditional hit rate is 30% for OLS, 22% for ridge, and 0% for buy-and-hold.

My results show that the ridge strategy has the highest maximum 12-month return over the sample period. Conversely, a buy-and-hold strategy has one of the lowest minimum 12-month return over the sample period, suggesting a dynamic strategy like ridge forecasts can help mitigate downside risk.

Finally, as a point forecasting tool, the multivariate OLS, ridge, and PCR have large root-mean-squared errors. Using the average of univariate regressions or the historical average will likely result in a more accurate point return forecast.

Table 3.8: Summary of portfolio performance in a long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns.

Period: 1/1996 - 1/2018	**Model performance**												
Model: 12m returns, directionality only	**Return mean**	**Return SD**	**IR**	**SR**	**Mean 12m return**	**Med 12m return**	**Max 12m return**	**Min 12m return**	**% long position**	**Hit rate (HR)**	**HR|pos. return**	**HR|neg. return**	**RMSE**
Univariate: Composite value	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	19.28%
Univariate: DY	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	19.32%
Univariate: PE	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	19.08%
Univariate: PB	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	19.37%
Univariate: CAPE	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	19.71%
Univariate: Composite cycle	11.0%	15.0%	0.74	0.59	10.4%	12.9%	61.5%	-35.3%	94%	65%	95%	9%	20.62%
Univariate: BCI	10.3%	15.0%	0.69	0.54	9.7%	12.7%	50.1%	-35.3%	94%	64%	94%	8%	20.57%
Univariate: LEI	9.6%	15.1%	0.63	0.49	8.9%	12.4%	45.9%	-47.0%	94%	64%	94%	7%	20.38%
Univariate: LMCI	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	19.28%
Univariate: CAY	7.4%	15.3%	0.48	0.34	6.7%	10.8%	43.4%	-56.3%	85%	61%	85%	15%	18.84%
Univariate: VRP	8.9%	15.3%	0.58	0.44	8.2%	12.4%	43.4%	-59.3%	100%	65%	99%	0%	19.01%
OLS	9.4%	15.0%	0.62	0.48	8.7%	11.5%	49.2%	-35.8%	73%	60%	76%	30%	20.77%
Ridge	11.0%	14.9%	0.73	0.59	10.4%	12.4%	61.8%	-35.3%	82%	63%	84%	22%	20.20%
Principal components	4.8%	15.3%	0.31	0.17	3.9%	7.3%	41.9%	-46.3%	84%	58%	83%	13%	20.98%
Average univariate OLS	8.9%	15.3%	0.58	0.44	8.2%	12.4%	43.4%	-59.3%	100%	65%	99%	0%	18.71%
Expanding historical average	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	18.89%

Table 3.8 shows backtest summary statistics over the full sample, but Figure 3.2 and Figure 3.3 illustrate the attractiveness of a dynamic strategy versus a buy-and-hold strategy over time. The plot of rolling 12-month returns show that OLS and ridge outperform a buy-and-hold strategy during the dot-com and global financial crisis downturns. However, OLS and ridge are slow entering equity again after the dot-com bust, so that the buy-and-hold strategy recovered over

2003. During normal periods without a drastic equity market downturn or rebound (i.e. 2011 to 2017), all strategies perform similarly.

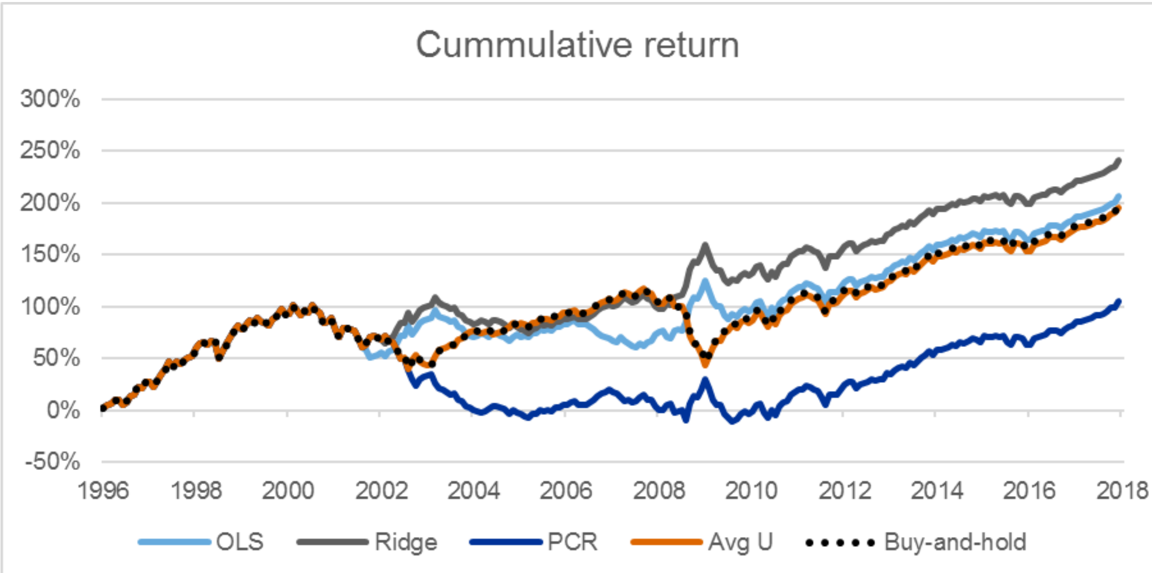

Figure 3.2: Cumulative return of the long-short equity strategy.

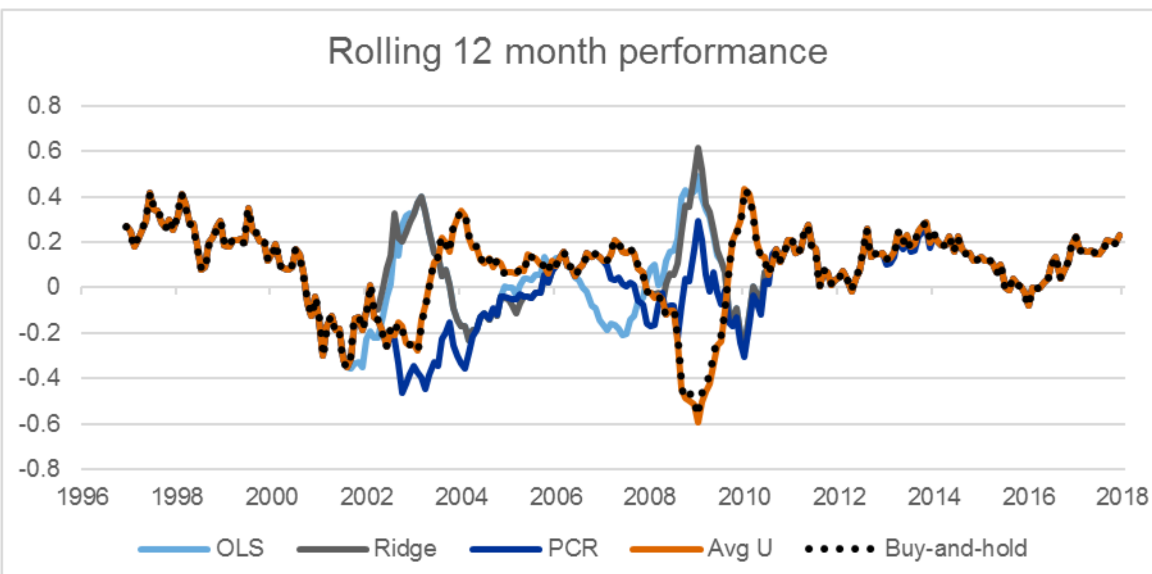

Figure 3.3: Rolling 12-month performance of the long-short equity strategy.

3.3.3 *Robustness check: change the portfolio implementation*

As a robustness check, I also examine a long-short strategy where the magnitude of the return forecast determines the magnitude of the long-short position. The returns from this robustness check are lower across the board than the 100% long-short strategy, because the positions taken are less aggressive. Also, since the historical average return is higher than the average out-of-sample forecast for most the models (see Table 3.6), the benchmark historical average strategy takes larger US equity long positions compared to the other strategies. As such, the benchmark has a relatively high average return over the sample period.

However, many of the takeaways from the previous section still hold true. OLS, ridge and average univariate beat the expanding window strategy on a risk-adjusted basis (i.e. larger IR and SR). Also, the minimum 12-month return is higher for OLS, ridge and average univariate than for the expanding window strategy. To summarize, the result that a multivariate method outperforms the expanding historical average strategy is robust to the choice of portfolio implementation. However, OLS replaces ridge as the top performer under this variation of portfolio implementation.

Table 3.9: Summary of portfolio performance in a long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast.

Period: 1/1996 - 1/2018	**Model performance**												
Model: 12m returns	**Return mean**	**Return SD**	**IR**	**SR**	**Mean 12m return**	**Med 12m return**	**Max 12m return**	**Min 12m return**	**% long position**	**Hit rate (HR)**	**HR|pos. return**	**HR|neg. return**	**RMSE**
Univariate: Composite value	6.7%	12.6%	0.53	0.36	6.5%	8.7%	39.8%	-48.5%	100%	65%	100%	0%	19.28%
Univariate: DY	5.6%	11.2%	0.50	0.30	5.3%	7.1%	37.8%	-39.3%	100%	65%	100%	0%	19.32%
Univariate: PE	6.7%	12.6%	0.54	0.36	6.5%	8.4%	38.1%	-54.6%	100%	65%	100%	0%	19.08%
Univariate: PB	7.5%	13.2%	0.57	0.40	7.3%	9.7%	42.3%	-47.1%	100%	65%	100%	0%	19.37%
Univariate: CAPE	6.7%	13.0%	0.52	0.35	6.6%	9.0%	40.5%	-48.0%	100%	65%	100%	0%	19.71%
Univariate: Composite cycle	6.9%	12.1%	0.57	0.39	6.3%	8.4%	39.2%	-36.6%	94%	65%	95%	9%	20.62%
Univariate: BCI	7.2%	12.2%	0.59	0.42	6.7%	8.8%	39.6%	-34.4%	94%	64%	94%	8%	20.57%
Univariate: LEI	6.4%	12.0%	0.53	0.35	5.8%	8.0%	39.0%	-38.3%	94%	64%	94%	7%	20.38%
Univariate: LMCI	7.7%	14.1%	0.55	0.39	7.3%	9.7%	43.4%	-52.0%	100%	65%	100%	0%	19.28%
Univariate: CAY	4.1%	8.9%	0.46	0.21	3.7%	1.9%	39.3%	-26.1%	85%	61%	85%	15%	18.84%
Univariate: VRP	7.1%	12.0%	0.59	0.41	6.7%	8.1%	41.2%	-32.6%	100%	65%	99%	0%	19.01%
OLS	6.8%	10.7%	0.63	0.43	6.9%	4.1%	38.7%	-22.7%	73%	60%	76%	30%	20.77%
Ridge	6.2%	10.7%	0.58	0.38	6.0%	3.7%	40.8%	-28.6%	82%	63%	84%	22%	20.20%
Principal components	4.4%	10.3%	0.42	0.21	4.2%	2.3%	40.7%	-35.8%	84%	58%	83%	13%	20.98%
Average univariate OLS	6.7%	11.1%	0.60	0.41	6.3%	7.5%	40.0%	-31.2%	100%	65%	99%	0%	18.71%
Expanding historical average	6.7%	12.0%	0.56	0.37	6.1%	8.7%	41.1%	-35.8%	100%	65%	100%	0%	18.89%

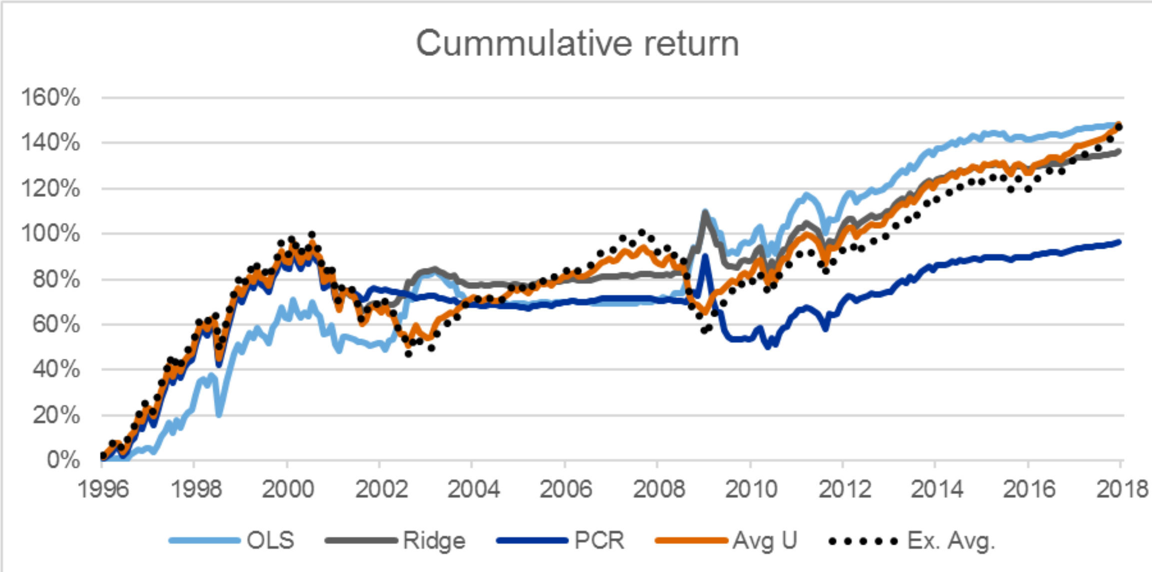

Figure 3.4: Cumulative return of the long-short equity strategy.

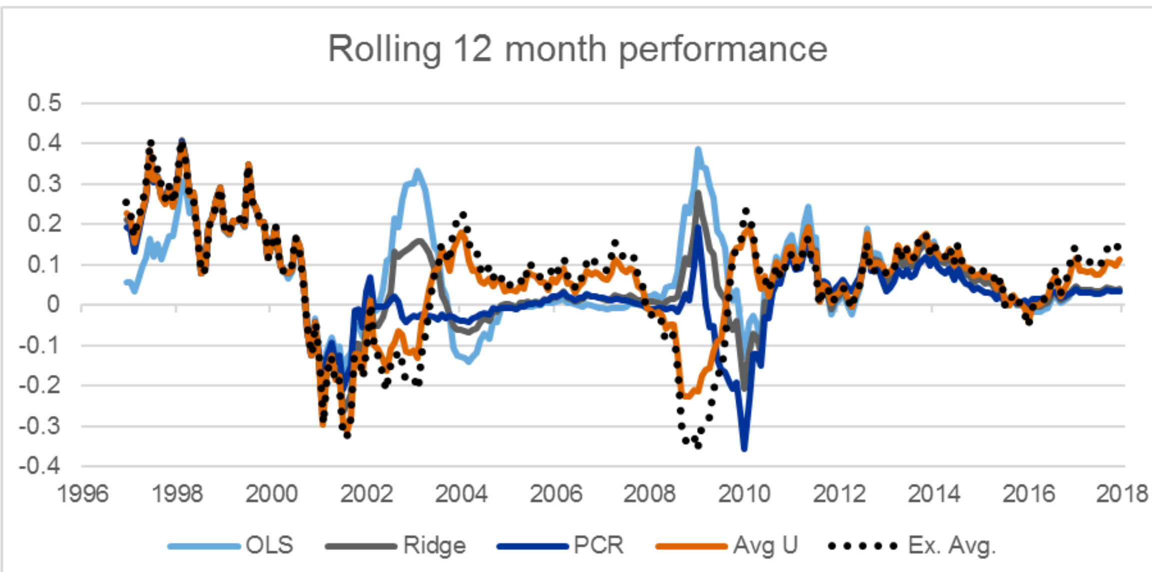

Figure 3.5: Rolling 12-month performance of the long-short equity strategy.

3.3.4 *Robustness check: subsample stability of models*

As another robustness check, I calculate Sharpe and information ratios during subsample periods. Table 3.10 illustrates that the subsample periods of 1996-2000 and 2011-2017 have the same SR and IR across the board, except for the PCR, because the strategies were 100% long equity.

Therefore, most of the performance differentiation comes from the 2001-2005 and 2006-2010 subsample periods.

During 2001-2005, all strategies had negative SR, including the buy-and-hold strategy. Buy-and-hold lost money during the dot-com bust, whereas OLS and ridge lost money because of their slow re-entry during the rebound. Overall, the SR and IR are close during this subsample period.
Then, during 2006-2010, the range of SRs and IRs are wide. A buy-and-hold strategy experiences financial loss during the market downturn, but recovers during the rebound. OLS and ridge, however, short equities during the downturn. Ridge has a 0.60 SR and a 0.72 IR during this subsample period, whereas buy-and-hold has a 0.04 SR and a 0.16 IR. In summary, ridge and OLS greatly outperform a buy-and-hold strategy from 2006 to 2010.

Table 3.10: Sharpe and information ratio across various subsample periods, in a long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns.

	Subsample Sharpe ratio					Subsample information ratio				
Model: 12m returns, directionality only	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**
Univariate: Composite value	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: DY	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: PE	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: PB	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: CAPE	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: Composite cycle	0.59	0.92	-0.08	0.55	0.98	0.74	1.05	0.06	0.67	1.18
Univariate: BCI	0.54	0.92	-0.08	0.36	0.98	0.69	1.05	0.06	0.49	1.18
Univariate: LEI	0.49	0.92	-0.08	0.17	0.98	0.63	1.05	0.06	0.30	1.18
Univariate: LMCI	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: CAY	0.34	0.92	-0.45	-0.05	0.99	0.48	1.05	-0.31	0.07	1.19
Univariate: VRP	0.44	0.92	-0.08	0.00	0.98	0.58	1.05	0.06	0.12	1.18
OLS	0.48	0.92	-0.22	0.24	0.98	0.62	1.05	-0.07	0.37	1.18
Ridge	0.59	0.92	-0.16	0.60	0.98	0.73	1.05	-0.01	0.72	1.18
Principal components	0.17	0.92	-1.29	0.00	0.95	0.31	1.05	-1.14	0.13	1.15
Average univariate OLS	0.44	0.92	-0.08	0.00	0.98	0.58	1.05	0.06	0.12	1.18
Expanding historical average	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18

If I let the magnitude of forecasts determine the magnitude of long-short positions, the subsample analysis still tells a similar story (Table 3.11), where OLS and ridge make the bulk of the risk-adjusted gains during the 2006-2010 subsample period. The importance of the 2006-2010 subsample period is robust to the choice of trading strategy.

Table 3.11: Sharpe and information ratio across various subsample periods, in a long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast.

	Subsample Sharpe ratio					Subsample information ratio				
Model: 12m returns	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**
Univariate: Composite value	0.36	0.77	-0.16	0.04	0.80	0.53	0.90	0.04	0.19	1.08
Univariate: DY	0.30	0.78	-0.40	-0.10	0.90	0.50	0.91	-0.18	0.09	1.27
Univariate: PE	0.36	0.76	0.01	-0.02	0.75	0.54	0.90	0.24	0.11	1.02
Univariate: PB	0.40	0.78	-0.08	0.13	0.82	0.57	0.92	0.10	0.26	1.08
Univariate: CAPE	0.35	0.79	-0.19	0.05	0.84	0.52	0.92	-0.02	0.19	1.13
Univariate: Composite cycle	0.39	0.88	-0.30	-0.02	0.91	0.57	1.01	-0.11	0.17	1.20
Univariate: BCI	0.42	0.88	-0.22	0.09	0.89	0.59	1.01	-0.04	0.28	1.18
Univariate: LEI	0.35	0.88	-0.36	-0.18	0.92	0.53	1.01	-0.17	0.02	1.21
Univariate: LMCI	0.39	0.91	-0.17	0.04	0.82	0.55	1.04	0.00	0.16	1.06
Univariate: CAY	0.21	0.87	-1.08	-0.22	0.45	0.46	1.02	-0.68	0.10	1.13
Univariate: VRP	0.41	0.88	-0.07	0.13	0.74	0.59	1.01	0.11	0.31	1.11
OLS	0.43	0.65	-0.02	0.52	0.48	0.63	0.79	0.22	0.70	0.82
Ridge	0.38	0.86	-0.37	0.19	0.52	0.58	0.99	-0.09	0.40	0.90
Principal components	0.21	0.84	-0.76	-0.30	0.53	0.42	0.97	-0.38	-0.10	1.02
Average univariate OLS	0.41	0.87	-0.26	0.07	0.81	0.60	1.00	-0.05	0.30	1.14
Expanding historical average	0.37	0.90	-0.24	-0.10	0.89	0.56	1.03	-0.05	0.09	1.21

3.3.5 *Robustness check: change the regression horizon*

Finally, I conduct the above analysis for a 1- and 3-month regression horizon. Multivariate methods perform better than the expanding historical average strategy, but my results show that ridge is only the top performer during a 12-month regression horizon.

Appendix C and D show the backtest results if I use a 1-month and 3-month regression horizon instead of a 12-month regression horizon. Using high Sharpe ratios as the criteria to compare methods, at least one of the multivariate models always outperforms the expanding historical average strategy, but the specific multivariate model varies depending on the combination of bet function (100% long-short, or varying magnitude) or the regression horizon (1, 3, or 12-month). The average of the univariate regressions is the best performer in one-half of the combinations (Table 3.12). In summary, the choice of regression horizon does not affect the result that a multivariate method will outperform the expanding average method, but the choice of regression horizon does affect which multivariate method will have the best performance.

Table 3.12: Compares OLS, ridge, PCR, average univariate OLS, and expanding historical average strategies. Reports the strategy with the highest Sharpe ratio, given the regression horizon and bet function.

Model with highest SR and its SR		
Period: 1/1996 - 12/2017	*Sign bet*	*Magnitude bet*
1 month regression horizon	Avg U. / 0.55	PCR / 0.22
3 month regression horizon	Avg U. / 0.55	Avg U. / 0.30
12 month regression horizon	Ridge / 0.59	OLS / 0.43

3.4 CONCLUSION

In this paper I generate out-of-sample forecasts using univariate regressions, multivariate regressions, and the average of univariate regressions. When I compare point forecasting accuracy with the expanding window average of historical returns, I find that the average of univariate regressions produces the lowest forecasting error. However, this does not necessarily result in the most profitable trading strategy. By using out-of-sample forecasts in a long-short equity strategy, where negative return forecasts correspond with a short equity position, I create a more attractive return stream. Specifically, during the 2006-2010 subsample period, a ridge regression strategy produces a 0.60 Sharpe ratio, versus the buy-and-hold strategy's 0.04 Sharpe ratio. In summary, I examine how multivariate forecasts or combinations of univariate forecasts can be used to time market downturns, improving risk-adjusted portfolio profits.

There are caveats. I neglect transaction costs. The models use macroeconomics indicators that are prone to revision, so that today's data may not have been available in real time. Furthermore, there are sometimes delays in data releases and portfolio implementation. To partially adjust for these potential factors that may bias my results upward, I put in a conservative one-month lag before the portfolio positions are implemented. Without this handicap, the performance of the forecasts and strategies are stronger, but the outcome would not be as realistic.

Equity timing is a notoriously difficult problem. This paper estimates a model iteratively to generate out-of-sample forecasts, but any paper contains research bias. After all, researchers are exposed to history. For example, a researcher who experienced the global financial crisis might find variables in hindsight that would have produced profitable strategies, but those variables may not have occurred to the researcher in real time. Only with a live track-record and the passage of time can researchers ascertain if these portfolio gains can be sustained.

BIBLIOGRAPHY

Chapter 1

[1] Black, F. (1995). Interest Rates as Options. Journal of Finance, 50(5), 1371-1376.

[2] Doh, T. & Choi, J. (2016). Measuring the Stance of Monetary Policy on and off the Zero Lower Bound. Economic Review, Federal Reserve Bank of Kansas City, issue Q III, pages 5-24.

[3] Dueker, M. (1999). Conditional Heteroscedasticity in Qualitative Response Models of Time Series: A Gibbs-Sampling Approach to the Bank Prime Rate. Journal of Business & Economic Statistics, 17(4), 466-472.

[4] Dueker, M. (2005). Dynamic Forecasts of Qualitative Variables. Journal of Business & Economic Statistics, 23(1), 96-104.

[5] Dueker, M. (2006). Kalman filtering with truncated normal state variables for Bayesian estimation of macroeconomic models. Economics Letters, 93(1), 58-62.

[6] Eichengreen, B., Watson, M., & Grossman, R. (1985). Bank Rate Policy Under the Interwar Gold Standard: A Dynamic Probit Model. The Economic Journal, 95(379), 725-745.

[7] Hakkio, C., & Kahn, G. (2014). Evaluating Monetary Policy at the Zero Lower Bound. Federal Reserve Bank of Kansas City Economic Review, 5-64.

[8] Hu, L., & Phillips, P. (2004). Dynamics of the federal funds target rate: A nonstationary discrete choice approach. Journal of Applied Econometrics, 19(7), 851-867.

[9] Johannsen, B., & Mertens, E. (2016). A Time Series Model of Interest Rates With the Effective Lower Bound. Finance and Economics Discussion Series, 2016(33), 1-46.

[10] Krippner, Leo. (2011). Modifying Gaussian term structure models when interest rates are near the zero lower bound. 089.

[11] Lombardi, M., & Zhu, Feng. (2014). A shadow policy rate to calibrate US monetary policy at the zero lower bound (BIS working papers; no. 452). Basel: Bank for International Settlements.

[12] Taylor, J. B. (1993). Discretion versus policy rules in practice. Carnegie-Rochester Confer. Series on Public Policy, 39(C), 195-214.

[13] Van Den Hauwe, Paap, & Van Dijk. (2013). Bayesian forecasting of federal funds target rate decisions. Journal of Macroeconomics, 37, 19-40.

[14] Wu, J., & Xia, F. (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. Journal of Money, Credit and Banking, 48(2-3), 253-291.

Chapter 2

[1] Asness, C., Moskowitz, T., & Pedersen, L. (2013). Value and Momentum Everywhere. Journal of Finance, 68(3), 929-985.

[2] Burnside, Craig, Martin Eichenbaum, Isaac Kleshchelski, and Sergio Rebelo (2006). "The Returns to Currency Speculation." National Bureau of Economic Research Working Paper No. 12489.

[3] Burnside, C., Eichenbaum, M., Kleshchelski, I., & Rebelo, S. (2011). Do Peso Problems Explain the Returns to the Carry Trade? The Review of Financial Studies, 24(3), 853-891.

[4] Burnside, C., Eichenbaum, M., & Rebelo, S. (2007). The Returns to Currency Speculation in Emerging Markets. American Economic Review, 97(2), 333-338.

[5] Burnside, C., Eichenbaum, M., & Rebelo, S. (2008). CARRY TRADE: THE GAINS OF DIVERSIFICATION. Journal of the European Economic Association, 6(2-3), 581-588.

[6] Burnside, C., Eichenbaum, M., & Rebelo, S. (2011). Carry Trade and Momentum in Currency Markets. 3(1), 511-535.

[7] Brunnermeier, M. K., Nagel, S. H., & Pedersen, L. (2008). Carry trades and currency crashes. NBER Macroeconomics Annual, 23, 313-347.

[8] Christiansen, C., Ranaldo, A., & Söderlind, P. (2011). The Time-Varying Systematic Risk of Carry Trade Strategies. Journal of Financial and Quantitative Analysis, 46(4), 1107-1125

[9] Engel, Charles. (1996) "The forward discount anomaly and the risk premium: a survey of recent evidence" Journal of Empirical Finance.

[10] Farhi, E., & Gabaix, X. (2016). Rare Disasters and Exchange Rates. Quarterly Journal Of Economics, 131(1), 1-52.

[11] Hodrick, Robert J. (1987). The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets. Harwood Academic.

[12] Kilian, & Taylor. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? Journal of International Economics, 60(1), 85-107.

[13] Meese, Richard, and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?." Journal of International Economics 14: 3-24.

[14] Meese, Richard, Kenneth Rogoff, and Jacob Frenkel. 1983. "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?." Exchange Rates and International Macroeconomics, 67-105. Chicago: University of Chicago Press and NBER, 67-105.

[15] Menkhoff, L., Sarno, L., Schmeling, M., & Schrimpf, A. (2012a). Carry Trades and Global Foreign Exchange Volatility. Journal of Finance, 67(2), 681-718.

[16] Menkhoff, L., Sarno, L., Schmeling, M., & Schrimpf, A. (2012b). Currency momentum strategies. Journal of Financial Economics, 106(3), 660-684.

[17] Okunev, J., & White, D. (2003). Do Momentum-Based Strategies Still Work in Foreign Currency Markets? Journal of Financial and Quantitative Analysis, 38(2), 425-447.

[18] Rogoff, K. (1996). The Purchasing Power Parity Puzzle. Journal of Economic Literature, 34(2), 647-668.

[19] Sarno, Lucio, Nonlinear Exchange Rate Models: A Selective Overview (May 2003). IMF Working Paper, Vol. , pp. 1-39, 2003.

[20] Taylor, A. (2002). A Century of Purchasing-Power Parity. Review of Economics and Statistics, 84(1), 139-150.

[21] Taylor, Alan M, & Taylor, Mark P. (2004). The Purchasing Power Parity Debate. Journal of Economic Perspectives, 18(4), 135-158.

[22] Taylor, M. P., Peel, D. A. and Sarno, L. (2001), Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution To the Purchasing Power Parity Puzzles. International Economic Review, 42: 1015–1042. doi:10.1111/1468-2354.00144

[23] Teräsvirta, Timo & van Dijk, Dick & Terasvirta, Timo & Hans Franses, Philip. (2002). Smooth Transition Autoregressive Models—A Survey of Recent Developments. Econometric Reviews. 21. 1-47. 10.1081/ETC-120008723.

Chapter 3

[1] Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia. The Review of Financial Studies, 22(11), 4463-4492.

[2] Campbell, J., & Shiller, R. (1988). Stock Prices, Earnings, and Expected Dividends. Journal of Finance, 43(3), 661-676.

[3] Cochrane, J. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. The Review of Financial Studies, 21(4), 1533-1575.

[4] Dueker, M. (2005). Dynamic Forecasts of Qualitative Variables. Journal of Business & Economic Statistics, 23(1), 96-104.

[5] Hero, A., & Rajaratnam, B. (2011). Large-Scale Correlation Screening. Journal of the American Statistical Association, 106(496), 1540-1552.

[6] Hull, B., & Qiao, X. (2017). A practitioner's defense of return predictability. Journal of Portfolio Management, 43(3), 60-76.

[7] Lettau, Martin, and Sydney Ludvigson. “Consumption, Aggregate Wealth, and Expected Stock Returns.” Journal of Finance, vol. 56, no. 3, 2001, pp. 815–849.

[8] Panopoulou, & Vrontos. (2015). Hedge fund return predictability; To combine forecasts or combine information? Journal of Banking and Finance, 56, 103-122.

[9] Pontiff, J., & Schall, L. (1998). Book-to-market ratios as predictors of market returns. Journal of Financial Economics, 49(2), 141-160.

[10] Rapach, D., Strauss, J., & Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. The Review of Financial Studies, 23(2), 821-862.

[11] Shiller, R. (2000). Irrational exuberance. Princeton, NJ: Princeton University Press.

[12] Welch, I., & Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. The Review of Financial Studies, 21(4), 1455-1508.

APPENDIX A

Maximum likelihood estimation procedure for nonlinear currency model in Section 2.3.2.

Let $\widehat{\Phi}$ denote the parameters for the model.

$$\widehat{\Phi} = \left\{\left(\hat{\beta}_M^{fair}, \hat{\beta}_C^{fair}, \hat{\beta}_V^{fair}\right), \left(\hat{\theta}_M, \hat{\theta}_C, \hat{\theta}_V\right), \hat{\gamma}, \hat{\sigma}\right\}$$

Let e_i denote the model residuals for a currency, for currencies i = AUD, CAD, CHF, EUR, GBP, JPY, NOK, NZD, SEK. There are nine currencies quoted against the USD.

$$\hat{e_i} = r_{i,t} - \Bigg(\left[\hat{\beta}_M^{fair} M_{i,t-1} + \hat{\beta}_C^{fair} C_{i,t-1} + \hat{\beta}_V^{fair} V_{i,t-1}\right] + \left(1 - exp\left(-\frac{exp(\hat{\gamma})}{var\left(V_{i,t-1}\right)} * (V_{t-1})^2\right)\right) * \left[\hat{\theta}_M M_{i,t-1} + \hat{\theta}_C C_{i,t-1} + \hat{\theta}_V V_{i,t-1}\right] \Bigg)$$

Let $\hat{e_j}$ denote the model residuals for a balanced panel of all nine currencies over the estimation period. Let n denote the total observations in the estimation period. The maximum likelihood estimation procedure involves choosing the parameters $\widehat{\Phi}$ that maximize the likelihood function.

$$\max_{\widehat{\Phi}} L\left(\widehat{\Phi}; \text{data}\right)$$

$$L\left(\widehat{\Phi}; \text{data}\right) = \frac{1}{n} \sum_{j=1}^{n} \ln\left(\frac{1}{(2\pi\hat{\sigma}^2)^{\frac{n}{2}}} exp\left(-\frac{\hat{e}_j^2}{2\hat{\sigma}^2}\right)\right)$$

In the restricted version of the model, the maximum likelihood estimation procedure involves choosing the parameters $\widehat{\Phi}$ that maximize the likelihood function above subject to parameter constraints: $\hat{\beta}_M^{fair} \geq 0$, $\hat{\theta}_M \leq 0$, $\hat{\beta}_C^{fair} \geq 0$, $\hat{\theta}_C \leq 0$, $\hat{\beta}_V^{fair} \leq 0$, $\hat{\theta}_V \leq 0$.

APPENDIX B

Performance graphics for currency strategies in Section 2.4 and 2.5.

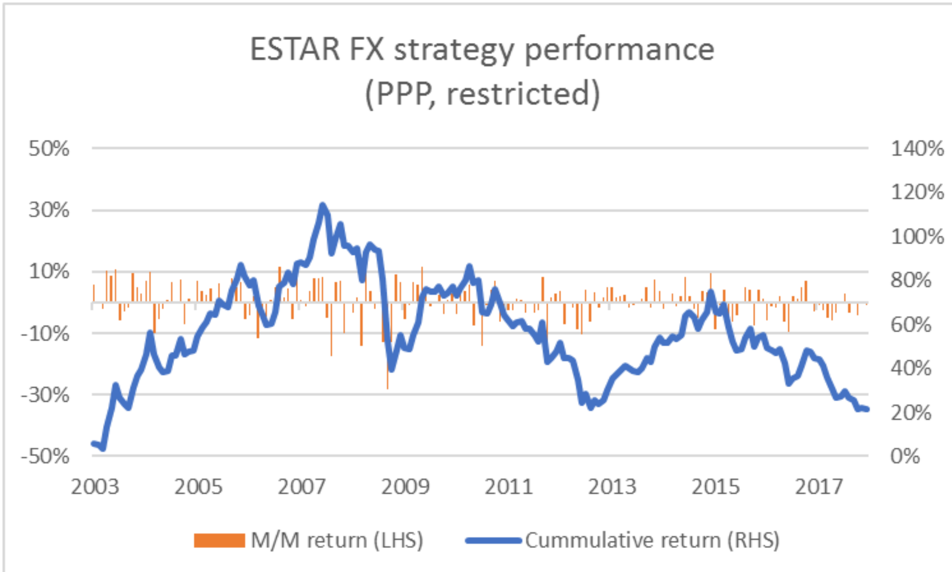

Figure B.1: Cumulative and month-over-month returns from January 2003 to December 2017, for a long top-three ranked and short bottom-three ranked currency strategy. Ranking comes from forecasts from an ESTAR model with restricted coefficients and absolute PPP as the value indicator.

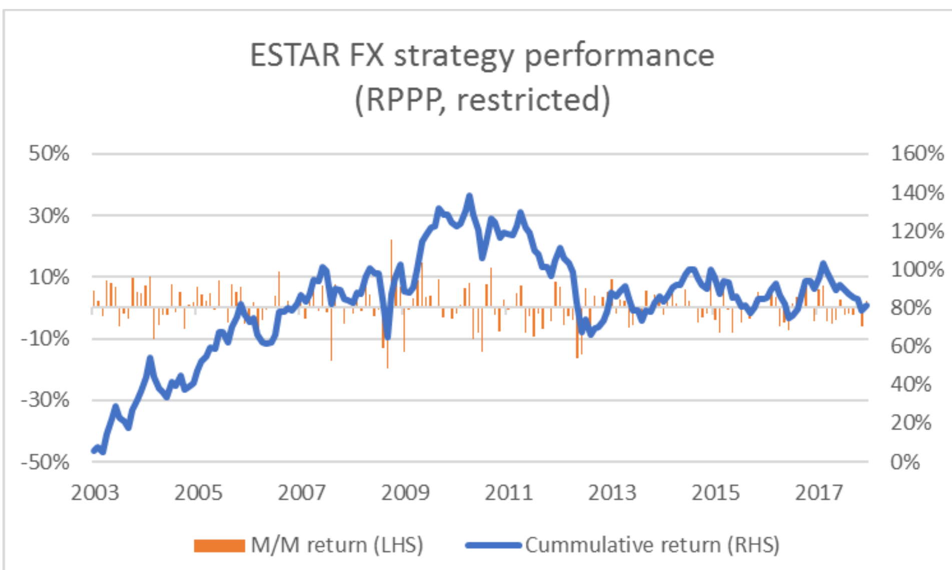

Figure B.2: Cumulative and month-over-month returns from January 2003 to December 2017, for a long top-three ranked and short bottom-three ranked currency strategy. Ranking comes from forecasts from an ESTAR model with restricted coefficients and relative PPP as the value indicator.

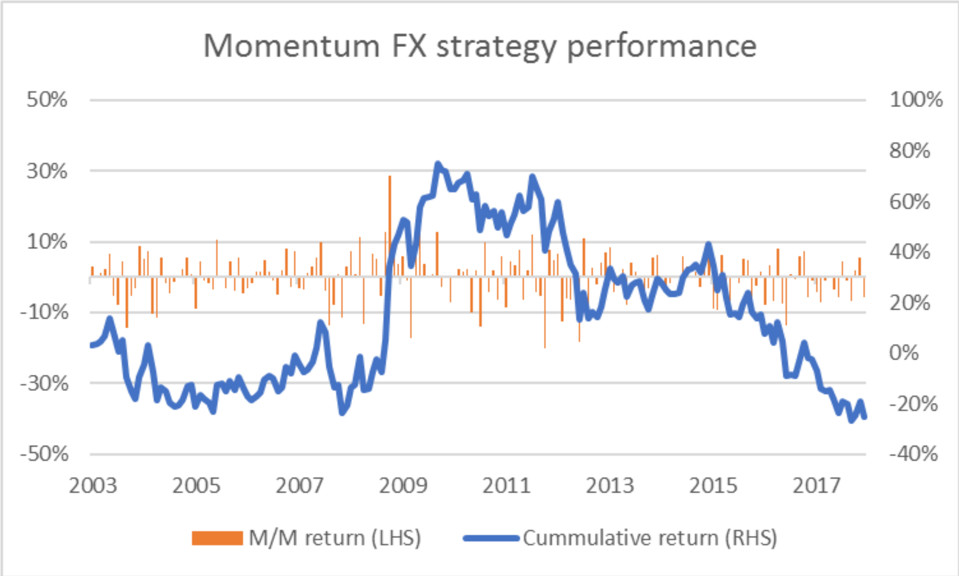

Figure B.3: Cumulative and month-over-month returns from January 2003 to December 2017, for a long top-three ranked and short bottom-three ranked currency strategy. Rank based on momentum.

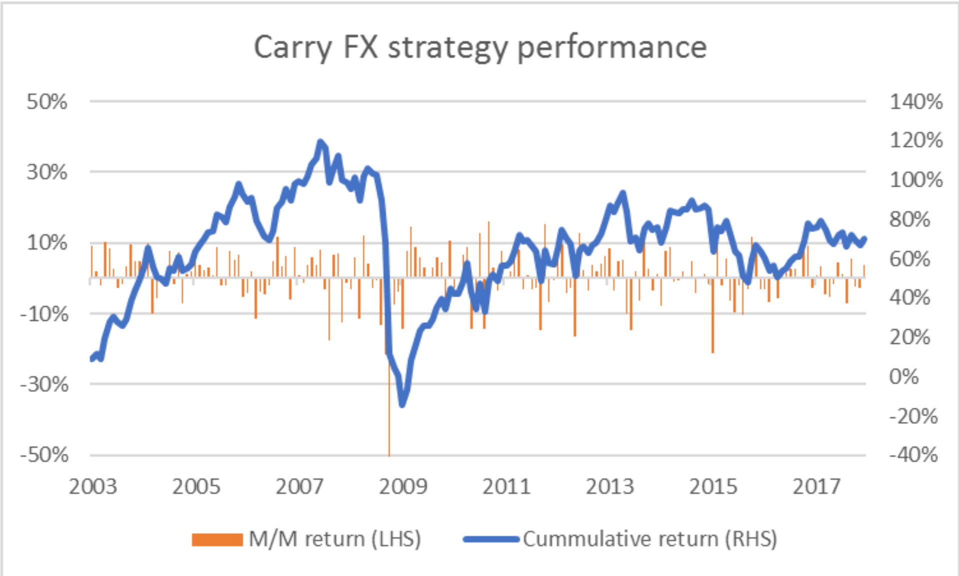

Figure B.4: Cumulative and month-over-month returns from January 2003 to December 2017, for a long top-three ranked and short bottom-three ranked currency strategy. Rank based on carry.

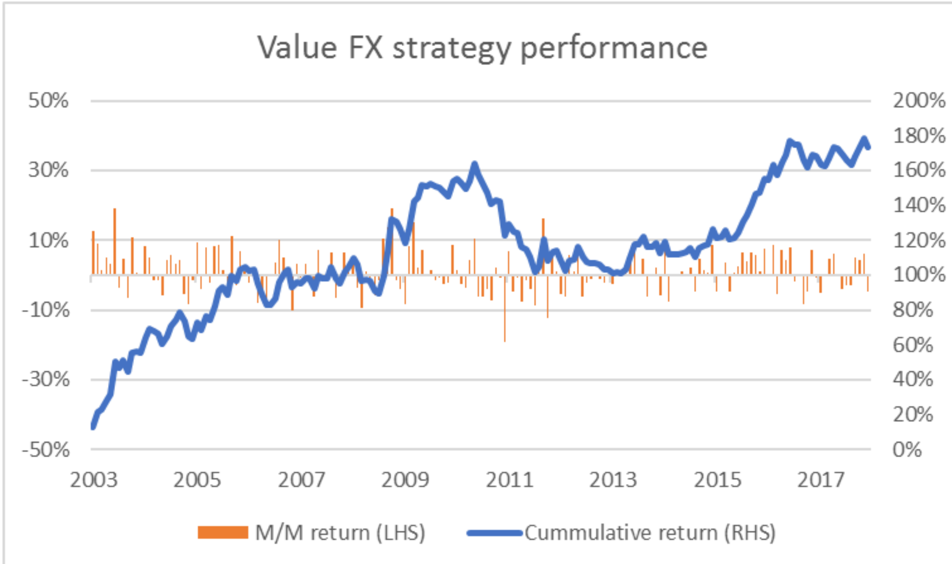

Figure B.5: Cumulative and month-over-month returns from January 2003 to December 2017, for a long top-three ranked and short bottom-three ranked currency strategy. Rank based on PPP value.

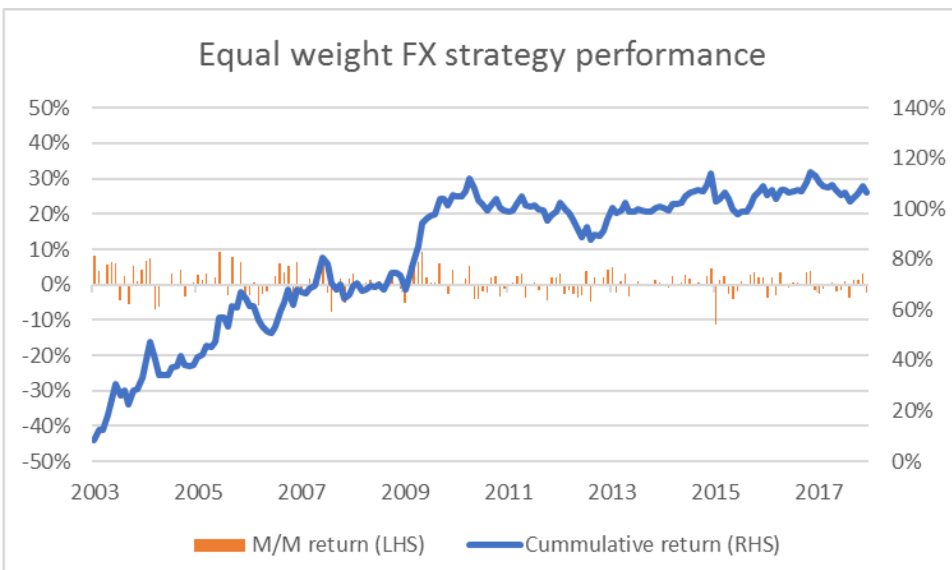

Figure B.6: Cumulative and month-over-month returns from January 2003 to December 2017, from averaging returns for a momentum, carry, and value long top-three ranked and short bottom-three ranked currency strategies.

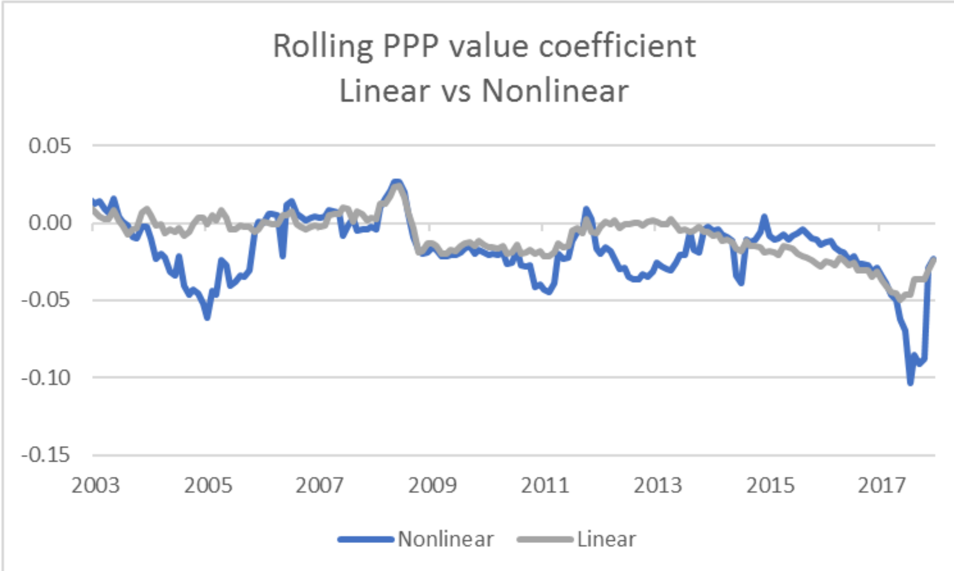

Figure B.7: Value coefficient estimated over a rolling window. “Nonlinear” denotes the value coefficient under the misvaluation regime in the PPP nonlinear model. “Linear” denotes the value coefficient in the linear regression with PPP as the value indictor.

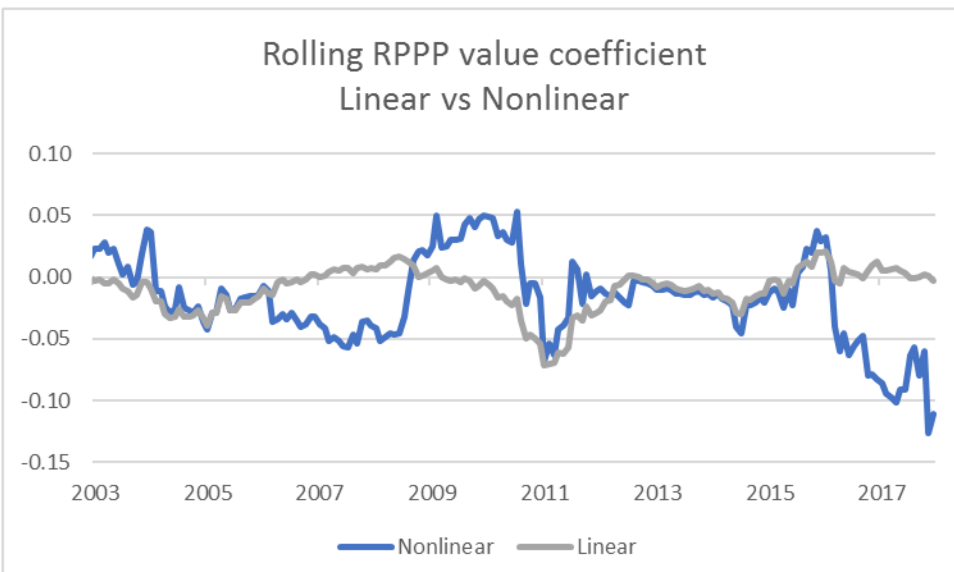

Figure B.8: Value coefficient estimated over a rolling window. “Nonlinear” denotes the value coefficient under the misvaluation regime in the relative PPP nonlinear model. “Linear” denotes the value coefficient in the linear regression with relative PPP as the value indictor.

APPENDIX C

Backtest results for equity timing model using a 1-month regression horizon (Section 3.3.5).

Table C.1: Summary statistics from a 1-month regression horizon model

Period: 1/1996 - 1/2018	Out-of-sample forecast summary statistics								
Model: 1m returns	**Mean**	**SD**	**Skew**	**Ex. Kurtosis**	**Min**	**Q1**	**Q2**	**Q3**	**Max**
Univariate: Composite value	0.9%	0.3%	-0.18	0.89	0.0%	0.8%	0.9%	1.1%	1.7%
Univariate: DY	0.7%	0.3%	1.06	1.21	0.0%	0.6%	0.7%	0.8%	1.6%
Univariate: PE	0.8%	0.4%	-0.19	-0.23	-0.3%	0.6%	0.8%	1.1%	1.7%
Univariate: PB	1.0%	0.3%	-0.01	0.59	0.1%	0.8%	1.0%	1.2%	2.0%
Univariate: CAPE	1.0%	0.3%	0.90	2.99	-0.1%	0.8%	0.9%	1.1%	2.3%
Univariate: Composite cycle	0.8%	0.8%	-4.63	24.32	-4.7%	0.8%	1.0%	1.1%	1.5%
Univariate: BCI	0.8%	0.8%	-4.94	27.46	-5.4%	0.8%	1.0%	1.2%	1.5%
Univariate: LEI	0.8%	0.7%	-4.35	22.93	-4.4%	0.8%	1.0%	1.1%	1.5%
Univariate: LMCI	1.1%	0.2%	0.36	-0.19	0.3%	0.9%	1.1%	1.2%	1.7%
Univariate: CAY	0.5%	0.4%	0.66	-0.34	-0.3%	0.2%	0.4%	0.7%	1.6%
Univariate: VRP	0.9%	1.4%	-5.37	69.62	-15.2%	0.4%	0.7%	1.4%	6.7%
OLS	0.2%	1.9%	-5.96	72.25	-22.5%	-0.5%	0.2%	1.0%	7.3%
Ridge	0.9%	0.4%	-5.18	57.22	-3.6%	0.7%	0.8%	1.1%	1.6%
Principal components	0.3%	1.8%	-5.50	64.66	-20.7%	-0.4%	0.2%	1.1%	6.9%
Average univariate OLS	0.8%	0.4%	-3.19	33.30	-3.2%	0.6%	0.8%	1.0%	2.1%
Expanding historical average	1.0%	0.2%	0.84	-0.58	0.7%	0.8%	0.9%	1.1%	1.4%

Table C.2: Correlation matrix from a 1-month regression horizon model

Period: 1/1996 - 1/2017	Out-of-sample forecast correlation matrix														
Model: 1m returns	**U1**	**U1a**	**U1b**	**U1c**	**U1d**	**U2**	**U2a**	**U2b**	**U3**	**U4**	**U5**	**OLS**	**Ridge**	**PCR**	**Avg U**	**Bench**
Univariate: Composite value (U1)	100%															
Univariate: DY (U1a)	59%	100%														
Univariate: PE (U1b)	20%	31%	100%													
Univariate: PB (U1c)	83%	60%	18%	100%												
Univariate: CAPE (U1d)	70%	53%	15%	83%	100%											
Univariate: Composite cycle (U2)	-3%	-5%	-34%	0%	3%	100%										
Univariate: BCI (U2a)	-2%	-6%	-35%	0%	3%	97%	100%									
Univariate: LEI (U2b)	-3%	-1%	-34%	0%	4%	96%	92%	100%								
Univariate: LMCI (U3)	37%	33%	7%	45%	51%	12%	14%	12%	100%							
Univariate: CAY (U4)	-11%	10%	5%	-1%	13%	8%	10%	9%	32%	100%						
Univariate: VRP (U5)	4%	3%	-15%	11%	19%	16%	25%	10%	14%	29%	100%					
OLS	13%	17%	-2%	13%	16%	39%	47%	34%	16%	23%	79%	100%				
Ridge	7%	4%	-24%	13%	12%	53%	60%	47%	19%	12%	73%	76%	100%			
Principal components (PCR)	1%	7%	-21%	8%	18%	44%	51%	37%	24%	40%	89%	87%	80%	100%		
Average univariate OLS (Av.U.)	16%	15%	-22%	21%	24%	62%	66%	58%	28%	19%	67%	72%	83%	77%	100%	
Expanding historical average (Bench)	-2%	3%	-19%	15%	32%	23%	22%	25%	27%	38%	33%	-3%	22%	31%	34%	100%
Fwd ret 1M	-15%	-5%	1%	-14%	-9%	8%	11%	4%	-8%	12%	13%	14%	4%	14%	12%	3%
Fwd ret 3M	-16%	-5%	0%	-13%	-8%	-1%	4%	-4%	-9%	9%	26%	25%	14%	23%	19%	2%
Fwd ret 6M	-17%	-11%	6%	-13%	-5%	-7%	-4%	-9%	-8%	16%	14%	14%	-3%	10%	7%	-3%
Fwd ret 9M	-18%	-14%	7%	-12%	-6%	-8%	-7%	-9%	-2%	19%	9%	11%	-4%	7%	1%	-10%
Fwd ret 12M	-21%	-19%	4%	-15%	-9%	-11%	-8%	-11%	-1%	20%	9%	10%	-5%	7%	1%	-15%

Table C.3: Summary of portfolio performance in a long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 1-month horizon.

Period: 1/1996 - 1/2018	Model performance												
Model: 1m returns, directionality only	**Return mean**	**Return SD**	**IR**	**SR**	**Mean 12m return**	**Med 12m return**	**Min 12m return**	**Max 12m return**	**% long position**	**Hit rate (HR)**	**HR|pos. return**	**HR|neg. return**	**RMSE**
Univariate: Composite value	8.8%	15.3%	0.58	0.43	8.2%	12.4%	43.4%	-56.3%	100%	65%	99%	0%	4.49%
Univariate: DY	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	4.49%
Univariate: PE	9.8%	15.2%	0.65	0.50	9.2%	12.4%	43.4%	-56.3%	98%	66%	99%	3%	4.48%
Univariate: PB	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	4.49%
Univariate: CAPE	8.8%	15.3%	0.58	0.43	8.2%	12.4%	43.4%	-56.3%	100%	65%	99%	0%	4.51%
Univariate: Composite cycle	10.5%	15.0%	0.70	0.55	9.9%	12.4%	45.6%	-35.3%	95%	64%	95%	7%	4.50%
Univariate: BCI	10.0%	15.0%	0.67	0.52	9.4%	12.4%	41.9%	-35.3%	94%	64%	95%	7%	4.48%
Univariate: LEI	10.5%	15.0%	0.70	0.55	9.9%	12.4%	45.9%	-35.3%	95%	64%	96%	5%	4.50%
Univariate: LMCI	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	4.50%
Univariate: CAY	7.5%	15.3%	0.49	0.35	6.8%	11.3%	43.4%	-56.3%	91%	62%	90%	9%	4.48%
Univariate: VRP	8.9%	15.3%	0.58	0.44	8.2%	12.1%	49.8%	-39.0%	93%	64%	94%	9%	4.52%
OLS	6.0%	15.2%	0.40	0.25	6.4%	8.9%	61.8%	-34.1%	59%	57%	62%	47%	4.66%
Ridge	9.0%	15.3%	0.59	0.45	8.3%	12.4%	41.9%	-51.4%	99%	65%	99%	1%	4.47%
Principal components	2.9%	15.5%	0.19	0.04	3.1%	1.2%	43.7%	-24.4%	58%	52%	58%	41%	4.61%
Average univariate OLS	10.5%	15.2%	0.69	0.55	9.8%	12.9%	43.6%	-44.9%	98%	66%	99%	3%	4.45%
Expanding historical average	9.0%	15.3%	0.59	0.45	8.3%	12.4%	43.4%	-56.3%	100%	65%	100%	0%	4.49%

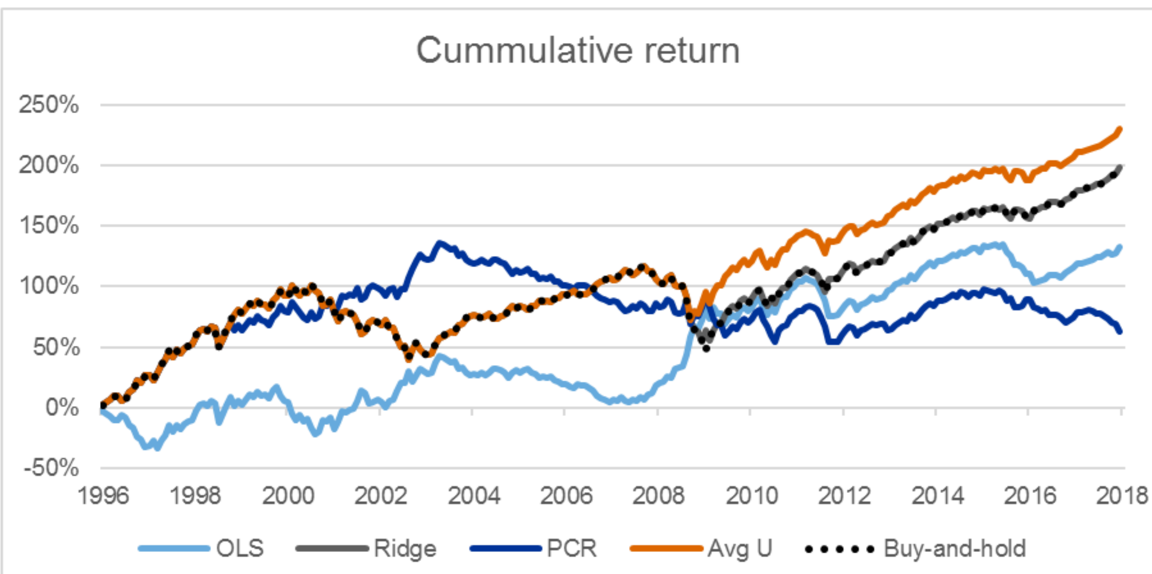

Figure C.1: Cumulative return of the long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 1-month horizon.

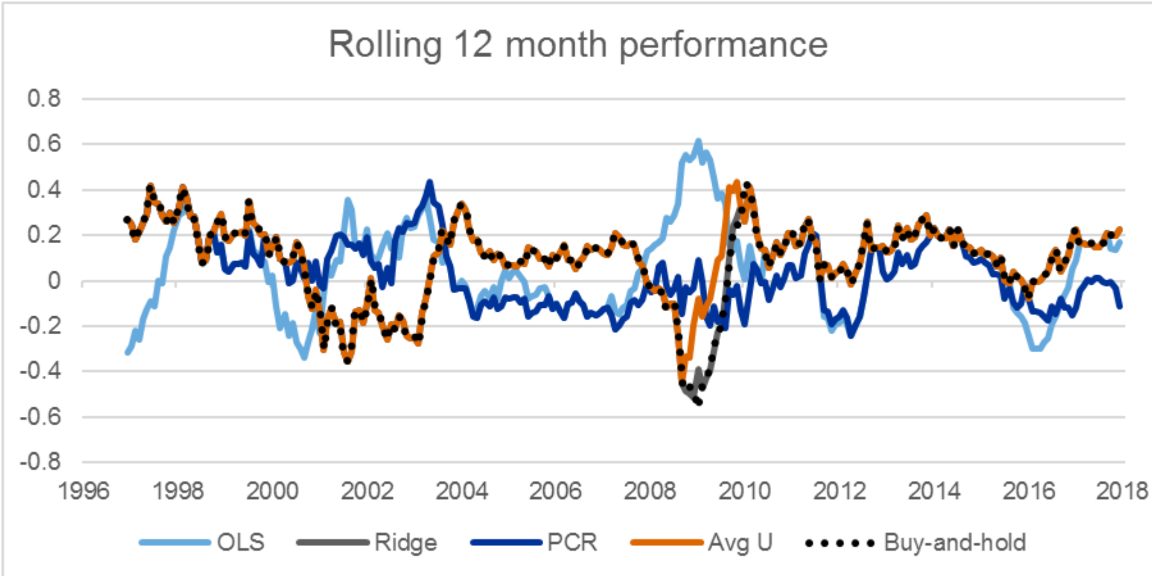

Figure C.2: Rolling 12-month performance of the long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 1-month horizon.

Table C.4: Summary of portfolio performance in a long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 1-month horizon.

Period: 1/1996 - 1/2018	**Model performance**												
Model: 1m returns	**Return mean**	**Return SD**	**IR**	**SR**	**Mean 12m return**	**Med 12m return**	**Min 12m return**	**Max 12m return**	**% long position**	**Hit rate (HR)**	**HR|pos. return**	**HR|neg. return**	**RMSE**
Univariate: Composite value	1.8%	3.4%	0.53	-0.11	1.8%	2.3%	11.8%	-13.0%	100%	65%	99%	0%	4.49%
Univariate: DY	1.5%	2.9%	0.51	-0.23	1.5%	1.8%	8.8%	-10.2%	100%	65%	100%	0%	4.49%
Univariate: PE	1.7%	3.2%	0.55	-0.14	1.7%	2.1%	12.1%	-16.3%	98%	66%	99%	3%	4.48%
Univariate: PB	2.1%	3.8%	0.57	-0.01	2.1%	2.6%	12.8%	-12.4%	100%	65%	100%	0%	4.49%
Univariate: CAPE	1.9%	3.8%	0.50	-0.07	1.9%	2.5%	11.9%	-11.8%	100%	65%	99%	0%	4.51%
Univariate: Composite cycle	2.3%	4.9%	0.47	0.02	2.1%	2.6%	21.4%	-20.5%	95%	64%	95%	7%	4.50%
Univariate: BCI	2.5%	4.9%	0.50	0.06	2.3%	2.8%	22.4%	-18.0%	94%	64%	95%	7%	4.48%
Univariate: LEI	2.0%	4.4%	0.45	-0.04	1.9%	2.5%	15.2%	-19.7%	95%	64%	96%	5%	4.50%
Univariate: LMCI	2.3%	3.9%	0.59	0.03	2.2%	2.8%	12.7%	-10.8%	100%	65%	100%	0%	4.50%
Univariate: CAY	1.3%	2.3%	0.55	-0.40	1.1%	0.8%	11.7%	-5.7%	91%	62%	90%	9%	4.48%
Univariate: VRP	3.9%	5.1%	0.75	0.33	3.9%	1.9%	24.4%	-8.4%	93%	64%	94%	9%	4.52%
OLS	3.2%	5.2%	0.61	0.20	3.5%	1.6%	29.3%	-6.7%	59%	57%	62%	47%	4.66%
Ridge	2.2%	3.4%	0.65	0.01	2.1%	2.3%	13.0%	-8.5%	99%	65%	99%	1%	4.47%
Principal components	3.4%	5.4%	0.63	0.22	3.5%	1.4%	23.1%	-6.0%	58%	52%	58%	41%	4.61%
Average univariate OLS	2.3%	3.2%	0.73	0.05	2.3%	2.1%	12.7%	-7.6%	98%	66%	99%	3%	4.45%
Expanding historical average	2.1%	3.6%	0.59	-0.02	2.0%	2.6%	12.8%	-10.0%	100%	65%	100%	0%	4.49%

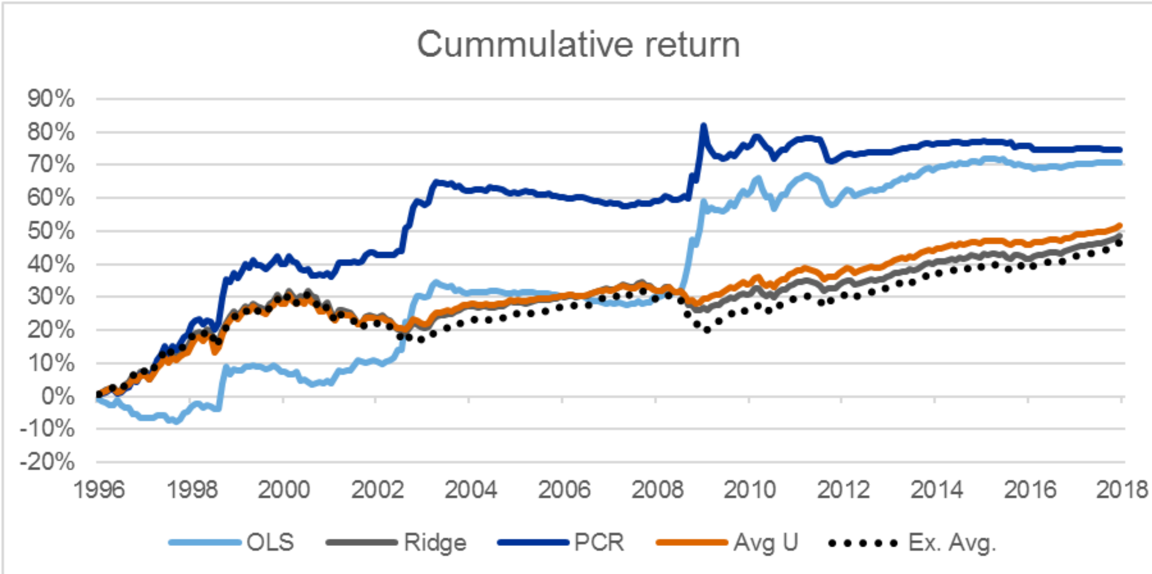

Figure C.3: Cumulative return of the long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts over a 1-month horizon.

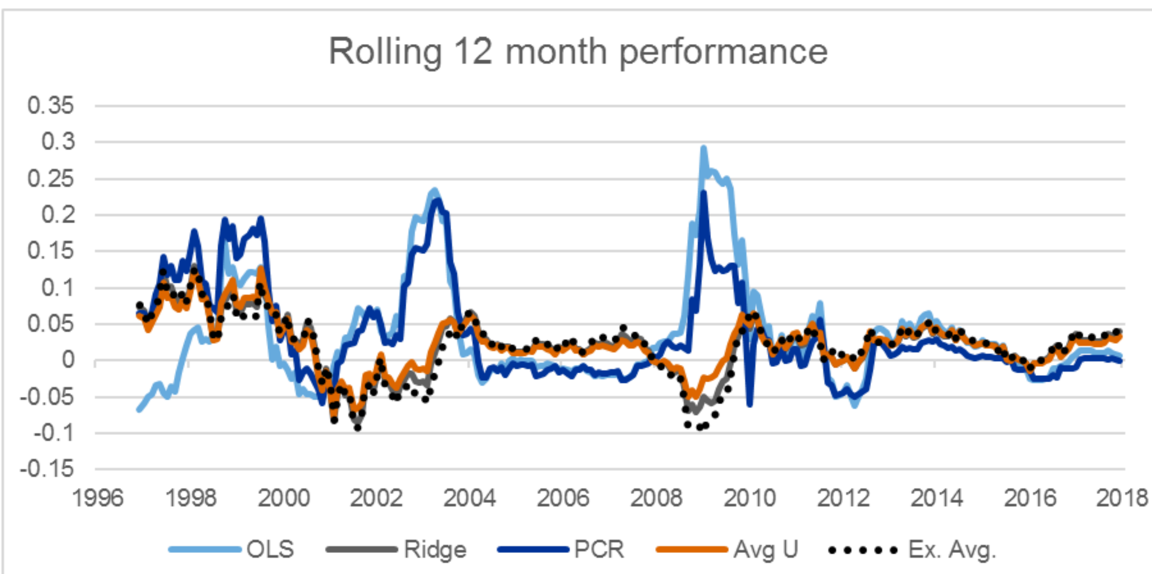

Figure C.4: Rolling 12-month performance of the long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 1-month horizon.

Table C.5: Sharpe and information ratio across various subsample periods, in a long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 1-month horizon.

Model: 1m returns, directionality only	Subsample Sharpe ratio 1/1996 - 12/2017	1/1996 - 12/2000	1/2001 - 12/2005	1/2006 - 12/2010	1/2011 - 12/2017	Subsample information ratio 1/1996 - 12/2017	1/1996 - 12/2000	1/2001 - 12/2005	1/2006 - 12/2010	1/2011 - 12/2017
Univariate: Composite value	0.43	0.85	-0.08	0.04	0.98	0.58	0.98	0.06	0.16	1.18
Univariate: DY	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: PE	0.50	0.74	0.34	0.04	0.98	0.65	0.87	0.48	0.16	1.18
Univariate: PB	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: CAPE	0.43	0.85	-0.08	0.04	0.98	0.58	0.98	0.06	0.16	1.18
Univariate: Composite cycle	0.55	0.92	-0.08	0.40	0.98	0.70	1.05	0.06	0.53	1.18
Univariate: BCI	0.52	0.92	-0.08	0.29	0.98	0.67	1.05	0.06	0.41	1.18
Univariate: LEI	0.55	0.92	-0.08	0.41	0.98	0.70	1.05	0.06	0.53	1.18
Univariate: LMCI	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: CAY	0.35	0.92	-0.64	0.04	1.11	0.49	1.05	-0.50	0.16	1.32
Univariate: VRP	0.44	0.81	0.44	0.25	0.29	0.58	0.93	0.59	0.37	0.47
OLS	0.25	-0.21	0.21	0.84	0.20	0.40	-0.09	0.36	0.97	0.39
Ridge	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Principal components	0.04	0.91	0.03	-0.39	-0.34	0.19	1.04	0.18	-0.27	-0.16
Average univariate OLS	0.55	0.92	-0.08	0.39	0.98	0.69	1.05	0.06	0.51	1.18
Expanding historical average	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18

Table C.6: Sharpe and information ratio across various subsample periods, in a long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 1-month horizon.

Model: 1m returns	Subsample Sharpe ratio 1/1996 - 12/2017	1/1996 - 12/2000	1/2001 - 12/2005	1/2006 - 12/2010	1/2011 - 12/2017	Subsample information ratio 1/1996 - 12/2017	1/1996 - 12/2000	1/2001 - 12/2005	1/2006 - 12/2010	1/2011 - 12/2017
Univariate: Composite value	-0.11	0.07	-0.57	-0.24	0.16	0.53	0.57	0.27	0.27	1.14
Univariate: DY	-0.23	0.18	-1.21	-0.52	0.00	0.51	0.66	-0.01	0.17	1.30
Univariate: PE	-0.14	-0.07	-0.41	-0.27	0.10	0.55	0.73	0.69	0.17	1.01
Univariate: PB	-0.01	0.16	-0.42	-0.11	0.23	0.57	0.60	0.31	0.36	1.11
Univariate: CAPE	-0.07	0.24	-0.58	-0.30	0.16	0.50	0.62	0.10	0.25	1.18
Univariate: Composite cycle	0.02	0.54	-0.80	-0.12	0.31	0.47	0.98	-0.05	0.16	1.22
Univariate: BCI	0.06	0.56	-0.70	-0.02	0.25	0.50	1.00	-0.01	0.26	1.19
Univariate: LEI	-0.04	0.54	-0.90	-0.31	0.32	0.45	0.97	-0.09	0.02	1.24
Univariate: LMCI	0.03	0.57	-0.56	-0.36	0.24	0.59	0.97	0.08	0.17	1.12
Univariate: CAY	-0.40	0.46	-3.04	-0.91	-0.93	0.55	0.99	-0.69	0.23	1.21
Univariate: VRP	0.33	0.92	0.24	0.30	-0.71	0.75	1.21	0.66	0.75	0.29
OLS	0.20	-0.24	0.55	0.63	-0.49	0.61	0.18	0.97	0.93	0.34
Ridge	0.01	0.66	-0.56	-0.52	-0.05	0.65	1.07	0.12	0.27	1.19
Principal components	0.22	0.81	0.48	0.15	-1.26	0.63	1.14	0.92	0.45	-0.16
Average univariate OLS	0.05	0.61	-0.54	-0.30	-0.09	0.73	1.03	0.28	0.54	1.13
Expanding historical average	-0.02	0.63	-0.65	-0.53	0.13	0.59	1.05	-0.01	0.13	1.23

APPENDIX D

Backtest results for equity timing model using a 3-month regression horizon (Section 3.3.5).

Table D.1: Summary statistics from a 3-month regression horizon model

Period: 1/1996 - 1/2018	Out-of-sample forecast summary statistics								
Model: 3m returns	**Mean**	**SD**	**Skew**	**Ex. Kurtosis**	**Min**	**Q1**	**Q2**	**Q3**	**Max**
Univariate: Composite value	2.8%	0.9%	0.58	0.91	0.8%	2.3%	2.7%	3.2%	5.5%
Univariate: DY	2.3%	0.9%	1.47	1.32	0.6%	1.8%	2.1%	2.3%	5.0%
Univariate: PE	2.6%	0.9%	-0.15	-0.46	-0.1%	1.9%	2.6%	3.3%	4.8%
Univariate: PB	3.2%	0.9%	0.65	1.00	1.0%	2.6%	3.1%	3.7%	6.3%
Univariate: CAPE	3.0%	1.1%	1.48	2.78	0.4%	2.5%	2.9%	3.1%	7.3%
Univariate: Composite cycle	2.4%	2.5%	-4.53	23.17	-14.1%	2.4%	2.9%	3.4%	4.2%
Univariate: BCI	2.4%	2.8%	-4.81	26.96	-18.7%	2.4%	2.9%	3.6%	4.4%
Univariate: LEI	2.4%	2.1%	-4.42	22.90	-11.3%	2.3%	2.8%	3.3%	4.5%
Univariate: LMCI	3.3%	0.7%	0.52	-0.36	1.4%	2.8%	3.2%	3.7%	4.9%
Univariate: CAY	1.4%	1.3%	0.56	-0.60	-0.9%	0.5%	1.0%	2.0%	4.4%
Univariate: VRP	2.8%	2.9%	-2.38	32.65	-25.2%	1.4%	2.2%	4.0%	17.2%
OLS	0.6%	4.5%	-3.63	36.73	-43.4%	-1.2%	0.5%	2.7%	18.0%
Ridge	2.0%	2.2%	-3.12	24.24	-17.3%	1.2%	1.8%	3.3%	7.5%
Principal components	0.9%	4.3%	-3.17	29.09	-38.3%	-0.7%	0.5%	3.1%	17.6%
Average univariate OLS	2.5%	1.1%	-0.96	9.56	-5.2%	2.0%	2.3%	3.0%	6.0%
Expanding historical average	3.0%	0.6%	0.84	-0.58	2.0%	2.5%	2.8%	3.4%	4.2%

Table D.2: Correlation matrix from a 3-month regression horizon model

Period: 1/1996 - 1/2017	Out-of-sample forecast correlation matrix														
Model: 3m returns	**U1**	**U1a**	**U1b**	**U1c**	**U1d**	**U2**	**U2a**	**U2b**	**U3**	**U4**	**U5**	**OLS**	**Ridge**	**PCR**	**Avg U**	**Bench**
Univariate: Composite value (U1)	100%															
Univariate: DY (U1a)	79%	100%														
Univariate: PE (U1b)	46%	12%	100%													
Univariate: PB (U1c)	86%	72%	37%	100%												
Univariate: CAPE (U1d)	83%	85%	12%	81%	100%											
Univariate: Composite cycle (U2)	-9%	12%	-35%	-4%	15%	100%										
Univariate: BCI (U2a)	-11%	8%	-37%	-5%	14%	97%	100%									
Univariate: LEI (U2b)	-11%	13%	-33%	-6%	14%	97%	92%	100%								
Univariate: LMCI (U3)	53%	65%	-4%	63%	69%	38%	37%	39%	100%							
Univariate: CAY (U4)	29%	53%	1%	34%	38%	17%	15%	20%	62%	100%						
Univariate: VRP (U5)	21%	32%	-21%	25%	37%	16%	22%	11%	35%	33%	100%					
OLS	8%	10%	-13%	20%	16%	48%	53%	40%	31%	18%	72%	100%				
Ridge	15%	32%	-27%	23%	36%	67%	69%	63%	58%	47%	77%	80%	100%			
Principal components (PCR)	8%	22%	-26%	17%	29%	57%	60%	52%	50%	42%	82%	85%	94%	100%		
Average univariate OLS (Av.U.)	33%	54%	-22%	40%	56%	61%	62%	58%	72%	60%	74%	69%	92%	86%	100%	
Expanding historical average (Bench)	34%	67%	-22%	28%	61%	41%	37%	44%	60%	59%	42%	2%	52%	39%	66%	100%
Fwd ret 1M	-5%	-3%	-3%	0%	0%	7%	12%	3%	-2%	6%	14%	18%	11%	14%	12%	1%
Fwd ret 3M	-1%	-3%	-1%	2%	-2%	2%	8%	-4%	-4%	7%	25%	27%	14%	19%	15%	0%
Fwd ret 6M	0%	0%	2%	6%	1%	-4%	3%	-10%	-4%	11%	15%	18%	3%	8%	8%	-4%
Fwd ret 9M	-1%	-3%	-1%	7%	0%	-6%	0%	-11%	-4%	13%	9%	14%	-2%	5%	5%	-8%
Fwd ret 12M	-7%	-8%	-9%	1%	-3%	-7%	-1%	-11%	-6%	12%	8%	14%	-3%	5%	4%	-12%

Table D.3: Summary of portfolio performance in a long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 3-month horizon.

Period: 1/1996 - 1/2018	**Model performance**												
Model: 3m returns, directionality only	**Return mean**	**Return SD**	**IR**	**SR**	**Mean 12m return**	**Med 12m return**	**Min 12m return**	**Max 12m return**	**% long position**	**Hit rate (HR)**	**HR|pos. return**	**HR|neg. return**	**RMSE**
Univariate: Composite value	9.0%	15.3%	0.59	0.45	9.0%	15.3%	59.1%	44.8%	100%	65%	100%	0%	8.16%
Univariate: DY	9.0%	15.3%	0.59	0.45	9.0%	15.3%	59.1%	44.8%	100%	65%	100%	0%	8.14%
Univariate: PE	9.1%	15.3%	0.60	0.45	9.1%	15.3%	59.7%	45.4%	100%	66%	100%	1%	8.13%
Univariate: PB	9.0%	15.3%	0.59	0.45	9.0%	15.3%	59.1%	44.8%	100%	65%	100%	0%	8.15%
Univariate: CAPE	9.0%	15.3%	0.59	0.45	9.0%	15.3%	59.1%	44.8%	100%	65%	100%	0%	8.23%
Univariate: Composite cycle	10.6%	15.0%	0.70	0.56	10.6%	15.0%	70.5%	55.9%	94%	64%	94%	8%	8.44%
Univariate: BCI	10.6%	15.0%	0.71	0.56	10.6%	15.0%	70.9%	56.4%	93%	64%	94%	9%	8.37%
Univariate: LEI	10.5%	15.0%	0.70	0.55	10.5%	15.0%	69.8%	55.3%	95%	64%	96%	5%	8.46%
Univariate: LMCI	9.0%	15.3%	0.59	0.45	9.0%	15.3%	59.1%	44.8%	100%	65%	100%	0%	8.24%
Univariate: CAY	6.3%	15.4%	0.41	0.27	6.3%	15.4%	41.1%	26.8%	89%	60%	88%	8%	8.11%
Univariate: VRP	8.1%	15.3%	0.53	0.39	8.1%	15.3%	53.0%	38.7%	96%	65%	97%	5%	7.90%
OLS	5.7%	15.3%	0.37	0.23	5.7%	15.3%	37.4%	23.1%	58%	57%	62%	48%	8.28%
Ridge	7.5%	15.4%	0.49	0.35	7.5%	15.4%	48.7%	34.5%	94%	64%	95%	7%	8.07%
Principal components	0.2%	15.6%	0.01	-0.13	0.2%	15.6%	1.4%	-12.6%	58%	50%	56%	38%	8.50%
Average univariate OLS	10.5%	15.1%	0.70	0.55	10.5%	15.1%	69.5%	55.1%	99%	66%	99%	2%	7.99%
Expanding historical average	9.0%	15.3%	0.59	0.45	9.0%	15.3%	59.1%	44.8%	100%	65%	100%	0%	8.17%

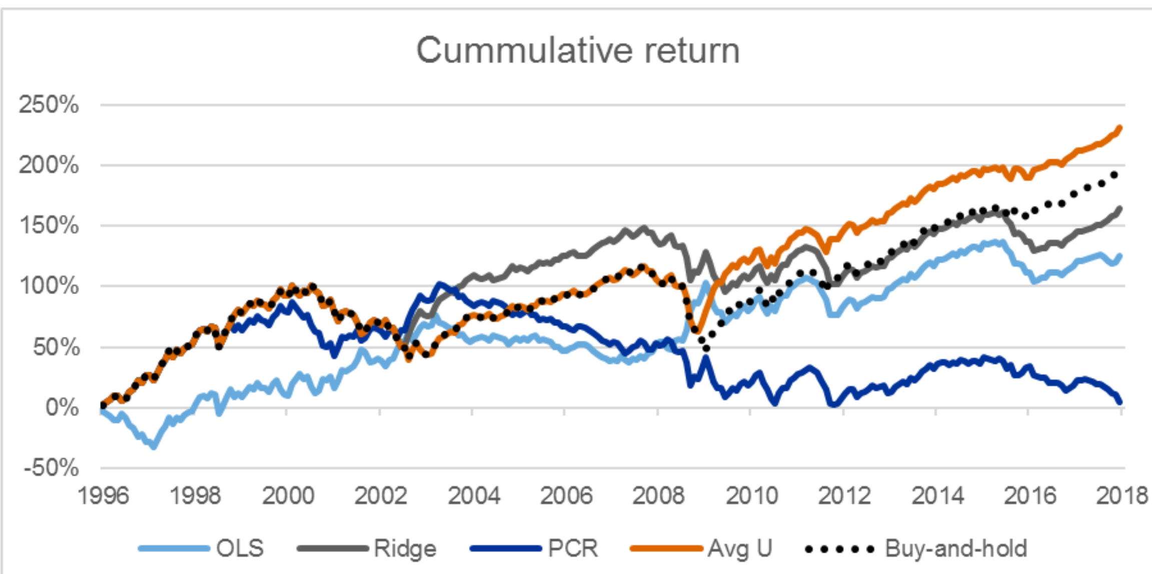

Figure D.1: Cumulative return of the long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 3-month horizon.

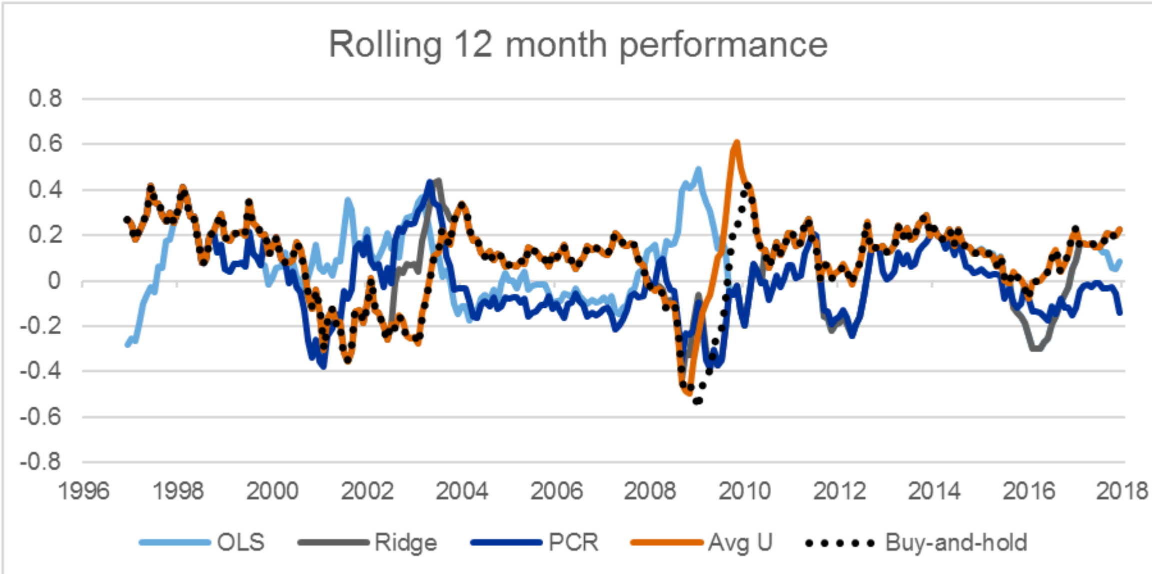

Figure D.2: Rolling 12-month performance of the long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 3-month horizon.

Table D.4: Summary of portfolio performance in a long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 3-month horizon.

Period: 1/1996 - 1/2018	**Model performance**												
Model: 3m returns	**Return mean**	**Return SD**	**IR**	**SR**	**Mean 12m return**	**Med 12m return**	**Min 12m return**	**Max 12m return**	**% long position**	**Hit rate (HR)**	**HR|pos. return**	**HR|neg. return**	**RMSE**
Univariate: Composite value	3.3%	6.3%	0.52	0.18	3.3%	4.4%	19.7%	-22.8%	100%	65%	100%	0%	8.16%
Univariate: DY	2.8%	5.6%	0.51	0.11	2.7%	3.5%	18.1%	-18.1%	100%	65%	100%	0%	8.14%
Univariate: PE	3.1%	5.7%	0.55	0.17	3.0%	4.0%	20.3%	-28.3%	100%	66%	100%	1%	8.13%
Univariate: PB	3.9%	6.9%	0.56	0.25	3.8%	4.9%	21.2%	-21.8%	100%	65%	100%	0%	8.15%
Univariate: CAPE	3.5%	7.1%	0.50	0.19	3.5%	4.5%	23.8%	-20.8%	100%	65%	100%	0%	8.23%
Univariate: Composite cycle	3.8%	7.0%	0.54	0.23	3.6%	4.5%	26.9%	-27.4%	94%	64%	94%	8%	8.44%
Univariate: BCI	4.3%	7.3%	0.59	0.29	4.0%	4.7%	30.4%	-26.3%	93%	64%	94%	9%	8.37%
Univariate: LEI	3.4%	6.9%	0.49	0.17	3.1%	4.1%	21.6%	-27.2%	95%	64%	96%	5%	8.46%
Univariate: LMCI	3.9%	6.9%	0.56	0.25	3.7%	4.9%	22.7%	-19.5%	100%	65%	100%	0%	8.24%
Univariate: CAY	2.2%	4.0%	0.55	0.01	2.0%	1.3%	19.8%	-9.3%	89%	60%	88%	8%	8.11%
Univariate: VRP	5.3%	7.7%	0.69	0.40	5.2%	3.2%	31.3%	-14.5%	96%	65%	97%	5%	7.90%
OLS	4.2%	6.9%	0.60	0.29	4.5%	2.4%	37.1%	-10.3%	58%	57%	62%	48%	8.28%
Ridge	3.9%	6.4%	0.61	0.27	3.8%	2.7%	25.5%	-13.4%	94%	64%	95%	7%	8.07%
Principal components	3.4%	7.4%	0.45	0.16	3.4%	0.7%	27.6%	-18.2%	58%	50%	56%	38%	8.50%
Average univariate OLS	3.9%	5.6%	0.69	0.30	3.7%	3.7%	22.0%	-13.2%	99%	66%	99%	2%	7.99%
Expanding historical average	3.6%	6.2%	0.58	0.22	3.3%	4.4%	21.9%	-17.5%	100%	65%	100%	0%	8.17%

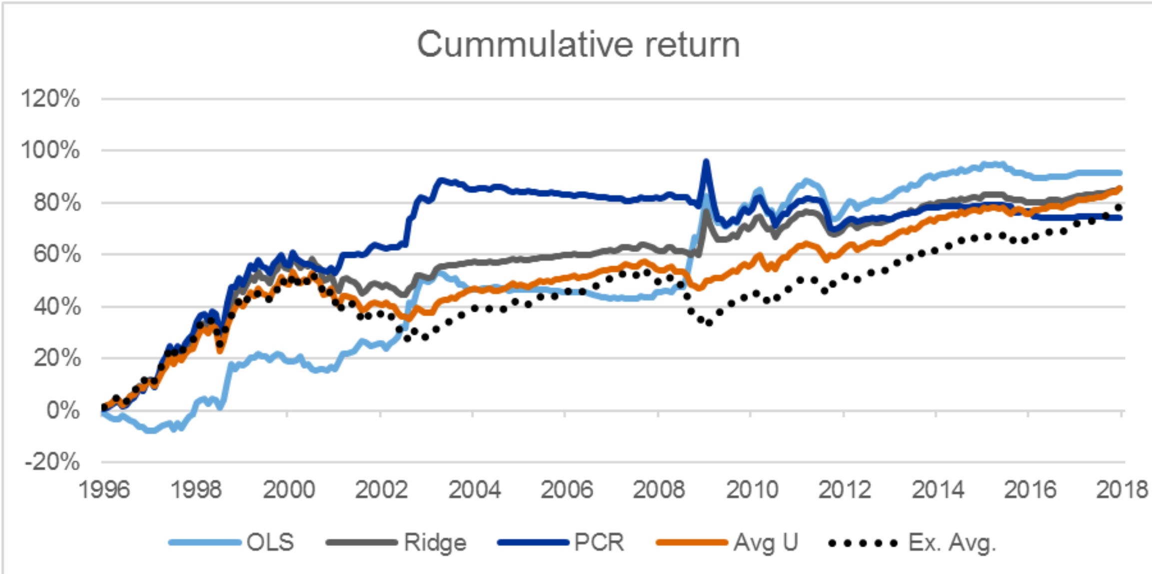

Figure D.3: Cumulative return of the long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 3-month horizon.

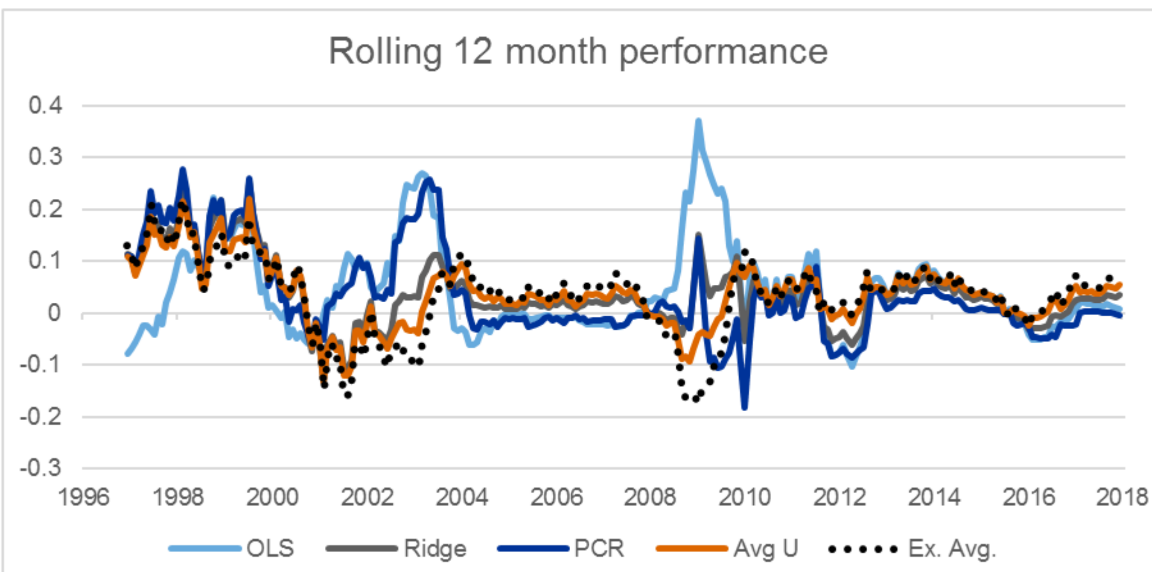

Figure D.4: Rolling 12-month performance of the long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 3-month horizon.

Table D.5: Sharpe and information ratio across various subsample periods, in a long-short equity strategy. I take a 100% long position if the model forecasts positive returns and a 100% short position if the model forecasts negative returns. Model forecasts are over a 3-month horizon.

	Subsample Sharpe ratio					Subsample information ratio				
Model: 3m returns, directionality only	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**
Univariate: Composite value	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: DY	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: PE	0.45	0.92	-0.06	0.04	0.98	0.60	1.05	0.09	0.16	1.18
Univariate: PB	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: CAPE	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: Composite cycle	0.56	0.92	-0.08	0.43	0.98	0.70	1.05	0.06	0.55	1.18
Univariate: BCI	0.56	0.92	-0.08	0.45	0.98	0.71	1.05	0.06	0.57	1.18
Univariate: LEI	0.55	0.92	-0.08	0.41	0.98	0.70	1.05	0.06	0.53	1.18
Univariate: LMCI	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18
Univariate: CAY	0.27	0.92	-0.85	0.04	0.95	0.41	1.05	-0.71	0.16	1.15
Univariate: VRP	0.39	0.81	0.19	0.26	0.29	0.53	0.93	0.34	0.38	0.47
OLS	0.23	0.17	0.13	0.50	0.11	0.37	0.30	0.28	0.63	0.29
Ridge	0.35	0.92	0.35	-0.11	0.28	0.49	1.05	0.49	0.01	0.46
Principal components	-0.13	0.49	0.03	-0.58	-0.43	0.01	0.61	0.18	-0.46	-0.24
Average univariate OLS	0.55	0.92	-0.08	0.41	0.98	0.70	1.05	0.06	0.53	1.18
Expanding historical average	0.45	0.92	-0.08	0.04	0.98	0.59	1.05	0.06	0.16	1.18

Table D.6: Sharpe and information ratio across various subsample periods, in a long-short equity strategy. I take a long position if the model forecasts positive returns and a short position if the model forecasts negative returns, where the magnitude of the position depends on the magnitude of the model forecast. Model forecasts are over a 3-month horizon.

	Subsample Sharpe ratio					Subsample information ratio				
Model: 3m returns	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**	**1/1996 - 12/2017**	**1/1996 - 12/2000**	**1/2001 - 12/2005**	**1/2006 - 12/2010**	**1/2011 - 12/2017**
Univariate: Composite value	0.18	0.42	-0.26	-0.06	0.54	0.52	0.67	0.20	0.23	1.11
Univariate: DY	0.11	0.49	-0.66	-0.26	0.54	0.51	0.73	-0.05	0.14	1.29
Univariate: PE	0.17	0.45	-0.05	-0.12	0.48	0.55	0.85	0.55	0.14	1.01
Univariate: PB	0.25	0.46	-0.16	0.05	0.59	0.56	0.68	0.25	0.32	1.10
Univariate: CAPE	0.19	0.50	-0.32	-0.11	0.58	0.50	0.69	0.07	0.21	1.17
Univariate: Composite cycle	0.23	0.73	-0.49	-0.04	0.66	0.54	0.99	-0.04	0.18	1.21
Univariate: BCI	0.29	0.77	-0.36	0.07	0.63	0.59	1.01	0.08	0.29	1.18
Univariate: LEI	0.17	0.69	-0.57	-0.17	0.67	0.49	0.94	-0.13	0.06	1.23
Univariate: LMCI	0.25	0.73	-0.33	-0.17	0.58	0.56	0.96	0.04	0.13	1.08
Univariate: CAY	0.01	0.71	-2.14	-0.48	-0.02	0.55	1.02	-0.76	0.19	1.20
Univariate: VRP	0.40	0.95	0.24	0.40	-0.35	0.69	1.15	0.54	0.68	0.24
OLS	0.29	0.18	0.54	0.57	-0.28	0.60	0.50	0.87	0.79	0.25
Ridge	0.27	0.86	-0.17	0.10	-0.22	0.61	1.08	0.31	0.41	0.57
Principal components	0.16	0.89	0.56	-0.32	-0.88	0.45	1.11	0.91	-0.08	-0.23
Average univariate OLS	0.30	0.78	-0.29	-0.01	0.39	0.69	1.02	0.19	0.49	1.11
Expanding historical average	0.22	0.79	-0.39	-0.26	0.59	0.58	1.03	-0.02	0.12	1.23