Empirical Studies on the Conditional CAPM and Equity Premium Prediction

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2021

Hyder Ali

Alliance Manchester Business School

Contents

Word Count: 82,466

List of Tables

List of Figures

Abstract

The University of Manchester
Hyder Ali
Doctor of Philosophy
Empirical Studies on the Conditional CAPM and Equity Premium Prediction
May 2021

This thesis consists of three self-contained essays that analyse two important subjects in Empirical Finance: the Conditional Capital Asset Pricing Model (CCAPM) and out-of-sample equity premium prediction. The first two essays concern the CCAPM model and analyse the choice of variables used to capture the time variation in the risk loadings. The lack of a theory to guide the choice of conditioning variables, and the rather large pool of potential variables that have been identified in the CCAPM literature, creates an empirical dilemma over how to optimally parameterise the model.

The first essay considers a dynamic model selection (DMS) approach where the choice of conditioning variables, selected from a large pool of state variables, is allowed to vary through time rather than remaining fixed. We find that estimating the CCAPM using the DMS method can improve the performance in some asset pricing tests, however, it still fails to explain the value and momentum anomalies. Using bootstrap methods to quantify the model uncertainty and instability, we find that the DMS selection of conditioning variables is subject to considerable estimation error. This provides strong motivation for our second essay, where we consider alternative forecasting approaches which try to address this variable-selection uncertainty (VSU). We implement combination of forecasts (CF), and combination of information (CI) approaches to capture the beta dynamics. CF combines forecasts generated from simple models, each incorporating a part of the whole information set, while CI brings the entire or selected information set into one single model to generate an ultimate forecast. Our findings suggest that CF approaches dominate the CI approaches in explaining the cross-section of assets returns. However, we also demonstrate that further improvements in results are possible by combining the CI and CF methods.

The topic of the third essay concerns the predictability, or otherwise, of the equity premium. In this essay, we use some of the techniques developed in earlier chapters of the thesis, such as CF and CI methods, in order to select the best conditioning variables for predicting market excess returns. In particular, we focus on the issue of parameter instability (PI) in predictive models caused by abrupt changes in financial market conditions which result in structural breaks in the underlying relationship between the variables in the model. Since standard forecasting models assume that the relationship between these variables remains constant over the entire period, any parameter instability, therefore, can lead to poor out-of-sample performance (e.g., Rapach and Wohar, 2006; Paye and Timmermann, 2006; Rapach et al., 2010). Here, we introduce a novel approach to predicting returns which uses a combining forecasts (CF) approach with a variance-covariance (VC) method that addresses PI and VSU. The essay has two main findings: i) by taking into account the correlation structure among forecast errors through our VC approach, the forecasting accuracy of univariate prediction of the equity premium significantly improves, and ii) by addressing PI and VSU simultaneously the VC approach can substantially improve the forecasting accuracy compared to existing approaches in equity premium such as CF (Rapach et al., 2010), CI approaches such as dimension reduction methods (Neely et al., 2014 and Kelly and Pruitt, 2013) and shrinkage methods such as Least Absolute Shrinkage and Selection Operator (LASSO), (Tibshirani, 1996), adaptive LASSO (Zou, 2006), and Elastic Net (Zou and Hastie, 2005).

Declaration

I, Hyder Ali, declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Dedication

I dedicate this thesis to my father, Aziz Ahmed Khawaja and my mentor, Professor Nisar Ahmed Siddiqui. Both of these individuals remained my inspiration throughout my PhD. Sadly, I lost both of them during my PhD journey.

”You are gone but your belief in me has made this journey possible...”

Acknowledgement

This PhD has been, undoubtedly, a fascinating journey of my life. This journey has been made possible by God through so many people, and the two most important of them are my supervisors, Dr. Alex Taylor and Prof. Ian Garrett. They gave me a chance to study at my dream university and provided me with immense guidance and knowledge. They helped me shape my ideas and develop a great understanding of my research area, thus developing the level of confidence that would surely make me a better researcher and person. Special thanks to Dr. Alex Taylor, who kept me motivated throughout my PhD path by advising me on both academic and health challenges. His detailed and constructive feedback on my writing drove me to write more effectively, to the point where I did not need to revise anything post viva, resulting in a pleasurable viva experience with no delay between the viva and receiving the title of "Dr."

I would also like to acknowledge the contribution of Prof. Tolga Cenesizoglu of HEC Montréal, Canada, who was my main supervisor in the first year of my PhD. He not only assisted me in refining my research ideas, but also supported me in overcoming my stress following the death of my father during the early stages of my PhD. I would like to thank other academic faculty members of AMBS, who provided constructive comments during my mid-year and annual reviews which helped me refine my research ideas and methodological skills. Mr. Mark Falzon deserves special recognition for his assistance with all administrative concerns during the degree programme. Moreover, I would also like to thank Sukkur IBA University, Pakistan and the University of Manchester, UK for giving me the financial support to undertake my PhD study.

Furthermore, I would like to thank my family for their selfless support. Especially I would like to express my gratitude to my wife, for encouraging and being with me in all the years it took to complete this journey. Without her support both morally and emotionally, this dream would have not come true. I would like to extend my thanks to my teachers, colleagues and friends whose guidance and teachings enabled me to attain this level that I am today.

Finally, I would like to thank my examiners, Prof. John Cotter of University College Dublin, Ireland and Prof. Stuart Hyde of the University of Manchester, UK for an intellectually stimulating and enjoyable viva. Their insightful comments taught me a lot, not only in terms of getting my work published but also in terms of understanding my strengths and weaknesses, which will serve me well throughout my life.

Chapter 1

Introduction

1.1 Motivation

A predictive regression is an extensively used econometric tool to assess the predictability of economic or financial variables by past values of one or more variables. An example of such a predictive regression is given in equation (1.1) which can be used to assess whether current information (X_t) can predict future stock returns (r_{t+1}).

$$r_{t+1} = \alpha + \beta X_t + u_{t+1} \tag{1.1}$$

When applied to aggregate market returns, these models provide forecasts of the equity premium. Such forecasts are extensively used in various empirical finance areas, most notably in leading asset pricing models (e.g., Campbell & Cochrane 1999, Bansal & Yaron 2004, Frazzini & Pedersen 2014).

In addition to directly providing estimates of the time-variation in the equity premium, the ever-growing literature on predictor variables has provided researchers with a large pool of potential conditioning variables to use in other asset pricing tests.[1] Relevant to this thesis is the case of the conditional CAPM (CCAPM), which is studied in Chapters 2, and 3. The primary requirement in this type of analysis is to accurately model time-varying betas (systematic risk, β) as a linear function of observable conditioning variables (X_t), where $\beta_t = b_0 + b_1 X_t$. As a result, the following regression model emerges:[2]

$$R_{t+1} = \alpha_0 + b_0 R_{m,t+1} + b_1 R_{m,t+1} X_t + u_{t+1} \tag{1.2}$$

where $E(u_{t+1}) = E(u_{t+1}[X_t R_{m,t+1}]) = 0$. The CCAPM implies that $\alpha_0 = 0$. This version of CCAPM given in equation (1.2) is also called CCAPM-IV because it uses predictor variables or instrumental variables (IVs). As mentioned, the CCAPM-IV approach draws heavily on the same predictors that have been reported as predictors of aggregate stock returns, the predictors identified by Goyal & Welch (2008) being a common choice.

A literature review evaluating the performance of predictive regressions suggests that using equity premium predictors can be criticised given the identification of predictor variables is econometrically challenging, leading to spurious identification of conditioning variables. These

[1]See Goyal & Welch (2008) for traditional predictor variables and some of the new predictors include technical indicators (Neely et al. 2014), investor sentiment and attention (Huang et al. 2015, Chen 2017), the short interest index (Rapach et al. 2016), and credit quality (Chava et al. 2015), among others.

[2]The conditional alpha (intercept) is also time-varying in some studies, as $\alpha_t = \alpha_0 + \alpha_1 X_t$ (e.g., Christopherson et al. 1998). As a result, the following regression model emerges, which makes the model, $R_{t+1} = \alpha_0 + \alpha_1 Z_t + b_0 R_{m,t+1} + b_1 R_{m,t+1} X_t + u_{t+1}$.

challenges include the issues with persistence leading to bias (Stambaugh 1999), inability to predict out-of-sample (Goyal & Welch 2008), model instability (Paye & Timmermann 2006, Rapach & Wohar 2006) and data mining and overfitting (Ferson et al. 2003). Moreover, recent advances in information technology allow hundreds of economic variables to be obtained in real-time. Using multiple economic predictors to forecast a target variable has become a recent trend in econometric research that seeks to exploit such data-rich environments (Wang et al. 2020). The lack of theoretical guidance on which subset of variables should be chosen from the prethora of predictors creates an empirical dilemma over how to optimally parameterise the model. We refer to this problem as *variable-selection uncertainty* (henceforth VSU). The VSU issue is important to address in the presence of many predictors because some recent studies have stressed that too many predictors can adversely affect a model's forecasting performance (e.g., Boivin & Ng 2006).[3]

Our strategies to address the issue of VSU from the CCAPM-IV perspective in Chapters 2 and 3 are motivated by equity premium prediction (EPP) literature. In the fourth chapter, I apply some of the techniques developed in Chapters 2 and 3 to make a contribution to the EPP literature. In the EPP literature review, I identify a significant research gap in out-of-sample EPP concerning the selection of both the predictor variables and optimal estimation window in out-of-sample forecasting regressions. Our motivation is based on the findings of some studies in macroeconomics and finance, relating the parameter instability to forecast failure (e.g., Stock & Watson 1996, Pesaran et al. 2006, Inoue & Rossi 2011). Parameter instability is due to structural breaks triggered by various factors like extreme events, significant changes in financial market conditions, presidential elections, regime switches in monetary policies, business changes, new technology, and significant changes in government regulations. In particular, in the predictive relationship between different economic variables and stock returns, Rapach & Wohar (2006) find clear evidence of structural breaks. As shown by Pesaran & Timmermann (2007) the performance of a forecasting model when structural breaks are present depends on the number of observations (window length) used to estimate the out-of-sample forecast. However, there is no clear consensus in the literature on the number of observations to be used in estimation, which is usually referred as *estimation window uncertainty* (EWU) (Pesaran & Timmermann 2007). Due to this issue, it is recommended rather than including all available observations for estimating the parameters, only the most recent observations be used (the so-called "rolling estimation" method). However, most of the existing forecasting strategies in EPP use an expanding window method, for example, Rapach et al. (2010) use combining forecast (CF) technique, whereas Neely et al. (2014) use combining information (CI) approach. Both of these strategies use a recursive expanding window, which uses all the observations available and, as a result, will be non-optimal in the presence of structural breaks. There are few studies which use rolling window approach, for example, most recently Li & Tsiakas (2017) and Yin (2020) use rolling window in implementing the shrinkage approaches such as Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (ENet) to predict equity premium.

In most of the above literature, the rolling window size is arbitrarily selected or supported by the results of past studies. However, the forecasting accuracy of the rolling window approach is found to be sensitive to the choice of window size (e.g., Pesaran & Timmermann 2007, Inoue et al. 2017). This implies that though the existing forecasting strategies (such as combining forecasts, combining information, and shrinkage methods) using either expanding window or rolling window approaches account for VSU but do not choose the window optimally, as a re-

[3] There are a variety of reasons why a model should be limited to a small number of predictor variables. The most apparent is that every single variable cannot explain the target variable, so any variable that is not related to it should be omitted (ontological sparsity). Second, even though all variables may explain the target variable, it is preferable to remove variables with minor effects, either to improve the final model's interpretability (epistemic sparsity) or to improve the model's predictive ability through reducing variance (predictive sparsity).

sult, fail to address the issue of EWU. In Chapter 4, we address this by using techniques to choose both the optimal variables and optimal window. Therefore our new forecasting approach addresses both VSU (choice of variable) and EWU (choice of window) simultaneously to improve the out-of-sample forecasts of the equity premium.

1.2 Contribution

This thesis contributes to the literature on explaining the cross-section of returns (Chapters 2 and 3) and forecasting the equity premium (Chapter 4). The following subsections provide an overview of each of the three studies.

1.2.1 Chapter 2: Conditional CAPM with Dynamic Model Selection (DMS) Approaches

In the first essay, we introduce a CCAPM model where the choice of conditioning variables, used to capture the variation in conditional betas, is allowed to vary through time and is selected from a large pool of potential state variables.[4] Under our approach, the subset of conditioning variables selected at time t is based on a pre-test procedure that uses past information to decide whether a predictor is 'in' or 'out'. Specifically, this approach selects the beta models that perform the best based on standard asset pricing criteria on past data at each point in time. We call this approach as *dynamically selected beta model* (DSBM).

In addition to our DSBM, we also apply some of the popular variable selection approaches found in the mainstream literature but to the best of our knowledge have not been applied to CCAPM-IV. These methods include: i) best subset selection, ii) sequential selection, and iii) shrinkage methods. For best subset selection the variables are chosen using various criteria including, adjusted R^2, Akaike information criterion (Akaike 1973), Bayesian information criterion (Schwarz 1978), and Mallows's C_P (Mallows 1973). The sequential selection approaches include forward selection, backward elimination, and stepwise regression. The shrinkage methods include the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996), Adaptive LASSO (Zou 2006), and Elastic Net (ENet) (Zou & Hastie 2005). To test the validity of CCAPM-IV using dynamic model selection approaches, we examine whether CCAPM-IV models based on time-varying conditioning information explain the cross-section of asset returns of 25 Size and Value portfolios. We use the two-pass regression framework of Fama & MacBeth (1973).

By evaluating conditional versions of the CAPM (CCAPM) through modelling a new type of time variation in conditional betas, our first essay contributes to the empirical asset pricing literature. Specifically, we complement the existing CCAPM literature that has focused on capturing beta dynamics using state variables to explain several patterns in the cross-section of stock returns like size, value, and momentum anomalies.[5] Our work is also closely related to studies such as Harvey (2001), and Cooper & Gubellini (2011), who find that the estimation of CCAPM is sensitive to choosing state variables. We also complement the literature by providing evidence against CCAPM (e.g., Lewellen & Nagel 2006) for its failure to explain the value and momentum anomalies. Finally, using bootstrap methods to quantify the model uncertainty and instability, we find that the DMS approaches of selecting conditioning variables are subject to considerable estimation error. This finding provides strong motivation for our second essay,

[4] We use 14 variables of Goyal & Welch (2008) for which monthly data are available from July 1926 to December 2018.

[5] A partial list includes Jagannathan & Wang (1996), Ferson & Harvey (1999), Lettau & Ludvigson (2001), Petkova & Zhang (2005), Cederburg & O'Doherty (2016) and others.

where we consider alternative forecasting approaches which try to address this variable-selection uncertainty (VSU).

1.2.2 Chapter 3: Conditional CAPM under Variable-selection Uncertainty (VSU)

In the literature on forecasting we generally find three approaches to account for VSU: the combination of forecasts (CF) (e.g., Bates & Granger 1969), the combination of information (CI) (e.g., Kelly & Pruitt 2013), and one that combines both CF and CI (e.g., Huang & Lee 2010). CF combines forecasts obtained from simple models where each incorporates a part of the whole information set, CI, on the other hand, brings the entire information set into one single model to generate a single optimal forecast (Huang & Lee 2010).[6] After reviewing the literature, I find that there are many alternative approaches to implement CI and CF. However, there is no clear consensus on which method is the best. Most studies are concerned with equity premium and macroeconomic prediction, but none of them examines the optimal approach suitable for modelling the time variation of betas in the CCAPM framework. Therefore we aim to compare various CI, CF, and a hybrid of CI and CF approaches in explaining the cross-section of asset returns within a CCAPM-IV framework. More specifically, the CF analysis combines point forecasts of betas estimated from univariate predictor-based regressions from a large pool of conditioning variables. At each point in time, these beta forecasts are weighted in various ways, including simple equally-weighted average and weighting schemes based on some criteria such as mean squared forecast error (MSFE).

In our CI approach, we use dimension reduction methods which include Principal Components (PCs) (Bai & Ng 2002) and Kelly & Pruitt (2013) three pass filter based on partial least squares (PLS). These approaches take the original pool of predictors and reduce it down to a small subset of variables known as factors. These factors are then used to fit the time-varying beta model. In approaches that combine CI and CF, our approaches include principal component combinations of Chan et al. (1999) and Huang & Lee (2010), variable selection and combination through shrinkage methods (Rapach & Zhou 2020). Moreover, we also consider the bootstrap aggregation (bagging or BAGG) technique which creates new training sets through bootstrap (e.g., Rapach & Strauss 2010). We draw B random samples with replacement from the original training set. For each bootstrap sample, we apply various CI, CF, and combinations of CI and CF approaches and obtain a forecast, and finally, we take an equally weighted average across B forecasts to obtain the final forecast.

Our second essay contributes to the literature in the following ways. First, to our knowledge, this is the first research to include a detailed comparison of various well-known approaches to dealing with VSU from a CCAPM perspective. Our out-of-sample results suggest that CF approaches dominate the CI approaches in explaining the cross-section of assets returns. Finally, consistent with studies as Hirano & Wright (2017) and Rapach & Zhou (2020), we show that a combination of conventional econometric methods and machine learning methods can outperform the individual methods. For example, we find the evidence on improved performance of CCAPM-IV with BAGG method where, in each pseudo sample, we first select the subset of variables based on the mean squared forecasting error (MSFE) in cross-validation sample, and then take a simple average of beta estimates across all pseudo samples. This method performs

[6] CI, is generally referred to dimension reduction, which is the transformation of data from a high dimensional space into a low-dimensional space such that any meaningful properties of the original data are preserved in the low-dimensional representation. However, we also include model selection approaches from Essay 1 under this category because either subset variable selection or dimension reduction would ultimately result in one model to generate the final forecast. The CF approach, on the other hand, always generates multiple forecasts for the same target variable and combines them into a composite forecast.

as well as the Fama & French (1993) three-factor model in explaining the cross-sectional returns of 25 Size-B/M, 30 industry and 10 momentum portfolios.

1.2.3 Chapter 4: Equity Premium Prediction under Variable-selection Uncertainty (VSU) and Parameter Instability (PI)

In the third essay, we propose a new combining forecasts (CF) approach based on a variance-covariance method that addresses *estimation window uncertainty* (EWU) and *variable-selection uncertainty* (VSU) simultaneously to improve the out-of-sample forecasts of the equity premium. A common strategy to handle EWU in the presence of structural breaks is to estimate breaking dates and use post-break observations for parameter estimation and forecast generation (see Bai & Perron 1998). However, Pesaran & Timmermann (2007) criticise this approach and show that due to limited post-break data, this approach introduces high estimation uncertainty which adversely affects the forecast accuracy measured as MSFE. They emphasise the importance of pre-break data in producing accurate forecasts and demonstrate that it can be useful to consider combining forecasts generated by the same model but over different estimation windows (Pesaran & Timmermann 2007, Pesaran et al. 2013, Tian & Anderson 2014, Tian & Zhou 2018). Despite improved forecasting performance, these approaches do not consider the correlation among forecasting errors and simply combine forecasts using simple average or weighted by MSFE. In chapter 4, we consider the possibility that a variance-covariance (VC) combination method which has been widely used in the CF literature (e.g., Bates & Granger 1969, Newbold & Granger 1974, Figlewski 1983, Cang & Yu 2014, and others), may improve forecasts of the equity premium.

The VC approach emphasises the consideration of correlation among forecasting errors, and the optimal weights are obtained as a solution to minimising the error variance-covariance matrix. It has been shown that VC can provide diversification effect and improve forecast accuracy (Bates & Granger 1969). Therefore, we aim to contribute to the literature by implementing the VC approach to obtain the optimal out-of-sample forecast for a particular economic predictor-based model based on multiple windows.

Moreover, Pesaran et al. (2013) show that estimation window uncertainty (EWU) and variable-selection uncertainty (VSU) are relevant problems for predicting macroeconomic and financial variables and introduced a new approach called average-average (AveAve). They argue that the two differently used approaches based on a simple average for accounting VSU (forecasts from various models, all estimated on a single window, are averaged, AveM) and EWU (calculated as the averages of forecasts generated from the same model over multiple windows, AveW) can be combined into one (AveAve). They show that out-of-sample "AveAve" forecasts outperform the AveM as well as the AveW forecasts. However, most of the equity premium literature considers the VSU and EWU as two different issues. For example, Rapach et al. (2010) account for VSU by taking a simple average across individual predictive models (AveM) and ignores EWU as all individual models were estimated using an expanding window. On the other hand, Tian & Zhou (2018) apply five alternative methods for directly dealing with EWU for various univariate and multivariate models based on Goyal & Welch (2008) predictors to forecast the equity premium; however, they do not consider VSU. This provides us with an opportunity to contribute the existing literature by implementing the VC approach to address the VSU and EWU issues simultaneously in forecasting out-of-sample equity premium.

Hence our third essay contributes to the existing literature by complementing the existing literature on methods for directly dealing with EWU in forecasting such as Pesaran & Timmermann (2007), Rossi (2013), Wang et al. (2020) and others. This is the first study to the best of our knowledge, to apply the VC approach for combining estimation windows of individual models.

We show that considering the correlation among forecast errors across estimation windows can significantly improve the forecasting accuracy of individual models. Secondly, for the first time in the forecasting literature, we introduce a panel combination approach based on VC approach to address the model uncertainty and parameter instability simultaneously. Based on out-of-sample forecasting results of the equity premium, we show that our new model not only outperforms the existing AveAve approach of Pesaran et al. (2013) but also existing approaches in equity premium such as CF (Rapach et al. 2010), dimension reduction methods (Kelly & Pruitt 2013, Neely et al. 2014) and shrinkage methods (Zhang et al. 2020).

1.3 Thesis structure

The thesis follows the journal format structure accepted by the Manchester Accounting and Finance Group, Alliance Manchester Business School, at the University of Manchester, United Kingdom. It facilitates the integration of chapters into a format acceptable for submission and publication in peer-reviewed scholarly journals. This thesis is therefore based around three empirical essays containing original studies in chapters 2, 3 and 4 on "Conditional CAPM (CCAPM) and Equity Premium Prediction." The chapters are self-contained, i.e. each chapter has a separate examination of literature, addresses distinct and original questions. There are independent equations, footnotes, charts, and figures and are numbered from each chapter's beginning. Throughout the thesis, page numbers, titles, and subtitles have a sequential order.

The thesis continues as follows. Chapter 2 presents empirical tests of the conditional CAPM, where we implement dynamic model selection approaches to capture the beta dynamics of assets. Chapter 3 extends Chapter 2 by implementing combining information (CI), combining forecasts (CF), and a hybrid of CI and CF approaches to capture beta dynamics. Chapter 4 introduces a new combing forecasts (CF) approach based on the variance-covariance (VC) method to address EWU and VSU issues for improving out-of-sample forecasts of the equity premium. Finally, Chapter 5 concludes and provide future directions for further research. The first person plural (we, our) is used in Chapters 2 to 4 instead of the first person singular (I, my), since I plan to consider some of the work for publication with my supervisor.

References

Akaike, H. (1973), Information theory and an extension of maximum likelihood principle, *in* 'Proceedings of Second International Symposium on Information Theory', pp. 267–281.

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. & Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Bansal, R. & Yaron, A. (2004), 'Risks for the long run: A potential resolution of asset pricing puzzles', *The Journal of Finance* **59**(4), 1481–1509.

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Boivin, J. & Ng, S. (2006), 'Are more data always better for factor analysis?', *Journal of Econometrics* **132**(1), 169–194.

Campbell, J. Y. & Cochrane, J. H. (1999), 'By force of habit: A consumption-based explanation of aggregate stock market behavior', *Journal of Political Economy* **107**(2), 205–251.

Cang, S. & Yu, H. (2014), 'A combination selection algorithm on forecasting', *European Journal of Operational Research* **234**(1), 127–139.

Cederburg, S. & O'Doherty, M. S. (2016), 'Does it pay to bet against beta? on the conditional performance of the beta anomaly', *The Journal of Finance* **71**(2), 737–774.

Chan, C. K., Kingsman, B. G. & Wong, H. (1999), 'The value of combining forecasts in inventory management–a case study in banking', *European Journal of Operational Research* **117**(2), 199–210.

Chava, S., Gallmeyer, M. & Park, H. (2015), 'Credit conditions and stock return predictability', *Journal of Monetary Economics* **74**, 117–132.

Chen, T. (2017), 'Investor attention and global stock returns', *Journal of Behavioral Finance* **18**(3), 358–372.

Christopherson, J. A., Ferson, W. E. & Glassman, D. A. (1998), 'Conditioning manager alphas on economic information: Another look at the persistence of performance', *The Review of Financial Studies* **11**(1), 111–142.

Cooper, M. J. & Gubellini, S. (2011), 'The critical role of conditioning information in determining if value is really riskier than growth', *Journal of Empirical Finance* **18**(2), 289–305.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. & MacBeth, J. D. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**(3), 607–636.

Ferson, W. E. & Harvey, C. R. (1999), 'Conditioning variables and the cross section of stock returns', *The Journal of Finance* **54**(4), 1325–1360.

Ferson, W. E., Sarkissian, S. & Simin, T. T. (2003), 'Spurious regressions in financial economics?', *The Journal of Finance* **58**(4), 1393–1413.

Figlewski, S. (1983), 'Optimal price forecasting using survey data', *The Review of Economics and Statistics* **65**(1), 13–21.

Frazzini, A. & Pedersen, L. H. (2014), 'Betting against beta', *Journal of Financial Economics* **111**(1), 1–25.

Goyal, A. & Welch, I. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *The Review of Financial Studies* **21**(4), 1455–1508.

Harvey, C. R. (2001), 'The specification of conditional expectations', *Journal of Empirical Finance* **8**(5), 573–637.

Hirano, K. & Wright, J. H. (2017), 'Forecasting with model uncertainty: Representations and risk reduction', *Econometrica* **85**(2), 617–643.

Huang, D., Jiang, F., Tu, J. & Zhou, G. (2015), 'Investor sentiment aligned: A powerful predictor of stock returns', *The Review of Financial Studies* **28**(3), 791–837.

Huang, H. & Lee, T.-H. (2010), 'To combine forecasts or to combine information?', *Econometric Reviews* **29**(5-6), 534–570.

Inoue, A., Jin, L. & Rossi, B. (2017), 'Rolling window selection for out-of-sample forecasting with time-varying parameters', *Journal of Econometrics* **196**(1), 55–67.

Inoue, A. & Rossi, B. (2011), 'Identifying the sources of instabilities in macroeconomic fluctuations', *Review of Economics and Statistics* **93**(4), 1186–1204.

Jagannathan, R. & Wang, Z. (1996), 'The conditional capm and the cross-section of expected returns', *The Journal of Finance* **51**(1), 3–53.

Kelly, B. & Pruitt, S. (2013), 'Market expectations in the cross-section of present values', *The Journal of Finance* **68**(5), 1721–1756.

Lettau, M. & Ludvigson, S. (2001), 'Consumption, aggregate wealth, and expected stock returns', *The Journal of Finance* **56**(3), 815–849.

Lewellen, J. & Nagel, S. (2006), 'The conditional capm does not explain asset-pricing anomalies', *Journal of Financial Economics* **82**(2), 289–314.

Li, J. & Tsiakas, I. (2017), 'Equity premium prediction: The role of economic and statistical constraints', *Journal of Financial Markets* **36**, 56–75.

Mallows, C. (1973), 'Some comments on cp', *Technometrics* **15**(1), 661–675.

Neely, C. J., Rapach, D. E., Tu, J. & Zhou, G. (2014), 'Forecasting the equity risk premium: the role of technical indicators', *Management Science* **60**(7), 1772–1791.

Newbold, P. & Granger, C. W. (1974), 'Experience with forecasting univariate time series and the combination of forecasts', *Journal of the Royal Statistical Society* **137**(2), 131–146.

Paye, B. S. & Timmermann, A. (2006), 'Instability of return prediction models', *Journal of Empirical Finance* **13**(3), 274–315.

Pesaran, M. H., Pettenuzzo, D. & Timmermann, A. (2006), 'Forecasting time series subject to multiple structural breaks', *The Review of Economic Studies* **73**(4), 1057–1084.

Pesaran, M. H., Pick, A. & Pranovich, M. (2013), 'Optimal forecasts in the presence of structural breaks', *Journal of Econometrics* **177**(2), 134–152.

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Petkova, R. & Zhang, L. (2005), 'Is value riskier than growth?', *Journal of Financial Economics* **78**(1), 187–202.

Rapach, D. E., Ringgenberg, M. C. & Zhou, G. (2016), 'Short interest and aggregate stock returns', *Journal of Financial Economics* **121**(1), 46–65.

Rapach, D. E. & Strauss, J. K. (2010), 'Bagging or combining (or both)? an analysis based on forecasting us employment growth', *Econometric Reviews* **29**(5-6), 511–533.

Rapach, D. E., Strauss, J. K. & 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.

Rapach, D. E. & Wohar, M. E. (2006), 'Structural breaks and predictive regression models of aggregate us stock returns', *Journal of Financial Econometrics* **4**(2), 238–274.

Rapach, D. E. & Zhou, G. (2020), 'Time-series and cross-sectional stock return forecasting: New machine learning methods', *Machine Learning for Asset Management: New Developments and Financial Applications* **1**(1), 1–33.

Rossi, B. (2013), *Advances in forecasting under instability*, Vol. 2 of *Handbook of Economic Forecasting*, Elsevier.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

Stambaugh, R. F. (1999), 'Predictive regressions', *Journal of Financial Economics* **54**(3), 375–421.

Stock, J. H. & Watson, M. W. (1996), 'Evidence on structural instability in macroeconomic time series relations', *Journal of Business and Economic Statistics* **14**(1), 11–30.

Tian, J. & Anderson, H. M. (2014), 'Forecast combinations under structural break uncertainty', *International Journal of Forecasting* **30**(1), 161–175.

Tian, J. & Zhou, Q. (2018), 'Improving equity premium forecasts by incorporating structural break uncertainty', *Accounting and Finance* **58**(1), 619–656.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58**(1), 267–288.

Wang, Y., Hao, X. & Wu, C. (2020), 'Forecasting stock returns: A time-dependent weighted least squares approach', *Journal of Financial Markets* **53**, 100568.

Yin, A. (2020), 'Equity premium prediction and optimal portfolio decision with bagging', *The North American Journal of Economics and Finance* **54**, 101274.

Zhang, H., He, Q., Jacobsen, B. & Jiang, F. (2020), 'Forecasting stock returns with model uncertainty and parameter instability', *Journal of Applied Econometrics* **35**(5), 629–644.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society* **67**(2), 301–320.

Chapter 2

Conditional CAPM with Dynamic Model Selection (DMS) Approaches

2.1 Introduction

2.1.1 Background

The capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) has long been a foundation of asset pricing theory and practice. It is, however, proven that the unconditional (or static) version of the CAPM does not explain the expected portfolio returns based on firm characteristics.[1] Despite the fact that the unconditional CAPM fails empirically, a conditional CAPM (CCAPM) that allows betas to vary over time might be able to explain the cross-section of average return (see Jagannathan & Wang 1996). Empirically, the performance of such a method relies on an excellent econometric framework that captures the time-variation of conditional betas (Ghysels 1998).

In the finance literature, estimating the time-varying betas have already been thoroughly studied by several different approaches.[2] However, to date, there is no convincing argument for either approach. In this paper, we follow Jagannathan & Wang (1996), Ferson & Harvey (1999), Lettau & Ludvigson (2001), Petkova & Zhang (2005), Cederburg & O'Doherty (2016) and others, and define the beta of an asset as a linear function of some observable *conditioning information variables* (CIVs). Since this approach uses predictor variables or instrumental variables (IVs), it is also referred to as CCAPM-IV. This essay contributes to the current literature by introducing a CCAPM-IV model where the choice of conditioning variables, used to model the time-varying betas, is allowed to vary through time and is selected from a large pool of potential state variables.

2.1.2 Motivation

The primary motivation for this study comes from the fact that the CCAPM-IV method relies heavily on the same predictors that are being reported as aggregate stock return predictors (Cooper & Gubellini 2011). The ever-growing literature on predictor variables has provided researchers with a large pool of possible conditioning variables in addition to explicitly providing

[1] This forms the basis of the anomalies literature. Well-known anomalies include size, book-to-market, momentum, beta, liquidity, profitability, growth, and others (e.g., Banz 1981, Fama & French 1992, Carhart 1997, Amihud 2002, Fama & French 2015, Hou et al. 2019).

[2] Some of the famous approaches include those using data-driven filters such as beta calculated from a 60-month rolling window as in Fama & MacBeth (1973), or a short window approach (Lewellen & Nagel 2006) and high-frequency data (Andersen et al. 2003), multivariate GARCH (Bollerslev et al. 1988), dynamic conditional correlation (DCC) (Engle 2002, Bali & Engle 2010), regime-switching model (Vendrame et al. 2018), mean-reverting stochastic process (Jostova & Philipov 2005), Kalman filter (Adrian & Franzoni 2009), and others.

estimates of the time-varying equity premium.[3]

However, using equity premium predictors can be criticised given the identification of predictor variables is econometrically challenging, leading to spurious identification of conditioning variables. These challenges include the issues with persistence leading to bias (Stambaugh 1999), inability to predict out-of-sample (Goyal & Welch 2008), model instability (Paye & Timmermann 2006, Rapach & Wohar 2006) and data mining and overfitting (Ferson et al. 2003, Harvey et al. 2016). The lack of theoretical guidance on which subset of variables should be chosen from the plethora of predictors creates an empirical dilemma over how to optimally parameterise the model. We refer to this problem as *variable-selection uncertainty* (henceforth VSU). The VSU issue is important to address because some recent studies have stressed that too many predictors can adversely affect a model's forecasting performance (e.g., Boivin & Ng 2006, Bai & Ng 2008).[4] The other motivation for including the small number of predictors is based on a practical viewpoint, provided that the collection and processing of information are often expensive and therefore contributes to rational inattention (e.g., Sims 2003, Abel et al. 2013, Luo & Young 2016, Gabaix 2019). This implies that rational investors only consider the most effective predictor variables as a consequence of rational inattention and neglect the rest. Given this motivation, we believe it is worth considering the approaches that can select a subset of predictors at a given time for forecasting asset betas.

2.1.3 Research Gaps and Objectives

The problem of *variable-selection uncertainty* (VSU) also applies to the tests of CCAPM-IV when predictors from the *equity premium prediction* (henceforth EPP) literature are used to model the time-variation in factor loadings. This is also evident from the standard CCAPM-IV approach, which generally uses a predetermined set of predictors. The dividend yield (DY), the short-term Treasury bill rate (TBL), the default premium (DEF), and the term premium (TMS) are four common variables used in several studies (e.g., Ferson & Harvey 1999, Petkova & Zhang 2005, Cai et al. 2015). Such an approach of using a fixed number of conditioning variables from a broad set of predictors to represent the information set has been criticised. For example, studies such as Ghysels (1998), Harvey (2001) and Cooper & Gubellini (2011) find that the performance of CCAPM-IV is sensitive to the researcher's selection of variables. This suggests that the existing CCAPM-IV approach faces the issue of VSU, which can be addressed by applying some of the strategies used in forecasting literature, in particular, EPP literature.

A review of forecasting literature suggests there have been many advances, especially in the area of EPP, that address the empirical challenges in using predictive regressions, in particular dealing with VSU. In the presence of various predictors, one of many approaches is the *variable selection* (henceforth VS) which directly selects the best predictors to carry out the forecast at each point in time.[5] VS approach implies that only subset of predictors are important at a point in time and all other predictors have weak prediction power to the target variable. Thus, it is interesting to apply some of these methods to the CCAPM-IV to address the issue of VSU. This study, therefore, aims to introduce a CCAPM model where the choice of conditioning variables,

[3] See Goyal & Welch (2008) for traditional predictor variables and some of the new predictors include technical indicators (Neely et al. 2014, Lin 2018), investor sentiment and attention (Huang et al. 2015, Ni et al. 2015, Coqueret 2020, Zhang et al. 2021), manager sentiment (Jiang et al. 2019), the short interest index (Rapach et al. 2016), bitcoin prices (Salisu et al. 2019), credit quality (Chava et al. 2015), among others.

[4] There are a variety of reasons why a model should be limited to a small number of predictor variables. The most apparent is that every single variable cannot explain the target variable, so any variable that is not related to it should be omitted (ontological sparsity). Second, even though all variables may explain the target variable, it is preferable to remove variables with minor effects, either to improve the final model's interpretability (epistemic sparsity) or to improve the model's predictive ability through reducing variance (predictive sparsity).

[5] See Zhang (2016) and Wang et al. (2020) for more details on VS approaches.

used to capture the variation in conditional betas, is allowed to vary through time and is selected from a large pool of potential state variables. Under our approach, the subset of conditioning variables selected at time t is based on a pretest procedure that uses past information to decide whether a predictor is 'in' or 'out'. Therefore, the main research question is whether conditional CAPM based on VS approaches can explain the cross-section of average returns. As this has been a challenge from previous implementations of the CCAPM (e.g., Harvey 2001, Cooper & Gubellini 2011).

Some of the popular VS approaches in linear models include the sequential selection, best subset selection, and shrinkage methods. These methods are also known as *dynamic model selection* (henceforth DMS) approaches where selected variables at a given time are used in a single model for estimation and forecasting. The sequential selection approach includes the methods such as forward selection, backward elimination, and stepwise regression (e.g., Efroymson 1960, Draper & Smith 1966, Smith 2018, Xu & Zhang 2001).[6] Under this approach, the candidate variables are evaluated at each step, one by one, usually using the t-statistics for the coefficients of the considered variables. The best subset variable selection approach tests all possible combination of the predictor variables and then select the best model according to some statistical criteria. Historically simplified criteria such as the determination coefficient (R^2), and its modified variant that penalises models with more parameters have been widely used. However, these methods have been shown to experience manifold weaknesses, particularly in a predictive context, where more advanced metrics are standard nowadays, such as information criteria, for instance, Akaike Information Criterion (AIC) (Akaike 1973), Bayesian Information Criterion (BIC) (Schwarz 1978), and Mallows's C_P (Mallows 1973).[7] To overcome the computational difficulties of the best subset problem, computationally convenient convex optimisation based shrinkage methods, also called penalised regression, have been proposed.[8] The most common shrinkage methods include the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996), Adaptive LASSO (Zou 2006), and Elastic Net (ENet) (Zou & Hastie 2005).

In addition to these standard methods, we also consider each model by its performance in pricing the assets. We call this approach as the *dynamically selected beta model* (henceforth DSBM). More specifically, this approach selects the beta models that perform the best based on standard asset pricing criteria on past data at each point in time. More precisely, our approach to select the optimal subset of variables at time t is as follows: i) we consider all possible sets of conditioning variables, ii) using past data (i.e. a training sample of data prior to time t), the set of variables that perform best in an asset pricing test is selected. Various asset pricing criteria are considered for selection of the optimal set, including time-series and cross-sectional tests. Our approach's natural consequence is that the conditioning information, as represented by the optimal subset of conditioning variables, will change over time as variables become more or less important.

2.1.4 Overview of Methodology

We use out-of-sample analysis to estimate time-varying betas with various *dynamic model selection* (DMS) approaches to prevent look-ahead bias. Our analysis examines whether the CCAPM model based on DMS approaches explains the cross-section of average asset returns. We use Fama & MacBeth (1973) two-step method, in which the factor loadings for each asset, i.e. the estimates of conditional betas in CCAPM, are obtained in the first step using time-series regressions. The first step involves regressing monthly excess asset returns on the market risk factor, in a model where the market beta varies with conditioning variables (see, e.g., Shanken 1990,

[6] See Morozova et al. (2015) for more details on sequential selection methods.
[7] See Kadane & Lazar (2004), Raffalovich et al. (2008), and Zhang (2016) for more details.
[8] See Kuhn & Johnson (2013) and Hastie et al. (2017) for more details on shrinkage methods.

Ferson & Harvey 1999, Cederburg & O'Doherty 2016). More specifically, $\beta_t = f(X_t)$, where X represents the subset of the full information set of investors (I), $X_t \subset I_t$. In our approach, we first select the subset of predictors, X_t^*, at each period based on some performance criteria, $X_t^* \subset X_t$ and then define $\beta_t = f(X_t^*)$. After obtaining the out-of-sample betas, we test the model by running a cross-sectional regression at each time t of the evaluation period, with the first-step betas obtained through different approaches serving as an explanatory variable. Our CCAPM cross-sectional tests are based on mainstream literature that evaluates the pricing abilities of a given model by looking at the significance of Fama & MacBeth (1973) parameter estimates.[9] In addition, we assess the performance of each model through various performance metrics such as sum of squared pricing errors (SSPE) (Adrian & Rosenberg 2008), and cross-sectional adjusted R^2 (Jagannathan & Wang 1996), and composite pricing errors (CPE) (Campbell & Vuolteenaho 2004).

2.1.5 Principal Results

We use the monthly excess returns on 25 size and value portfolios of Fama & French (1993) to perform the tests for a sample period from July 1926 to December 2018. The conditioning information variables used in this study are taken from Goyal & Welch (2008), we select the 14 variables for which monthly data are available from July 1926 to December 2018. The cross-sectional results for out-of-sample periods August 1936 to December 2018 and August 1968 to December 2018 show that all the DMS approaches do not handle the equity premium properly since the excess return on the zero-beta portfolio (constant from Fama & MacBeth (1973) second stage regression) is significant and large in magnitude. In addition, these approaches yield significant pricing errors measured as composite pricing errors (CPE). However, in terms of explaining cross-sectional variation measured as R^2, DSBM outperforms all the standard DMS approaches including sequential selection, best subset selection, and shrinkage methods.

Lewellen & Nagel (2006) criticise the cross-sectional tests and argue in favour of time-series tests for testing the suitability of CCAPM by directly evaluating the ability of a model in explaining the anomalies unexplained by unconditional CAPM. Following them, we use time-series analysis to assess the performance of DMS approaches compared to the unconditional CAPM. Specifically, our time-series tests compare the unconditional and conditional performance of size, value, and momentum portfolios.[10] Consistent with the findings of Lewellen & Nagel (2006), results show that the size premium is insignificant at 5% for all the models, including both DMS and unconditional CAPM. However, the average pricing errors for 'VMG' and 'WML' are significant for all the models, including DSBM, implying a failure of DMS approaches in explaining the value and momentum anomalies.

Finally, to test the robustness of DMS approaches, we follow mainstream literature and use the bootstrap method, which is the standard approach for quantifying the model uncertainty and instability (see, e.g., Sauerbrei & Schumacher 1992, De Bin et al. 2016, Petropoulos et al. 2018). Results show that the selection of conditioning variables under all approaches is subject to considerable estimation error. In other words, DMS approaches in our application fail to fully address the issue of VSU. These findings are consistent with recent criticism of DMS approaches regarding their inability to properly address the variable-selection uncertainty and to achieve model stability (Smith 2018, Petropoulos et al. 2018, Makridakis et al. 2020, and others).

[9] We use t-statistics to test the significance using Newey & West (1987) heteroskedasticity and autocorrelation consistent standard errors.

[10] The test portfolios are from Kenneth R. French's database. 'Small' ('Big') indicates the average of the five low (high) portfolios sorted by market-cap, and the difference of 'Small' and 'Big' is denoted as 'SMB'. Likewise, 'Growth' ('Value') is the average of the five low (high) portfolios sorted by book-to-market, and 'VMG' indicates the returns on 'Value' minus 'Growth' portfolio. 'Losers' ('Winners') are the bottom (top) decile of Fama-French momentum sorted portfolios, and 'Winners minus Loosers (WML)' is their difference.

2.1.6 Contribution

This essay contributes to the current empirical asset pricing literature by evaluating conditional versions of the CAPM (CCAPM) through modelling a new form of time variance in conditional betas. More specifically, we complement the existing CCAPM literature that has focused on capturing beta dynamics using state variables to explain unconditional alphas for portfolios sorted by firm characteristics such as size and value.[11] Our research is also related to studies such as Harvey (2001), and Cooper & Gubellini (2011), who find that the estimation of CCAPM is sensitive to choosing state variables. Finally, we also complement the literature by providing evidence against CCAPM (e.g., Lewellen & Nagel 2006) for its failure in explaining the value and momentum anomalies.

The remaining structure of this chapter is as follows. An overview of the literature on the CCAPM, and model selection approaches is provided in Section 2.2. The econometric methodology is discussed in Section 2.3. Section 2.4 addresses the implementation of model selection approaches from the CCAPM perspective. Section 2.5 provides an overview of data and benchmark models. The discussion of the empirical results is given in Section 2.6. Section 2.7 reports the results of various robustness tests. The conclusions are drawn in the Section 2.8.

2.2 Literature Review

This section gives an overview of the CAPM literature as well as potential issues with the CCAPM-IV. We also go through some well-known variable selection strategies that have been used in the forecasting literature to deal with variable-selection uncertainty (VSU).

2.2.1 Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) has been pivotal in empirical asset pricing for more than fifty years now. Though the static version of the CAPM managed to endure intense econometric investigation for years, recent tests suggest that this static single-beta model is insufficient to explain expected returns on stocks.[12] The failure of the CAPM may be due to two potential reasons. In the first explanation, many scholars agree that the beta is not the sole indicator of systemic risk and that asset prices are subject to a variety of risks other than the market beta risk, including the leverage ratio, size, book-to-market equity, momentum, liquidity, and others (e.g., Banz 1981, Fama & French 1992, Carhart 1997, Amihud 2002). The multi-factor models emerged as a result of this. The Fama & French (1993) three-factor (FF3F) and the Carhart (1997) four-factor model are the most prominent multi-factor factors. The former incorporated size and value as two additional factors to the standard CAPM, while Carhart (1997) added the momentum factor to the market, size, and value factors.[13]

The second potential explanation is generally linked to the fact that the standard CAPM is a one-period model. This suggests that the standard or unconditional CAPM is founded on

[11] A partial list includes Jagannathan & Wang (1996), Ferson & Harvey (1999), Lettau & Ludvigson (2001), Petkova & Zhang (2005), Cederburg & O'Doherty (2016) and others.

[12] For example, see Banz (1981), Fama & French (1992), Carhart (1997), Amihud (2002), Fama & French (2015), Hou et al. (2019) and others.

[13] Some other multi-factor models include the q-factor model of Hou et al. (2015), which include four factors (viz., market, size, investment, and return on equity). Later, Hou et al. (2019) added a growth factor to the q-model. Fama & French (2015) proposed the five-factor model, which is based on market, size, value, profit and investment factors. Later, Fama & French (2018) added momentum to their five-factor model. Stambaugh & Yuan (2017) proposed a four-factor model, including the market and size factors plus two mispricing factors (viz., management and performance).

the premise that assets' betas and risk aversion of investors do not change over time. In the real world, however, it is well understood that when the economy is in a slump, investors are much more risk-averse, and when investment opportunities increase during a period of economic growth, their risk aversion declines. Consequently, risk premia should fluctuate during the business cycle, since increased risk aversion during economic downturns, according to Fama & French (1989) and Cochrane (1999), necessitates a higher risk premium, resulting in the predictability of equity premium. Therefore, some researchers believe that the conditional CAPM, which includes time-varying beta and market risk premium, may explain the cross-section of stock returns that the unconditional CAPM cannot. For example, Jagannathan & Wang (1996) shows that the CAPM holds if beta and risk premium are allowed to vary over time. Using a value-weighted index and the average returns of 100 stock portfolios, they show that the conditional CAPM can explain more than 50% of the cross-sectional variation in average returns. In contrast, the unconditional CAPM can only explain 1%. Specifically, Jagannathan & Wang (1996) define conditional CAPM as:

$$E[R_{it}|I_{t-1}] = \gamma_{0t-1} + \gamma_{1t-1}\beta_{it-1} \tag{2.1}$$

where they define $E[R_{it}]$ as expected return on asset i at time t which is conditional on investors' information set I, available to to them at time $t-1$. γ_{0t-1} represents the expected return on "zero-beta" portfolio, γ_{1t-1} indicates the expected market risk premium at time $t-1$, and β_{it-1} is conditional β which is given as:

$$\beta_{it-1} = \frac{Cov(R_{it}, R_{mt}|I_{t-1})}{Var(R_{mt}|I_{t-1})} \tag{2.2}$$

In order to define cross-sectional variation they take unconditional expectations from both sides of equation (2.1) and get:

$$E[R_{it}] = \gamma_0 + \gamma_1\overline{\beta_i} + Cov(\gamma_{1t-1}, \beta_{it-1}) \tag{2.3}$$

where γ_1 represents the expected market risk premium and $\overline{\beta_i}$ is expected beta. Conditional CAPM is equivalent to unconditional CAPM if the co-variance in equation (2.3) is zero. Jagannathan & Wang (1996) suggest that the risk premium and conditional betas are highly correlated because, in a recession, low-performing firms' financial leverage is expected to increase sharply relative to other firms, increasing their systematic risk (beta).

Over the years, many empirical studies show that conditional CAPM helps explaining the unconditional alphas (e.g., Lettau & Ludvigson 2001, Petkova & Zhang 2005, Adrian & Franzoni 2009, Bali & Engle 2010, Cai et al. 2015, Vendrame et al. 2018). More recently, Cederburg & O'Doherty (2016) show that CCAPM helps to explain the beta premium (betting against beta, BAB). We follow this line of inquiry in this study and examine whether a simple conditional CAPM can explain the cross-section of average returns.

Despite the fact that multiple studies report the success of CCAPM, its implementation in practice is difficult. CCAPM has two big concerns. First, it is unclear how the set of predictive variables should be selected. Second, there is no consensus among academics on how to model the time variation in beta and risk premium. Consequently, the empirical success of a particular strategy in overcoming the problems associated with unconditional CAPM is solely determined by the method used to estimate time-varying betas (Ghysels 1998). To date, we find many approaches to capture beta-dynamics. Some of the famous approaches include those using data-driven filters such as beta calculated from a 60-month rolling window as in Fama & MacBeth (1973), or a short window approach (Lewellen & Nagel 2006) and high-frequency data (Andersen et al. 2003), multivariate GARCH (Bollerslev et al. 1988), dynamic conditional correlation (DCC) (Engle 2002, Bali & Engle 2010), regime-switching model (Vendrame et al.

2018), mean-reverting stochastic process (Jostova & Philipov 2005), Kalman filter (Adrian & Franzoni 2009), and others.

Another method to incorporate time variation in beta is called an *instrumental variable* (IV) approach, where beta is defined as a linear function of some predetermined instrumental variables (Jagannathan & Wang 1996, Lettau & Ludvigson 2001, Ferson & Harvey 1999, Petkova & Zhang 2005, Cederburg & O'Doherty 2016).[14] In this study we follow this approach and focus on the models where betas are function of some observable variables. The primary motivation for selecting the IV approach is the fact that other approaches have been criticised for concealing the driving force behind betas. The IV method, on the other hand, can offer more economic insight into beta movement by identifying the predictor variables that cause fluctuations in systemic risk. That is why this approach is economically more fascinating as it not only reproduces how investors actually behave, but it also offers strong economic implications without which an asset pricing theory is void (Cochrane 2005).

2.2.2 Challenges of CCAPM-IV approach

Since there is no guidance on choosing the correct set of conditioning information variables to model beta dynamics, the IV approach of CCAPM is extremely difficult to test (Cochrane 2009). The existing studies using IV approach rely on the *conditioning information variables* (CIVs), which obviously leads to the researcher's discretion on the choice of predictors. For example, the four most common CIVs frequently used in literature to capture beta dynamics include the dividend yield, the term premium, the default premium, and the short-term Treasury bill rate. Cooper & Gubellini (2011) find out how sensitive CCAPM tests are to changes in CIVs that a researcher is likely to consider. They extended the dataset by adding more relevant CIVs to four of the variables mentioned above and tested the CCAPM with a different combination of variables. Through simulations, they show that the combinations of various variables produce different results. Thus, consistent with the findings of Ghysels (1998) and Harvey (2001), they conclude that the performance of CCAPM-IV is sensitive to the researcher's selection of variables. However, the standard methods of CCAPM based on observable variables use a predetermined set of CIVs.

The severity of issues with the IV approach is also evident from findings of Simin (2008), who analyses the performance of unconditional and conditional models using both individual security and portfolio return. His conditional model given in equation (2.4) follows the IV approach, using predictor variables (Z_t) to allow both alpha $\alpha_t = \alpha_0 + \alpha_1 Z_t$ and beta $\beta_t = b_0 + b_1 Z_t$ to vary over time.

$$R_{t+1} = \alpha_0 + \alpha_1 Z_t + b_0 R_{p,t+1} + b_1 R_{p,t+1} Z_t + u_{t+1}, \qquad E(u_{t+1}) = E(u_{t+1}[Z_t R_{p,t+1}]) = 0 \quad (2.4)$$

The author concludes that for step-ahead prediction, conditional models appear to yield higher mean squared errors than unconditional models, implying that variable-selection uncertainty under the IV approach will devastate the CCAPM's step-ahead predictability.

The other point to consider is the nature of predictors that are used as CIVs. The standard IV approach relies heavily on the same predictors that are being reported as aggregate stock return predictors (Cooper & Gubellini 2011). However, using equity premium predictors can be criticised given the identification of predictor variables is econometrically challenging, leading to spurious identification of conditioning variables. These challenges include the issues with persistence leading to bias (Stambaugh 1999), inability to predict out-of-sample (Goyal & Welch

[14]Some studies use aggregate variables such as interest rates (e.g., Ferson & Harvey 1999), consumption-to-wealth ratio (e.g., Lettau & Ludvigson 2001), and other use firm specific variables (e.g., Avramov & Chordia 2006, Bauer et al. 2010) to model beta dynamics.

2008), model instability (Paye & Timmermann 2006, Rapach & Wohar 2006) and data mining and overfitting (Ferson et al. 2003).

In summary, the researchers under IV approach generally make two assumptions: i) predetermined CIVs are the only relevant variables representing investors' information set, and ii) the investors' information set captured through preselected variables does not change over time. Therefore it is interesting to apply some of the methods to the CCAPM-IV where the chosen subset of variables can optimally vary over time.

2.2.3 Dynamic Model Selection Strategies to overcome VSU

This section discusses various strategies used in forecasting literature to address the challenges related to variable-selection uncertainty (VSU) due to the lack of theory on relevant predictors to be used in a model. One of the approaches to deal with VSU in the presence of many predictors is variable selection (VS), which can be defined as the process of selecting a subset of relevant variables in a data set. The key objective of variable selection methods is removing irrelevant or closely redundant variables without losing too much information.[15] Thus, variable selection helps simplify forecasting models and reduce the susceptibility to over-fitting problems.

In forecasting literature, we find many approaches to select a subset of variables. Sequential selection (Draper & Smith 1966), best subset selection (Hocking & Leslie 1967), and shrinkage methods (Tibshirani 1996) are well-known approaches for selecting and estimating the linear model parameters. Sequential selection methods were introduced in the 1960s and have since gained popularity in many fields due to computational convenience (Morozova et al. 2015). Under this approach, to select subset variables from a large pool, one needs to perform a series of steps. The candidate variables are evaluated at each step, one by one, usually using the t-statistics for the coefficients of the considered variables. Forward selection, backward elimination, and stepwise regression are examples of sequential selection methods.

Another approach is best subset variable selection which consists of testing all possible combinations of the predictor variables and then selecting the best model according to some statistical criteria. In order to choose between alternatives, conventionally simplistic criteria such as the coefficient of determination (R^2) and its modified variant that penalises models with more parameters have been widely used. However, these criteria suffer from many weaknesses, especially in a predictive context, where nowadays more advanced measures, such as information criteria including Akaike information criteria (AIC) (Akaike 1973) and the Bayesian information criterion (BIC) (Schwarz 1978) are the procedures that have gained more attention from the list of model selection procedures (Burnham & Anderson 2002).

To overcome the computational difficulties of the best subset problem, computationally friendlier convex optimisation based shrinkage methods, also called penalised regression, have been proposed.[16] These methods impose regularisation constraints on the objective function by adding a penalty term. There are four common regularisation constraints: *least absolute shrinkage and selection operator* (LASSO) introduced by Tibshirani (1996), improved version of LASSO known as adaptive LASSO proposed by Zou (2006), Ridge from Hoerl & Kennard (1970), and finally Elastic Net (ENet) developed by Zou & Hastie (2005). The implementation of shrinkage methods is based on penalised regression, in which the penalty function simultaneously induces coefficient shrinkage and variable selection.

[15] Comprehensive details on variable selection approaches can be found in Kuhn & Johnson (2013) and Hastie et al. (2017).

[16] More information on penalised regression approaches are given in Kuhn & Johnson (2013), Hastie et al. (2017), and Gu et al. (2020).

The primary motivation for using shrinkage methods is grounded from a statistical perspective. For example, if a set of important predictors include some variables that are strongly correlated (positively), the LASSO will choose one and discard the others, preventing overfitting and improving predictive accuracy (see, e.g., Tibshirani 1996). These methods have been applied in many fields to improve the forecasting accuracy. For example, equity premium (Buncic & Tischhauser 2017), crude oil (Zhang et al. 2019), macroeconomic variables (Baybuza 2018), tourism (Lourenço et al. 2021), health care (Xiao et al. 2019), weather (Al-Obeidat et al. 2020), energy consumption (González-Briones et al. 2019), corporate failure (Pereira et al. 2016), and others.

2.3 Econometric Methodology

This section discusses the econometric framework to estimate and test the conditional CAPM. Before discussing the models and tests, we first introduce notation, and the process of sample splitting. Next, we move to the econometric framework of dynamic model selection (DMS) approaches. Finally, we discuss cross-sectional tests for evaluating the performance of various models.

2.3.1 Notations

There are N assets indexed by $i = 1, \ldots., N$, K represents total predictors indexed by $k = 1, \ldots., K$ and M indicates the total available models indexed by $j = 1, \ldots., M$. T represents total observations indexed by $t = 1, \ldots., T$. The initial training sample is indicated by m and W represents the estimation window. S indicates out-of-sample observations for final evaluation of model, which is given as, total observations (T) less initial training sample (W). $R_{i,t}$ indicates the excess returns for asset i at time t, and $R_{m,t}$ indicates excess market returns at time t. I_t indicates the vector of investors' information set, $X_{j,t}$ represents the vector of explanatory variables in model j ($X_t \subseteq I_t$). α_{itj} and β_{itj} represent pricing error and beta for asset i at time t with model j, respectively.

2.3.2 Sample splitting

We use out-of-sample analysis to estimate time-varying betas with various dynamic model selection approaches to prevent look-ahead bias. This involves dividing the total sample into training and testing. The model parameters are estimated using the training sample and then applied to unseen data to obtain out-of-sample forecasts. Specifically, we use two sample-splitting approaches summarised in Figure (2.1). In the first approach, we divide the total sample of T observations into two portions: i) W as a training sample, and ii) $S = T - W$ to evaluate the out-of-sample performance. We use a rolling window approach, with a fixed window of w observations that rolls over each time up to the last observation of the sample. Figure (2.2) shows sample-split with rolling window framework.

Figure 2.1: Approaches for Splitting Sample

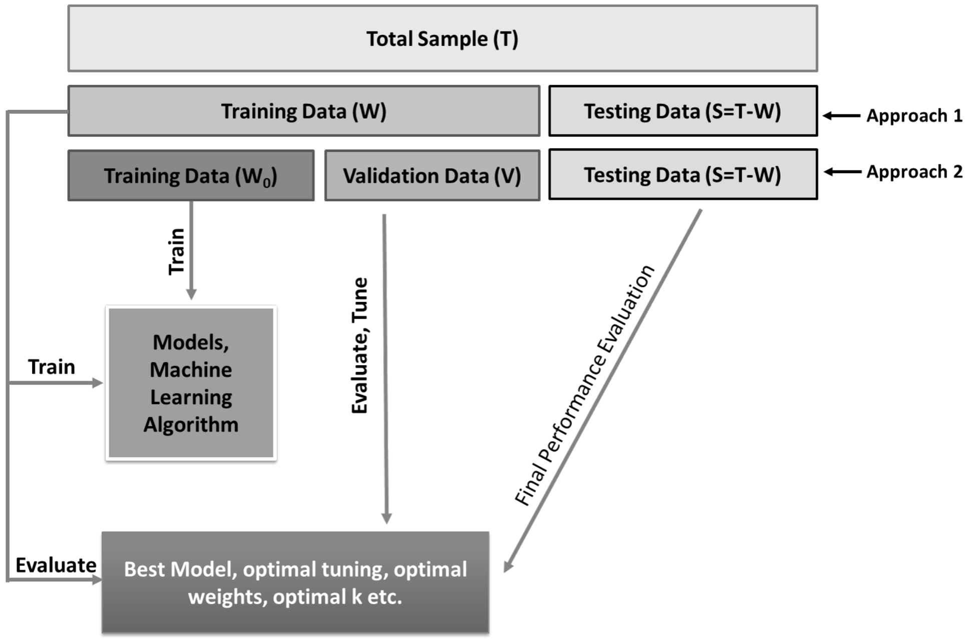
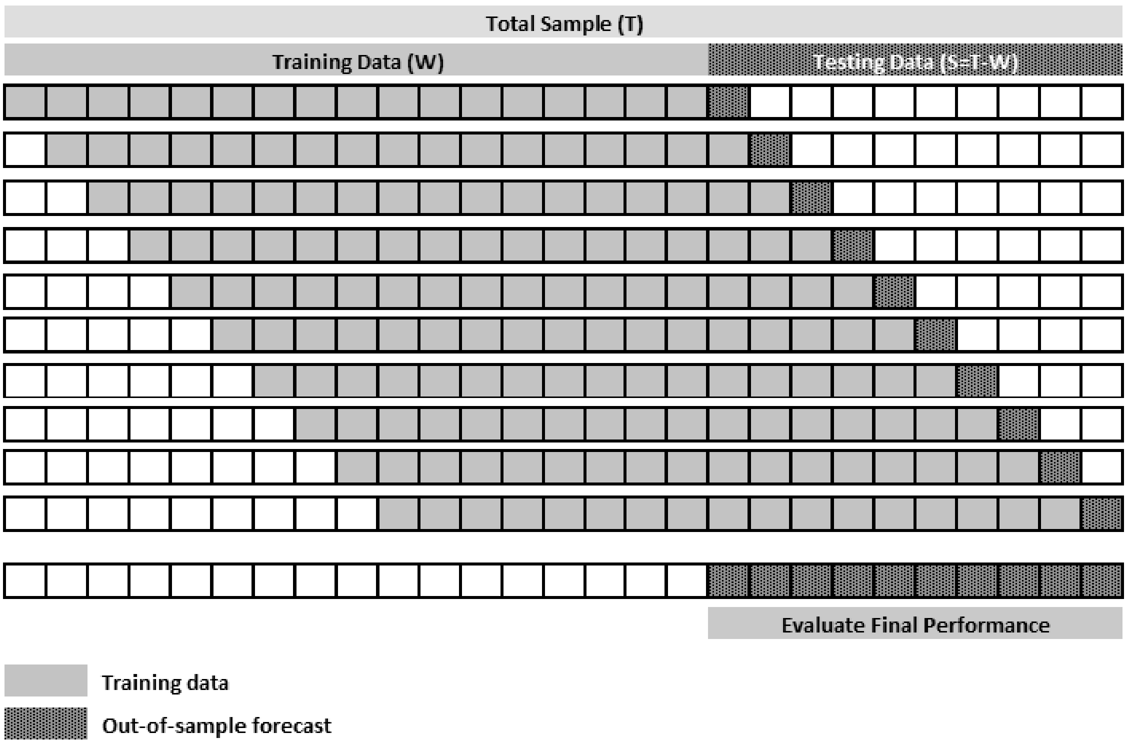


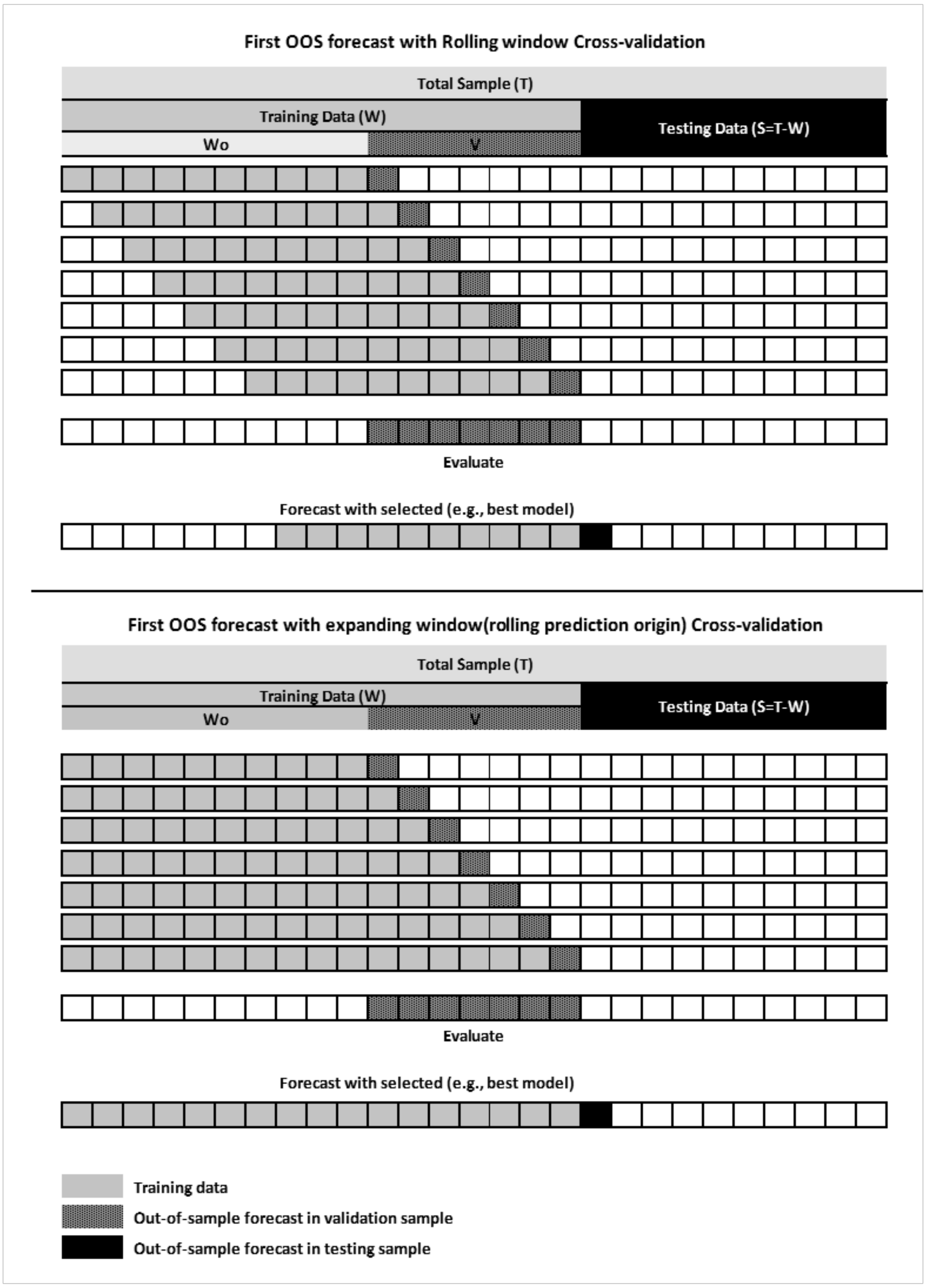
Figure 2.2: Sample-split - Training and Testing datasets



In our second approach, instead of splitting our total sample into two parts, we divide it into three parts: i) training (W_0), ii) validation (V), and iii) testing (S). The validation sample is used to assess the performance of a given model by evaluating its ability to predict the future. This sample-splitting method is widely used in machine learning methods to estimate hyperparameters for methods such as LASSO and Elastic Net, which are also part of our analysis.

See Gu et al. (2020) for more details on the implementation of this approach in machine learning methods. Note that the validation sample is not used to assess the final performance of the model. Instead, it only helps to optimise the hyperparameters or identify the best model for making the out-of-sample forecast. Thus, the third subsample (testing sample) consisting of $S = T - W_0 - V$ observations, which is not used for estimation or validation, is simply out of the sample and is instead used in evaluating the predictive performance of the given model. Note that we call our second approach as *cross-validation* (henceforth CV) which is summarised in Figure (2.3).

Figure 2.3: Walk Forward Cross-validation with Rolling and Expanding window



2.3.3 Econometric Framework

Assuming that, in a dynamic economy, the hedging motives of risk averse investors are negligible, the conditional version of Black (1972) CAPM is described by Jagannathan & Wang (1996) as:

$$E_t[R_{i,t+1}] := E[R_{i,t+1}|I_t] = \lambda_{0,t} + \beta_{i,t}\lambda_{1,t}, \tag{2.5}$$

where $R_{i,t+1}$ denotes the return on asset i in period $t+1$, I_t represents the information set available to investors at the end of period t. In this version of conditional CAPM, $\lambda_{0,t}$ denotes the conditional expected return on a "zero beta" portfolio, while $\lambda_{1,t}$ represent the conditional market risk premium. $\beta_{i,t}$ is the conditional beta of asset i based on the given information set I_t, which is defined as:

$$\beta_{it} = \frac{Cov(R_{i,t+1}, R_{m,t+1}|I_t)}{Var(R_{m,t+1}|I_t)} \tag{2.6}$$

where $R_{m,t+1}$ denotes the return on the market portfolio in period $t+1$.

We use Fama & MacBeth (1973) two-step method to estimate the β and λ parameters of (2.5). The factor loadings for each asset, i.e. the estimates of conditional betas in CCAPM, are obtained in the first step using time-series regressions. The next step requires estimating a cross-sectional regression at each period of excess asset returns on the first step's conditional betas.

2.3.3.1 First-pass Regressions – Estimating Conditional Betas

To estimate conditional betas, we follow the CCAPM-IV approach (e.g., Shanken 1990, Ferson & Harvey 1999, Petkova & Zhang 2005, Cederburg & O'Doherty 2016) and model the portfolio beta as a function of some observable instrumental variables (IVs). Our main analysis of estimating conditional betas uses the following time-series model:

$$R_{i,t+1} = a_i^{IV} + (\gamma_{i,0} + \gamma'_{i,1}X_t)R_{m,t+1} + \varepsilon_{i,t+1}, \tag{2.7}$$

where t indexes months, $R_{i,t+1}$ and $R_{m,t+1}$ are the excess returns on asset i and the market during period $t+1$, respectively, and $X_t \subseteq I_t$ is a vector of L instruments which represents the broader set of investors' information, I_t. It is thus assumed that the conditional portfolio beta is a linear function of some observable variables known at time t, $\beta_{i,t}^{IV} = \gamma_{i,0} + \gamma'_{i,1}X_t$, and the conditional portfolio alpha is constant.[17] Past studies such as Ferson & Harvey (1999), Petkova & Zhang (2005), Cederburg & O'Doherty (2016) and others use a predetermined set of instruments. However rather than selecting a prior a subset of variables from the large set of potential conditioning variables X_t we use dynamic model selection approaches to be discussed in section (2.4). This approach naturally leads to model where the betas are driven by a time-varying set of conditioning variables. In addition, when X_t is the null information set, the CCAPM model given in equation (2.7) reduces to the unconditional (static) CAPM, which restricts the beta of the portfolio to be constant.

To obtain the conditional asset betas based on a given DMS approach, we use following equation

$$\hat{\beta}_{i,t}^{DMS} = \hat{\gamma}_{i,0,t} + \hat{\gamma}'_{i,1,t}X_t^* \tag{2.8}$$

[17] We also estimate model (2.7) with time-varying alphas for the robustness of results, where conditional portfolio alpha is defined as a function of same observable variables used to for modelling betas. Results show that the impact of allowing alphas to vary over time, is negligible on our main findings.

where $\hat{\gamma}_{i,0,t}$ and $\hat{\gamma}_{i,1,t}$ are the estimates of $\gamma_{i,0}$ and $\gamma_{i,1}$, respectively obtained from equation (2.7) by regressing $R_{i,2:t}$ on a constant, $R_{m,2:t}$ and $R_{m,2:t}.X^*_{1:t-1}$. Where $X^*_t \subseteq X_t$ indicates a vector of L predictors identified through model selection approaches to be discussed in section (2.4). We use a rolling-window approach (e.g., Fama & MacBeth 1973) which employs a window of fixed length w (60 months in our case) to estimate the market beta of asset i. Specifically, to have $\hat{\beta}_i^{DMS}$ at time t, we simply estimate equation (2.7) by using the observations within the estimation window $[t-w+2 \quad t]$ for R_i and R_m and $[t-w+1 \quad t-1]$ for X^*. To generate a beta forecast for the next period, we move the window forward by one step while keeping the window size fixed, adding one new observation and dropping the farthest one. The process continues until we obtain the final forecast at time T, which effectively generates the sequence of S out-of-sample beta estimates.

2.3.3.2 Second-pass Regressions and Cross-sectional Tests of CCAPM

After obtaining the out-of-sample betas, we test the model by running a cross-sectional regression at each time t of the evaluation period, with the first-step betas obtained through DMS approaches serving as an explanatory variable.

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}\hat{\beta}_{i,t}^{DMS} + \alpha_{i,t+1} \tag{2.9}$$

where $\hat{\beta}_{i,t}^{DMS}$ is the conditional β of asset i based on a given DMS approach, $\lambda_{0,t+1}$ represents the expected excess return on a 'zero beta' portfolio and $\lambda_{1,t+1}$ denotes the expected market risk premium. This will generate $S \times 1$ out-of-sample estimates of $\hat{\lambda}_0$ and $\hat{\lambda}_1$, and $S \times N$ estimates of pricing errors $\hat{\alpha}$.

To test the model, we first get the time-series averages of excess zero-beta rate (λ_0), risk premium (λ_1) and pricing errors (α_i) as:

$$\bar{\hat{\lambda}}_0 = \frac{1}{S}\sum_{t=1}^{S}\hat{\lambda}_{0,t} \tag{2.10}$$

$$\bar{\hat{\lambda}}_1 = \frac{1}{S}\sum_{t=1}^{S}\hat{\lambda}_{1,t} \tag{2.11}$$

$$\bar{\hat{\alpha}}_i = \frac{1}{S}\sum_{t=1}^{S}\hat{\alpha}_{i,t} \tag{2.12}$$

Our main asset pricing test is based on testing whether DMS models imply a reasonable risk-free rate (zero-beta rate, R_{zb}) and thus adequately fits the equity premium. Here R_{zb} represents the return on an asset with zero sensitivity to risk factors (conditional beta, in our case) of any given pricing model. If a model's R_{zb} is equivalent to the prevailing risk-free rate, that model fits the equity premium well (Black 1972). The fitted constant in the cross-sectional regression given in equation (2.9) indicates the difference between the implied R_{zb} and the observed risk-free rate (i.e., $\hat{\lambda}_0 = R_{zb} - R_f$). There are two versions of tests in the literature. The first version constraints the zero-beta rate to the return on a risk-free asset by setting the cross-sectional constant to zero. The second version, however, relaxes this restriction and estimate the model with a constant.[18] If unrestricted constant of a given model is statistically insignificant (i.e., $\hat{\lambda}_0 = R_{zb} - R_f \equiv 0$), we can say that the model adequately fits the equity premium (see, e.g., Jagannathan & Wang 1996, Cochrane 2005). We follow the second version and estimate the cross-sectional regression with a constant and test whether the average excess zero-beta rate is

[18] Note that under the restricted version the investor's wealth is assumed to be allocated between equities and Treasuries, whereas the unrestricted version implies that the wealth is only allocated in equities.

insignificant. Following the asset pricing theory, we also test whether the estimated risk-premium is significant. Specifically, we use the following t-statistics to test DMS models;

$$t-\text{statistic}_{\bar{\hat{\lambda}}_k} = \frac{\bar{\hat{\lambda}}_k}{\hat{\sigma}_{\lambda_k}}, \qquad k = 0, 1 \tag{2.13}$$

where $\bar{\hat{\lambda}}_k$ and $\hat{\sigma}_{\lambda_k}$ are averages and standard errors of $\hat{\lambda}_0$ and $\hat{\lambda}_1$, respectively. Note that to compute t-statistics, we use Newey & West (1987) consistent standard errors for heteroskedasticity and autocorrelation.

2.3.3.3 Performance Evaluation

We also consider four different measures for assessing and comparing the performance of various models under consideration. First, we assess each model's ability to generate insignificant pricing errors for individual assets using a significance level of 1% and 5%. The total number of mispriced assets (MPA) out of a total of N assets is our performance metric. A model with a lower value of MPA indicates a better pricing ability. Next, we follow Adrian & Rosenberg (2008) and use the sum of square pricing errors (SSPE) which is defined as:

$$SSPE = \bar{\hat{\alpha}}' \, \bar{\hat{\alpha}} \tag{2.14}$$

SSPE does not take into account the number of assets, so we use the root mean square pricing errors (RMSPE), which is computed as:

$$RMSPE = \sqrt{SSPE/N} \tag{2.15}$$

Next, following Jagannathan & Wang (1996), we use adjusted R^2:

$$Adjusted\ R^2 = 1 - \frac{(S-1)(1-R^2)}{(S-K-1)} \tag{2.16}$$

$$R^2 = \frac{var_c(\overline{R}) - var_c(\bar{\hat{\alpha}})}{var_c(\overline{R})} \tag{2.17}$$

where $\overline{R} = 1/S \sum_{t=1}^{S} R_t$ and var_c is cross sectional variance. The metrics such as SSPE and adjusted R^2, give all test assets equal weight, but some assets are actually less volatile than others (Campbell & Vuolteenaho 2004). To overcome this issue, two additional metrics are considered, both of which test whether all the pricing errors from cross-sectional regressions are jointly zero ($H_0 : \hat{\alpha} = 0$). The first measure is JA (joint alpha test) which is a χ^2-statistic and can be given as:

$$JA = \bar{\hat{\alpha}}' cov(\bar{\hat{\alpha}})^{-1} \bar{\hat{\alpha}} \quad \sim \quad \chi^2_{N-P} \tag{2.18}$$

where N, P, $\bar{\hat{\alpha}} = \frac{1}{S}\sum_{t=1}^{S} \hat{\alpha}_t$, and $\hat{\alpha}_t = [\hat{\alpha}_{1,t}, \hat{\alpha}_{2,t}, ..., \hat{\alpha}_{N,t}]'$ denote number of assets, number of factors in a given model, the average pricing errors and vector of estimated errors, respectively. According to the joint χ^2 test, if the pricing theory holds, the pricing errors generated by a model should be close to or equal to zero. The higher the statistic value, the greater the pricing errors produced by the model. JA value is compared to the critical value to test the significance of pricing errors. If JA exceeds the χ^2_{N-P} 5% critical value, the pricing errors are significant. In order to estimate the variance-covariance matrix of pricing errors $\bar{\hat{\alpha}}$, denoted as $cov(\bar{\hat{\alpha}})$ and a version accounting for autocorrelation, denoted as $\widetilde{cov}(\bar{\hat{\alpha}})$, we estimate following equations:

$$cov(\bar{\hat{\alpha}}) = \frac{1}{S^2} \sum_{t=1}^{S} (\hat{\alpha}_t - \bar{\hat{\alpha}})(\hat{\alpha}_t - \bar{\hat{\alpha}})' \tag{2.19}$$

$$\widetilde{cov}(\overline{\hat{\alpha}}) = \frac{1}{S^2}\sum_{t=1}^{S}(\hat{\alpha}_t - \overline{\hat{\alpha}})(\hat{\alpha}_t - \overline{\hat{\alpha}})' + \frac{1}{S^2}\sum_{j=1}^{q}\sum_{t=j+1}^{S}(1 - \frac{j}{q+1})(\hat{\alpha}_t - \overline{\hat{\alpha}})(\hat{\alpha}_{t-j} - \overline{\hat{\alpha}})' \qquad (2.20)$$

where $q = \lfloor(4(S/100)^{2/9})\rfloor$ and $\lfloor x \rfloor$ denotes larger integer not greater than x.

Our second measure for joint alpha test is the Composite Pricing Error (CPE), which was used by Campbell & Vuolteenaho (2004) and is defined as:

$$CPE = \overline{\hat{\alpha}}'\hat{\Omega}^{-1}\overline{\hat{\alpha}} \quad \sim \quad \chi^2_{N-P} \qquad (2.21)$$

where $\hat{\Omega}$ represents a diagonal matrix with main diagonal carrying the variances of estimated returns. Under this measure of an aggregate pricing error, assets with more volatile alphas receive less weight. The null hypothesis that pricing errors produced by a given model are jointly zero is rejected, if CPE exceeds the 5% critical value.

2.4 Dynamic Model Selection (DMS) Approaches

In this section, we discuss the implementation of various *dynamic model selection* (DMS) approaches to model the beta dynamics. We classify these approaches into two groups which are discussed in two different subsections. The first group consists of our new model selection approach that, at each point in time, chooses the model that currently has the best asset pricing performance. The second group consists of standard model selection approaches that are not based on asset pricing criteria and include sequential selection, best subset selection and shrinkage methods.

2.4.1 Dynamically Selected Beta Model (DSBM)

Under this approach, we use asset pricing criteria to select the best subset of variables from all possible combinations of conditioning variables. This selection of the optimal subset is performed at each point in time t. The optimal information set changes over time; therefore, we call this approach as *dynamically selected beta model* (henceforth DSBM). This approach of exhaustive search can be time-consuming and computationally intensive. However, this limitation is not a major concern for our research because the Goyal & Welch (2008) database of conditioning variables is considered as comprehensive, and has 14 variables so the number of potential combinations can still be managed. To implement our approach, we first estimate all possible models using K predictors as:

$$M_0 = 2^K - 1 \qquad (2.22)$$

where M_0 denotes the number of possible models. Here a model is defined as the particular subset of K predictors. Minus 1 represents the model when $k = 0$, which means a model with a constant only. Note that the total available models M_0, are indexed by $j = 1, \ldots, M_0$. X_{jt} represents the vector of explanatory variables in model j. X_t^* indicates a vector of explanatory variables in the best model among all models.

The first step of our procedure is to screen out models which exhibit high multicollinearity. Multicollinearity is the condition where X variables are highly correlated and often causes difficulty for estimating regression parameters (Hocking 2013). Due to multicollinearity, the standard error of the coefficients may be unreasonably high (McClendon 2002), suggesting that the coefficients for one or more independent predictors may not be significant. In other words, multicollinearity causes certain variables statistically insignificant when they should be significant. The other problem is that a parameter estimate may achieve a different sign than expected (Efron et al. 2004). The negative impacts of multicollinearity in terms of forecast accuracy on model selection and model averaging has been identified by studies such as Jayakumar (2014), Daoud (2017) and others.

One approach to address or reduce the problem of multicollinearity is to drop redundant variables directly from the regression model (Bowerman & O'Connell 1993, Lin 2008). However, to remove a variable, one needs to detect the multicollinearity, which is often done through *variance inflation factor* (VIF). By calculating the impact of multicollinearity on an estimated regression coefficient in terms of increase in its variance, the VIF quantifies the severity of multicollinearity in an OLS regression (O'brien 2007).[19] To show the degree of multicollinearity, each indicator has a VIF value, and a large value indicates that a variable needs to be either removed or substituted. A standard thumb rule is that the VIF greater than 5 indicates high multicollinearity (Daoud 2017). Following this, at each period we calculate the VIF factor and exclude all the models with the $VIF > 5$. We indicate the selected models as M, where $M \subseteq M_0$ by only including M_j if $VIF < 5$. These models are indexed by $j = 1, \ . \ . \ . \ , \ M$.

After screening models using above VIF constraint, we then select the best model by evaluating the asset pricing performance of each model. Specifically, we consider following three different approaches.

2.4.1.1 DSBM-I

We fit the following CCAPM model in the training sample using data up to time t for each combination of variables (X_j).

$$R_{i,t+1} = a_{i,j}^{IV} + (\gamma_{0,i,j} + \gamma_{1,i,j}' X_{t,j}) R_{m,t+1} + \varepsilon_{i,j,t+1} \tag{2.23}$$

For CCAPM to hold, pricing errors ($a_{i,j}^{IV}$) for all the assets should be zero. This can be tested with the following null hypothesis:

$$H_0: \quad a_{i,j}^{IV} = 0, \quad \forall i$$

To test our H_0, we use the following test:

$$F = \left[\frac{(T - N - K)}{N}\right] [1 + \hat{\mu}_k' \hat{\Omega}_k^{-1} \hat{\mu}_k]^{-1} \ \hat{a}_0 \hat{\Sigma}^{-1} \hat{a}_0' \sim F_{T-N-K}^{N} \tag{2.24}$$

where T is total number of observations, N represents the number of assets, K indicates the number of parameters. $\hat{\mu}$ is a $K \times 1$ vector of the average values of the K explanatory variables, $\hat{\Omega}$ is a $K \times K$ covariance matrix of K and $\hat{\Sigma}$ is $N \times N$ residual covariance matrix.

Note that we may have more than one models which pass the test, i.e., producing insignificant pricing errors (α) for all the assets. We denote these models with P, and in case none of the models passes the joint tests, then we take all the models to next step ($P = M$) where for each model we compute average R^2 across the N test assets:

[19] $VIF = \frac{1}{1-R_i^2}$, we can calculate the VIF for each variable in the model, and the process is to regress the variable, assuming that it is i^{th} variable against all other predictors.

$$\overline{R}_j^2 = \frac{1}{N}\sum_{i=1}^{N} R_{ij}^2 \tag{2.25}$$

where

$$R_{ij}^2 = \frac{Var[\hat{\beta}_{ijt}R_{mt}]}{Var[R_{it}]} = \frac{Explained\ Variance}{Total\ Variance} \tag{2.26}$$

Next, we select the best model (X^*) as one that achieves the maximum average R^2. We call this approach as DSBM-I.

2.4.1.2 DSBM-II

Our next approach is to choose the model based on cross-sectional measures of performance. Specifically, we run following cross-sectional regression at each time t.

$$R_{i,t+1} = \lambda_{0,j,t+1} + \lambda_{1,j,t+1}\hat{\beta}_{i,j,t} + \alpha_{i,j,t+1} \tag{2.27}$$

where $\hat{\beta}_{i,j,t}$ represent the conditional fitted betas, $\hat{\beta}_{i,j,t} = \hat{\gamma}_{0,i,j} + \hat{\gamma}'_{1,i,j}X_{t,j}$, obtained from time-series regression given in equation (2.23) for each model j. According to theory, the average cross-sectional intercept ($\bar{\hat{\lambda}}_0$) should be zero. To test this, we use t-statistics given in equation (2.13) to see whether average cross-sectional intercept ($\bar{\hat{\lambda}}_0$) is insignificant at 5% level. Similar to the first approach, we denote shortlisted models as P and when none of the models produce insignificant $\bar{\hat{\lambda}}_0$, then $P = M$. In the next step, we compute the cross-sectional adjusted R^2 given in equation (2.16) for each of P models. Finally, we select the best model (X^*) as one that achieves the maximum adjusted R^2. We call this approach as DSBM-II.

2.4.1.3 DSBM-III

In our final approach, we estimate the *mean squared pricing error* (MSPE) based on asset pricing errors over cross-validation sample.[20] Given that we leave V observations in validation sample, the MSPE can be given as:

$$\text{CV–MSPE}_{i,j} = \frac{1}{V}\sum_{t=1}^{V} \hat{\alpha}_{i,j,t}^2 \tag{2.28}$$

where $\hat{\alpha}_{i,j,t}$ is one-month ahead conditional pricing error based on the predicted beta obtained with model j.[21] A model with minimum CV-MSPE is selected as the best. We call this approach as DSBM-III.

2.4.1.4 Testing DSBM models

Let $X_t^* \subset X_t$ be the subset variables identified through DSBM-I, DSBM-II, or DSBM-III. The DSBM beta for asset i using estimated parameters of $\hat{\gamma}_{0,i}$ and $\hat{\gamma}_{1,i}$ from (2.7) based on observation up to time t:

$$\hat{\beta}_{i,t}^{DSBM} = \hat{\gamma}_{0,i,t} + \hat{\gamma}'_{1,i,t}X_t^* \tag{2.29}$$

[20] See section (2.3.2) for details on cross-validation and sample-split

[21] Note that under this approach, a model is allowed to vary across assets as well. Whereas in DSBM-I and DSBM-II the same model is selected to price each asset.

In the next step, we run a cross-sectional regression at each period t:

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}\hat{\beta}_{i,t}^{DSBM} + \alpha_{i,t+1} \tag{2.30}$$

Finally, we evaluate the performance of each model (DSBM-I, DSBM-II, and DSBM-III) using tests discussed in section (2.3.3.2) and (2.3.3.3).

2.4.2 Non Asset Pricing Model Selection Approaches

In this section, we discuss some of the other DMS approaches that are widely used in forecasting literature but have not been applied to CCAPM. These approaches include best subset selection, stepwise selection and shrinkage methods.

2.4.2.1 Best Subset Selection/Exhaustive search

Under best subset selection or exhaustive search approach, we first estimate all possible combinations of variables as discussed in the previous section. All methods under this approach are based on estimation results of equation (2.31) for all possible M models using training data up to time t.

$$R_{i,t+1} = a_{i,j}^{IV} + (\gamma_{0,i,j} + \gamma'_{1,i,j}X_{j,t})R_{m,t+1} + \varepsilon_{i,t+1}, \tag{2.31}$$

Given that j model contains p predictors estimated with T observations. RSS_j indicates the residual sum of squares and TSS_j is the total sum of squares for model j estimated through equation (2.31).

Performance metrics such as adjusted R^2, and information criteria such as AIC, BIC and Mallow's C_p are the most widely used criteria to evaluate the quality of a multivariate regression model and to compare different models. We use all these criteria in this study.

2.4.2.1.1 Information Criteria The Akaike information criteria (AIC) (Akaike 1973) and the Bayesian information criterion (BIC) (Schwarz 1978) are the procedures that have gained more attention from the list of model selection procedures based on information criteria. By enforcing penalties, information criteria quantify the loss of information. AIC and BIC are defined as:

$$AIC = T\ln(RSS/T) + 2(p+1) \tag{2.32}$$

$$BIC = T\ln(RSS/T) + (p+1)\ln(T) \tag{2.33}$$

Both AIC and BIC try to balance the complexity of the model and its goodness of fit. Usually, RSS would become smaller or at least the same with more explanatory variables in a regression model, suggesting that the first component of AIC and BIC would become smaller. More predictors, however, will increase the model's complexity and thus, the second component of AIC and BIC would become larger. During the model selection process, a model with the smallest value of either AIC or BIC is selected.

Hurvich & Tsai (1989) provide a refined version of the AIC estimate for small samples:

$$AICc = AIC + 2\frac{(p+1)(p+2)}{T-p-2} \tag{2.34}$$

Burnham & Anderson (2002) argue that AICc should be used whenever $T/p < 40$.

We also use Mallows C_p which is a variant of AIC developed by Mallows (1973). It compares a subset model with p predictors with a full model with K predictors, $(K > p)$. The C_p statistic can be given as:

$$C_{p,j} = \frac{RSS_{p,j}}{MSE_K} + 2(p+1) - T \tag{2.35}$$

where $RSS_{p,j}$ represents the residual sum of squares for model j consisting of p predictors and MSE_K is used for the mean squared errors of a full model containing all K predictors. Similar to AIC, BIC, and AICc, a model with lowest value of C_p is selected.

2.4.2.1.2 Adjusted R^2 For an ordinary least squares model with p variables, estimated with T observations, the adjusted R^2 can be given as:

$$Adjusted\ R_j^2 = 1 - \frac{RSS_j/(T-p-1)}{TSS_j/(T-1)} \tag{2.36}$$

where TSS represents the total sum of squares and RSS indicates the residual sum of squares.[22] Unlike information criteria such as AIC and BIC where a lower value is preferred, a model with highest value of adjusted R^2 is selected.

2.4.2.2 Sequential Selection

The subset variable selection requires the estimation of all possible models (2^K). In the case of a very large K, the implementation of the best subset selection becomes difficult due to high computational requirement. For example, with 30 variables, we would have to estimate more than 1.07 billion models every time to find the right one, which could lead to overfitting. Moreover, as the search space grows (e.g., 16384 models with 14 predictors), it becomes more likely to find models that appear to be successful on training samples but have little or no predictive ability on future data. To overcome these issues, sequential selection techniques, which look at a much smaller number of models, are preferable to the best subset selection method. Four widely used variants of the sequential selection technique include *forward stepwise selection* (FSS), *backward elimination* (BE), *stepwise regression* (SReg), and *univariate selection* (US).

2.4.2.2.1 Forward stepwise selection (FSS) FSS starts with a model that includes no predictors, and then one predictor that improves the model's fit at each step is included in the model. A variable, once added, cannot be removed. To implement FSS, we follow following steps.

(i) We begin a single variable in the model.[23]

(ii) Check p-value of all the predictors currently not part of the model and identify a predictor with with lowest p-value less than α_{crit}.

(iii) Continue the process until all significant predictors are added in the model.

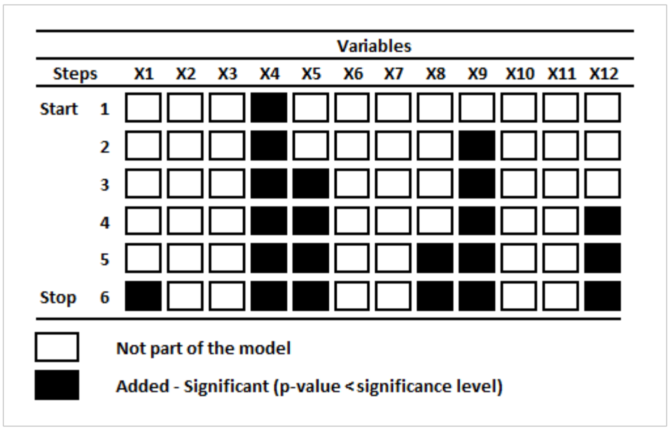
The α_{crit} is sometimes called the "p-to-remove or add". We set the $\alpha_{crit} = 10\%$.[24]

[22] $TSS = \sum_{t=1}^{T}(y_t - \bar{y})^2$ and $RSS = \sum_{t=1}^{T}(y_t - \hat{y}_t)^2$, where $\bar{y}$ is mean across T observations.

[23] Generally, FSS starts with no variables in the model, however, in our case we want to link beta dynamics with CIVs, so we want to have at least one variable in the model; otherwise, our model will become unconditional CAPM. We choose the first variable which has the highest t-value

[24] We also set α_{crit} to 5% and 15%, however, results at 10% are better.

Figure 2.4: Forward Selection



2.4.2.2.2 Backward stepwise elimination (BSE) BSE starts with a full model having all K variables, and then the least important variable is iteratively eliminated at each step. Once a predictor is eliminated, it cannot return to the model. To implement BSE, we follow following steps.

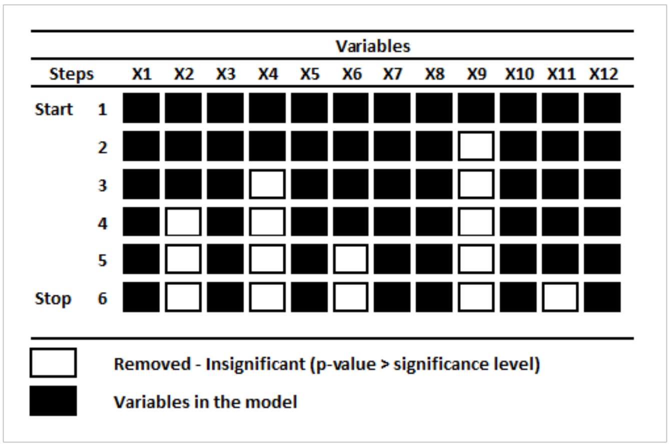
(i) We begin with a full model containing all K predictors.

(ii) Remove the least important predictor defined as one with highest p-value (insignificant), p-value greater than $\alpha_{crit} = 10\%$.

(iii) Refit the model and go to (ii)

(iv) Stop when there are no more insignificant predictors, when all p-values are less than $\alpha_{crit} = 10\%$.

Figure 2.5: Backward stepwise elimination (BSE)



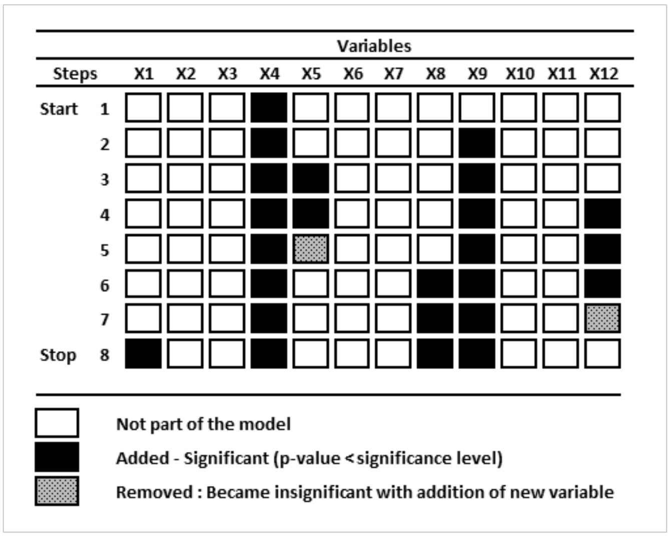
2.4.2.2.3 Stepwise regression The next method is stepwise regression ($SReg$) which combines both BE and FSS. $SReg$ is a variant of the forward selection method, with the exception that variables that are in the model are not always retained. Similar to FSS, one variable at-a-time is added. However, the model is re-estimated after a new variable is added, and a variable already in the model can be omitted if it has become irrelevant due to the addition of a new predictor. To implement $SReg$, we follow following steps.

(i) We begin with the most significant univariate model.

(ii) Repeat: Perform an FSS stage. Next, execute a BSE phase after each variable inclusion. Variables that were omitted in previous FSS steps should be reconsidered in subsequent FSS steps.

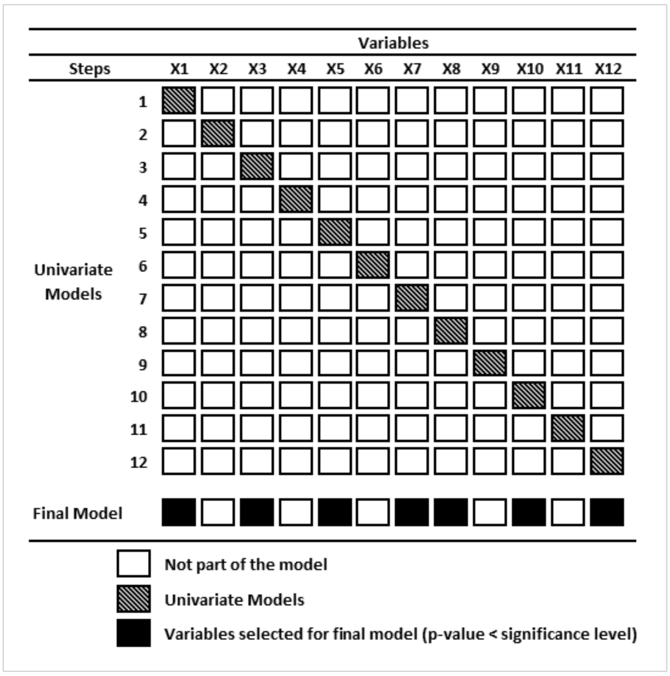
(iii) Stop if there are no more significant variables to add, when all p-values are more than $\alpha_{crit} = 10\%$.

Figure 2.6: Stepwise Regression



2.4.2.2.4 Univariate Selection The next method is the univariate selection (US), where we first estimate all univariate models. In the next step, we select the variables with $p < \alpha_U$. Figure (2.7) summarises the process of selecting variables based on univariate models at a given time.

Figure 2.7: Univariate Selection



In addition to these standard approaches, we also implement the sequential selection approaches where rather than identifying the significant predictors in traditional way we use asset pricing criteria. More specifically, we use cross-sectional t-statistics on zero-beta rate as measure to select more significant predictors in forward selection framework.

2.4.2.3 Shrinkage Methods

The shrinkage methods also called 'parameterisation penalties' or 'penalised regressions' are aimed at overcoming the problems associated with *kitchen sink* (KS) regression by using regularisation constraints. The main motivation behind these methods is the empirical evidence against the KS regressions for weak out-of-sample performance due to overfitting (e.g., Goyal & Welch 2008). In other words, if additional "irrelevant" predictors are part of the model, the simple predictive regression model can be ineffective. Regularisation constraints may therefore be used to prevent overfitting and to simplify the model through simultaneous coefficient shrinkage and variable selection. The most popular method of parameter parsimony is parameterisation penalty and is included in the objective function to maximise out-of-sample efficiency. It is thus necessary to give the new objective function as follows:

$$\mathcal{L}(\gamma;.) = \underbrace{\mathcal{L}(\gamma)}_{\text{Loss function}} + \underbrace{\phi(\gamma;.)}_{\text{Penalty}} \tag{2.37}$$

The "kitchen sink" OLS model, containing all the predictors (K) is given as:

$$R_{t+1} = a_0 + \gamma_0 R_{m,t+1} + \sum_{j=1}^{K} \gamma_j (R_{m,t+1}.X_{jt}) + e_{t+1} \tag{2.38}$$

The loss function is the residual sum of squares (RSS) of equation (2.38), in which all K predictors are part of the model. To obtain the penalty term $\phi(\gamma;.)$ we consider following four commonly used penalties.

2.4.2.3.1 LASSO The penalty term under LASSO estimate proposed by Tibshirani (1996) takes the following form:

$$\phi(\gamma;\lambda) = \lambda \sum_{j=1}^{K} |\gamma_j|$$

where hyperparameter λ is the degree of parameterisation penalty. Following literature (e.g., Gu et al. 2020), λ is optimised adaptively using MSFE in cross-validation sample. The LASSO estimator can be obtained through the following optimisation problem

$$\hat{\gamma}^{lasso} = \underset{\gamma_j}{\text{minimise}}\ RSS + \lambda \sum_{j=1}^{K} |\gamma_j| \tag{2.39}$$

where $RSS(a_0, \gamma_0, \gamma_j)$ is the sum of squared residuals of equation (2.38), in which all K predictors are part of the model. λ is the tuning parameter which shrinks the coefficients to 0 for large λ, $\hat{\gamma}^{lasso}$ is the $K \times 1$ vector of estimated parameters of γ_j using LASSO. The conditional beta estimate made at time t with LASSO can be given as:

$$\hat{\beta}_{i,t}^{LASSO} = \gamma_{0,t}^{lasso} + \sum_{j=1}^{K} \hat{\gamma}_{j,t}^{lasso}.X_{j,t} \tag{2.40}$$

2.4.2.3.2 Adaptive LASSO Zou (2006) shows that the LASSO constraint does not choose the correct subset of variables and proposes a new approach called adaptive LASSO, which applies weights to all of the LASSO parameters. The penalty term under adaptive LASSO estimate takes the following form:

$$\phi(\gamma;\lambda) = \lambda \sum_{j=1}^{K} \omega_j |\gamma_j|$$

Suppose $\hat{\omega} = 1/|\ \hat{\lambda}\ |$ and $\hat{\lambda}$ is a consistent estimator for λ, Zou (2006) defines the adaptive lasso estimates by:

$$\hat{\gamma}^{AdLasso} = \underset{\gamma_j}{\text{minimise}}\ RSS + \lambda \sum_{j=1}^{K} \hat{\omega}_j |\gamma_j| \tag{2.41}$$

The conditional beta estimate made at time t with adaptive LASSO can be given as:

$$\hat{\beta}_{i,t}^{AdLasso} = \gamma_{0,t}^{AdLasso} + \sum_{j=1}^{K} \hat{\gamma}_{j,t}^{Adlasso}.X_{j,t} \tag{2.42}$$

2.4.2.3.3 Ridge Next shrinkage approach is Ridge introduced by Hoerl & Kennard (1970). which instead of setting coefficients of irrelevant parameters to zero, tries to compress them as close to zero as possible. Unlike LASSO and adaptive LASSO where coefficients are set to zero, Ridge compresses the coefficients of the irrelevant parameters to as close to zero as possible. The penalty term of Ridge can be given as:

$$\phi(\gamma;\lambda) = \lambda \sum_{j=1}^{K} |\gamma_j|^2,$$

The Ridge estimator can be obtained through following optimisation problem

$$\hat{\gamma}^{Ridge} = \underset{\gamma_j}{\text{minimise}}\ RSS + \lambda \sum_{j=1}^{K} |\gamma_j|^2 \tag{2.43}$$

The main difference between LASSO and Ridge is that in case of highly correlated variables, LASSO usually selects one at random, while Ridge shrinks them toward each other (Zou & Hastie 2005). The conditional beta estimate made at time t with Ridge can be given as:

$$\hat{\beta}_{i,t}^{Ridge} = \gamma_{0,t}^{Ridge} + \sum_{j=1}^{K} \hat{\gamma}_{j,t}^{Ridge}.X_{j,t} \tag{2.44}$$

2.4.2.3.4 Elastic Net The final shrinkage method that we consider is the Elastic Net (ENet) developed by Zou & Hastie (2005). The penalty term in ENet is the linear combination of Ridge and LASSO. The ENet penalty can be given as:

$$\phi(\gamma;\lambda,\rho) = \lambda(1-\rho) \sum_{j=1}^{K} |\gamma_j| + \lambda\rho \sum_{j=1}^{K} |\gamma_j|^2,$$

The Ridge estimator can be obtained through following optimisation problem

$$\hat{\gamma}^{ENet} = \underset{\gamma_j}{\text{minimize}}\ RSS + \lambda(1-\rho) \sum_{j=1}^{K} |\gamma_j| + \lambda\rho \sum_{j=1}^{K} |\gamma_j|^2 \tag{2.45}$$

where ρ is a non-negative hyperparameter which is used to balance the penalty terms of LASSO and Ridge. Note that if the $\rho = 0$ ($\rho = 1$), then equation (2.45) reduces to the LASSO (Ridge). The ENet promotes simple models for intermediate values of $\rho = 0.5$ through both shrinkage and selection. Similar to λ, the value of ρ is adaptively optimised using MSFE in the validation sample. The conditional beta estimate made at time t with ENet can be given as:

$$\hat{\beta}_{i,t}^{ENet} = \gamma_{0,t}^{ENet} + \sum_{j=1}^{K} \hat{\gamma}_{j,t}^{ENet}.X_{j,t} \tag{2.46}$$

2.4.2.4 Testing non-asset pricing methods

Let $\hat{\beta}_{i,t}^{NAP}$ be the beta estimates made at time t with a given non-asset pricing method (NAP). The cross-sectional regression at each time t can be given as:

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}\hat{\beta}_{i,t}^{NAP} + \alpha_{i,t+1} \tag{2.47}$$

Finally, we evaluate the performance of each model using tests discussed in section (2.3.3.2) and (2.3.3.3).

2.5 Data and Benchmark Models

2.5.1 Data

Following a large body of empirical research on explaining cross-sectional variation in expected returns, we use the 25 size and book-to-market portfolios to perform the tests for a sample period from July 1926 to December 2018. In addition, for robustness tests, we also use 25 size and momentum portfolios, 30 industry portfolios, and 10 momentum portfolios. The returns of all these portfolios and market factor are calculated in excess of risk-free rate. The data on portfolio returns, risk-free rate, and market factor are taken from Kenneth French's website.[25]

The conditioning information variables are from Goyal & Welch (2008), who provide detailed descriptions of the data and their sources. The dataset includes 14 variables considered as relevant in predicting equity premium in past empirical studies.[26] These variables include stock characteristics (the dividend yield (DY), the dividend-price ratio (DP), the dividend-payout ratio (DE), the earning-price ratio (EP), the book-to-market ratio (BM), the net equity expansion (NTIS), and the stock variance (SVAR)), interest rate related variables (the Treasury bill rate (TBL), the long-term return (LTR), the long-term yield (LTY), the term spread (TMS), the defaults-return spread (DFR), and the default-yield spread (DFY)), and inflation (INFL) to represent the macroeconomy. The description of these variables is given in Table 2.1. We use monthly data for all these variables spanning from July 1926 to December 2018.

[Insert Table 2.1 about here]

2.5.2 Benchmark Models

To analyse the performance of the various DMS approaches (see section 2.4) relative to the standard asset pricing models, we consider three standard conditional CAPM and two multi-factor models. Standard CCAPM approaches include models where beta dynamics is captured through: i) 60-month rolling window (Fama & MacBeth 1973), ii) short window (Lewellen &

[25] We use updated version of dataset available at
https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[26] We use updated version of dataset available at http://www.hec.unil.ch/agoyal/

Nagel 2006), and iii) predetermined conditioning variable (Ferson & Harvey 1999). Multi-factor models include: i) the Fama & French (1993) three-factor model, and ii) the four-factor model of Carhart (1997). For all these models, we use Fama & MacBeth (1973) two-stage regressions, where time-series regressions are estimated using 60-month rolling window.

2.5.2.1 Standard Conditional CAPM Models

2.5.2.1.1 CCAPM with Rolling window Under this approach, we follow Fama & MacBeth (1973), and use a 60-month rolling window to model the beta dynamics. The time-series regression to obtain time-varying beta β_t^{RW} can be given as:

$$R_{i,t+1} = \alpha_i + \beta_{i,t}^{RW}(R_{m,t+1}) + \varepsilon_{i,t+1}, \tag{2.48}$$

where

$$\beta_{i,t}^{RW} = \frac{Cov(R_{i,t+1}, R_{m,t+1})}{Var(R_{m,t+1})} \tag{2.49}$$

2.5.2.1.2 CCAPM with Short-window Under this approach, we follow Lewellen & Nagel (2006) and use daily return data for estimating a separate CAPM regression at each month t to obtain a time-series of non-overlapping CCAPM regression parameters covering the entire sample. Following Lewellen & Nagel (2006), Boguth et al. (2011), Cederburg & O'Doherty (2016), and others, we use the following regression model that includes the lags of the excess market return to mitigate the effects of asynchronous trading:

$$R_{i,d} = \alpha_i + \beta_{i,0}R_{m,d} + \beta_{i,1}R_{m,d-1} + \beta_{i,2}\left[\frac{R_{m,d-2} + R_{m,d-3} + R_{m,d-4}}{3}\right] + e_{i,d} \tag{2.50}$$

where $R_{i,d}$ is excess daily return of asset i at day d, $R_{m,d}$ is excess market returns. To reduce the effect of asynchronous trading we follow Lewellen & Nagel (2006) and use lags of the excess market return. The estimate of asset beta for month t can be given as:

$$\hat{\beta}_{i,t}^{sw} = \hat{\beta}_{i,0} + \hat{\beta}_{i,1} + \hat{\beta}_{i,2} \tag{2.51}$$

2.5.2.1.3 CCAPM with predetermined IVs Under this standard approach, we follow Ferson & Harvey (1999) and predetermine the set of conditioning information variables and use the same variables at each period to estimate out-of-sample betas. We first run following time series regression at each time with a 60-month rolling window:

$$R_{i,t+1} = a_i^{IV} + (\gamma_{i,0} + \gamma_{i,1}' X_t)R_{m,t+1} + \varepsilon_{i,t+1} \tag{2.52}$$

where X_t is set of 12 conditioning variables from Goyal & Welch (2008) for kitchen sink model.[27] Following Ferson & Harvey (1999), we use 4 standard variables including term spread, default yield, t-bill rate and spread between 3 months and one-month t-bill rate.

$$\hat{\beta}_{i,t}^{PIV} = \hat{\gamma}_{i,0,t} + \hat{\gamma}_{i,1,t}' X_t \tag{2.53}$$

where $\hat{\beta}_{i,t}^{PIV}$ is estimated conditional beta for either Ferson & Harvey (1999) (PIV=FH) or kitchen sink model (PIV=KS), $\hat{\gamma}_{i,0,t}$ and $\hat{\gamma}_{i,1,t}$ are the ordinary least squares (OLS) estimates of $\gamma_{i,0}$ and $\gamma_{i,1}$, respectively, in (2.52).

[27] Following Elliott et al. (2013), Zhang et al. (2020), and others, we exclude the log dividend–earnings ratio and the long-term yield from 14 predictors to avoid multicollinearity.

2.5.2.2 Multi-factor Models

The following time-series regression is estimated at each period for multi-factor models (FF3F and Carhart) using a 60-month rolling window:

$$R_{i,t+1} = a_i + \beta'_{i,t} F_{t+1} + \epsilon_{i,t+1} \tag{2.54}$$

where $R_{i,t+1}$ represents the excess return on portfolio i at time $t+1$, F_{t+1} represents the $K \times 1$ vector of factors at time $t+1$, $\beta_{i,t}$ represents the $K \times 1$ vector of factor loadings for portfolio i, respectively.

Cross-sectional regression for benchmark models

In the second the pass for all benchmark models mentioned above, we test the model by estimating the following cross-sectional regression at each time:

$$R_{t+1} = \hat{\beta}_t \lambda_{t+1} + \alpha_{t+1} \tag{2.55}$$

where $\hat{\beta}_t = [1\ \hat{\beta}'_{t,1};\ ..;1\ \hat{\beta}'_{t,N}]$ is a $N \times (K+1)$ matrix that includes a $N \times 1$ vector of ones in its first column and estimated factor loadings in other columns. $\lambda_{t+1} = [\lambda_{0_{t+1}}\ \lambda_{1_{t+1}}\ ...\ \lambda_{K_{t+1}}]'$ represents a $(K+1) \times 1$ vector where λ_t is the conditional zero-beta rate and $\lambda_{k,t+1}$ is the conditional risk premium on the k^{th} factor in period $t+1$. Note that for all CCAPM $K = 1$, for FF3F (Carhart) $K = 3$ ($K = 4$). The asset pricing tests and performance evaluation metrics for benchmark models are already discussed in section (2.3.3.2 and 2.3.3.3).

2.6 Empirical Results

In this section, we discuss the empirical results of the various dynamic model selection (DMS) approaches, applied to the CCAPM. This section consists of four subsections: *(i)* cross-sectional results of DMS approaches, *(ii)* comparison of results from *(i)* with benchmark models, *(iii)* evaluating the ability of DMS approaches in explaining the size, value and momentum anomalies, and *(iv)* evaluating the extent to which DMS approaches address the *variable-selection uncertainty* (VSU).

2.6.1 Cross-sectional Results for DMS approaches

In this section, we discuss the results of the cross-sectional tests of the CCAPM based on various DSM approaches, which include: *(i)* dynamically selected beta models (DSBMs), and *(ii)* standard DMS approaches, including the sequential selection, best subset selection, and shrinkage methods. Our CCAPM cross-sectional tests are based on mainstream literature that evaluates the pricing abilities of a given model by looking at the significance of Fama & MacBeth (1973) parameter estimates.[28] In addition, we assess the performance of models through SSPE, RMSPE, cross-sectional adjusted R^2, a number of mispriced assets at 1% and 5%, joint alpha tests which include JA and CPE (see section (2.3.3.3) for details). The test assets are the 25 assets sorted by size and book-to-market ratio. The conditional betas are estimated with a 60-month rolling window and the out-of-sample period is August 1936 to December 2018.[29]

Table 2.2 reports the results of the standard Fama & MacBeth (1973) methodology for testing the CCAPM based on various dynamic model selection (DMS) approaches. The findings

[28] We use t-statistics to test the significance using Newey & West (1987) heteroskedasticity and autocorrelation consistent standard errors.

[29] Note that our sample starts from July 1926 to December 2018. The in-sample period is ten years because some methods use five years of data as a validation sample. To compare all the models, our out-of-sample analysis is based on August 1936 to December 2018.

in Table 2.2 clearly contradict the underlying theory. The constant (excess zero-beta rate) is significantly different from zero for all the DMS approaches, even though all the models are estimated on excess returns. In this case, the theory behind all the models would predict the constant to be zero. Moreover, the null hypothesis that pricing errors for all the assets are jointly zero is also rejected for all the models based on JA and CPE. Since all the DMS models fail the standard CCAPM test, so we compare the performance of these models based on various metrics discussed in section (2.3.3.3).

[Insert Table 2.2 about here]

Panel A consists of the best subset variable selection approaches, which evaluate all possible combinations of the predictor variables and then choose the best model according to some criteria. Panel A is further divided into (A.1) Asset pricing and (A.2) Traditional DMS methods. Under panel A.1, we report the results of three dynamically selected beta model (DSBM), which chooses the model at each point in time that currently has the best asset pricing performance (see section (2.4.1) for details). Results suggest that the best performing model is DSBM-III which achieves an adjusted R^2 of 39.84%, which is significantly higher than DSBM-I and DSBM-II, which achieve an adjusted R^2 of 29.13% and 35.12%, respectively. DSBM-III also outperforms the DSBM-I and DSBM-II by achieving lower pricing errors measured as SSPE and RMSPE. Panel A.2 reports the results of traditional methods, which include AIC, AICc, BIC, Mallow's C_p, and adjusted R^2 (see section (2.4.2.1) for details). Results reported in Panel A.2 of Table 2.2 suggest that AICc outperforms all the traditional best subset variable selection methods by achieving an R^2 of 24.70%. However, this performance is significantly lower than all the DSBM models where the best model, DSBM-III, achieves an adjusted R^2 of 39.84%.

Panel B of Table 2.2 reports the results of sequential selection approaches, which do not require estimation of all possible combinations of variables. Instead, they explore a much more limited range of models, making them desirable alternatives to the best subset selection. We consider four widely used variants of the sequential selection techniques, including forward stepwise selection (FSS), backward elimination (BE), stepwise regression (SReg), and univariate selection (US) (see section (2.4.2.2) for details). Results given in Panel B.1 of Table 2.2 suggest that the stepwise regression (SReg) approach performs the best among sequential selection approaches by achieving an adjusted R^2 of 17.78%. However, this performance is significantly lower than the best subset variable approaches given in Panel A of Table 2.2. In addition to these standard approaches, we also implement the sequential selection approaches where we use asset pricing criteria to identify the significant predictors. Specifically, we use a cross-sectional t-statistics on the excess zero-beta rate as a measure to select more significant predictors in the forward selection (FSS-APC) and univariate (US-APC) framework. The reported results in Panel B.2 of Table 2.2 suggest that the introduction of asset pricing criteria significantly improves the sequential selection approach's performance. Model FSS-APC achieves an R^2 of 33.47%, which is not only significantly higher than all the sequential selection approaches but also significantly higher than the traditional best subset approaches reported in Panel A.2 of Table 2.2. However, DSM-III still significantly outperforms all the models considered so far.

Panel C of Table 2.2 reports the results of shrinkage methods which include LASSO, adaptive LASSO, Ridge, and Elastic Net (ENet) (see section (2.4.2.3) for details). Results suggest that the ENet achieves an adjusted R^2 of 13.67%, which remains the highest among the four shrinkage methods. However, these methods perform poorly compared to all the DMS approaches considered in this study. For example, adjusted R^2 of 13.67% achieved by ENet is about three times lower than the DSBM-III, which achieves an adjusted R^2 of 39.84%.

To summarise the results on DMS approaches, we find that the model selection approaches based on asset pricing criteria outperform DMS with standard approaches. The best performing model among all DMS approaches is the DSBM-III which selects the best models based on MSPE of an individual asset in CV sample. DSBM-III achieves an adjusted R^2 of 39.84%, and it also produces lower pricing errors by achieving the lowest values for measures such as SSPE, RMSPE, JA, and CPE.[30] The better performance of DSBM-III compared to standard methods may be attributed to the fact that the traditional models use the residuals of time-series regression to select beta models. On the hand, DSBM-III chooses the model that best prices the assets (i.e., consider the conditional pricing errors).

2.6.2 Comparing Results with Benchmark Models

In this section, we provide a direct comparison of the best performing DSM approach, DSBM-III, to the various standard CCAPM and multi-factor models. Standard CCAPM approaches include models where beta dynamics is captured through: i) 60-month rolling window (Fama & MacBeth 1973), ii) short window (Lewellen & Nagel 2006), iii) four standard variables (Ferson & Harvey 1999), and iv) kitchen sink model containing all variables. While, multi-factor models include: i) Fama & French (1993) three-factor model, and ii) the four-factor model of Carhart (1997). See section (2.5.2) for more details on benchmark models.

Results in Table 2.3 show that consistent with our previous findings for DMS approaches, cross-sectional constant (excess zero-beta rate) is highly significant for all the benchmark models.[31] However, results show that our DSBM-III outperforms the CCAPM benchmark models presented in Panel A of Table 2.3. The CCAPM benchmark models generally do a poor job. Starting with the zero-beta rate, all the benchmark models produce a significant zero-beta rate with a minimum of 0.8404 achieved by CAPM (β_{FH}). On the other hand, DSBM-III produces a zero-beta rate of 0.7124, which is significant but significantly lower than the CCAPM benchmark models.[32] Next, we can see that the estimated risk premium for all the CCAPM benchmark models is negative. However, DSBM-III produces a positive risk premium of 0.3919. In terms of asset pricing performance, our DSBM-III achieves an adjusted R^2 of 39.84%, which is higher than the best performing CCAPM benchmark, CCAPM (β_{RW}), which achieves an adjusted R^2 of 33.65%. Moreover, compared to our DSBM-III model, all the CCAPM models produce significantly larger errors measured as SSPE and RMSPE. However, all the models reject the null hypothesis of producing zero pricing errors ($H_0 : \overline{\hat{\alpha}} = 0$) for all the portfolios measured by measures of joint alpha (JA) and composite pricing errors (CPE). There are two main conclusions from these findings. First, conditioning information can play a significant role in capturing the beta dynamics compared to models where beta is a function of time, i.e., using a rolling window (β_{RW}) or short window (β_{SW}) to model time-variation in betas. Second, time-varying IVs can improve the performance of CCAPM-IV with the predetermined IVs.

[Insert Table 2.3 about here]

We can also examine the pricing performance of each model by plotting the average monthly estimated excess return with a given model, and the average monthly realised excess return for

[30] We also use cross-sectional (CS) excess zero-beta rate, CS R^2, and combination of excess zero-beta rate and CS R^2 as alternative criteria in cross-validation sample and find consistent results.

[31] Note that the models are estimated with a constant (Excess zero-beta rate). In unreported results, we also estimate the model without constant assuming that $R_{zr} = R_f$ and find consistent results, i.e. same ranking of models.

[32] The differences between zero-beta rates of DSBM-III and CCAPM benchmark models are significant at 5% level.

each asset. If a model perfectly fits the returns, then the test assets should lie on the 45^0 line in Figure 2.8. For a given test asset, the coordinates below (above) the 45^0 line correspond to the negative (positive) pricing errors, $\hat{\alpha}_i$. The Figure 2.8 provides a graphical illustration of the findings presented in Table 2.3. The CCAPM benchmark models fit the data poorly as there is a substantial deviation of assets returns from the 45^0 line. In contrast, the DSBM-III has asset returns that are reasonably close to the 45^0 line. However, dispersion is still high, which suggests that even after allowing IVs to vary over time, conditional CAPM does not fully explain the cross-section of asset returns for size and book-to-market sorted portfolios. On the other hand, if we compare the performance of our DSBM-III model with FF3F and Carhart models, we see that asset returns are relatively close to the 45^0 line, which implies that factor models fit the data well compared to all given models.

Panel B in Table 2.3 reports the results of Fama & French (1993) three-factor model (FF3F) and Carhart (1997) four-factor model (Carhart). Consistent with the findings of CCAPM benchmark models, multi-factor models produce a negative risk premium ($\bar{\hat{\lambda}}_{MKT}$) with -0.3113 and -0.2785 achieved by FF3F and Carhart, respectively. However, the HML factor is highly significant and positive, and the SMB factor is positive but significant at 10% level for both the models. On comparison of the performance of the DSBM-III and multi-factor models, we find that consistent with findings through graphical illustration in Figure 2.8, the FF3F (Carhart) performs better than our DSBM-III by achieving an adjusted R^2 of 72.10% (77.26%), which is about 32% (37%) higher than our DSBM-III model for FF3F (Carhart) model. In summary, results suggest that factor models outperform our DSBM-III model; however, the role of additional degrees in high explanatory power cannot be ignored (Campbell & Vuolteenaho 2004). More specifically, there are three (four) freely estimated parameters in the FF3F (Carhart).

2.6.3 Ability of DSBM to explain anomalies

Lewellen & Nagel (2006) criticise the cross-sectional tests and argue in favour of time-series tests for testing the suitability of CCAPM by directly evaluating the ability of model in explaining the anomalies unexplained by unconditional CAPM. The main difference between time-series and cross-sectional tests is that the former uses realised market returns to estimate the conditional pricing errors of a given asset i, $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\beta}_{i,t}.R_{m,t+1})$. On the other hand, the pricing errors for cross-sectional tests are the residuals of second pass regressions, $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}.\hat{\lambda}_{1,t+1})$.

Following Lewellen & Nagel (2006) and Cederburg & O'Doherty (2016), we also use time-series analysis to assess the performance of DMS approaches compared to unconditional CAPM and CCAPM benchmark models.[33] More specifically, in our time-series tests, we compare the unconditional and conditional performance of size, value, and momentum portfolios.[34] Our analysis is based on testing whether the average conditional alphas of portfolio, $\bar{\hat{\alpha}}^{IV}_{SMB}$, $\bar{\hat{\alpha}}^{IV}_{VMG}$, and $\bar{\hat{\alpha}}^{IV}_{WML}$, are equal to zero as implied by the CCAPM. We also test whether average conditional alphas produced by a given model are significantly higher than the corresponding unconditional alphas ($\hat{\alpha}^U_i$). This is done through testing the null hypothesis that $\bar{\hat{\alpha}}^C_i \leq \hat{\alpha}^U_i$, where i indicates SMB, VMG, and WML.

[33] Note that DSBM-III outperforms all DMS approaches in time-series tests. Therefore we provide a comparison between DSBM-III and benchmark models.

[34] The test portfolios are from Fama French database. 'Small' and 'Big' represent the simple average across the five low-market-cap portfolios and the five high-market-cap portfolios, respectively. 'SMB' is the difference between 'Small' and 'Big'. 'Growth' and 'Value' represent the simple average across the five low-B/M portfolios and the five high-B/M portfolios, respectively. 'VMG' is the difference between 'Value' and 'Growth'. 'Winners' ('Losers') represent the top (bottom) decile of Fama-French momentum sorted portfolios. 'WML' is the difference between 'Winners' and 'Losers'.

Table 2.4 reports the average pricing errors of SMB, VMG, and WML portfolios produced by our DSBM-III model, CCAPM benchmark models, and unconditional CAPM. Conditional alpha is estimated as $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\beta}_{i,t}.R_{m,t+1})$, where $\hat{\beta}_{i,t}$ represents the forecast of conditional beta made at time t, which results either from our DSBM-III model or benchmark CCAPM model. To obtain the estimates of unconditional alpha $(\alpha^{UC}_{i,OLS})$, we regress the excess portfolio return on excess market return.

[Insert Table 2.4 about here]

Results for the out-of-sample period August 1936 to December 2018 show that the size premium is insignificant at 5% for all the models. This is consistent with the findings of Lewellen & Nagel (2006) that both conditional and unconditional models can explain the size premium. However, the average pricing errors for 'VMG' and 'WML' are significant for all the models, including DSBM-III are highly significant. Interestingly, DSBM-III produces 18.22% lower pricing errors for 'VMG' portfolio than unconditional CAPM, which is significant at 5% significance level. However, there is no significant difference between unconditional and conditional alphas for 'WML' portfolio. The significant reduction in pricing errors for 'VMG' is some success compared to other CCAPM benchmark models as there is no significant difference between pricing errors produced by unconditional CAPM and CCAPM benchmark models.

Next, following Lewellen & Nagel (2006), Boguth et al. (2011), Cederburg & O'Doherty (2016), and others, we decompose the difference between the unconditional $(\hat{\alpha}^U)$ and the average conditional alpha $(\hat{\alpha}^C)$ into the market-timing (MT) and the volatility-timing (VT).

$$\hat{\alpha}^u_i - \overline{\hat{\alpha}}^C_i \approx (1 + \frac{\overline{R}^2_{m,t}}{\sigma^2_m})Cov(\hat{\beta}^C_{i,t}, R_{m,t}) - \frac{\overline{R}_{mt}}{\sigma^2_m}Cov(\hat{\beta}^C_{i,t}, R^2_{m,t}) \tag{2.56}$$

the market-timing (MT) is estimated as $(1+\frac{\overline{R}^2_{m,t}}{\sigma^2_m})Cov(\hat{\beta}^C_{i,t}, R_{m,t})$, where $\overline{R}_{m,t}$ and $\hat{\sigma}^2_m$ are average market risk premium and its unconditional variance respectively. $Cov(\hat{\beta}^C_{i,t}, R_{m,t})$ is covariance between conditional beta of asset i and realised market excess returns. The volatility-timing (VT) is estimated as $\frac{\overline{R}_{m,t}}{\sigma^2_m}Cov(\hat{\beta}^C_{i,t}, R^2_{m,t})$. From this decomposition it is clear that if CCAPM holds i.e. $\overline{\hat{\alpha}}^C_i = 0$, the unconditional alphas should be explained by MT and VT.

The results in Table 2.5 show that consistent with Lewellen & Nagel (2006), the covariance components are insufficient to explain the unconditional alpha, confirming the failure of time-varying IVs to explain the value and momentum anomalies.[35] The results show that the 18.22% reduction in alpha for value premium using our DSBM-III model is coming from market timing which is 0.0789.[36] Moreover, results suggest no evidence of volatility timing bias which is also consistent with Lewellen & Nagel (2006). However, unlike Lewellen & Nagel (2006), our findings suggest that there is positive covariance between conditional betas of value-stock and the market risk premium, while the betas of growth stocks hold the opposite.

[Insert Table 2.5 about here]

[35] Note that the alpha bias given in equation (2.56) would not exactly be equal to the difference of MT and VT because our conditional alphas and betas are estimated out-of-sample.

[36] Note that the difference between unconditional alpha and average pricing errors under DSBM-III for VMG is 0.099 which ≈ MT

2.6.4 Does Dynamic Model Selection (DMS) overcome VSU?

The time-series and cross-sectional results suggest that although the best performing model, DSBM-III, outperforms the standard CCAPM approaches in monthly returns but it fails to outperform factor models in cross-sectional tests. Most importantly, according to standard time-series tests (e.g., Lewellen & Nagel 2006), value and momentum anomalies remain unexplained by all the DMS approaches. Our main objective to introduce DMS approaches is to overcome the problem of *variable-selection uncertainty* (VSU). However, often, DMS approaches such as sequential selection and best-subset methods have been criticised for high instability (see, e.g., Miller 2002, Petropoulos et al. 2018). Here, "instability" refers to a model's sensitivity to minor changes in the data, modifying the chosen predictors (Gifi 1990). Although DMS suggests that we have to choose a different model (set of variables) that best represent the investor information set at a given time, we expect a selected model not to change completely in the next period. This is important because it is unrealistic that information set changes entirely from one period to another.

Considering this, we perform various tests to evaluate the VSU and model instability. The frequency-based measures to quantify the model stability are more common in literature (e.g., Sauerbrei et al. 2015, Petropoulos et al. 2018). These metrics look at the frequency of selection of subset variables, or a particular model. We first evaluate the model stability over time using the original sample based on equation (2.57) that measures the variability in selected variables over time in each out-of-sample period.

$$M_{STB} = \frac{\sum_{t=2}^{S} I_t}{S-1} \tag{2.57}$$

where M_{STB} is model stability measure, I_t is an indicator function taking a value of one if a model selected at time t has at least one variable from variables selected at time $t-1$ and zero otherwise, and S indicates the out-of-sample observations. A high value of M_{STB} for any model selection approach would indicate model stability. Results suggest that the best performing approach is $DSBM-III$ which achieves a score of 67.5% for M_{STB}. Since our indicator is very optimistic as it takes the value of one in case of one single variable from variables at $t-1$ is part of the model at time t, so we expect a value for M_{STB} closer to one for any given approach. However, a value of 67.5% suggests that there is 32.5% probability that the variables selected at time $t+1$ would be 100% different from the variables selected at time t. This suggests that all the dynamic model selection approaches exhibit model instability.

Next, we evaluate the model instability through bootstrap method, widely used in literature to quantify the model uncertainty and instability (e.g., De Bin et al. 2016, Petropoulos et al. 2018). The bootstrap simulates multiple realisations of training data assuming that it is a representation of the population. In other words, one can obtain B bootstrapped samples with repeated random 'sampling with replacement' of the original dataset of size T. Each bootstrap sample is, therefore, the same size as the original dataset, but includes some sample replicates, while others are replaced. After getting bootstrapped samples, a given DMS method is then implemented for each pseudo sample, resulting in selection of different models due to minor data changes. The important point to note here is that the inclusion and exclusion of predictors can provide useful information about the relevance of particular predictors in explaining the target variable. It is expected that the relevant variables will almost always be the part of the selected models across bootstrapped samples, whereas less related to the target variable would rarely be included in the model across samples. The proportion of times a particular predictor is part of the models is known as the "inclusion frequency," suggesting the importance of each variable and can take a value between 0 and 1, where zero indicates that the variable was never selected.

In contrast, a value of one suggests the variable was always selected in each pseudo sample. The variable with strong (weak) effect, in the ideal case, have inclusion frequencies equal to one (zero).

In our analysis, we evaluate the inclusion frequency of variables with weak and strong effects. Specifically, we consider whether strongest (weakest) variables are always included (excluded) in a subset of variables selected by a particular approach at each period. To implement this, we generate 100 bootstrapped samples with 5 years of training dataset. Next, we select the subset of variables in each sample with a given model selection approach and calculate the inclusion frequencies for each variable.

$$IF_j = \frac{\sum_{b=1}^{B} I_{jb}}{B} \tag{2.58}$$

where IF_j indicates the inclusion frequency of variable j, I_{jb} is an indicator function taking a value of one if variable j is included in the selected subset of variables under given approach in bootstrapped sample b, and zero otherwise.

To implement our approach, we are interested in extreme values of inclusion frequencies. We aim to compare this with selected variables with original data. We see whether a variable with highest (lowest) inclusion frequency is included (excluded) from selected variables based on original data. We use three measures to evaluate the model stability: i) false inclusion ($FalseI$) - the inclusion of variable with lowest IF, ii) false exclusion ($FalseE$) - exclusion of a variable with highest IF, and iii) false inclusion and false exclusion ($FalseIE$) - false inclusion of variable with the weakest effect and false exclusion of variable with the strongest effect, simultaneously. At each period for a given model selection approach, we would assign a value of one only if false inclusion or exclusions are made, and zero otherwise. We repeat this process for each out-of-sample period which provides a time-series of values for overall S out-of-sample observations, and next, we calculate the correct selection of variables over the entire period as:

$$MSIF_g = \frac{\sum_{t=1}^{S} I_t}{S} \tag{2.59}$$

where $MSIF_g$ is model stability based on inclusion frequency for one of the three measures ($FalseI$, $FalseE$, or $FalseIE$), I_t is an indicator function that takes a value of one if variable inclusion condition under each measure is satisfied, zero otherwise. A high value of $MSIF$ for any model selection approach would indicate model instability. Figure 2.9 shows the values for three false inclusion/exclusion for various approaches. The values can be interpreted as the ratio of false inclusion/exclusion periods to total out-of-sample periods. Figure 2.9 suggests that all DMS approaches exhibit model instability as values for all the three measures are high. However, the best performing model is DSBM-III which achieves a value of 34% for $MSIF_{FalseIE}$; this means that about 34% of overall out-of-sample periods, DSBM-III has both false inclusion and false exclusion. The worst model is a univariate variable selection which includes the variable based on 5% significance level, which achieves a value of 65% for $MSIF_{FalseIE}$. On average all models do false exclusion (inclusion) for about 60% (56%) of overall periods.

[Insert Figure 2.9 about here]

So far, our approach uses the input-based definition of stability which does not consider the outcome. This may be criticised because in our case, some variables such as DP and DY are highly correlated. There is a possibility that DP is part of the model selected using original data.

Still, DY has the highest inclusion frequency. Our earlier analysis would consider this as false exclusion when this may not be true, as the impact on the outcome (forecast) may be negligible. Considering this, next, we use a different definition of stability, which is based on the variability of the output generated by a model selection approach concerning data sampling (e.g., Nogueira et al. 2017). To implement this, following Sauerbrei et al. (2015), we use a cross-validation approach, dividing the sample into *training* for estimating model parameters and *validation* for assessing the predictive performance of individual models. Next, we estimate all possible combinations of variables $M = 2^{14} - 1 = 16383$ models indexed by j ($j = 1, \ 2, \ . \ . \ . \ . \ M$) and Let ($M_j^*$) be the optimal model selected with particular approach.

After generating 100 bootstrap samples based on training data, the model inclusion frequencies under each model selection approach are calculated. Let $\mathbb{F}(M_j)$ be the model selection frequency which is the proportion of bootstrap samples ($B = 100$) in which model (M_j) was selected. The main idea behind calculating the model inclusion frequency is to form a weighted average forecast where each model is weighted according to model inclusion frequency. Petropoulos et al. (2018) call this approach as Bootstrap model combination (BMC) which can be given as:

$$\hat{\beta}_{i,t}^{BMC} = \sum_{j=1}^{M} \omega_j \hat{\beta}_{i,j,t} \tag{2.60}$$

where $\omega_j = \mathbb{F}(M_j)$ and $\sum_{j=1}^{M} = 1$, and $\hat{\beta}_{i,j,t}$ represent the beta estimates with model j using original data.[37]

These estimates are used to calculate prediction errors in the cross-validation sample as:

$$\hat{\alpha}_{i,t+1}^{BMC} = R_{i,t+1} - \hat{\beta}_{i,t}^{BMC} R_{m,t+1} \tag{2.61}$$

Given that V and $\hat{\alpha}_i^{BMC}$ are CV observations and one-step ahead pricing errors based on BMC, the $MSFE^{CV}$ can be given as:

$$MSFE_{CV}^{BMC} = \frac{1}{V} \sum_{t=1}^{V} (\hat{\alpha}_t^{BMC})^2 \tag{2.62}$$

Similarly, MSFE in validation sample is computed for beta forecasts based on optimally selected forecasts given as $MSFE_{CV}^{*}$. Next, we compare these two MSFE to see whether the selected model (subset variables) performs better than BMC. The main idea behind making a comparison of performance between these two is that the BMC forecasts have considered the model uncertainty into account which makes BMC a perfect benchmark against a single selected model (see, e.g., Buchholz et al. 2008, Sauerbrei et al. 2015). To quantify the stability, we use a measure similar to the model stability based on inclusion frequency ($MSIF_g$) given in equation (2.59), but here indicator function would take a value of one if the best model performs poorly than BMC and zero otherwise. Figure (2.10) shows that all the dynamic model selection approaches produce large prediction errors compared to BMC in all the out-of-sample periods. Moreover, instead of selecting the best model under any given approach, we compare the $MSFE - CV$ of each of individual 16383 models with BMC and find that BMC always outperforms individual models. This is consistent with Petropoulos et al. (2018) that model selection approaches do not overcome the issue of variable-selection uncertainty (VSU).

[Insert Figure 2.10 about here]

[37] For more details on $\omega_j = \mathbb{F}(M_j)$, see Buchholz et al. (2008).

Finally, we evaluate the sensitivity of model selection across assets. The main motivation for such analysis is that the standard IV approach assumes that the predetermined predictors can price all the assets equally. But from our results, we find that DSBM-III, which allows choosing a different set of variables for each asset performs better than DSBM-I and DSBM-II.[38] To perform this analysis, we investigate the relative importance of individual predictors in DSBM-III betas of individual assets. Following Gu et al. (2020), we calculate the increase in out-of-sample MSFE by excluding a particular predictor from information set. Given K predictors, the importance factor can be given as:

$$\Delta MSFE_{i,j} = (MSFE_{i,K-1} - MSFE_{i,K}) \tag{2.63}$$

where $\Delta MSFE_{i,j}$ is an increase in MSFE of asset i due to absence of jth predictor, $MSFE_{i,K}$ is MSFE using all predictors and $MSFE_{i,K-1}$ is MSFE without jth predictor. Now the importance factor can be calculated as:

$$\Phi_{i,j} = \frac{\Delta MSFE_{i,j}}{\sum\limits_{j=1}^{K} \Delta MSFE_{i,j}} \tag{2.64}$$

where $\Phi_{i,j}$ is the variable importance of each predictor j for asset i, which is normalised to sum to one. Here, the importance factor indicates the weights of individual predictor for a given asset. We obtain the weights of individual assets across B bootstrapped sample and finally calculate the summary statistics (minimum, maximum, mean, and standard deviation) across assets. This would provide an idea about the sensitivity of variables across assets. For example, if a particular variable is equally important to all the assets, then the standard deviation should be zero. On the other hand, if the importance (weights) vary across assets, we can say that variables are not equally important for different assets. Figure 2.11 plots the average summary statistics. The range of each predictor shows that the importance of individual predictors varies across assets. For example, the variable LTY has a standard deviation of 4.7% with a wider range. The minimum value for a single asset is 0.25% and maximum value is 16.79%. This means that on average, the variable is not equally important for all the assets. This is an important finding of the CCAPM-IV approach which suggests that existing CCAPM-IV approach not only assumes that predetermined variables on average work for all periods but it also assumes that a single model on average works equally well for all the assets. This is also evident from our findings of significantly better performance of DSBM-III compared to DSBM-I and DSBM-II.

[Insert Figure 2.11 about here]

To summarise, our VSU and model instability analysis suggests that the DMS approaches do not fully account for VSU. The variable and model inclusion frequencies vary, implying that a small change in data will alter the selected variables. This is also consistent with the recent criticism on model selection approaches due to the assumption that there is a true model and we can identify it.[39] This may never be the case as shown by many studies that model selection approaches fail to account for model uncertainty and stability (see, e.g., Smith 2018, Petropoulos et al. 2018, Makridakis et al. 2020).

[38] DSBM-I and DSBM-II use the best identified model at a given time to price all the assets.

[39] The quest of selecting the best model is often viewed as the “holy grail” in forecasting (Makridakis et al. 2020).

2.7 Robustness Tests

In this section, we discuss the findings from various robustness tests that we perform to see whether our original results are affected by additional tests.

First, following Ferson & Harvey (1999), Christopherson et al. (1998), among others, we estimate the CCAPM-IV with time-varying alphas.

$$\alpha_{i,t} = a_{0,i} + a'_{1,i}X_t \tag{2.65}$$

and given β is also a linear function of conditioning variables:

$$\beta_{i,t} = \gamma_{0,i} + \gamma'_{1,i}X_t \tag{2.66}$$

By putting equations (2.65) and (2.66) in equation (2.7) we get following econometric model:

$$R_{i,t+1} = a_{0,i} + a'_{1,i}X_t + (\gamma_{0,i} + \gamma'_{1,i}X_t)R_{m,t+1} + \varepsilon_{i,t+1} \tag{2.67}$$

By using equation (2.67), we redo all the asset pricing tests. The findings show that the results are very similar to those of the main tests based on the assumption that alpha is constant. These findings indicate that enabling alphas to vary over time has a negligible effect on our primary results.

Next, we analyse the ability of all the models considered for benchmark tests in pricing additional test portfolios, namely the 10 momentum sorted portfolios, 25 portfolios formed on size and momentum, and 30 industry portfolios. Results in Table 2.6 suggest that with these additional assets, the findings are consistent with our main results.

[Insert Table 2.6 about here]

Next, we apply alternative asset pricing criteria to see whether these can change the results. In addition to adjusted R^2, we use Composite Pricing Errors (CPE) given in equation (2.21). Moreover, we also follow Boguth et al. (2011) and use a two-step IV approach that can offer more clear evidence of the relationship between the conditioning variables and the beta of the portfolio. The first step involves the estimation of monthly betas using daily returns, $\hat{\beta}^{sw}_{i,t}$, with the model given in equation (2.50). Next, we regress the estimated monthly betas on a set of IVs,

$$\hat{\beta}^{sw}_{i,t} = \phi_{i,j,0} + \phi'_{i,j,1}X_{j,t-1} + e_{i,j,t} \tag{2.68}$$

Next, we choose the model based on R^2 from the regression model in equation (2.68). A model with the highest adjusted R^2 is one that explains the most beta dynamics, the optimal conditioning information would be given as X^*.

$$\hat{\beta}^{sw}_{i,t} = \phi_{i,0} + \phi'_{i,1}X^*_{t-1} + e_{i,t} \tag{2.69}$$

Next, we get the conditional fitted betas using estimates of $\phi_{i,0}$ and $\phi_{i,1}$

$$\tilde{\beta}_{i,t} = \hat{\phi}_{i,0} + \hat{\phi}'_{i,1}X^*_{t-1}, \tag{2.70}$$

Next, by defining β_i as a function of fitted betas $\beta_{i,t} = \gamma_{0i} + \gamma_{1i}\tilde{\beta}_{i,t}$ and by putting value of β_i in (2.7) we get:

$$R_{i,t} = a_i^{IV2} + \gamma_{i,0}R_{m,t} + \gamma_{i,1}\tilde{\beta}_{i,t}R_{m,t} + \varepsilon_{i,t} \tag{2.71}$$

Next, we get the conditional beta using estimates of $\hat{\gamma}_{i,0}$ and $\hat{\gamma}_{i,1}$

$$\hat{\beta}_{i,t}^{IV2} = \hat{\gamma}_{0i} + \hat{\gamma}_{1i}\tilde{\beta}_{i,t} \tag{2.72}$$

Results reported in Table 2.7 suggests that these additional asset pricing criteria and beta estimation methods do not make any significant impact on the conclusions of our original results.

[Insert Table 2.7 and Table 2.8 about here]

Finally, we analyse the performance of different models considered in the alternative sample period. Following mainstream literature, we use a post-1963 period. The sub-sample results are consistent with the full-sample period as our DSBM-III outperforms the CCAPM benchmark models.

2.8 Conclusion

This study examines the performance of a conditional CAPM where we use various dynamic model selection (DMS) approaches to estimate conditional market betas. We contribute to the literature by introducing a new model selection approach that at each point in time, chooses the model that currently has the best asset pricing performance. Our main empirical results suggest that all the DMS approaches, in particular, shrinkage methods such as LASSO, adaptive LASSO, and ENet perform poorly in explaining cross-sectional variation in expected returns. All the models yield significant pricing errors based on joint alpha (JA) and composite pricing errors (CPE). However, the DSM approaches based on asset pricing criteria perform better than traditional approaches such as sequential selection, best subset selection and shrinkage methods by achieving higher R^2. One potential reason for the traditional methods' poor performance is their reliance on residuals from time-series regression of CCAPM-IV as their primary objective function. However, according to the CAPM theory, when the returns are measured in excess of risk-free rate, the intercept term a_i^{IV} indicates the expected abnormal return, which should be zero. Therefore, the DMS approach based on asset pricing tests ensures that a model that minimises the pricing errors is selected to capture beta dynamics.

However, consistent with Lewellen & Nagel (2006), we find the failure of DMS approaches in explaining the value and momentum anomalies. Using bootstrap methods to quantify the model uncertainty and instability, we find that the DMS approaches do not fully account for variable-selection uncertainty (VSU). These findings are in line with recent criticism of DMS approaches regarding their failure to fully account for variable-selection uncertainty and to achieve model stability (see, Smith 2018, Petropoulos et al. 2018, Makridakis et al. 2020, and others). Based on these findings, in the next chapter we consider alternative forecasting approaches such as combining information (e.g., Kelly & Pruitt 2013, Neely et al. 2014) and combining forecasts (e.g., Bates & Granger 1969, Timmermann 2006, Rapach et al. 2010) to address the issue of VSU.

References

Abel, A. B., Eberly, J. C. & Panageas, S. (2013), 'Optimal inattention to the stock market with information costs and transactions costs', *Econometrica* **81**(4), 1455–1481.

Adrian, T. & Franzoni, F. (2009), 'Learning about beta: Time-varying factor loadings, expected returns, and the conditional capm', *Journal of Empirical Finance* **16**(4), 537–556.

Adrian, T. & Rosenberg, J. (2008), 'Stock returns and volatility: Pricing the short-run and long-run components of market risk', *The Journal of Finance* **63**(6), 2997–3030.

Akaike, H. (1973), Information theory and an extension of maximum likelihood principle, *in* 'Proceedings of Second International Symposium on Information Theory', pp. 267–281.

Al-Obeidat, F., Spencer, B. & Alfandi, O. (2020), 'Consistently accurate forecasts of temperature within buildings from sensor data using ridge and lasso regression', *Future Generation Computer Systems* **110**, 382–392.

Amihud, Y. (2002), 'Illiquidity and stock returns: cross-section and time-series effects', *Journal of Financial Markets* **5**(1), 31–56.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Avramov, D. & Chordia, T. (2006), 'Asset pricing models and financial market anomalies', *The Review of Financial Studies* **19**(3), 1001–1040.

Bai, J. & Ng, S. (2008), 'Forecasting economic time series using targeted predictors', *Journal of Econometrics* **146**(2), 304–317.

Bali, T. G. & Engle, R. F. (2010), 'Resurrecting the conditional capm with dynamic conditional correlations', *Available at SSRN 1298633* .

Banz, R. W. (1981), 'The relationship between return and market value of common stocks', *Journal of Financial Economics* **9**(1), 3–18.

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Bauer, R., Cosemans, M. & Schotman, P. C. (2010), 'Conditional asset pricing and stock market anomalies in europe', *European Financial Management* **16**(2), 165–190.

Baybuza, I. (2018), 'Inflation forecasting using machine learning methods', *Russian Journal of Money and Finance* **77**(4), 42–59.

Black, F. (1972), 'Capital market equilibrium with restricted borrowing', *The Journal of Business* **45**(3), 444–455.

Boguth, O., Carlson, M., Fisher, A. & Simutin, M. (2011), 'Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas', *Journal of Financial Economics* **102**(2), 363–389.

Boivin, J. & Ng, S. (2006), 'Are more data always better for factor analysis?', *Journal of Econometrics* **132**(1), 169–194.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A capital asset pricing model with time-varying covariances', *Journal of Political Economy* **96**(1), 116–131.

Bowerman, B. L. & O'Connell, R. T. (1993), *Forecasting and time series: An applied approach (3rd edition)*, Duxbury Press.

Buchholz, A., Holländer, N. & Sauerbrei, W. (2008), 'On properties of predictors derived with a two-step bootstrap model averaging approach—a simulation study in the linear regression model', *Computational Statistics and Data Analysis* **52**(5), 2778–2793.

Buncic, D. & Tischhauser, M. (2017), 'Macroeconomic factors and equity premium predictability', *International Review of Economics and Finance* **51**, 621–644.

Burnham, K. P. & Anderson, D. R. (2002), *Model selection and multimodel inference: A practical information-theoretic approach*, Vol. 2, Springer.

Cai, Z., Ren, Y. & Yang, B. (2015), 'A semiparametric conditional capital asset pricing model', *Journal of Banking and Finance* **61**, 117–126.

Campbell, J. Y. & Vuolteenaho, T. (2004), 'Bad beta, good beta', *American Economic Review* **94**(5), 1249–1275.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of Finance* **52**(1), 57–82.

Cederburg, S. & O'Doherty, M. S. (2016), 'Does it pay to bet against beta? on the conditional performance of the beta anomaly', *The Journal of Finance* **71**(2), 737–774.

Chava, S., Gallmeyer, M. & Park, H. (2015), 'Credit conditions and stock return predictability', *Journal of Monetary Economics* **74**, 117–132.

Christopherson, J. A., Ferson, W. E. & Glassman, D. A. (1998), 'Conditioning manager alphas on economic information: Another look at the persistence of performance', *The Review of Financial Studies* **11**(1), 111–142.

Cochrane, J. (2005), Financial markets and the real economy, Report, National Bureau of Economic Research.

Cochrane, J. H. (1999), 'New facts in finance', *Economic Perspectives Federal Reserve Bank of Chicago* **23**(3), 36–58.

Cochrane, J. H. (2009), *Asset pricing: Revised edition*, Princeton University Press.

Cooper, M. J. & Gubellini, S. (2011), 'The critical role of conditioning information in determining if value is really riskier than growth', *Journal of Empirical Finance* **18**(2), 289–305.

Coqueret, G. (2020), 'Stock-specific sentiment and return predictability', *Quantitative Finance* **20**(9), 1531–1551.

Daoud, J. I. (2017), 'Multicollinearity and regression analysis', *Journal of Physics* **949**(1), 012009.

De Bin, R., Janitza, S., Sauerbrei, W. & Boulesteix, A. (2016), 'Subsampling versus bootstrapping in resampling-based model selection for multivariable regression', *Biometrics* **72**(1), 272–280.

Draper, N. & Smith, H. (1966), *Applied regression analysis*, Vol. 3, John Wiley and Sons.

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004), 'Least angle regression', *The Annals of Statistics* **32**(2), 407–499.

Efroymson, M. (1960), *Multiple regression analysis in Mathematical Methods for Digital Computers*, Vol. 1, Wiley Publishers.

Elliott, G., Gargano, A. & Timmermann, A. (2013), 'Complete subset regressions', *Journal of Econometrics* **177**(2), 357–373.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business and Economic Statistics* **20**(3), 339–350.

Fama, E. F. & French, K. R. (1989), 'Business conditions and expected returns on stocks and bonds', *Journal of Financial Economics* **25**(1), 23–49.

Fama, E. F. & French, K. R. (1992), 'The cross-section of expected stock returns', *The Journal of Finance* **47**(2), 427–465.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. & French, K. R. (2015), 'A five-factor asset pricing model', *Journal of Financial Economics* **116**(1), 1–22.

Fama, E. F. & French, K. R. (2018), 'Choosing factors', *Journal of Financial Economics* **128**(2), 234–252.

Fama, E. F. & MacBeth, J. D. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**(3), 607–636.

Ferson, W. E. & Harvey, C. R. (1999), 'Conditioning variables and the cross section of stock returns', *The Journal of Finance* **54**(4), 1325–1360.

Ferson, W. E., Sarkissian, S. & Simin, T. T. (2003), 'Spurious regressions in financial economics?', *The Journal of Finance* **58**(4), 1393–1413.

Gabaix, X. (2019), *Behavioral inattention in Handbook of Behavioral Economics*, Vol. 2, Elsevier.

Ghysels, E. (1998), 'On stable factor structures in the pricing of risk: do time-varying betas help or hurt?', *The Journal of Finance* **53**(2), 549–573.

Gifi, A. (1990), *Nonlinear multivariate analysis*, Wiley-Blackwell.

González-Briones, A., Hernández, G., Pinto, T., Vale, Z. & Corchado, J. M. (2019), 'A review of the main machine learning methods for predicting residential energy consumption'.

Goyal, A. & Welch, I. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *The Review of Financial Studies* **21**(4), 1455–1508.

Gu, S., Kelly, B. & Xiu, D. (2020), 'Empirical asset pricing via machine learning', *The Review of Financial Studies* **33**(5), 2223–2273.

Harvey, C. R. (2001), 'The specification of conditional expectations', *Journal of Empirical Finance* **8**(5), 573–637.

Harvey, C. R., Liu, Y. & Zhu, H. (2016), '... and the cross-section of expected returns', *The Review of Financial Studies* **29**(1), 5–68.

Hastie, T., Tibshirani, R. & Friedman, J. (2017), *The elements of statistical learning: data mining, inference, and prediction*, Springer Science and Business Media.

Hocking, R. R. (2013), *Methods and applications of linear models: regression and the analysis of variance*, John Wiley and Sons.

Hocking, R. R. & Leslie, R. (1967), 'Selection of the best subset in regression analysis', *Technometrics* **9**(4), 531–540.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Hou, K., Mo, H., Xue, C. & Zhang, L. (2019), 'Which factors?', *Review of Finance* **23**(1), 1–35.

Hou, K., Xue, C. & Zhang, L. (2015), 'Digesting anomalies: An investment approach', *The Review of Financial Studies* **28**(3), 650–705.

Huang, D., Jiang, F., Tu, J. & Zhou, G. (2015), 'Investor sentiment aligned: A powerful predictor of stock returns', *The Review of Financial Studies* **28**(3), 791–837.

Hurvich, C. M. & Tsai, C.-L. (1989), 'Regression and time series model selection in small samples', *Biometrika* **76**(2), 297–307.

Jagannathan, R. & Wang, Z. (1996), 'The conditional capm and the cross-section of expected returns', *The Journal of Finance* **51**(1), 3–53.

Jayakumar, D. S. (2014), 'Identification of multicollinearity and it's effect in model selection', *Electronic Journal of Applied Statistical Analysis* **7**(1), 153–179.

Jiang, F., Lee, J., Martin, X. & Zhou, G. (2019), 'Manager sentiment and stock returns', *Journal of Financial Economics* **132**(1), 126–149.

Jostova, G. & Philipov, A. (2005), 'Bayesian analysis of stochastic betas', *Journal of Financial and Quantitative Analysis* **40**(4), 747–778.

Kadane, J. B. & Lazar, N. A. (2004), 'Methods and criteria for model selection', *Journal of the American Statistical Association* **99**(465), 279–290.

Kelly, B. & Pruitt, S. (2013), 'Market expectations in the cross-section of present values', *The Journal of Finance* **68**(5), 1721–1756.

Kuhn, M. & Johnson, K. (2013), *Applied predictive modeling*, Vol. 26, Springer.

Lettau, M. & Ludvigson, S. (2001), 'Consumption, aggregate wealth, and expected stock returns', *The Journal of Finance* **56**(3), 815–849.

Lewellen, J. & Nagel, S. (2006), 'The conditional capm does not explain asset-pricing anomalies', *Journal of Financial Economics* **82**(2), 289–314.

Lin, F.-J. (2008), 'Solving multicollinearity in the process of fitting regression model using the nested estimate procedure', *Quality and Quantity* **42**(3), 417–426.

Lin, Q. (2018), 'Technical analysis and stock return predictability: An aligned approach', *Journal of Financial Markets* **38**, 103–123.

Lintner, J. (1965), 'Security prices, risk, and maximal gains from diversification', *The Journal of Finance* **20**(4), 587–615.

Lourenço, N., Gouveia, C. M. & Rua, A. (2021), 'Forecasting tourism with targeted predictors in a data-rich environment', *Economic Modelling* **96**.

Luo, Y. & Young, E. R. (2016), 'Long-run consumption risk and asset allocation under recursive utility and rational inattention', *Journal of Money, Credit and Banking* **48**(2-3), 325–362.

Makridakis, S., Hyndman, R. J. & Petropoulos, F. (2020), 'Forecasting in social settings: The state of the art', *International Journal of Forecasting* **36**(1), 15–28.

Mallows, C. (1973), 'Some comments on cp', *Technometrics* **15**(1), 661–675.

McClendon, M. J. (2002), *Multiple regression and causal analysis*, Waveland Press.

Miller, A. (2002), *Subset selection in regression*, CRC Press.

Morozova, O., Levina, O., Uusküla, A. & Heimer, R. (2015), 'Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in russia', *BMC Medical Research Methodology* **15**(1), 1–17.

Neely, C. J., Rapach, D. E., Tu, J. & Zhou, G. (2014), 'Forecasting the equity risk premium: the role of technical indicators', *Management Science* **60**(7), 1772–1791.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Ni, Z.-X., Wang, D.-Z. & Xue, W.-J. (2015), 'Investor sentiment and its nonlinear effect on stock returns—new evidence from the chinese stock market based on panel quantile regression model', *Economic Modelling* **50**, 266–274.

Nogueira, S., Sechidis, K. & Brown, G. (2017), 'On the stability of feature selection algorithms', *The Journal of Machine Learning Research* **18**(1), 6345–6398.

O'brien, R. M. (2007), 'A caution regarding rules of thumb for variance inflation factors', *Quality and Quantity* **41**(5), 673–690.

Paye, B. S. & Timmermann, A. (2006), 'Instability of return prediction models', *Journal of Empirical Finance* **13**(3), 274–315.

Pereira, J. M., Basto, M. & da Silva, A. F. (2016), 'The logistic lasso and ridge regression in predicting corporate failure', *Procedia Economics and Finance* **39**, 634–641.

Petkova, R. & Zhang, L. (2005), 'Is value riskier than growth?', *Journal of Financial Economics* **78**(1), 187–202.

Petropoulos, F., Hyndman, R. J. & Bergmeir, C. (2018), 'Exploring the sources of uncertainty: Why does bagging for time series forecasting work?', *European Journal of Operational Research* **268**(2), 545–554.

Raffalovich, L. E., Deane, G. D., Armstrong, D. & Tsao, H.-S. (2008), 'Model selection procedures in social research: Monte-carlo simulation results', *Journal of Applied Statistics* **35**(10), 1093–1114.

Rapach, D. E., Ringgenberg, M. C. & Zhou, G. (2016), 'Short interest and aggregate stock returns', *Journal of Financial Economics* **121**(1), 46–65.

Rapach, D. E., Strauss, J. K. & 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.

Rapach, D. E. & Wohar, M. E. (2006), 'Structural breaks and predictive regression models of aggregate us stock returns', *Journal of Financial Econometrics* **4**(2), 238–274.

Salisu, A. A., Isah, K. & Akanni, L. O. (2019), 'Improving the predictability of stock returns with bitcoin prices', *The North American Journal of Economics and Finance* **48**, 857–867.

Sauerbrei, W., Buchholz, A., Boulesteix, A. & Binder, H. (2015), 'On stability issues in deriving multivariable regression models', *Biometrical Journal* **57**(4), 531–555.

Sauerbrei, W. & Schumacher, M. (1992), 'A bootstrap resampling procedure for model building: application to the cox regression model', *Statistics in Medicine* **11**(16), 2093–2109.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

Shanken, J. (1990), 'Intertemporal asset pricing: An empirical investigation', *Journal of Econometrics* **45**(1-2), 99–120.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **19**(3), 425–442.

Simin, T. (2008), 'The poor predictive performance of asset pricing models', *Journal of Financial and Quantitative Analysis* **43**(2), 355–380.

Sims, C. A. (2003), 'Implications of rational inattention', *Journal of Monetary Economics* **50**(3), 665–690.

Smith, G. (2018), 'Step away from stepwise', *Journal of Big Data* **5**(1), 1–12.

Stambaugh, R. F. (1999), 'Predictive regressions', *Journal of Financial Economics* **54**(3), 375–421.

Stambaugh, R. F. & Yuan, Y. (2017), 'Mispricing factors', *The Review of Financial Studies* **30**(4), 1270–1315.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58**(1), 267–288.

Timmermann, A. (2006), 'Forecast combinations', *Handbook of Economic Forecasting* **1**, 135–196.

Vendrame, V., Guermat, C. & Tucker, J. (2018), 'A conditional regime switching capm', *International Review of Financial Analysis* **56**, 1–11.

Wang, Z., Zhu, Z. & Yu, C. (2020), 'Variable selection in macroeconomic forecasting with many predictors', *Available at arXiv e-prints* .

Xiao, J., Ding, R., Xu, X., Guan, H., Feng, X., Sun, T., Zhu, S. & Ye, Z. (2019), 'Comparison and development of machine learning tools in the prediction of chronic kidney disease progression', *Journal of Translational Medicine* **17**(1), 1–13.

Xu, L. & Zhang, W.-J. (2001), 'Comparison of different methods for variable selection', *Analytica Chimica Acta* **446**(1-2), 475–481.

Zhang, H., He, Q., Jacobsen, B. & Jiang, F. (2020), 'Forecasting stock returns with model uncertainty and parameter instability', *Journal of Applied Econometrics* **35**(5), 629–644.

Zhang, Y., Chu, G. & Shen, D. (2021), 'The role of investor attention in predicting stock prices: The long short-term memory networks perspective', *Finance Research Letters* **38**, 101484.

Zhang, Y., Ma, F. & Wang, Y. (2019), 'Forecasting crude oil prices with a large set of predictors: Can lasso select powerful predictors?', *Journal of Empirical Finance* **54**, 97–117.

Zhang, Z. (2016), 'Variable selection with stepwise and best subset approaches', *Annals of Translational Medicine* **4**(7), 136–142.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society* **67**(2), 301–320.

Table 2.1: List of Predictor Variables

Variable	Symbol	Definition
i) Stock characteristics		
Dividend-price ratio	DP	The difference between log of dividends (12-month moving sum of dividends paid on the S&P 500 index) and the log of stock prices (S&P 500 index price)
Dividend-yield	DY	The difference between log of a 12-month moving sum of dividends and the log of lagged prices (S&P 500 index)
Earning-price ratio	EP	The difference between log of earnings (12-month moving sum of earnings on the S&P 500 index) and the log of stock prices (S&P 500 index)
Book-to-market ratio	BM	The book value to market value ratio for the Dow Jones Industrial Average
Dividend-payout ratio	DE	The difference between log of dividends (12-month moving sum of dividends paid on the S&P 500 index) and the log of earnings (12-month moving sum of earnings on the S&P 500 index)
Stock variance	SVAR	Sum of squared daily stock market returns on the S&P 500
	RVOL	Neely et al. (2014) show that SVAR produces outliers and they use equity risk premium volatility, (RVOL) as an alternative measure for stock variance (SVAR) which is based on the moving standard deviation estimator, $\hat{\sigma}_t = \frac{1}{12}\sum_{i=1}^{12} \lvert r_{t+1-1} \rvert$ and subsequently converted to $\widehat{Vol}_t \equiv \sqrt{\frac{\pi}{2}}\sqrt{12}\hat{\sigma}_t$
Net equity expansion	NTIS	Ratio of 12-month moving sums of net equity issues by NYSE-listed stocks to the year-end NYSE market capitalization
ii) Interest rate related		
Treasury bill rate	TBL	Three-month Treasury bill rate
Long term return	LTR	Long-term government bond return
Long term yield	LTY	Yield on long-term government bond
Term-spread	TMS	The difference between the long-term government bond yield (LTY) and the Treasury bill rate (TBL)
Default yield spread	DFY	The difference between yields on BAA- and AAA-rated corporate bonds
Default return spread	DFR	The difference between return on long-term corporate bond and return on long-term government bond
iii) Macroeconomic		
Inflation	INFL	One-month lag of growth in the consumer price index Following Neely et al. (2014), we also use two lags for inflation for robustness to account for the delay in CPI releases

Table 2.2: Out-of-sample Cross-sectional Results of DMS approaches

This table reports the results from the second step of the Fama & MacBeth (1973) methodology for *dynamic model selection* (DMS) approaches. Panel A reports the results of best subset selection approaches where under asset pricing approach (A.1) we report results of three models, we call them DSBM-I, DSBM-II, and DSBM-III (See section (2.4.1) for details). Traditional methods given in A.2 include AIC, AICc, BIC, adjusted R^2, and Mallow's C_p (see section (2.4.2.1) for details). Under stepwise methods given in B, rather than estimating all possible models we use forward (FWD) selection, backward elimination (BE), stepwise regression (SReg), and univariate selection (US) (see section (2.4.2.2) for details). Panel C reports the results of shrinkage methods which include LASSO, Adaptive LASSO, Ridge, and Elastic Net (2.4.2.3) for details). Results consists of price of risk and performance evaluation. Under price of risk, $\overline{\lambda}_0$ is average return on zero beta portfolio, $\overline{\lambda}_1$ is estimated risk premium, to test the null hypothesis that the price of risk is equal to zero, the Newey & West (1987) t-statistics is reported below the coefficient estimates. [*], [**] and [***] asterisks denote the significance of coefficients at a 10%, 5% and 1% level, respectively. The performance evaluation criteria include SSPE, RMSFE, Adj. R^2, mispriced asset (MA), JA, and CPE are sum of square pricing errors, root mean square pricing error, Jagannathan & Wang (1996) adjusted R^2, joint alpha test, and composite pricing error. Under MA, at 1% and MA 5% indicate the number of portfolios for which pricing errors are significant at 1% and 5% levels respectively. JA and CPE statistics are used for testing the null hypothesis that pricing errors are jointly zero. [$**$] indicates that the null hypothesis is rejected at 5% critical value based on the normal distribution for JA, and for CPE, we use bootstrap distribution. Monthly observations are used to estimate all the models. The test assets are the 25 assets sorted by size and book-to-market ratio. The out-of-sample period is August 1936 to December 2018.

	Price of Risk		**Performance Evaluation Criteria**						
						Mispriced Assets		**Joint Test**	
	$\overline{\lambda}_0$	$\overline{\lambda}_1$	**SSPE**	**RMSPE**	**Adj. R^2**	**at 5%**	**at 1%**	**JA**	**CPE**
A. Best Subset Selection									
A.1 Asset Pricing									
A.1.1 DSBM-I	0.8423***	0.2335	0.6032	0.15533	0.29131	11	8	71.2271**	0.0202**
	3.1251	0.8781							
A.1.2 DSBM-II	0.7582***	0.3235	0.5812	0.1525	0.3512	10	7	69.3827**	0.0197**
	3.0321	1.0528							
A.1.3 DSBM-III	0.7125***	0.3919	0.5694	0.1509	0.3984	9	7	67.1012**	0.0185**
	2.9415	1.1381							
A.2 Traditional									
A.2.1 AIC	0.9209***	0.1028	0.7113	0.1687	0.2006	12	9	81.2967**	0.0209**
	3.2000	0.8273							
A.2.2 AICc	0.8997***	0.1491	0.6682	0.1634	0.2470	11	8	74.4219**	0.0277**
	3.1874	0.8770							
A.2.3 BIC	0.9019***	0.1266	0.6895	0.1661	0.2246	11	9	76.7924**	0.0287**
	3.1975	0.8653							
A.2.4 Mallow's C_p	0.9376***	0.0957	0.9982	0.1998	0.1813	12	9	95.0373**	0.0204**
	3.2031	0.8182							
A.2.5 Adjusted R^2	0.9426***	0.0836	0.9984	0.1998	0.1811	12	9	95.1949**	0.0215**
	3.2283	0.8037							
B. Stepwise Selection									
B.1 Traditional									
B.1.1 FWD Selection (FSS)	0.9803***	0.0469	1.0209	0.2021	0.1626	12	9	95.8931**	0.0221**
	3.2769	0.4952							
B.1.2 BKWD Elimination (BE)	0.9609***	0.0757	1.0154	0.2015	0.1671	12	9	80.2408**	0.0216**
	3.2661	0.6420							
B.1.3 Stepwise Regression (SReg)	0.9547***	0.0795	1.0023	0.2002	0.1778	12	9	86.1940**	0.0210**
	3.2544	0.7447							
B.1.4 Univariate Selection (US)	1.0117***	-0.0791	1.0245	0.2024	0.1597	12	9	79.0407**	0.0220**
	3.2811	-0.4822							
B.2 Asset Pricing									
B.2.1 FWD Selection (FSS-APC)	0.7795***	0.2838	0.5916	0.1538	0.3347	10	7	69.1499**	0.0244**
	3.0517	1.0286							
B.2.2 Univariate Selection (US-APC)	0.8196***	0.2578	0.6138	0.1567	0.3097	11	7	73.2844**	0.0243**
	3.0993	0.9617							
C. Shrinkage Methods									
C.1 LASSO	1.0930***	-0.1394	1.0734	0.2072	0.1196	12	9	65.6347**	0.0225**
	3.4203	-0.6987							
C.2 Adaptive LASSO	1.0910***	-0.1302	1.0541	0.2053	0.1353	12	9	71.9532**	0.0227**
	3.3315	-0.6287							
C.3 Ridge	1.1171***	-0.1602	1.0798	0.2078	0.1143	12	9	71.9878**	0.0229**
	3.4368	-0.8035							
C.4 Elastic Net	1.0594***	-0.1106	1.0525	0.2052	0.1367	12	9	80.2285**	0.0216**
	3.3234	-0.5810							

Table 2.3: Out-of-sample Cross-sectional Results of DSBM and Benchmark Models

This table reports the results from the second step of the Fama & MacBeth (1973) methodology for our DSBM-III and benchmark asset pricing models. DSBM-III, is based on MSFE over CV sample using conditional pricing errors, $\hat{\alpha}_{i,t+1}$. Panel A reports the results of various CCAPM benchmark models where models differ on the basis of capturing time-variation in betas. Model CAPM (β_{RW}), and CAPM (β_{SW}) report the results of CAPM where the time variation in beta is captured through 60 monthly rolling window and short window approach of Lewellen & Nagel (2006), respectively. CAPM (β_{FH}) reports the results of CAPM with beta defined as predetermined set of four conditioning variables used by Ferson & Harvey (1999). These variables include term spread, default yield, t-bill rate and spread between 3 months and one-month t-bill rate. CAPM (β_{KS}) reports the results of CAPM with beta defined as function of all the 12 predictors of Goyal & Welch (2008). Panel B reports the results of factor models which include Fama & French (1993) three-factor and Carhart (1997) four-factor models. See Table (2.2) for details on tests.

	Prices of Risk		Cross-sectional Explanation: Performance Evaluation Measures						
						Mispriced Assets		Joint Test	
	$\bar{\lambda}_0$	$\bar{\lambda}_1$	SSPE	RMSFE	Adj. R^2	at 5%	at 1%	JA	CPE
DSBM-III	0.7124*** 2.9415	0.3919 1.1381	0.5694	0.1509	0.3984	9	7	67.1012**	0.0185**
Panel A - CCAPM Benchmark Models									
A.1 CAPM (RW)	0.9022*** 3.4158	-0.2494 -0.6508	0.5904	0.1537	0.3365	11	8	72.4341**	0.0209**
A.2 CAPM(SW)	0.9499*** 4.0056	-0.2486 -1.2921	0.7203	0.1697	0.1828	12	7	80.9653**	0.0263**
A.3 CAPM(FH)	0.8404*** 5.4505	-0.0586 -0.4341	0.6112	0.1564	0.3130	11	7	69.4144**	0.0236**
A.4 CAPM(KS)	0.9250*** 5.6163	-0.0599 -1.1625	0.7149	0.1691	0.1965	12	10	81.2772**	0.0222**

Panel B - Factor Models

	Prices of Risk					Performance Evaluation Measures						
									Mispriced Assets		Joint Test	
	$\bar{\lambda}_0$	$\bar{\lambda}_{MKT}$	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	SSPE	RMSFE	Adj. R^2	at 5%	at 1%	JA	CPE
B.1 FF3F	0.9917*** 5.8432	−0.3113* -1.6833	0.1070 1.1602	0.3338*** 3.5625		0.2482	0.0996	0.7210	7	3	65.3967**	0.0063**
B. Carhart	0.9509*** 5.3710	-0.2785 -1.4590	0.1288 1.4063	0.3281*** 3.5430	0.1487 0.7889	0.2023	0.0900	0.7726	6	4	71.1373**	0.0052**

Table 2.4: Explaining the Size, Value and Momentum Anomalies

This table reports the pricing errors and R^2 for size, value and momentum anomalies produced by unconditional CAPM and various CCAPM including DSBM for sample period August 1936 to December 2018. 'Small' and 'Big' represent the simple average across the five low-market-cap portfolios and the five high-market-cap portfolios, respectively. 'SMB' is the difference between 'Small' and 'Big'. 'Growth' and 'Value' represent the simple average across the five low-B/M portfolios and the five high-B/M portfolios, respectively. 'VMG' is the difference between 'Value' and 'Growth'. 'Winners' ('Losers') represent the top (bottom) decile of Fama-French momentum sorted portfolios. 'WML' represents the difference between 'Winners' and 'Losers'. DSBM-III selects the best model for each asset based on CV-MSFE. See section (2.5.2) for details on the benchmark models. The pricing error ($\hat{\alpha}_U$) and R^2_U for unconditional model are obtained from OLS regressions of portfolio excess return on the market excess return. $\overline{\hat{\alpha}}_i$ indicate the average conditional pricing error for given asset i where monthly conditional alpha is estimated as $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\beta}_{i,t}.R_{m,t+1})$, where $\hat{\beta}_{i,t}$ is the forecast of beta made at time t, which results from the DSBM-III. The significance of the pricing errors is given with 't-statistics'. R^2_C indicates the R^2 for conditional model, which is given as $R^2_C = Var[ER_{it+1}]/Var[R_{it+1}]$, where ER_{it+1} is explained returns and can be obtained as $\hat{\beta}_{it}R_{m,t+1}$. $\Delta\hat{\alpha}_i$ and ΔR^2_i indicate the change in pricing errors ($[\overline{\hat{\alpha}}^C_i - \hat{\alpha}^U_i]/\hat{\alpha}^U_i$) , and R^2 ($[R^2_{C,i} - R^2_{U,i}]/R^2_{U,i}$) relative to unconditional CAPM, respectively. A negative (positive) value for $\Delta\hat{\alpha}_i$ (ΔR^2_i) indicate the improvement compared to unconditional model. $\Delta\hat{\alpha}_i$ (ΔR^2_i) in bold indicates that pricing errors (R^2) are significantly lower (higher) than unconditional CAPM at 5% level based on bootstrapped p-values.

Model	**Performance**	**Size**			**Value**			**Momentum**		
	Measure	Small	Big	SMB	Value	Growth	VMG	Winner	Loser	WML
Unconditional CAPM	$\hat{\alpha}^U_i$	0.0485	0.0143	0.0342	0.2871***	−0.2062***	0.4934***	0.4575***	−0.8813***	1.3389***
	tstat	0.3722	1.6951	0.1138	2.7803	−2.6837	4.5522	4.4753	−5.9598	6.3297
	R^2_i	0.6783	0.9335	0.1038	0.7485	0.8526	0.0037	0.7306	0.6713	0.0397
DSBM-III	$\overline{\hat{\alpha}}_i$	0.0570	0.0238	0.0332	0.2241**	−0.1793**	0.4035***	0.4437***	−0.8460***	1.2897***
	tstst	0.4624	0.6611	0.1121	2.4173	−2.3374	3.8204	4.4309	−5.8114	5.9478
	$\Delta\hat{\alpha}_i$			-0.0297			**-0.1822**			-0.0367
	R^2_i	0.7391	0.9595	0.1541	0.8034	0.8671	0.1914	0.7803	0.7084	0.1432
	ΔR^2_i			0.4857			**50.8579**			**2.6093**
CAPM (β^{RW})	$\overline{\hat{\alpha}}_i$	0.1207	0.0447	0.0760	0.2746***	−0.1784**	0.4531***	0.4590***	−0.8534***	1.3125***
	tstst	0.9632	1.2270	0.5630	3.0709	−2.2716	4.3649	4.4374	−5.8887	6.2431
	$\Delta\hat{\alpha}_i$			1.2207			-0.0816			-0.0197
	R^2_i	0.6964	0.9373	0.1320	0.7784	0.8554	0.1076	0.7184	0.6777	0.0394
	ΔR^2_i			0.2720			28.1417			-0.0061
CAPM (β^{SW})	$\overline{\hat{\alpha}}_i$	0.0534	0.0719	-0.0184	0.2716***	−0.1866***	0.4582***	0.4630***	−0.8507***	1.3137***
	tstst	0.4512	0.0499	1.5566	3.2340	−2.8218	3.7954	3.8143	−4.7972	5.4418
	$\Delta\hat{\alpha}_i$			-0.4627			-0.0713			-0.0188
	R^2_i	0.6201	0.9335	0.0397	0.7765	0.8324	0.0400	0.7616	0.7089	0.0344
	ΔR^2_i			-0.6172			9.8470			-0.1325
CAPM (β^{FH})	$\overline{\hat{\alpha}}_i$	0.0914	0.0655	0.0259	0.2693***	−0.1784**	0.4477***	0.4405***	−0.8576***	1.2981***
	tstst	0.7194	0.7876	0.1871	3.4919	−2.2901	4.1396	3.5230	−5.6981	5.6907
	$\Delta\hat{\alpha}_i$			-0.2442			-0.0927			-0.0304
	R^2_i	0.6880	0.9366	0.0897	0.7781	0.8453	0.0636	0.7104	0.6810	0.0598
	ΔR^2_i			-0.1358			16.2396			0.5075
CAPM (β^{KS})	$\overline{\hat{\alpha}}_i$	0.0205	0.0780	-0.0575	0.3012***	−0.1927***	0.4940***	0.4476***	−0.8983***	1.3459***
	tstst	0.1213	1.5932	-0.3107	3.0944	−2.9955	3.4414	2.2848	−3.8309	3.6414
	$\Delta\hat{\alpha}_i$			2.6461			0.0619			0.0053
	R^2_i	0.4479	0.8869	0.0228	0.6339	0.7370	0.0055	0.4008	0.2116	0.0132
	ΔR^2_i			-0.7805			0.4999			-0.6668

Table 2.5: Alpha Bias Decomposition

This table reports the results from the decomposition of the unconditional alphas for size, value and momentum anomalies. Market timing and volatility timing effects are two components of the unconditional alpha bias, $\alpha^{UC}_{i,OLS} - \overline{\alpha}^C_i$. The market-timing bias is estimated as $(1 + \frac{\overline{R}^2_{mt}}{\hat{\sigma}^2_m})Cov(\hat{\beta}^C_{i,t}, R_{m,t})$, where $\overline{R}_{mt}$ and $\hat{\sigma}^2_m$ are average market risk premium and its unconditional variance respectively. $Cov(\hat{\beta}^C_{i,t}, R_{m,t})$ is covariance between conditional beta and realised market risk premium. The volatility-timing bias is estimated as $\frac{\overline{R}_{mt}}{\sigma^2_{m,t}}Cov(\hat{\beta}^C_{i,t}, R^2_{m,t})$. Unconditional alpha is obtained by regressing the excess return of the portfolio on the market excess return. Conditional alpha is estimated as $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\beta}_{i,t}.R_{m,t+1})$, where $\hat{\beta}_{i,t}$ is the beta forecast at time t, resulting either from the DSBM-III or conditional benchmark model. Note that the beta estimates for our all models are out-of-sample betas, therefore the alpha bias decomposition, $\alpha^{UC}_{i,OLS} - \overline{\alpha}^C_i$, can only be approximated to the difference between market timing and volatility timing. For more details see Cederburg & O'Doherty (2016).

Model	**Anomaly**	**Portfolio**	**Market Timing** $(1 + \frac{\overline{R}^2_{mt}}{\hat{\sigma}^2_m})Cov(\hat{\beta}^C_{i,t}, R_{m,t})$	−	**Volatility Timing** $\frac{\overline{R}_{mt}}{\sigma^2_m}Cov(\hat{\beta}^C_{i,t}, R^2_{m,t})$	=	**Total**	≈	$\hat{\alpha}^U_i$	−	$\overline{\alpha}^C_i$
DSBM-III	**Size**	Small	-0.0094	−	-0.0006	=	-0.0088	≈	0.0485	−	0.0570
		Big	0.0450	−	0.0024	=	0.0427	≈	0.0643	−	0.0238
		SMB	-0.0544	−	-0.0030	=	-0.0514	≈	-0.0158	−	0.0332
	Value	High	0.0612	−	-0.0075	=	0.0688	≈	0.2872	−	0.2242
		Low	-0.0177	−	0.0046	=	-0.0222	≈	-0.2062	−	-0.1793
		HML	0.0789	−	-0.0121	=	0.0910	≈	0.4934	−	0.4035
	Momentum	Winner	0.0092	−	-0.0041	=	0.0134	≈	0.4575	−	0.4438
		Loser	-0.0240	−	0.0097	=	-0.0336	≈	-0.8814	−	-0.8460
		WML	0.0332	−	-0.0138	=	0.0470	≈	1.3389	−	1.2898
CAPM (β^{RW})	**Size**	Small	-0.0557	−	0.0260	=	-0.0817	≈	0.0485	−	0.1207
		Big	0.0290	−	0.0077	=	0.0213	≈	0.0643	−	0.0447
		SMB	-0.0847	−	0.0183	=	-0.1030	≈	-0.0158	−	0.0760
	Value	High	0.0071	−	-0.0024	=	0.0094	≈	0.2872	−	0.2947
		Low	-0.0242	−	0.0037	=	-0.0280	≈	-0.2062	−	-0.1785
		HML	0.0313	−	-0.0061	=	0.0374	≈	0.4934	−	0.4731
	Momentum	Winner	-0.0291	−	-0.0099	=	-0.0192	≈	0.4575	−	0.4590
		Loser	0.0086	−	0.0106	=	-0.0020	≈	-0.8814	−	-0.8535
		WML	-0.0377	−	-0.0205	=	-0.0172	≈	1.3389	−	1.3125
CAPM (β^{SW})	**Size**	Small	-0.0317	−	-0.0019	=	-0.0298	≈	0.0485	−	0.0535
		Big	0.0117	−	-0.0017	=	0.0134	≈	0.0143	−	0.0719
		SMB	-0.0434	−	-0.0002	=	-0.0432	≈	0.0342	−	-0.0184
	Value	High	-0.0765	−	-0.0081	=	-0.0683	≈	0.2872	−	0.2716
		Low	-0.0956	−	-0.0048	=	-0.0908	≈	-0.2062	−	-0.1866
		HML	0.0192	−	-0.0033	=	0.0225	≈	0.4934	−	0.4583
	Momentum	Winner	-0.0142	−	-0.0067	=	-0.0075	≈	0.4575	−	0.4630
		Loser	0.0296	−	0.0061	=	0.0234	≈	-0.8814	−	-0.8507
		WML	-0.0438	−	-0.0128	=	-0.0310	≈	1.3389	−	1.3138
CAPM (β^{FH})	**Size**	Small	0.0407	−	-0.0008	=	0.0415	≈	0.0485	−	0.0914
		Big	0.0207	−	0.0102	=	0.0105	≈	0.0643	−	0.0655
		SMB	0.0200	−	-0.0110	=	0.0310	≈	-0.0158	−	0.0259
	Value	High	0.0185	−	0.0005	=	0.0181	≈	0.2872	−	0.2693
		Low	-0.0407	−	-0.0185	=	-0.0223	≈	-0.2062	−	-0.1784
		HML	0.0593	−	0.0385	=	0.0208	≈	0.4934	−	0.4477
	Momentum	Winner	0.0100	−	-0.0110	=	0.0210	≈	0.4575	−	0.4406
		Loser	0.0002	−	0.0227	=	-0.0225	≈	-0.8814	−	-0.8576
		WML	0.0098	−	-0.0337	=	0.0435	≈	1.3389	−	1.2982
CAPM (β^{KS})	**Size**	Small	0.0874	−	0.0650	=	0.0224	≈	0.0485	−	0.0205
		Big	-0.0053	−	0.0111	=	-0.0163	≈	0.0643	−	0.0780
		SMB	0.0927	−	0.0539	=	0.0388	≈	-0.0158	−	-0.0575
	Value	High	-0.0240	−	-0.0082	=	-0.0158	≈	0.2872	−	0.3013
		Low	-0.0210	−	-0.0070	=	-0.0140	≈	-0.2062	−	-0.1927
		HML	-0.0030	−	-0.0012	=	-0.0019	≈	0.4934	−	0.4940
	Momentum	Winner	0.0070	−	-0.0027	=	0.0097	≈	0.4575	−	0.4476
		Loser	0.0687	−	0.0526	=	0.0161	≈	-0.8814	−	-0.8984
		WML	-0.0617	−	-0.0553	=	-0.0064	≈	1.3389	−	1.3460

Table 2.6: Out-of-sample Cross-sectional Results of DSBM and Benchmark Models with other portfolios

This table reports the out-of-sample Fama & MacBeth (1973) coefficient estimates and performance measures for additional test portfolio which include 25 assets sorted by size and momentum, 25 assets sorted by size and book-to-market and 30 industry, and 25 assets sorted by size and book-to-market, 30 industry, and 10 momentum portfolios. See Table (2.3) for details on models and performance metrics.

Model	Assets	Prices of Risk		Cross-sectional Explanation: Performance Evaluation Measures						
							Mispriced Assets		Joint Test	
		$\bar{\hat{\lambda}}_0$	$\bar{\hat{\lambda}}_1$	**SSPE**	**RMSFE**	**Adj. R^2**	**at 5%**	**at 1%**	**JA**	**CPE**
DSBM-III	25 Size and Momentum	0.4595** 2.3792	0.3426 1.5417	2.0862	0.2889	0.1949	12	11	116.0729**	0.0613**
	25 Size and BM, 30 IND	0.6165*** 3.5464	0.1567 0.5874	1.1423	0.1441	0.1785	13	7	159.5574**	0.0388**
	25 Size and BM, 30 IND, 10 MOM	0.6372*** 3.1137	0.1350 1.2327	1.9781	0.1744	0.1262	18	12	172.1341**	0.0764**
Panel A - CCAPM Benchmark Models										
A.1 CAPM (RW)	25 Size and Momentum	0.5059*** 2.7713	0.2840 1.2364	2.3874	0.3090	0.1391	14	12	118.5957**	0.0651**
	25 Size and BM, 30 IND	0.6606*** 3.8362	0.1024 0.4913	1.1588	0.1452	0.1223	14	8	167.9551**	0.0409**
	25 Size and BM, 30 IND, 10 MOM	0.7070*** 3.8856	-0.0301 -0.3423	2.5803	0.1992	0.0468	20	14	194.4556**	0.0809**
A.2 CAPM(SW)	25 Size and Momentum	0.6123*** 3.6257	0.0397 1.2031	2.4686	0.3142	0.0777	16	14	165.5445**	0.0675**
	25 Size and BM, 30 IND	0.7001*** 4.8370	0.0699 0.4915	1.4042	0.1598	0.0502	14	10	214.4788**	0.0501**
	25 Size and BM, 30 IND, 10 MOM	0.6999*** 4.9341	-0.0083 -0.0804	2.4658	0.1948	0.0329	19	12	198.9120**	0.1030**
A.3 CAPM(FH)	25 Size and Momentum	0.5398*** 3.1663	0.1284 1.3896	2.3364	0.3057	0.1471	16	12	112.3635**	0.0622**
	25 Size and BM, 30 IND	0.6434*** 4.4068	0.1140 0.6918	1.2065	0.1481	0.1264	13	9	130.4471**	0.0420**
	25 Size and BM, 30 IND, 10 MOM	0.7372*** 3.5946	-0.0411 -0.1945	2.6107	0.2004	0.0287	20	14	201.3725**	0.1046**
A.4 CAPM(KS)	25 Size and Momentum	0.6653*** 3.7725	0.0705 1.2199	2.4182	0.3110	0.1286	15	13	127.2059**	0.0642**
	25 Size and BM, 30 IND	0.7841*** 4.9017	0.0055 0.1309	1.2755	0.1523	0.1171	12	8	170.0615**	0.0474**
	25 Size and BM, 30 IND, 10 MOM	0.7604*** 5.1499	0.0016 0.0395	2.1104	0.1802	0.0984	17	11	184.8403**	0.0774**64

Table (2.6) Continue

Panel B - Factor Models		Prices of Risk					Performance Evaluation Measures						
										Mispriced Assets		Joint Test	
		$\bar{\hat{\lambda}}_0$	$\bar{\hat{\lambda}}_{MKT}$	$\bar{\hat{\lambda}}_{SMB}$	$\bar{\hat{\lambda}}_{HML}$	$\bar{\hat{\lambda}}_{MOM}$	SSPE	RMSFE	Adj. R^2	at 5%	at 1%	JA	CPE
B.1 FF3F	25 Size and Momentum	1.3447*** 5.6497	-0.6091^{**} -2.3580	0.2379** 2.4273	-0.1483 -0.8686		0.6258	0.1582	0.7662	10	5	108.1681**	0.0142**
	25 Size and BM, 30 IND	0.8214*** 5.9002	-0.1401 -0.8373	0.1095 1.2603	0.2662*** 2.9163		0.8569	0.1248	0.3867	14	5	185.3015**	0.0297**
	25 Size and BM, 30 IND, 10 MOM	0.8859*** 6.4137	-0.2060 -1.2393	0.0976 1.1391	0.2418*** 2.6940		1.7579	0.1645	0.2491	13	8	167.8736**	0.0566**
B.2 Carhart	25 Size and Momentum	1.0577*** 5.3883	-0.3450 -1.6390	0.2639*** 2.7505	0.0026 0.0199	0.6468*** 4.9909	0.2647	0.1029	0.9011	8	6	87.1427**	0.0062**
	25 Size and BM, 30 IND	0.7675*** 5.6894	-0.0867 -0.5184	0.1440* 1.6432	0.2629*** 2.9265	0.1915 1.1958	0.7325	0.1154	0.4757	10	4	162.4989**	0.0259**
	25 Size and BM, 30 IND, 10 MOM	0.7935*** 5.8655	-0.1093 -0.6605	0.1334 1.5353	0.2696*** 3.0394	0.4136*** 3.6178	0.8140	0.1119	0.6523	12	5	193.2309**	0.0274**

Table 2.7: Out-of-sample Cross-sectional Results for other asset pricing criteria

This table reports the out-of-sample Fama & MacBeth (1973) coefficient estimates and performance measures for CCAPM with alternate asset pricing criteria for 25 assets sorted by size and book-to-market. Panel A reports results of the base case CCAPM with asset pricing criteria based on MSFE in cross-validation sample. Panel B reports some other criteria which include maximum cross-sectional adjusted R^2 (B.1.1), models B.1.2 and B.1.3 indicate models using minimum CPE using pricing errors based on realised market returns, $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\beta}_{i,t} R_{m,t+1})$, and estimated risk premium, $\hat{\alpha}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1})$, respectively. Model B.1.4 is based on minimising the excess zero-beta rate, $\overline{\hat{\lambda}}_0$. Model B.2 indicates the IV approach of Boguth et al. (2011), which is based on two-steps. First, we obtain the monthly asset betas using daily returns with a short window approach which are then regressed on all possible combinations of lagged conditioning variables ($M = 2^K - 1$, models in total). In the second step, we use the fitted conditional betas obtained from step one in return regression (See section (2.7) for details). Finally a model with maximum Adj. R^2 is selected. See Table (2.2) for details on performance metrics.

Model	Cross-Sectional Explanation								
	Price of Risk		Performance Evaluation Criteria						
						Mispriced Assets		Joint Test	
	$\overline{\hat{\lambda}}_0$	$\overline{\hat{\lambda}}_1$	SSPE	RMSFE	Adj. R^2	at 5%	at 1%	JA	CPE
A. Base Case									
A.1 DSBM-III (MSFE-CV (APC))	0.7124***	0.3919	0.5694	0.1509	0.3984	9	7	67.1011***	0.0184***
	2.9415	1.1381							
B. Altenrative asset pricing criteria									
B.1 IV1									
B.1.1 Maximum Cross-sectional Adj. R^2	0.7602***	0.3221	0.5832	0.1527	0.3484	10	7	69.8127**	0.01979**
	3.0332	1.0462							
B.1.2 Minimum CPE (TS)	0.8212***	0.2578	0.6149	0.1568	0.3087	11	7	71.4290**	0.0243**
	3.1546	0.9217							
B.1.3 Minimum CPE (CS)	0.7828***	0.2825	0.6025	0.1552	0.3102	10	7	71.0077**	0.0201**
	3.1635	0.9617							
B.1.4 Minimum Excess Zero Beta rate	0.6556***	0.3165	0.5972	0.1546	0.3314	10	7	70.0315**	0.01987**
	2.7793	0.9881							
B.2 IV2									
B.2.1 Maximum R^2	0.7731***	0.2954	0.6012	0.1551	0.3212	10	7	70.7248**	0.0200**
	3.1234	0.9416							

Table 2.8: OOS Cross-sectional Performance of DSBM and Benchmark Models for different samples

This table reports the results from the second stage of Fama & MacBeth (1973) methodology for alternate sample periods for 25 assets sorted by size and momentum, 25 assets sorted by size and book-to-market. See Table (2.3) for details on models and performance metrics.

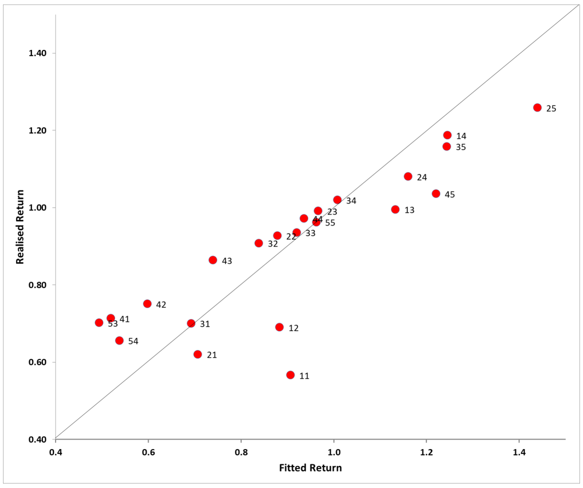
Model	Sample	Cross-sectional Explanation								
		Prices of Risk		Performance Evaluation Measures						
							Mispriced Assets		Joint Test	
		$\bar{\hat{\lambda}}_0$	$\bar{\hat{\lambda}}_1$	SSPE	RMSFE	Adj. R^2	at 5%	at 1%	JA	CPE
DSBM-III	Aug 1936-Dec 2018	0.7124*** 2.9415	0.3919 1.1381	0.5694	0.1509	0.3984	9	7	67.1012**	0.0185**
	Aug 1936-June 1963	0.6522*** 2.7367	0.4136 1.3681	0.5454	0.1477	0.4132	9	7	65.7823**	0.0179**
	Aug 1973-Dec 2018	0.8622*** 3.3864	0.0919 0.3812	0.5910	0.1538	0.3323	11	8	71.6653**	0.0198**
Panel A - CCAPM Benchmark Models										
A.1 CAPM (RW)	Aug 1936-Dec 2018	0.9022*** 3.4158	-0.2494 -0.6508	0.5904	0.1537	0.3365	11	8	72.4341**	0.0209**
	Aug 1936-June 1963	0.9251*** 3.3829	-0.1705 -0.5864	0.5051	0.1421	0.3926	8	4	68.8447**	0.0198**
	Aug 1973-Dec 2018	0.9471*** 4.0192	-0.0858 -0.3786	0.5884	0.1534	0.3157	9	7	73.9596**	0.0223**
A.2 CAPM(SW)	Aug 1936-Dec 2018	0.9499*** 4.0056	-0.2486 -1.2921	0.7203	0.1697	0.1828	12	7	80.9653**	0.0263**
	Aug 1936-June 1963	0.8486*** 4.5204	-0.0369 -0.1913	0.7094	0.2040	0.2010	11	7	80.8347**	0.0252**
	Aug 1973-Dec 2018	1.0895*** 4.8367	-0.2319 -1.2163	0.7442	0.1725	0.1345	13	10	78.5178**	0.0299**
A.3 CAPM(FH)	Aug 1936-Dec 2018	0.8404*** 5.4505	-0.0586 -0.4341	0.6112	0.1564	0.3130	11	7	69.4144**	0.0236**
	Aug 1936-June 1963	0.6935*** 3.4245	0.1000 0.4532	0.6069	0.1558	0.3324	11	6	63.0578**	0.0227**
	Aug 1973-Dec 2018	0.7833*** 3.6495	-0.0055 -0.0236	0.6199	0.1575	0.2790	12	7	62.8487**	0.0244**
A.4 CAPM(KS)	Aug 1936-Dec 2018	0.9250*** 5.6163	-0.0599 -1.1625	0.7149	0.1691	0.1965	12	10	81.2772**	0.0222**
	Aug 1936-June 1963	1.0698*** 3.3766	-0.0574 -0.7271	0.6835	0.1653	0.2033	12	9	80.8241**	0.0212**
	Aug 1973-Dec 2018	0.9652*** 5.8705	-0.0573 -1.2534	0.7257	0.1704	0.1746	12	10	82.7957**	0.0236**

Table (2.8) Continue

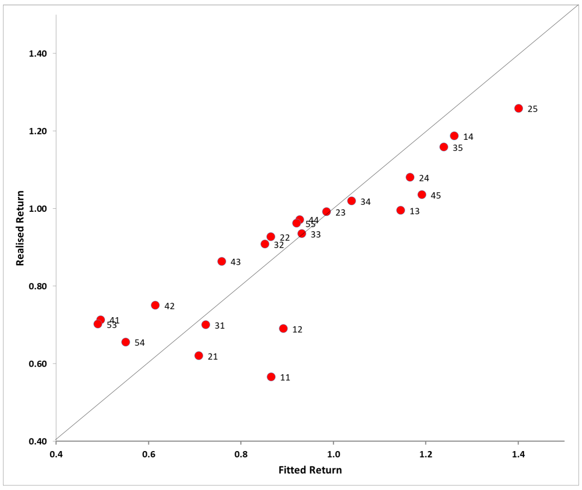
Panel B - Factor Models													
		Prices of Risk					Performance Evaluation Measures						
										Mispriced Assets		Joint Test	
		$\overline{\lambda}_0$	$\overline{\lambda}_{MKT}$	$\overline{\lambda}_{SMB}$	$\overline{\lambda}_{HML}$	$\overline{\lambda}_{MOM}$	**SSPE**	**RMSPE**	**Adj. R^2**	at 5%	at 1%	JA	CPE
B.1 FF3F	Aug 1936-Dec 2018	0.9917*** 5.8432	−0.3113* -1.6833	0.1070 1.1602	0.3338*** 3.5625		0.2482	0.0996	0.7210	7	3	65.3967**	0.0063**
	Aug 1936-June 1963	0.9312*** 3.3001	0.0024 0.0078	-0.0349 -0.2548	0.3497** 2.0552		0.4458	0.1335	0.6020	6	2	83.0417**	0.0097**
	Aug 1973-Dec 2018	1.0961*** 5.9598	−0.4603** −2.2853	0.1136 1.1147	0.2998*** 3.1498		0.2549	0.1010	0.7035	7	4	62.4926**	0.0081**
B.2 Carhart	Aug 1936-Dec 2018	0.9509*** 5.3710	-0.2785 -1.4590	0.1288 1.4063	0.3281*** 3.5430	0.1487 0.7889	0.2023	0.0900	0.7726	6	4	71.1373**	0.0052**
	Aug 1936-June 1963	0.8955*** 3.1852	0.0273 0.0911	0.0023 0.0173	0.3440*** 2.0627	0.1526 0.5238	0.3431	0.1171	0.6937	3	2	59.9188**	0.0077**
	Aug 1973-Dec 2018	1.0291*** 5.3259	-0.4017 -1.9177	0.1297 1.2820	0.2944*** 3.1140	0.1633 0.7722	0.2262	0.0951	0.7369	8	4	71.2839**	0.0072**

Figure 2.8: Actual (Realised) vs Fitted Returns

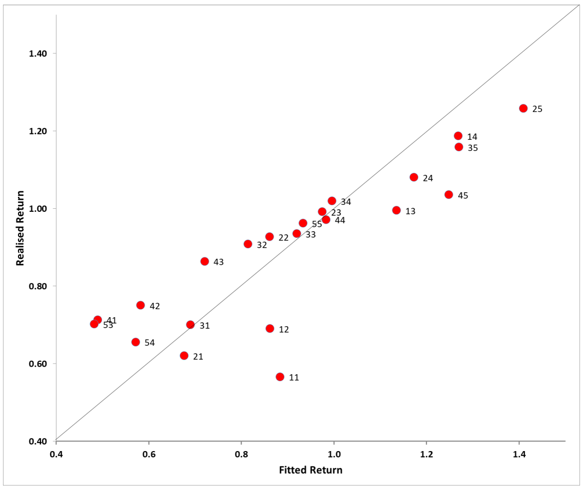
(a) DSBM-III



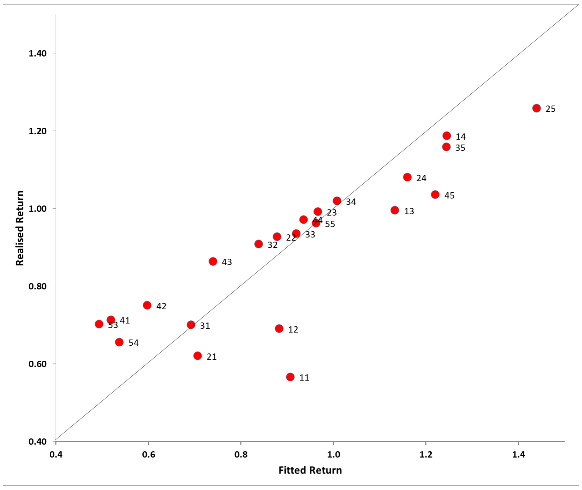
(b) CAPM (β_{RW})



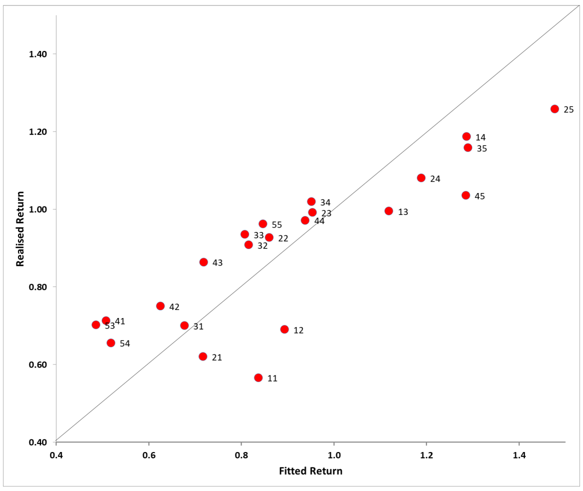
(c) CAPM (β_{SW})



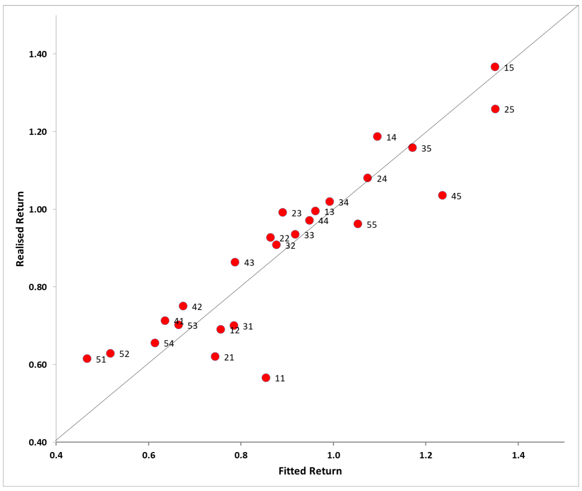
(d) CAPM (β_{FH})



(e) CAPM (β_{KS})

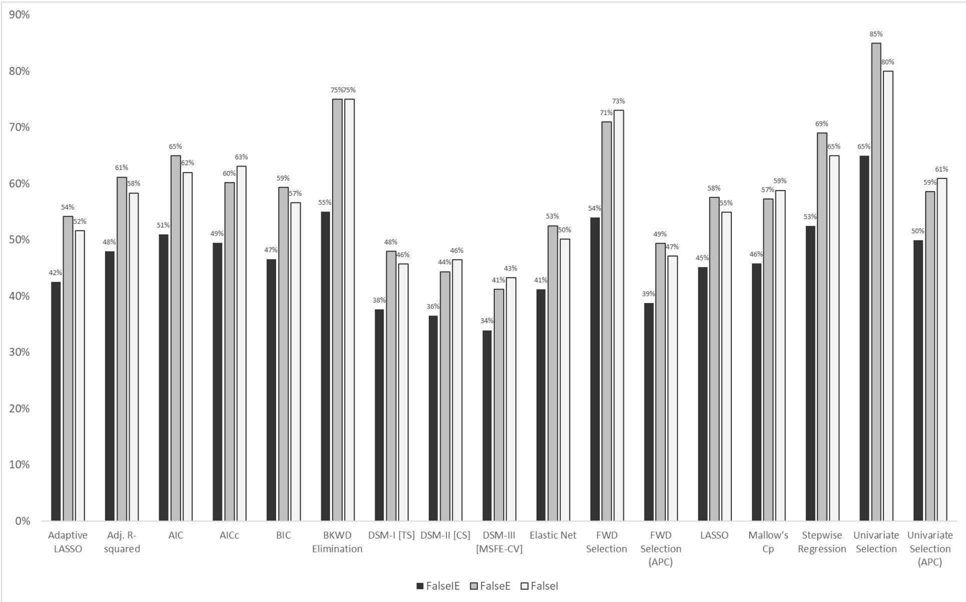


(f) FF3F



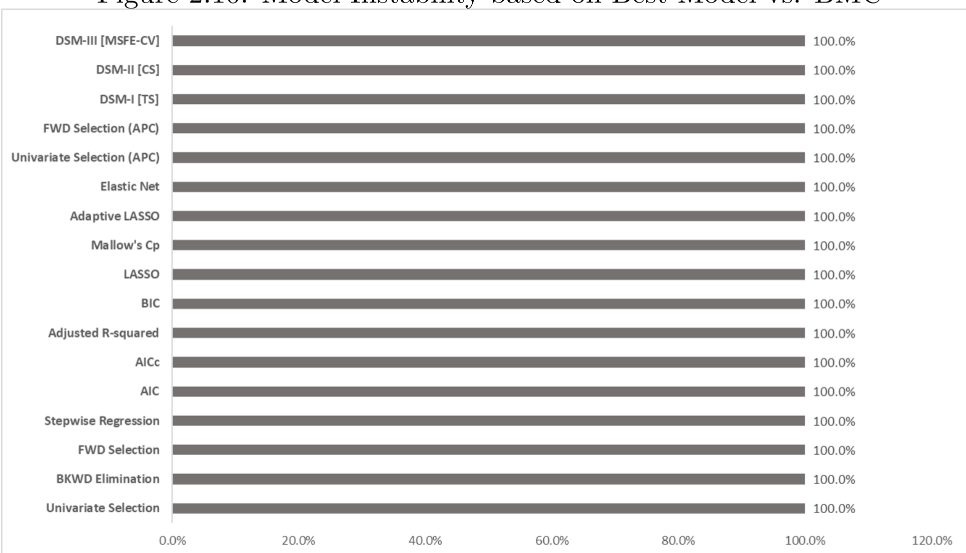
Note: This figure compares the average fitted excess returns of 25 size and book-to-market sorted portfolios and their realised (actual) excess returns for various models. The two numbers indicate the individual portfolios where the first, and second digit indicate the size and the book-to-market quantile, respectively.

Figure 2.9: Model Instability based on Inclusion Frequency



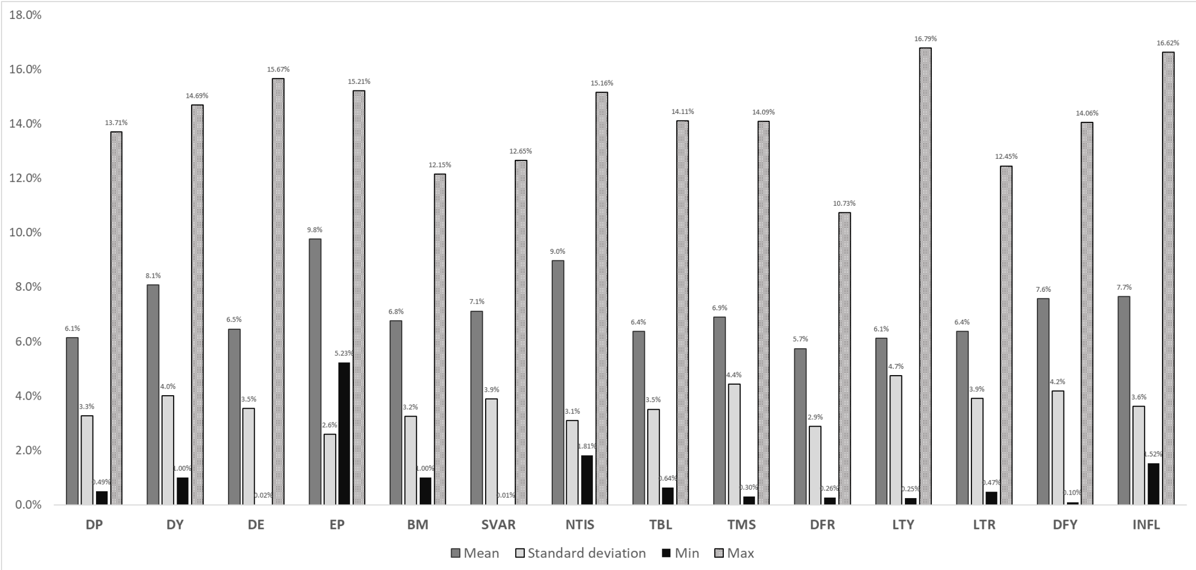
Note: This figure shows the model instability measure based on bootstrapped samples indicating the proportion of overall OOS periods when there is false inclusion, false exclusion of variables, or both under each model selection approach. A higher value suggests that there is high instability.

Figure 2.10: Model Instability based on Best Model vs. BMC



Note: This figure shows the model instability measure based on bootstrapped samples comparing the performance of best subset variables with bootstrap model combination (BMC). Bars indicate the proportion of overall OOS periods when MSFE for best selected variables is lower than MSFE of BMC. A higher value suggests that there is high instability.

Figure 2.11: Variable selection uncertainty (VSU) across assets



Note: This figure shows the VSU across assets. The values indicate the average importance, standard deviation in importance, minimum and maximum importance values of each variable across assets.

Chapter 3

Conditional CAPM (CCAPM) under Variable-selection Uncertainty (VSU)

3.1 Introduction

3.1.1 Background

The conditional CAPM (CCAPM) is well established in theory to hold perfectly, period by period; on the other hand, the unconditional CAPM misprices stocks (e.g., Jagannathan & Wang 1996). The main argument in favour of CCAPM is that a stock's conditional alpha (or pricing error) could be zero if its beta varies and is strongly correlated with the equity premium or market volatility (see Lewellen & Nagel 2006).[1] However, empirically, the performance of such a method relies on an excellent econometric framework that captures the time-variation of conditional betas (Ghysels 1998).

One of the approaches to incorporate time variation in beta is called an *instrumental variable* (IV) approach (CCAPM-IV), where beta is defined as a function of some selected instrumental variables (e.g., Jagannathan & Wang 1996, Ferson & Harvey 1999, Lettau & Ludvigson 2001, Petkova & Zhang 2005, Cederburg & O'Doherty 2016).[2] While these studies provide enough evidence on the success of CCAPM-IV, however, its implementation in practice is rather complex. One of the challenges is "variable-selection uncertainty," which arises from the inability to determine the best set of predictive variables. Cochrane (2009), for example, points out that these models are difficult to test, since proxies for the information set are needed to properly capture the dynamics of betas. Moreover, studies such as Ghysels (1998), Harvey (2001) and Cooper & Gubellini (2011) find that the performance of CCAPM-IV is sensitive to the researcher's selection of variables. In this paper, we use various approaches from the forecasting literature to model asset betas in a CCAPM-IV context to deal with variable-selection uncertainty.

3.1.2 Motivation

The main motivation for this study comes from the findings of our first essay, where we applied various dynamic model selection (DMS) approaches: *i) best subset selection* based on adjusted

[1] If the conditional CAPM holds, Lewellen & Nagel (2006) show that the unconditional alpha of a given asset can be explained by the covariance between time-varying betas (β_t) and the market risk premium (Rm_t), $\alpha_U = Cov(\beta_t, Rm_t)$.

[2] Some of the other famous approaches to model time variation in beta include those using data-driven filters such as beta calculated from a 60-month rolling window as in Fama & MacBeth (1973), or a short window approach (Lewellen & Nagel 2006) and high-frequency data (Andersen et al. 2003), multivariate GARCH (Bollerslev et al. 1988), dynamic conditional correlation (DCC) (Engle 2002, Bali & Engle 2010), regime-switching model (Vendrame et al. 2018), mean-reverting stochastic process (Jostova & Philipov 2005), Kalman filter (Adrian & Franzoni 2009), and others.

R^2, Akaike information criterion (Akaike 1973), Bayesian information criterion (Schwarz 1978), Mallows's C_P (Mallows 1973), *ii) sequential selection* approaches based on forward selection, backward elimination, and stepwise regression, *iii) the shrinkage* methods based on the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996), Adaptive LASSO (Zou 2006), and Elastic Net (ENet) (Zou & Hastie 2005), and iv) our newly introduced *dynamically selected beta model* (DSBM) that at each point in time, selects the model that currently has the best asset pricing performance. Consistent with Lewellen & Nagel (2006), we found that CCAPM based on DMS approaches cannot explain the value and momentum anomalies. Moreover, using bootstrap methods to quantify the model uncertainty and instability, we find that the DMS approaches of selecting conditioning variables is subject to considerable estimation error. These findings provide strong motivation for our second essay, where we consider alternative forecasting approaches which try to address *variable-selection uncertainty* (VSU).

3.1.3 Research Gaps and Objectives

The CCAPM-IV approach is challenging to implement due to *variable-selection uncertainty* (VSU), which arises from a lack of guidance on the predictor variables to include in the model to better reflect the investors' information set. Since this problem is so prevalent in empirical economics and finance, researchers have developed a variety of approaches to address VSU over time; however, their application to CCAPM-IV is less common. Given this gap, the primary objective of this study is to apply various forecasting approaches to CCAPM-IV in order to address VSU. In the forecasting literature, we generally find three approaches to account for VSU: the combination of forecasts (CF) (e.g., Bates & Granger 1969), the combination of information (CI) (e.g., Kelly & Pruitt 2013), and one that combines both CI and CF (e.g., Huang & Lee 2010). CF combines forecasts obtained from simple models where each model incorporates a portion of the entire information set, CI, on the other hand, integrates the whole information set into one single model to produce a single optimal forecast.[3] From a literature review, we find that there are many alternative approaches to implement CF and CI. However, there is no clear consensus on which method is the best. Most studies are concerned with equity premium and macroeconomic prediction, but none of them examines the optimal approach suitable for modelling the time variation of betas in the CCAPM framework. Therefore we aim to compare various CF, CI, and a hybrid of CI and CF approaches in explaining the cross-section of asset returns within a CCAPM-IV framework.

The CF approaches are motivated by the pioneering work of Bates & Granger (1969) and have been followed by many researchers in various empirical applications showing that the CF approach can deal with model uncertainty issues and often perform superior to their counterparts (e.g., Clemen 1989, Timmermann 2006, Elliott & Timmermann 2016). Following Rapach et al. (2010), our CF analysis combines point forecasts of betas estimated from univariate predictor-based regressions from a large pool of conditioning variables. At each point in time, these beta forecasts are weighted in various ways, including simple equally-weighted average and weighting schemes based on some criteria such as mean squared error (MSE).

The CI approaches are motivated by the availability of various high dimensional datasets that researchers these days use, for example, the Federal Reserve Economic Data (FRED) database (e.g., McCracken & Ng 2016). Some studies suggest that dimension reduction methods outper-

[3] CI, is generally referred as dimension reduction, which is the means of transforming data from a high-dimensional space to a low-dimensional space while preserving all of the original data's meaningful properties. However, we also include model selection approaches from Essay 1 under this category because either subset variable selection or dimension reduction would ultimately result in one model to generate the final forecast. The CF approach, on the other hand, always generates multiple forecasts for the same target variable and combines them into a composite forecast.

form individual models, model selection, and forecast combination methods (e.g., Ludvigson & Ng 2007, Bai & Ng 2008, Kelly & Pruitt 2013, Neely et al. 2014, Tu & Lee 2019). In our CI approach, we use dimension reduction methods which include Principal Components (PCs) (Bai & Ng 2002) and Kelly & Pruitt (2013) three pass filter based on partial least squares (PLS). These approaches take the original pool of predictor variables and reduce it down to a small subset of variables known as factors. These factors are then used to fit the time-varying beta model.

There exist other methods that combine CI and CF by first identifying the relevant predictors through various variable selection approaches (CI) and then combine the forecasts of those using CF approaches. We find two groups of studies in this category. First, using the traditional CI and CF methods, for example, Huang & Lee (2010) use principal components to extract relevant factors of forecasts (CI) and combine them using various weighting schemes (CF). Moreover, Kourentzes et al. (2019) first select predictors with classical approaches such as AIC, BIC, and adjusted R^2 and then combine the forecasts of selected predictors. The second group of studies combining CI and CF are mainly motivated by the success of machine learning techniques in forecasting.[4] These studies employ hybrid approaches, which incorporate traditional econometrics and machine learning techniques. Hirano & Wright (2017), for example, suggested a split-sample (SPLT) approach to address the variable-selection uncertainty. They demonstrated that choosing a model through AIC through SPLT and adding a bootstrap aggregation (bagging) stage improves prediction performance significantly. Liu & Xie (2019) extended the work of Hirano & Wright (2017) by replacing model selection with model averaging and applied the bagging step. Most recently, Rapach & Zhou (2020) extend the machine learning methods where they first use the elastic net method to preselect the individual predictor variables and then apply combining forecast approaches of Rapach et al. (2010). They find that combining elastic net and simple CF methods enhances forecast accuracy; the authors claim that to date, this strategy is one of the best for predicting out-of-sample equity premium.

Following the above studies, our hybrid of CI and CF approaches include principal component combinations of Chan et al. (1999) and Huang & Lee (2010), variable selection and combination (Rapach & Zhou 2020). Moreover, we consider the bootstrap aggregation (bagging or BAGG) method, which requires a training phase that involves bootstrapping new training sets. B random samples are taken from the original training dataset, with replacement. Our approach follows a two-step process where we first apply various CI approaches to get a forecast for each bootstrap sample. In the final step, we take a simple average across B forecasts to obtain a combined forecast. More specifically, our implementation of bagging follows the approach of Inoue & Kilian (2008), Rapach & Strauss (2010), and Borup & Schütte (2020). For each sample, we first select a subset of variables, such that only the statistically significant variables based on t-statistics are included. In the next step, we use these variables in a single multivariate CCAPM-IV time-series regression to model asset betas. We also use hybrid methods in the BAGG framework, which combine traditional econometrics and machine learning. We first follow Hirano & Wright (2017) to use a split-sample (SPLT) for model selection using asset pricing criteria. Second, we follow Liu & Xie (2019) and use model averaging approaches to obtain a composite forecast in each sample and then take an average across B samples.

3.1.4 Summary of Methodology

Our all methods to estimate time-varying betas are based on out-of-sample analysis to prevent look-ahead bias. Our analysis examines whether CCAPM models based on CI, CF and hybrid of CI and CF explain the cross-section of asset returns. Our analysis is based on Fama & MacBeth (1973) two-step method, in which the factor loadings for each asset, i.e. the estimates

[4] See Gu et al. (2020) for more information on machine learning approaches and their application in financial forecasting.

of conditional betas in CCAPM, are obtained in the first step using time-series regressions. The first step involves regressing monthly excess asset returns on the market risk factor, in a model where the market beta varies with conditioning variables (see, e.g., Shanken 1990, Ferson & Harvey 1999, Cederburg & O'Doherty 2016). More specifically, $\beta_t = f(X_t)$, where X represents the subset of the full information set of investors (I), $X_t \subset I_t$. Under dimension reduction approach, we first construct relevant factors F_t from X_t using various approaches, and then define $\beta_t = f(F_t)$. Under CF framework, the combined beta forecast based on individual betas j is given as: $\beta_t = f(\omega_j, \beta_{jt})$, where $\beta_{jt} = f(X_{jt})$. Finally, in our hybrid of CI and CF approaches, we first preselect the subset of variables $X_t^* \subset X_t$, next we obtain the estimates of individual betas, $\beta_{jt} = f(X_{jt}^*)$, which are then used to form combined forecast of beta, $\beta_t = f(\omega_j, \beta_{jt})$. Note that while the CI method yields one beta estimate for each asset, the CF framework yields one point forecast for each model, which we then combine to obtain a combined forecast. After obtaining the out-of-sample betas, we test the model by running a cross-sectional regression at each time t of the evaluation period, with the first-step betas (obtained through different approaches) serving as an explanatory variable. Our CCAPM cross-sectional tests are based on mainstream literature that evaluates the pricing abilities of a given model by looking at the significance of Fama & MacBeth (1973) parameter estimates.[5] In addition, we assess the performance of each model through various performance metrics such as the sum of squared pricing errors (SSPE) (Adrian & Rosenberg 2008), cross-sectional adjusted R^2 (Jagannathan & Wang (1996)), and composite pricing errors (CPE) (Campbell & Vuolteenaho 2004).

3.1.5 Principal results

We use the monthly excess returns on 25 size and value portfolios of Fama & French (1993) to perform the tests for a sample period from July 1926 to December 2018. The conditioning information variables used in this study are taken from Goyal & Welch (2008); we select the 14 variables for which monthly data are available from July 1926 to December 2018. The cross-sectional results for out-of-sample periods August 1936 to December 2018 and August 1968 to December 2018 show that all the approaches considered in this study do not handle the equity premium properly since the excess return on the zero-beta portfolio (constant from Fama & MacBeth (1973) second stage regression) is significant and large in magnitude. However, CF approaches outperform CI approaches in explaining the cross-section of asset returns measured as adjusted R^2. The improved performance of CF is in line with previous research that claims that the CF is effective in lowering forecast errors and significantly reducing model uncertainty (Bates & Granger 1969, Timmermann 2006, Rapach et al. 2010). However, hybrid models that combine CF and CI methods improve the performance of CCAPM beyond CF. These results are consistent with Huang & Lee (2010), Kourentzes et al. (2019), Rapach & Zhou (2020), and others that show the superior performance of hybrid approaches compared to individual CI and CF approaches. Moreover, such studies claim that hybrid approaches help to reduce variable-selection uncertainty (VSU). Thus, the reduction in VSU, we believe, is the primary reason for the improved results of CCAPM-IV.

3.1.6 Contribution

Our second essay contributes to the empirical asset pricing literature in the following ways: to the best of our knowledge, this is the first research to include a detailed comparison of various well-known approaches identified by literature to deal with VSU from a CCAPM perspective. Consistent with the studies such as Hirano & Wright (2017) and Rapach & Zhou (2020), we show that a combination of traditional econometric and machine learning approaches can outperform the individual methods. For example, we find the evidence on improved performance

[5] We use t-statistics to test the significance using Newey & West (1987) heteroskedasticity and autocorrelation consistent standard errors.

of CCAPM-IV with BAGG method, where in each pseudo sample, we first select the subset of variables based on the mean squared forecasting error (MSFE) in cross-validation sample, and then take a simple average of beta estimates across all pseudo samples. This method performs as well as the Fama & French (1993) three-factor model in explaining the cross-sectional returns of 25 Size-B/M, 30 industry and 10 momentum portfolios.

The remaining structure of this chapter is as follows. Section 3.2 provides an overview of the literature on various approaches to address VSU. The econometric methodology is discussed in Section 3.3. Section 3.4 addresses the implementation of CI, CF, and hybrid of CI and CF approaches to CCAPM-IV. An overview of data and benchmark models is provided in Section 3.5. Section 3.6 reports the empirical results. Section 3.7 reports the results of various robustness tests. The conclusions are drawn in the section 3.8.

3.2 Literature Review

This section provides an overview of a variety of strategies that have been used in forecasting literature to address the challenges related to variable-selection uncertainty (VSU).[6] From a time-series forecasting perspective, these approaches can be classified into (i) using a single model to generate a forecast; and (ii) combining forecasts obtained from different models. Huang & Lee (2010) name these approaches as combination of information (CI) and combination of forecasts (CF). To generate a fundamental forecast, CI integrates the entire information into a single, highly comprehensive model. CF, on the other hand, combines forecasts generated by simple models that each use a piece of the whole information set. The third strategy to address the VSU combines the CI, and CF approaches to improve forecasting accuracy (e.g., Huang & Lee 2010, Tu & Lee 2019). These approaches are discussed in the following sections.

3.2.1 Combining Information (CI)

The first category is combining information (CI) which includes approaches that are based on a single model to generate a forecast obtained through either subset variable selection or dimension reduction. Subset variable selection approaches are already discussed in Chapter 2, so here we only discuss dimension reduction approaches.

3.2.1.1 Dimension reduction methods

Addressing the issue of the increased number of predictors generally requires strategies for reducing the impact of estimation error caused by trying to include more predictors in the model. One approach is to use a few linear combinations of predictors in the forecasting model that have been carefully selected. One of the most famous approaches is to employ dynamic factor models (DFMs), which has produced a wide body of literature over the last two decades. Many studies show improvement in forecasting accuracy using DFM. For example, Ludvigson & Ng (2007) show the better out-of-sample performance of quarterly equity premium forecasts based on dynamic factors extracted from 172 financial and 209 macroeconomic predictors. Moreover, Neely et al. (2014) also demonstrate the forecasting gains for equity premium based on dynamic factors extracted from a set of popular technical indicators and Goyal & Welch (2008) predictors. DFMs can be classified into supervised and unsupervised (Tu & Lee 2019). The term "supervision" refers to the process of training predictors to forecast a variable. One of the most famous unsupervised factor models is principal component regression, PCR. A two-step process is used in PCR, in which the first step involves grouping predictors into a small number of linear

[6] For brevity, we do not provide the literature on conditional CAPM as it has already been given in Chapter 2.

combinations that best maintain the predictors' covariance structure. In the second stage, standard predictive regression is used with a few leading components. The ultimate statistical goal, predicting a target variable, is not included in the dimension reduction stage of PCR, which is a disadvantage. In other words, the PCR only accounts for the variance of the chosen predictors and does not use the forecast target information explicitly.

To overcome this problem, supervised factor models such as partial least squares (PLS) are used, which are designed to minimise dimension by directly exploiting predictor-target covariation.[7] The application of PLS methods can be found in Kelly & Pruitt (2013), who introduce a new dimension reduction technique based on PLS that considers the relationship between the target variable and predictors. Their approach consists of three steps, which they refer to as a three-pass filter (3PF), and they demonstrate out-of-sample forecasting gains for equity premium prediction using factors derived from a collection of disaggregated valuation ratios (see section (3.4.1) for the implementation of 3PF).

3.2.1.2 Hybrid of variable selection and dimension reduction

Some approaches within CI group combine the variable selection approach and dimension reduction. The idea is to overcome the criticisms on PCR that it is unsupervised as it does not consider the target variable in factor construction. Under this approach, the supervision refers to variable selection, also known as subset selection, which identifies the best predictors for the target variable. The principal components are then constructed using the selected predictors and used in the predictive regression model. This approach is called "targeted PCR". Two widely used targeting methods proposed in the literature are hardthresholding and softthresholding. Hardthresholding begins by running single linear regressions on the target for each predictor (Bai & Ng 2008). The algorithm then selects only those predictor variables with t-statistics greater than a certain threshold level (e.g., 10%). Hardthresholding is limited by the fact that it only considers the bivariate relationship between the variable X_i and the variable of interest y, ignoring the association between X_i and X_j. As a result, strongly collinear predictors are likely to be selected. Softthresholding is intended to mitigate this issue by ranking the variables in order of their significance, which is referred to as the position of inclusion (Bai & Ng 2008). More complex variable selection methods, such as Least Angle Regression (Efron et al. 2004), LASSO (Tibshirani 1996, Zou 2006), Elastic Net (Zou & Hastie 2005), and so on, have recently appeared in the literature. All of these softthresholding methods attempt to rank the predictors and choose a subset suggested by their rankings.[8]

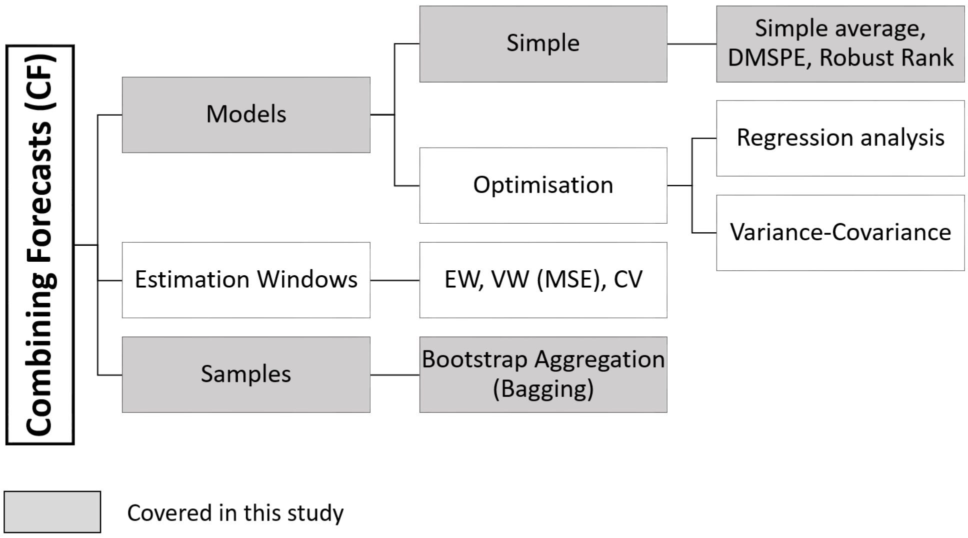
3.2.2 Combining Forecasts (CF)

Since the highly regarded publications of Barnard (1963) and Bates & Granger (1969), a stream of studies on combining various forecasts appeared in the forecasting literature. More details on combining forecasts (CF) methods can be found in Clemen (1989), Timmermann (2006), Rapach & Zhou (2013), and Timmermann (2018). A literature review suggests that CF approaches can be classified into three groups based on combining forecasts: i) across models, ii) across estimation windows, and iii) across samples. These approaches are summarised in Figure 3.1. Considering the objective of this study, that of addressing the variable-selection uncertainty (VSU), we only consider CF across models and samples.

[7] See Tu & Lee (2019), who provide a comparison between supervised and unsupervised factor models.

[8] See Bair et al. (2006) and Bai & Ng (2008), who provide detailed analysis on the success of the "targeted PCR".

Figure 3.1: Classification of CF literature



3.2.2.1 Combining across models

One CF approach is to combine the forecasts obtained through univariate predictor-based models, which limit the number of total models to the number of predictors (K). This approach generally avoids multicollinearity issues, where highly correlated predictors in the same model can result in overfitting. From the historical perspective, this approach has been used to combine forecasts when underlying information is unknown, for example, expert forecasts where the only thing known is the forecasts. Note that under this approach, the forecasts remain the same to all various techniques used to combine point forecasts. The only difference is the way weights are obtained. We generally find two approaches to obtain combining weights: i) equal weights (simple average) and ii) value weights. The value weight can further be divided into based on simple criteria such as mean square forecast error (MSFE) and optimisation techniques requiring estimation of the error-covariance matrix. Empirical evidence suggests that CF based on simple average often outperforms the optimal forecast combination based on sophisticated weighting schemes. This empirical reality has been named the "forecast combination puzzle". Many studies, such as Smith & Wallis (2009), and Graefe et al. (2014) have attempted to explain this puzzle. These studies demonstrate that the impact of the error on weight estimation can be large, offering an empirical explanation for the forecast puzzle.

Since simple weighting schemes such as equally weighted forecast is more straightforward, the issues regarding the value-weighted CF are severe in forecasting literature. Therefore, following Rapach et al. (2010), in this study, we only use simple approaches and does not focus on optimisation techniques of CF.

3.2.2.2 Combining across estimation windows

Pesaran & Timmermann (2007) and Pesaran et al. (2013) argue that CF across models fails to address the structural break problem. They demonstrate that CF across models is based on the assumption that the underlying data generation process and the models are consistent over time. Considering this, Pesaran & Timmermann (2007) propose a CF approach that combines the forecasts obtained from the same model but calculated over different estimation windows. This method is particularly useful when the existence of the breaks, as well as the number of them, is unclear. Pesaran et al. (2013) proposed a new method called average-average (AveAve) which is based on combining the two averaging methods: i) CF across models and ii) CF across estimation windows. They show that the AveAve approach outperforms the simple combination approach across models using a single-window, i.e. rolling or expanding. Note that in this paper, our objective is to address the variable-selection uncertainty by combining forecasts across models estimated with the same estimation window. Therefore, we do not apply the CF across estimation windows in our analysis.

3.2.2.3 Combining across samples – Bagging

Another form of combination is "bootstrap aggregation" (also known as bagging or BAGG), which involves creating a combined forecast for a given model using a collection of bootstrap-based training samples (multiple versions of a predictor). Breiman (1996) suggested bagging as a tool for smoothing instabilities from modelling techniques that include hardthresholding and pre-testing in order to increase forecast accuracy. In bagging, the modelling procedure is implemented to bootstrap samples many times, and the final forecast is determined by combining the forecasts obtained from the bootstrap samples. Bühlmann & Yu (2002) illustrate how bagging reduces prediction variance and, as a result, increases accuracy. Stock & Watson (2012) use the t-statistics to derive a shrinkage representation for bagging, demonstrating that it is asymptotically identical to shrinking the unrestricted coefficient estimate to zero. The t-statistic determines the magnitude of shrinkage.

For predicting economic and financial variables, bagging is becoming a common forecasting strategy. For example, Inoue & Kilian (2008) apply various bagging strategies to forecast US inflation using many predictors. On the other hand, Rapach & Strauss (2010) apply bagging for forecasting the unemployment growth using 30 predictors. When bagging is applied to a pre-test technique that selects variables based on individual t-statistics, they notice that the forecasts are very competitive when compared to CF across univariate predictor-based models. Bagging's validity in reducing forecast errors in the presence of time-series dependence is justified by Jin et al. (2014). Their findings suggest that bagging significantly reduces the mean squared error (MSE) for "unstable" predictors in predicting out-of-sample equity premium. More details on bagging can be found in Petropoulos et al. (2018) and Yin (2020).

3.2.2.4 Summary on effectiveness of combining forecasts (CF)

There are many reasons why the Combining Forecasts (CF) approach outperforms the best model chosen from a number of alternatives. Prediction models are basic approximations for data-generating processes that are almost always much more complex than one expects. As a result, it is unlikely that a single model forecast would cover all other models. Even if a particular model consistently outperforms other models by producing lower prediction errors, there is a possibility to obtain diversification gains by assigning weights to other models (Bates & Granger 1969).

Timmermann (2006) summarises the three main reasons for preferring CF over an individual model selected from a pool of models.[9] First, combining the forecasts of individual models can help to minimise the risk of selecting the wrong model (Hendry & Clements 2004). Moreover, combining forecasts (CF) based on models that use different sets of conditioning information may provide more reliable forecasts than a single model that tries to integrate all of the information, similar to the concept of portfolio diversification (Huang & Lee 2010). Second, CF

[9] The importance of CF approaches in combining forecasts across different estimation windows can be found in Pesaran & Timmermann (2007), Pesaran et al. (2013), Tian & Anderson (2014), and others. Moreover, the benefits of CF across different samples is also reported by many studies. For example, Petropoulos et al. (2018) show that combining forecasts across pseudo-samples through bagging framework can address the model uncertainty.

can be more resistant to unknown structural breaks that tend to support one model over the other at various points in time (Clark & McCracken 2010). This is also consistent with the arguments of some studies (e.g., Stock & Watson 2004, Aiolfi & Timmermann 2006) that no single model will outperform the others over all the periods. The rankings of models in terms of producing lower forecast errors keep changing over time, making the CF approach more suitable to address the issue of model misspecification. Moreover, some predictive models may well be able to respond rapidly to events like economic downturns and times of high uncertainty about the outlook of the economy, while others take longer. During or, ideally, prior to such events, a rational decision-maker might choose to use predictions from more reliable models with more accurate parameter estimates (Elliott & Timmermann 2005). Therefore, a rational investor will consider a time-varying CF scheme in light of changing economic conditions and the performance of forecasting models, making CF more appealing. Finally, if some models have omitted variable bias, CF might be able to average out such unknown biases and avoid choosing a single bad model (Panopoulou & Vrontos 2015).

The advantages of CF over the last five decades can be summarised as:[10]

- it aggregates information about various predictors since one forecasting approach is based on variables or information that has not considered by the other forecasts (Bates & Granger 1969, Chan et al. 1999, Graefe et al. 2014);
- it allows for the identification of the underlying mechanism, as various forecasting models are capable of capturing different aspects of the information available for forecasting (Reeves & Lawrence 1982, Clemen 1989);
- it considers the relative accuracy of individual approaches as well as the covariance of forecast errors across methods (Winkler & Makridakis 1983);
- it enhances forecasting performance by producing lower forecasting errors (Makridakis & Winkler 1983, Rapach et al. 2010, Cang & Yu 2014);
- it reduces the variability of accuracy for various measures of variance (Makridakis & Winkler 1983, Mahmoud 1989, Hibon & Evgeniou 2005);
- it allows for the reduction of uncertainty and is simpler and less costly than depending on a single approach (Winkler 1989, Hibon & Evgeniou 2005, Bordignon et al. 2013);
- it could lead to more normally distributed errors (Barrow & Kourentzes 2016)

3.2.2.5 Some challenges of combining forecasts (CF)

Despite a proven record of CF, there are still some challenges that can affect the forecasting performance. Some of the challenges include decisions about the selection of performance criteria, weighting scheme, and model evaluation sample. The other issue is related to the choice between a simple average and a more sophisticated weighting scheme. However, the most challenging part and perhaps the most concerning point is to decide whether to combine all the available forecasts. Though, in a value-weighted approach, different weighting methods have been suggested to reduce the effect of inaccurate predictions by assigning a low weight to a model with poor forecasting results. However, since unweighted combinations work very well (Timmermann 2006) and the majority of studies use equally weighted CF, effectively without excluding any forecast. A weakness in CF methods is that they presume that all of the predictions to be combined are important. However, it appears that a worse forecast could be given more weight than a better forecast, thus weakening the weighted forecast. Kourentzes et al. (2019) go into

[10] For more details, see, Timmermann (2006), Kolassa (2011), Elliott & Timmermann (2016).

great detail about this issue, arguing that choosing an appropriate pool of forecasts is crucial to the model development. They suggest a technique known as forecast pooling, which states that only a subset of the total set of forecasts should be combined. This is also consistent with the results of Aiolfi & Timmermann (2006), who look into how forecast pools are constructed using forecasts from arbitrarily selected top-performing quantiles or clustering methods. According to the authors, pooling selected forecasts generate more accurate forecasts compared to combining all the forecasts, but they also acknowledge that the pooling strategies are sensitive to the choice of quantiles.

The proposal of pooling forecasts naturally leads to a strategy that combines CI and CF by choosing a subset of models or forecasts and then combining them into a single forecast. The following section provides an overview of the literature using forecasting methods which are a hybrid of CI and CF approaches.

3.2.3 Combining CI and CF

In this section, we present an overview of forecasting literature that employs a combination of CI and CF methods, demonstrating that the hybrid of CI and CF produces better forecasts than either approach alone.

3.2.3.1 Pooling – Hybrid of variable selection and CF

A collection of predictors, in an ideal world, will contain all of the essential information needed to predict a variable of interest. However, if a predictor is strongly correlated with another, it runs the risk of adding noise to the data rather than predictive power, making forecasts less accurate. Moreover, forecasts could be biased towards the portion of the target variable that this group explains if a particular category of predictive variables is heavily represented in the selected set of predictors. Therefore, with a smaller but more balanced set of predictor variables, more accurate forecasts could be made.

Pooling is a forecasting strategy that uses a subset of the available forecasts rather than all of them. The aim is to minimise forecast errors even further while increasing computational efficiency. According to Timmermann (2006), the cost of implementing increased parameter estimation error should be balanced against the value of adding forecasts. He uses three easy trimming rules relying on the out-of-sample mean squared prediction errors (MSPE) of models: the top 75%, 50% and 25% models. Aiolfi & Favero (2005) also notice that trimming 80% of forecasts based on the model's R^2 boost forecasting accuracy in the context of stock return forecasting. Bjørnland et al. (2012) also provide evidence on the effectiveness of combining the inflation forecasts of the top 5% of models.

Aiolfi & Timmermann (2006) combine all forecasts from each quartile and then conduct a weighted combination of the combined quartile forecasts. Instead of decreasing the number of base forecasts employed, they reduce the number of forecasts requiring the estimation of combining weights. Most recently, Kourentzes et al. (2019) analyse forecast pooling strategies and find that they improve forecast accuracy. The authors suggest an algorithm for automatically generating forecast pools, regardless of their origin, and show that it outperforms model selection and the standard CF methods under various conditions.

3.2.3.2 Combining Forecast Principal Component (CFPC)

Combining forecast principal component (CFPC) is a form of supervised factor models (Tu & Lee 2019). This approach differs from other supervised methods like PLS, which computes factors directly. Under CFPC, forecasts are computed first, and then the principal components of the forecasts are estimated as a means for CF. The applications of combining forecasts using principal components can be found in Chan et al. (1999), Stock & Watson (2004), Huang & Lee (2010), Tu & Lee (2019), and others. The main difference between PCR and CFPC lies in the input used for extracting factors. The principal components are determined directly from $X's$ in PCR, without taking into account their association with the variable of interest, y. Because of the unsupervised nature of the PCR approach, Bai & Ng (2008) propose that a subset of X variables ("targeted predictors") that seem to be useful in forecasting be selected first, and then the subset be used to extract factors. CFPC, on the other hand, computes the principal components from a collection of individual forecasts ($\hat{y_1}$, $\hat{y_2}$, . . . , $\hat{y_N}$) that contains data on $X's$ as well as all previous y values. This helps us understand why studies like Tu & Lee (2019) find CFPC to be more powerful than PCR.

3.2.3.3 Combining CF and Machine Learning

Given the success of machine learning (ML) techniques in forecasting, some studies call traditional methods, in particular, linear statistical models, ineffective and advocate the use of machine learning methods (see Gu et al. 2020). On the other hand, some studies find evidence against machine learning methods. For example, Makridakis et al. (2018) use a broad subset of 1045 monthly time series from the M3 Competition for comparing the performance of conventional statistical approaches and machine learning across different forecasting horizons.[11] After contrasting the post-sample accuracy of common ML methods to that of conventional statistical methods, the authors discover that traditional methods outperform ML methods in all accuracy measures and forecasting horizons tested.[12] On this note, some studies have shown that combining machine learning techniques with linear statistical models will outperform the two approaches separately.[13]

Rapach & Zhou (2020) extend the machine learning methods where they first use the elastic net method to preselect the individual predictor variables and then apply combining forecast approaches of Rapach et al. (2010). They find that combining elastic net and simple CF methods enhances forecast accuracy; the authors claim that this strategy is one of the best for predicting out-of-sample equity premium to date. The split-sample approach and its model averaging extensions are two other hybrid approaches. Hirano & Wright (2017), for example, suggested a split-sample (SPLT) approach to address the variable-selection uncertainty. They demonstrated that choosing a model with AIC through SPLT and adding a bootstrap aggregation (bagging) stage improves prediction performance significantly. Liu & Xie (2019) extended the work of Hirano & Wright (2017) by replacing model selection with model averaging and applied the bagging step.

[11] Also see Xie et al. (2020) who provide a detailed review of comparison between conventional econometric methods and machine learning methods.

[12] For a discussion of why ML models are less reliable than statistical models, see Makridakis et al. (2018).

[13] Note that our focus is on linear models, so we consider shrinkage and bagging as ML methods.

3.3 Econometric Methodology

This section discusses the econometric framework to estimate and test the conditional CAPM. Before discussing the models and tests, we first introduce notation, and the process of sample splitting. Next, we move to the econometric framework of various approaches considered in this study. Finally, we discuss cross-sectional tests for evaluating the performance of various models.

3.3.1 Notations

There are N assets indexed by $i = 1, \ldots, N$, K represents total predictors indexed by $k = 1, \ldots, K$ and M indicates the total available models indexed by $j = 1, \ldots, M$. T represents total observations indexed by $t = 1, \ldots, T$. The initial training sample is indicated by W and w represents the estimation window. S indicates out-of-sample observations for final evaluation of model, which is given as, total observations (T) less initial training sample (W). The excess returns for asset i at time t, are indicated as $R_{i,t}$, and $R_{m,t}$ indicates excess market returns at time t. I_t indicates the vector of investors' information set, $X_{j,t}$ represents the vector of explanatory variables in model j ($X_t \subseteq I_t$). α_{itj} and β_{itj} represent pricing error and beta for asset i at time t with model j, respectively.

3.3.2 Sample splitting

We use out-of-sample analysis to estimate time-varying betas with various CI, CF and combination of CI and CF approaches to prevent look-ahead bias. This involves dividing the total sample into training and testing. The model parameters are estimated using the training sample and then applied to unseen data to obtain out-of-sample forecasts. Specifically, we use two sample-splitting approaches, where in the first approach, we divide the total sample of T observations into two portions: i) W as a training sample, and ii) $S = T - W$ to evaluate the out-of-sample performance. We use a rolling window approach, with a fixed window of w observations that rolls over each time up to the last observation of the sample. In our second approach, instead of splitting our total sample into two parts, we divide it into three parts: i) training (W_0), ii) validation (V), and iii) testing (S). The validation sample is used to assess the performance of a given model by evaluating its ability to predict the future. Note that the validation sample is not used to assess the final performance of the model. Instead, it only helps to identify the best model or obtaining combining weights for making the out-of-sample forecast. Thus, the third subsample (testing sample) consisting of $S = T - W_0 - V$ observations, which is not used for estimation or validation, is simply out of the sample and is instead used in evaluating the predictive performance of the given model.

3.3.3 Econometric Framework

Assuming that, in a dynamic economy, the hedging motives of risk averse investors are negligible, the conditional version of Black (1972) CAPM is described by Jagannathan & Wang (1996) as:

$$E_t[R_{i,t+1}] := E[R_{i,t+1}|I_t] = \lambda_{0,t} + \beta_{i,t}\lambda_{1,t}, \tag{3.1}$$

where $R_{i,t+1}$ denotes the return on asset i in period $t+1$, I_t represents the information set available to investors at the end of period t. In this version of conditional CAPM, $\lambda_{0,t}$ denotes the conditional expected return on a "zero beta" portfolio, while $\lambda_{1,t}$ represent the conditional market risk premium. $\beta_{i,t}$ is the conditional beta of asset i based on the given information set I_t, which is defined as:

$$\beta_{it} = \frac{Cov(R_{i,t+1}, R_{m,t+1}|I_t)}{Var(R_{m,t+1}|I_t)} \tag{3.2}$$

where $R_{m,t+1}$ denotes the return on the market portfolio in period $t+1$.

We use Fama & MacBeth (1973) two-step method to estimate the β and λ parameters of (3.1). The first step is based on time-series regressions to obtain the factor loadings for each asset, i.e. the estimates of conditional betas in CCAPM. The next step requires estimating a cross-sectional regression at each period of excess portfolio returns on the first step's conditional betas.

3.3.3.1 First-pass regressions – Estimating conditional betas

To estimate conditional betas, we follow the CCAPM-IV approach (e.g., Shanken 1990, Ferson & Harvey 1999, Petkova & Zhang 2005, Cederburg & O'Doherty 2016) and model the portfolio beta as a function of some observable instrumental variables (IVs). Our main analysis of estimating conditional betas uses the following time-series model:

$$R_{i,t+1} = a_i^{IV} + (\gamma_{i,0} + \gamma_{i,1}' X_t) R_{m,t+1} + \varepsilon_{i,t+1}, \tag{3.3}$$

where t indexes months, $R_{i,t+1}$ and $R_{m,t+1}$ are the excess returns on asset i and the market during period $t+1$, respectively, and $X_t \subseteq I_t$ is a vector of L instruments which represents the broader set of investors' information, I_t. It is thus assumed that the conditional portfolio beta is a linear function of some observable variables known at time t, $\beta_{i,t}^{IV} = \gamma_{i,0} + \gamma_{i,1}' X_t$, and the conditional portfolio alpha is constant.[14] Past studies such as Ferson & Harvey (1999), Petkova & Zhang (2005), Cederburg & O'Doherty (2016) and others use a predetermined set of instruments. However rather than selecting a prior a subset of variables from the large set of potential conditioning variables X_t, we use combining information (CI), combining forecast (CF), and combination of CI and CF methods to be discussed in section 3.4. Following subsections provide an overview of the process for obtaining out-of-sample forecast of asset betas with CI and CF respectively.

3.3.3.1.1 Combining Information In order to obtain the conditional portfolio betas based on a given CI approach, we use following equation

$$\hat{\beta}_{i,t}^{CI} = \hat{\gamma}_{i,0,t} + \hat{\gamma}_{i,1,t}' X_t^* \tag{3.4}$$

where $\hat{\gamma}_{i,0,t}$ and $\hat{\gamma}_{i,1,t}$ are the estimates of $\gamma_{i,0}$ and $\gamma_{i,1}$, respectively obtained from equation (3.3) by regressing $R_{i,2:t}$ on a constant, $R_{m,2:t}$ and $R_{m,2:t}.X_{1:t-1}^*$. Where $X_t^* \subseteq X_t$ indicates a vector of L predictors identified either through model selection approaches or factor construction to be discussed in section (3.4.1).

3.3.3.1.2 Combining forecasts The conditional portfolio betas based on a given CF approach can be given as:

$$\hat{\beta}_{i,t}^{CF} = \sum_{j=1}^{M} \omega_{j,t}^* \hat{\beta}_{i,j,t} \tag{3.5}$$

where $\hat{\beta}_{i,t}^{FC}$ is the weighted average beta for asset i across M models. $\omega_{j,t}^*$ indicates the combining weights for each model j which are obtained at time t through various methods discussed in section (3.4.2). $\hat{\beta}_{i,j,t} = \hat{\gamma}_{i,j,0,t} + \hat{\gamma}_{i,j,1,t} X_{j,t}$ is the out-of-sample beta for asset i at time t using predictor X_j. The estimates, $\hat{\gamma}_{i,j,0,t}$ and $\hat{\gamma}_{i,j,1,t}$ are the estimates of $\gamma_{i,j,0}$ and $\gamma_{i,j,1}$, respectively

[14]We also estimate model (3.3) with time-varying alphas for the robustness of results, where conditional portfolio alpha is defined as a function of same observable variables used to for modelling betas. Results show that the impact of allowing alphas to vary over time, is negligible on our main findings.

obtained from equation (3.3) by regressing $R_{i,2:t}$ on a constant, $R_{m,2:t}$ and $R_{m,2:t}.X_{j,1:t-1}$.

We use a rolling-window approach (e.g., Fama & MacBeth 1973) which employs a window of fixed length w (60 months in our case) to estimate the market beta of asset i. Specifically, to have either $\hat{\beta}_i^{CI}$ or $\hat{\beta}_i^{CF}$ at a given period, we simply estimate equation (3.3) using constructed factors or individual predictors j by using the observations within the estimation window $[t-w+2 \;\; t]$ for R_i and R_m and $[t-w+1 \;\; t-1]$ for X. To generate a beta forecast in the next period, we roll the window one step forward where one new observation is added, and the most distant one is dropped. The process continues until we obtain the final forecast at time T, which effectively generates the sequence of S out-of-sample beta estimates.

3.3.3.2 Second-pass regressions and Cross-sectional tests of CCAPM

After obtaining the out-of-sample betas, we test the model by running a cross-sectional regression at each time t of the evaluation period, with the first-step betas obtained either through CI, CF, or hybrid of CI and CF approaches serving as an explanatory variable.

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}\hat{\beta}_{i,t}^{C} + \alpha_{i,t+1} \tag{3.6}$$

where $\hat{\beta}_{i,t}^{C}$ is forecasted β of asset i based on either CI, CF, or hybrid of CI and CF approach, $\lambda_{0,t+1}$ represents the expected excess return on a 'zero beta' portfolio and $\lambda_{1,t+1}$ denotes the expected market risk premium. This will generate $S \times 1$ out-of-sample estimates of $\hat{\lambda}_0$ and $\hat{\lambda}_1$, and $S \times N$ estimates of pricing errors $\hat{\alpha}$.

To test the model, we first get the time-series averages of excess zero-beta rate (λ_0), risk premium (λ_1) and pricing errors (α_i) as:

$$\bar{\hat{\lambda}}_0 = \frac{1}{S}\sum_{t=1}^{S}\hat{\lambda}_{0,t} \tag{3.7}$$

$$\bar{\hat{\lambda}}_1 = \frac{1}{S}\sum_{t=1}^{S}\hat{\lambda}_{1,t} \tag{3.8}$$

$$\bar{\hat{\alpha}}_i = \frac{1}{S}\sum_{t=1}^{S}\hat{\alpha}_{i,t} \tag{3.9}$$

Our main asset pricing test is based on testing whether a given model implies a reasonable risk-free rate (zero-beta rate, R_{zb}) and thus adequately fits the equity premium. Here R_{zb} represents the return on an asset with zero sensitivity to risk factors (conditional beta, in our case) of any given pricing model. If a model's R_{zb} is equivalent to the prevailing risk-free rate, that model fits the equity premium well (Black 1972). The fitted constant in the cross-sectional regression given in equation (3.6) indicates the difference between the implied R_{zb} and the observed risk-free rate (i.e., $\hat{\lambda}_0 = R_{zb} - R_f$). There are two versions of tests in the literature. The first version constraints the zero-beta rate to the return on a risk-free asset by setting the cross-sectional constant to zero. The second version, however, relaxes this restriction and estimate the model with a constant.If unrestricted constant of a given model is statistically insignificant (i.e., $\hat{\lambda}_0 = R_{zb} - R_f \equiv 0$), we can say that the model adequately fits the equity premium (see, e.g., Jagannathan & Wang 1996, Cochrane 2005). We follow the second version and estimate the cross-sectional regression with a constant and test whether the average excess zero-beta rate is insignificant. Following the asset pricing theory, we also test whether the estimated risk-premium is significant. Specifically, we use the following t-statistics to test a given model:

$$t-\text{statistics}_{\bar{\hat{\lambda}}_k} = \frac{\bar{\hat{\lambda}}_k}{\hat{\sigma}_{\lambda_k}}, \qquad k = 0, 1 \tag{3.10}$$

where $\overline{\hat{\lambda}}_k$ and $\hat{\sigma}_{\lambda_k}$ are averages and standard errors of $\hat{\lambda}_0$ and $\hat{\lambda}_1$, respectively. Note that to compute t-statistics, we use Newey & West (1987) consistent standard errors for heteroskedasticity and autocorrelation.

Vendrame et al. (2018) argue that although the cross-sectional tests can provide the evidence on the performance of a given model in explaining cross-sectional returns, the evidence is nevertheless incomplete and should be complemented by evaluating anomalies separately. Following them, we evaluate the size, value, and momentum anomalies. If our models explained these anomalies, then the loadings from the three factors should not be priced. This means anomalies are already explained by models in the first pass regressions.

To implement this, we run S cross-sectional regressions with the three factors on net returns:

$$R^*_{i,t+1} = \lambda_{0,t+1} + \lambda_{s,t+1}\hat{\beta}^{SMB}_{i,t} + \lambda_{h,t+1}\hat{\beta}^{HML}_{i,t} + \lambda_{mom,t+1}\hat{\beta}^{MOM}_{i,t} + \alpha_{it+1} \tag{3.11}$$

where R^* represents the risk adjusted returns given by $R^*_{i,t+1} = R_{i,t+1} - (R_{m,t+1}\beta^C_{i,t})$, where $R_{i,t+1}$ is excess return on portfolio i at time $t+1$, $\beta^C_{i,t}$ is obtained through given model and $R_{m,t+1}$ is realised risk premium at time $t+1$. In addition, we consider the net returns based on estimated risk premium, $R^*_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}^C_{i,t}\hat{\lambda}_{1,t+1})$.

The loadings for size ($\beta^{SMB}_{i,t}$), value ($\beta^{HML}_{i,t}$), and momentum ($\beta^{MOM}_{i,t}$) are estimated through Carhart (1997) four-factor model with 60 monthly rolling window:

$$R_{i,t} = a_i + \beta^{mkt}_{i,t}R_{m,t} + \beta^s_{i,t}SMB_t + \beta^h_{i,t}HML_t + \beta^{mom}_{i,t}MOM_t + \epsilon_{i,t} \tag{3.12}$$

where $R_{i,t}$ is excess return on portfolio i at time t and $R_{m,t}$, SMB_t, HML_t, MOM_t are excess market returns, size, value, and momentum factor, respectively. The time-series test is then performed on out-of-sample means of $\lambda_{s,t+1}$, $\lambda_{h,t+1}$, and $\lambda_{mom,t+1}$. In order to test whether the loadings on size, value, and momentum anomalies are significant, we use t-statistics given in equation (3.10).

3.3.3.3 Performance Evaluation

We also consider four different measures for assessing and comparing the performance of various models under consideration. First, we assess each model's ability to generate insignificant pricing errors for individual assets using a significance level of 1% and 5%. The total number of mispriced assets (MPA) out of a total of N assets is our performance metric. A model with a lower value of MPA indicates a better pricing ability. Next, we follow Adrian & Rosenberg (2008) and use the sum of square pricing errors (SSPE) which is defined as:

$$SSPE = \overline{\hat{\alpha}}' \; \overline{\hat{\alpha}} \tag{3.13}$$

SSPE does not take into account the number of assets, so we use the root mean square pricing errors (RMSPE), which is computed as:

$$RMSPE = \sqrt{SSPE/N} \tag{3.14}$$

Next, following Jagannathan & Wang (1996), we use adjusted R^2:

$$Adjusted\ R^2 = 1 - \frac{(S-1)(1-R^2)}{(S-K-1)} \tag{3.15}$$

$$R^2 = \frac{var_c(\overline{R}) - var_c(\overline{\hat{\alpha}})}{var_c(\overline{R})} \tag{3.16}$$

where $\overline{R} = 1/S\sum_{t=1}^{S} R_t$ and var_c is cross sectional variance. The metrics such as SSPE and adjusted R^2, give all test assets equal weight, but some assets are actually less volatile than others (Campbell & Vuolteenaho 2004). To overcome this issue, two additional metrics are considered, both of which test whether all the pricing errors from cross-sectional regressions are jointly zero ($H_0 : \hat{\alpha} = 0$). The first measure is JA (joint alpha test) which is a χ^2-statistic and can be given as:

$$JA = \overline{\hat{\alpha}}' cov(\overline{\hat{\alpha}})^{-1}\overline{\hat{\alpha}} \quad \sim \quad \chi^2_{N-P} \tag{3.17}$$

where N, P, $\overline{\hat{\alpha}} = \frac{1}{S}\sum\limits_{t=1}^{S} \hat{\alpha}_t$, and $\hat{\alpha}_t = [\hat{\alpha}_{1,t}, \hat{\alpha}_{2,t}, ..., \hat{\alpha}_{N,t}]'$ denote number of assets, number of factors in a given model, the average pricing errors and vector of estimated errors, respectively. According to the joint χ^2 test, if the pricing theory holds, the pricing errors generated by a model should be close to or equal to zero. The higher the statistic value, the greater the pricing errors produced by the model. JA value is compared to the critical value to test the significance of pricing errors. If JA exceeds the χ^2_{N-P} 5% critical value, the pricing errors are significant. In order to estimate the variance-covariance matrix of pricing errors $\overline{\hat{\alpha}}$, denoted as $cov(\overline{\hat{\alpha}})$ and a version accounting for autocorrelation, denoted as $\widetilde{cov}(\overline{\hat{\alpha}})$, we estimate following equations:

$$cov(\overline{\hat{\alpha}}) = \frac{1}{S^2}\sum_{t=1}^{S} (\hat{\alpha}_t - \overline{\hat{\alpha}})(\hat{\alpha}_t - \overline{\hat{\alpha}})' \tag{3.18}$$

$$\widetilde{cov}(\overline{\hat{\alpha}}) = \frac{1}{S^2}\sum_{t=1}^{S} (\hat{\alpha}_t - \overline{\hat{\alpha}})(\hat{\alpha}_t - \overline{\hat{\alpha}})' + \frac{1}{S^2}\sum_{j=1}^{q}\sum_{t=j+1}^{S} (1 - \frac{j}{q+1})(\hat{\alpha}_t - \overline{\hat{\alpha}})(\hat{\alpha}_{t-j} - \overline{\hat{\alpha}})' \tag{3.19}$$

where $q = \lfloor (4(S/100)^{2/9}) \rfloor$ and $\lfloor x \rfloor$ denotes larger integer not greater than x.

Our second measure for joint alpha test is the Composite Pricing Error (CPE), which was used by Campbell & Vuolteenaho (2004) and is defined as:

$$CPE = \overline{\hat{\alpha}}'\hat{\Omega}^{-1}\overline{\hat{\alpha}} \quad \sim \quad \chi^2_{N-P} \tag{3.20}$$

where $\hat{\Omega}$ represents a diagonal matrix with main diagonal carrying the variances of estimated returns. Under this measure of an aggregate pricing error, assets with more volatile alphas receive less weight. The null hypothesis that pricing errors produced by a given model are jointly zero is rejected, if CPE exceeds the 5% critical value. To supplement our analysis, we follow Andronoudis et al. (2019) and consider the magnitude of the pricing errors across models. More specifically, we use the square root of the CPE indicated as PEM and Hansen & Jagannathan (1997) distance measure (HJ):

$$PEM = \sqrt{CPE} \tag{3.21}$$

$$HJ = \sqrt{\overline{\hat{\alpha}}'[(\overline{R}'\ \overline{R})^{-1}]\overline{\hat{\alpha}}} \tag{3.22}$$

For the weighting matrix, HJ relies on the moment matrix of expected asset excess returns.

3.4 Combining Information and Combining Forecasts

In this section, we discuss the various approaches widely used in forecasting literature to address VSU. Our objective is to compare these methods from the CCAPM perspective. The approaches can be classified into three categories: i) combining information (CI), ii) combining forecast (CF), and iii) combining CI and CF.

3.4.1 Combining Information

Under this approach, we apply dimension reduction methods.[15]

3.4.1.1 Dimension Reduction

Dimension reduction approaches are used to address the problem associated with high dimensional data which has become common and is of increasing importance in finance domain. Dimension reduction methods are often known as dynamic factor models (DFM) or diffusion indices (DI). The DI technique integrates the information found in a big number of predictors into a handful of estimated factors to avoid the negative consequences of having too many parameters, such as over-fitting and poor forecast performance. The resulting more parsimonious data set usually contributes to forecasting accuracy (e.g., Stock & Watson 2004, Bai & Ng 2008). These methods can be classified into supervised and unsupervised. The key distinction is that supervised approaches take into account the association between explanatory variable (X) and the target variable (y). In contrast, unsupervised methods extract the factors which explain the cross-section of the predictors (X) without considering their relevance to the target variable (y). We use both supervised and unsupervised methods in our analysis. The supervised method includes 'Principal Components Regression (PCR)', and two standard methods from supervised include the PCR with targeted predictors and partial least squares (PLS) where we consider the three-pass filter of Kelly & Pruitt (2013). In order to estimate out-of-sample forecasts of beta based on diffusion indices (DI), we use following model.

$$R_{t+1} = a^{IV} + \gamma_0 R_{m,t+1} + \gamma'_{DI}(R_{m,t+1}.\widehat{F}_t) + u_{t+1} \tag{3.23}$$

where a^{IV} represent the intercept, γ_0 is coefficient on excess market returns, and γ_{DI} is a q-vector of slope coefficients on factors.

3.4.1.1.1 Principal Components Regression (PCR) The first dimension reduction method we use is PCR. The steps are as follows:

$$x_t = \Lambda F_t + \upsilon_t, \tag{3.24}$$

where $x_t = (x_{1t}, \ . \ . \ . \ , \ x_{Kt})'$ is $(K \times 1)$ vector of predictors, r is true number of factors, Λ is $K \times r$ and F_t is $r \times 1$ vector of common factors. The latent common factors $F = (F_1 \ F_2 \ ... \ F_T)'$ can be obtained by using the principal component methodology:

$$\widehat{F} = X\widehat{\Lambda}/K \tag{3.25}$$

where K represents the size of x_t, $X = (x_1 \ x_2 \ ... \ x_T)'$, and factor loading $\widehat{\Lambda}$ is set to $\sqrt{K}$ times the eigenvectors corresponding to the r largest eigenvalues of $X'X$ (see Bai & Ng 2002).[16] In order to estimate out-of-sample forecasts of beta based on factors obtained through principal components, we use equation (3.23). Next, the out-of-sample beta forecast made at time t using estimated parameters of $\hat{\gamma}_0$ and $\hat{\gamma}_{DI}$ from (3.23) using data up to time t can be given as:

$$\hat{\beta}_t^{CI\text{-}PCR} = \hat{\gamma}_0 + \hat{\gamma}'_{DI}\widehat{F}_t \tag{3.26}$$

[15] CI approaches also include variable selection methods such as stepwise selection methods, best subset selection, and shrinkage methods. We applied these methods in Chapter 2.

[16] If there is no information on the true number of factors, one can estimate it by minimising some information criteria such as AIC and BIC (e.g., Bai & Ng 2002). However, following Neely et al. (2014), we use the adjusted R^2 criterion to select optimal factors and also use a small number of factors, i.e. 1, 2, and 3.

3.4.1.1.2 Principal Components with targeted predictors Recent research has found that using too many predictors to compute the factors will damage a model's forecasting accuracy because the PCR approach does not account for the target variable's properties. For example, when the factors are calculated using a subset of variables having high potential to predict the target variable y, Boivin & Ng (2006) and Bai & Ng (2008) show that the forecasting performance is significantly improved over regular PCR forecasts. The chosen subset of predictors is referred to as "targeted predictors," and this refinement of the DI approach is referred to as "targeted DI forecasts" (Bai & Ng 2008).

To find a subset of targeted predictors, various techniques can be used. In this study, we use both hard and softthresholding rules to target the predictors, as recommended by Bai & Ng (2008). We also introduce asset pricing criteria for identifying the target predictors.

3.4.1.1.2.1 Hardthresholding After controlling for lags of the target variable, hardthresholding approach evaluates the bivariate relationship between each variable (X_j) and the target variable (y), independent of other predictors. A predictor is added to the pool of targeted predictors if there is a statistically significant relationship with target variable. Specifically, our approach is based on Bai & Ng (2008) where we estimate equation (3.27) for each $j = 1, \ldots, K$ using OLS with four lags, $p = 4$. Let the t-statistics associated with X_j is denoted by t obtained through OLS regression. We next sort $t_1, t_2, \ldots, t_K$ in descending order and identify the predictors with t-statistics greater than a certain level of significance, α.

$$R_{t+1} = a_0 + \sum_{d=1}^{p} a_d' R_{t-d+1} + \gamma_0 R_{m,t+1} + \gamma_j'(R_{m,t+1}.X_{jt}) + e_{t+1} \tag{3.27}$$

The equation (3.27) yields a t-value on γ_j for each of the possible predictors X_j, defined as t_j. For any given threshold, $\mho$, we can define a vector of targeted predictors as $X_t^* \subset X_t$ by only including X_j if $\mid t_j \mid > \mho$. We use threshold of $\mho = 1.96$, which is critical value of two-tailed at 5 percent level.[17] The factors can now be estimated using equation (3.24) and remaining process to estimate out-of-sample beta remain same as discussed in (3.4.1.1.1). We indicate the beta under this method as $\hat{\beta}^{CI\text{-}TPCR}$ (HT).

3.4.1.1.2.2 Softthresholding The hardthresholding method simply decides if the predictors are significant without taking into account their correlations. As a consequence, hardthresholding has a tendency to pick strongly correlated predictors. Softthresholding methods resolve this flaw by performing subset selection and shrinkage on the entire range of predictors at the same time. Following Li & Chen (2014), we use LASSO method for identifying the subset of relevant predictors $X_t^* = \{X_{jt} \in X_t \mid \hat{\gamma}^{LASSO} \neq 0\}$. The factors can then be estimated from selected predictors X_t^* using equation (3.24) and remaining process to estimate out-of-sample beta remain same as discussed in (3.4.1.1.1).[18] We indicate the beta under this method as $\hat{\beta}^{CI\text{-}TPCR}$ (ST).

3.4.1.1.2.3 Targeted predictors with Asset Pricing Criteria The targeted predictor methods such as hardthresholding and softthresholding are designed for a single target variable. In our application, we model the beta of individual asset separately which targets

[17] As there is no clear indication what the exact threshold should be, we also use different threshold values, $\mho = [2.58,\ 1.65,\ 1.28]$. These are the two-tailed 1, 10 and 20 percent levels respectively. However, we do not find big difference in results with these threshold values.

[18] We also use Elastic Net (ENet), Least Angle Regression (LARS) and stepwise selection methods. However, we do not find a significant difference in results. Note that these methods are discussed in Chapter 2, so we do not provide details of variable selection approaches.

different predictors for each asset. Here we apply cross-sectional asset pricing criteria to target the predictors which not only identify the predictors that are related to cross-sectional returns but also restricts the predictors to be same for all the assets which helps in identifying the important predictors overtime.[19] In order to implement our approach, we follow Fama & MacBeth (1973) two step process where in the first step we estimate time-series regression, $R_{i,t+1} = a^{IV}_{i,j} + (\gamma_{i,j,0} + \gamma'_{i,j,1}X_{j,t})R_{m,t+1} + \varepsilon_{i,t+1}$, for each predictor j using observation up to time t as training sample. Next, we get $(T \times N)$ estimates of the fitted conditional betas, $\hat{\beta}_{i,j,t} = \hat{\gamma}_{i,j,0} + \hat{\gamma}'_{i,j,1}X_{j,t}$. Next, we run following cross-sectional regression at each time t $(t = 1,\ 2,\ \ldots,T)$:

$$R_{i,t+1} = \lambda_{0j,t+1} + \lambda_{1j,t+1}\hat{\beta}_{i,j,t} + \alpha_{ij,t+1} \tag{3.28}$$

where $\lambda_{0,j,t+1}$ represents the excess zero-beta rate and according to theory this should be zero. We test this with t-statistics using equation (3.10) discussed in section (3.3.3.2). This provides one t-value for each of the predictor, denoted as t_j. For a threshold of $\mho = 1.96$, which is critical value of two-tailed at 5 percent level, we can create a $X^*_t \subset X_t$ by only including X_j if $\mid t_j \mid < \mho$. The factors can now be estimated using equation (3.24), and the remaining process to estimate out-of-sample betas remains the same as discussed in (3.4.1.1.1). Note that our application is similar to hardthresholding, where we select the predictors based on the t-value of the bivariate relationship between the target variable and predictor. We indicate the beta under this method as $\hat{\beta}^{CI\text{-}TPCR}$ (APC).

3.4.1.1.3 Three pass filter of Kelly & Pruitt (2013) The three-pass regression filter of Kelly & Pruitt (2013) identifies the relevant factors to explain a variable of interest, i.e. asset returns in our case. This method can be interpreted as a series of three regressions. In the first pass regression given in equation (3.29), K time-series regressions are estimated, one for each predictor.[20] The predictor (X_j) is the dependent variable in these first pass regressions, the proxies (Z) are the regressors, and the estimated coefficients define the predictor's sensitivity to the factors represented by the proxies.[21]

$$X_{j,t} = \phi_{0,j} + Z'_t\phi_j + e_{j,t} \tag{3.29}$$

The second pass consists of estimating T different cross-sectional regressions of the predictors $X_{j,t}$ on coefficients $\hat{\phi}'_j$ obtained from the first-stage.

$$X_{j,t} = \phi_{0,t} + \hat{\phi}'_j\mathbf{F}_t + \varepsilon_{j,t} \tag{3.30}$$

This gives time series of $\hat{F}_t$ which are then used in the following CCAPM equation.

$$R_{t+1} = a^{IV} + \gamma_0 R_{m,t+1} + \gamma'_{KP}(R_{m,t+1}.\widehat{F}_t) + u_{t+1} \tag{3.31}$$

where a^{IV} is the intercept, γ_0 represent coefficient on excess market returns, and γ_{KP} is a q-vector of slope coefficients. The forecasted beta at time t using estimated parameters of $\hat{\gamma}_0$ and $\hat{\gamma}_{KP}$ from (3.31) can be given as:

$$\hat{\beta}^{CI\text{-}KP}_t = \hat{\gamma}_{0,t} + \hat{\gamma}'_{KP,t}\widehat{F}_t \tag{3.32}$$

[19] We also use t-statistics of time-series of conditional alpha $\alpha_{i,j,t+1} = R_{i,t+1} - \hat{\beta}_{i,j,t}R_{m,t+1}$, which selects different predictors for each asset. We do not find much difference in results. Like, hardthresholding, we also use various threshold levels for both time-series and cross-sectional criteria.

[20] Predictors need to be standardised to have unit variance.

[21] Following Kelly & Pruitt (2015), we use one proxy and setting it equal to the target variable, $Z = R_i$.

3.4.1.1.4 Single Index using asset pricing criteria Under this approach, we create a factor (single-index) based on asset pricing performance. We first define a single index of predictors as:

$$X_t^{SI} = \sum_{j=1}^{K} \omega_j X_{jt} \tag{3.33}$$

For each $X_{j,t}$, we have T observations, but we only use $(T-1)$ observations as a training sample for obtaining optimal loadings for a single index. We initially set equal weights, $1/K$, that we later solve through optimisation. The time-series model based on equally-weighted index can be given as:

$$R_{t+1} = a_0 + \gamma_0 R_{m,t+1} + \gamma_1 (R_{m,t+1}.X_t^{SI}) + e_{t+1} \tag{3.34}$$

Next we obtain the fitted conditional betas, $\hat{\beta}_t^{SI} = \hat{\gamma}_0 + \hat{\gamma}_1 X_t^{SI}$, and compute the pricing errors as:

$$\hat{\alpha}_{t+1}^{SI} = R_{t+1} - (\hat{\beta}_t^{SI} R_{m,t+1}) \tag{3.35}$$

Next we obtain the asset pricing performance measure, mean squared error (MSE) of conditional pricing errors as:[22]

$$MSE^{SI} = \frac{1}{T-1} \sum_{s=1}^{T-1} (\hat{\alpha}_s^{SI})^2 \tag{3.36}$$

where $\hat{\alpha}_s^{SI}$ is one-step ahead conditional pricing error given in equation (3.35) estimated at time s $(s = 1,\ 2,\ .\ .\ .\ .\ T-1)$. Note that the MSE is function of single index which depends upon weight of individual predictor (ω_j). If we assume that weights remain same overtime then our optimisation problem would be:

$$\underset{w_j}{\text{minimise}} \quad MSE^{SI} \tag{3.37}$$

The optimal weights $(\hat{\omega}^*)$ now can be used to construct the optimal single index as:

$$X_t^{SI^*} = \sum_{j=1}^{K} \hat{\omega}_j^* X_{jt} \tag{3.38}$$

Now we use this optimal index as only conditioning information variable in CCAPM equation as:[23]

$$R_{t+1} = a_0 + \gamma_0 R_{m,t+1} + \gamma_1 (R_{m,t+1}.X_t^{SI^*}) + e_{t+1} \tag{3.39}$$

Now by using these estimates we can estimate the conditional beta as:

$$\hat{\beta}_t^{SI} = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t} X_t^{SI^*} \tag{3.40}$$

3.4.1.1.5 Cross-sectional Beta Premium (CSBP) In this method, to construct common factors, we use the estimated cross-sectional beta premium (CSBP) based on each model. The idea behind this approach is to create common factors by relating conditional betas obtained with conditioning information.

Assume we have M predictors $(j = 1,\ 2,\ ...\ ,M)$, N assets $(i = 1,\ 2,\ ...\ ,N)$ and T observations in total for one-month ahead returns for asset i, $R_{i,t+1}$, one-month ahead returns

[22] We also use cross-sectional adjusted R^2 as objective function. Additionally, we use a restricted version where we use two conditions, $\sum_{j=1}^{K} w_j = 1$ and $0 \leq w_j \leq 1$

[23] Note that the index consists of T observations, where $T-1$ observations are used as training data and last observation is held for out-of-sample beta forecast.

for market, $R_{m,t+1}$, and predictors $X_{j,t}$.[24] We use following steps to reach out-of-sample beta forecast based on cross-sectional beta premium (CSBP).

Step 1: Estimate time-series regression
By using T observation as training sample, following time-series regression is estimated for each univariate predictor-based model j:

$$R_{i,t+1} = a_{i,j}^{IV} + (\gamma_{i,j,0} + \gamma'_{i,j,1} X_{j,t}) R_{m,t+1} + \varepsilon_{i,t+1}, \tag{3.41}$$

Step 2: Estimate fitted conditional betas
Next, we obtain $T \times N$ estimates of conditional betas using following equation.

$$\hat{\beta}_{i,j,t} = \hat{\gamma}_{i,j,0} + \hat{\gamma}'_{i,j,1} X_{j,t} \tag{3.42}$$

Step 3: Estimate cross-sectional regression
Next, we estimate following cross-sectional regression at each time t ($t = 1,\ 2,\ \dots, T$):

$$R_{it+1} = \lambda_{0j,t+1} + \lambda_{1j,t+1} \hat{\beta}_{i,j,t} + \alpha_{ij,t+1} \tag{3.43}$$

This provides $T \times 1$ loadings on estimated betas (cross-sectional beta premium, $\hat{\lambda}_1$) for each model j. With M models we will have a panel of $T \times M$ cross-sectional beta premium.

Step 4: Estimate factors of cross-sectional beta premium
Next, we use panel of $T \times M$ cross-sectional beta premium to construct P factors, denoted as F_t, using principal component method discussed in section (3.4.1.1.1).[25]

Step 5: Use factors as lagged predictor
Given that we have T observations of $R_{i,t+1}$ and $F_{j,t+1}$ ($j = 1,\ 2,\ \dots, P$), and we want to use cross-sectional beta premium as lagged predictor, so we use $T-1$ observations for estimation of factor regression (3.44) where we use $2:T$ observations of $R_{i,t+1}$ and $1:T-1$ observations of $F_{j,t+1}$ and hold T^{th} observation of factor for out-of-sample beta estimation.

$$R_{i,t+1} = a_i^{IV} + (\gamma_{i,0} + \gamma'_{i,CSBP} F_t) R_{m,t+1} + \varepsilon_{i,t+1}, \tag{3.44}$$

Step 6: Estimate out-of-sample beta
Finally, our beta estimate at time t can be given as:

$$\hat{\beta}_{i,t}^{CI\text{-}CSBP} = \hat{\gamma}_{i,0} + \hat{\gamma}'_{i,CSBP} \widehat{F}_t \tag{3.45}$$

3.4.1.2 Testing Dimension Reduction (DR) models

Let $\hat{\beta}_{i,t}^{DR}$ be the beta estimates made at time t with a given dimension reduction (DR) method. The cross-sectional regression at each time t can be given as:

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1} \hat{\beta}_{i,t}^{DR} + \alpha_{i,t+1} \tag{3.46}$$

Finally, we evaluate the performance of each model using tests discussed in section (3.3.3.2) and (3.3.3.3).

[24] Note that here we set $M = K$, however, the same method can be applied to any given number of models
[25] We also use targeted predictors and KP's 3PF and find consistent results

3.4.2 Combining Forecasts

This section discusses the framework to obtain optimal weights for combining the betas of individual assets. Note that all CF methods require individual point forecasts and their combining weights, so our each CF method only differs in the way it obtains the combining weights. Given that out-of-sample CF of beta is given as:

$$\hat{\beta}_{i,t}^{CF} = \sum_{j=1}^{M} \omega_{j,t}^{*} \hat{\beta}_{i,j,t} \tag{3.47}$$

All the CF methods discussed below are aimed at obtaining $\omega_{j,t}^{*}$ to be applied to beta forecast $\hat{\beta}_{i,j,t}$. The CF literature suggests that the combining weights for forecasting models depend upon the performance of individual models in terms of forecasting accuracy. This is normally measured using forecasting errors of individual models. From this perspective, the forecasting errors would be given as the residuals, $\varepsilon_{i,t+1}$ obtained from the time-series regression, $R_{i,t+1} = a_{i,j}^{IV} + (\gamma_{i,j,0} + \gamma'_{i,j,1} X_{j,t}) R_{m,t+1} + \varepsilon_{i,t+1}$. However, from asset pricing perspective, the forecasting accuracy depends upon the conditional pricing errors, $\hat{\alpha}_{i,t+1} = R_{i,t+1} - \hat{\beta}_{i,t} R_{m,t+1}$.[26] Considering this, our all CF methods are based on asset pricing criteria using time-series of conditional alphas as main input for performance evaluation of individual models.[27]

To implement our CF approaches, we follow three simple steps. First for each portfolio i, we estimate following CCAPM model for individual predictor-based model j using data up to time t:

$$R_{i,t+1} = a_{i,j}^{IV} + (\gamma_{i,j,0} + \gamma'_{i,j,1} X_{j,t}) R_{m,t+1} + \varepsilon_{i,t+1}, \tag{3.48}$$

Next, by using estimates of $\hat{\gamma}_{i,j,0}$ and $\hat{\gamma}'_{i,j,1}$, obtain conditional betas as:

$$\hat{\beta}_{i,j,t} = \hat{\gamma}_{i,j,0} + \hat{\gamma}'_{i,j,1} X_{j,t} \tag{3.49}$$

Next, by using the estimates of betas, we calculate conditional alphas as:

$$\hat{\alpha}_{i,j,t+1} = R_{i,t+1} - \hat{\beta}_{i,j,t} R_{m,t+1} \tag{3.50}$$

Our CF approaches are based on following methods.

3.4.2.1 Simple average

Under simple CF methods, following Rapach et al. (2010), we use simple average, trimmed mean, and median forecast models. The combining weights under this approach can be defined as:

$$\omega_j = \frac{1}{M}, \qquad j = 1,\ 2, \ldots,\ M \tag{3.51}$$

The equally-weighted beta forecast for any given asset i, can be given as:

$$\hat{\beta}_{i,t}^{EW} = \frac{1}{M} \sum_{j=1}^{M} \hat{\beta}_{ij,t} \tag{3.52}$$

The simple average can be sensitive to outliers. Following Rapach et al. (2010), we also consider the trimmed mean combination forecast which sets $\omega_{j,t} = 0$ for the smallest and largest individual

[26] We also use standard forecasting approaches which use OLS residuals as main criteria for forecasting accuracy. The idea is to make a comparison and support our argument that using OLS residuals as main criteria is not ideal for combining weights for betas.

[27] We also use cross-sectional pricing errors (cross-sectional residuals) of individual assets as alternative asset pricing criteria for robustness and find consistent results with time-series approach.

forecasts and then the combining weights for remaining $M-2$ forecast can be given as: $\omega_{j,t} = 1/(M-2)$.[28]

3.4.2.2 Value weight based on asset pricing criteria

Under this approach the combining weights are the function of some asset pricing criteria. We use R^2 as performance measure of each model j, which can be given as:[29]

$$R_j^2 = 1 - \frac{Var[\hat{\alpha}_{j,t}]}{Var[R_t]} = 1 - \frac{Unexplained\ Variance}{Total\ Variance} \tag{3.53}$$

where $\hat{\alpha}_{j,t}$ is conditional pricing errors given in equation (3.50). The combining weights and beta estimates can be given as:

$$\omega_{jt}^{VW-APC} = \frac{R_j^2}{\sum_{j=1}^{M} R_j^2} \tag{3.54}$$

$$\hat{\beta}_{i,t}^{CF-VWAPC} = \sum_{j=1}^{M} \omega_{jt}^{VW-APC} \hat{\beta}_{ij,t} \tag{3.55}$$

3.4.2.3 Discounted mean square forecast error (DMSFE) method

Under DMSFE method, more weight is assigned to the recent forecasts than distant ones. Following Winkler & Makridakis (1983), the combining weights under DMSFE can be given as:

$$\omega_j^{DMSFE} = \frac{1/\sum_{t=1}^{T} \theta^{T-t-1} \hat{\alpha}_{jt}^2}{1/\sum_{j=1}^{M}\sum_{t=1}^{T} \theta^{T-t-1} \hat{\alpha}_{jt}^2} \tag{3.56}$$

where θ indicates the discounting factor with $0 < \theta \leq 1$, $\hat{\alpha}_{jt}$ is the j^{th} forecast error obtained through (3.50), whereas T represents the number of observations and M indicates the the number of individual forecasts, respectively. The coefficient θ gives less (more) weight to the distant (recent) forecasting errors. Following Rapach et al. (2010) we set $\theta = 0.9$ and $\theta = 1$. The beta forecast can be given as:

$$\hat{\beta}_{i,t}^{CF-DMSFE} = \sum_{j=1}^{M} \omega_{jt}^{DMSFE} \hat{\beta}_{ij,t} \tag{3.57}$$

3.4.2.4 Robust weighting scheme

Aiolfi & Timmermann (2006) propose a robust weighting method that inversely weighs forecast models to their rank based on some performance criteria such as MSFE.

$$\omega_j^{RobR} = \frac{Rank_j^{-1}}{\sum_{j=1}^{M} Rank_j^{-1}} \tag{3.58}$$

[28] As robustness we also use the median and the mode of $\hat{\beta}_{i,j,t}$ and do not find much difference in results.
[29] We also use other measures such as Root Mean Squared Pricing Error (RMSPE) and find similar results.

where a rank of 1 is given to the best model, a rank of 2 is received by the second best model, and so on. The out-of-sample beta forecast can be given as:

$$\hat{\beta}_{i,t}^{CF-RobR} = \sum_{j=1}^{M} \omega_{jt}^{RobR} \hat{\beta}_{ij,t} \tag{3.59}$$

3.4.2.5 Testing CF models

Let $\hat{\beta}_{i,t}^{CF}$ be the beta estimates made at time t with a given combining forecasts (CF) method. The cross-sectional regression at each time t can be given as:

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}\hat{\beta}_{i,t}^{CF} + \alpha_{i,t+1} \tag{3.60}$$

Finally, we evaluate the performance of each model using tests discussed in section (3.3.3.2) and (3.3.3.3).

3.4.3 Combining CI and CF

These forecasting schemes combine the CI and CF approaches. These methods include forecast pooling, combining forecasts principal components (CFPC), CF of univariate PCR, CF with cross-sectional beta premium, and combining machine learning (ML) and traditional CI and CF methods.

3.4.3.1 Forecast pooling

Forecast pooling is a technique that uses only a portion of the available forecasts rather than all of them. The main objective of pooling strategies is to minimise forecast errors even further while increasing computational efficiency (Aiolfi & Timmermann 2006). The main motivation for this strategy is that, rather than using a sophisticated weighting scheme to allocate lower weights to weak forecasts, it is preferable to exclude them and pool the remaining forecasts using a simple average (Kourentzes et al. 2019). The primary issue in this approach is which forecast assessment criteria should be used to identify the best forecasts to combine. The traditional approaches to forecast pooling are discussed first, followed by our asset pricing criteria for selecting appropriate forecasts. Finally, we introduce a method that automatically chooses the best subset of forecasts for pooling.

3.4.3.1.1 Pooling top quantiles In our first approach, we follow Timmermann (2006), who claims that the advantage of incorporating additional forecasts must be balanced against the cost of increasing parameter estimation error. He recommends to construct a pooled forecast using only top q% models based on the pseudo out-of-sample mean squared forecast error (MSFE).[30] Given that V represents the number of observations in validation sample, and $\hat{\alpha}_{j,t}$ indicates the one-step ahead conditional pricing errors of individual forecast j, at time t of validation sample, the MSFE in CV sample, $MSFE^{CV}$, for forecast j can be given as:

$$MSFE_j^{CV} = \frac{1}{V}\sum_{t=1}^{V}(\hat{\alpha}_{j,t})^2 \tag{3.61}$$

Following Timmermann (2006), we consider combining the top 75%, top 50% and top 25% models based on $MSFE^{CV}$, we call them as $\hat{\beta}_t^{CF\text{-}Q3}$, $\hat{\beta}_t^{CF\text{-}Q2}$, and $\hat{\beta}_t^{CF\text{-}Q1}$, respectively.

[30] We split T observations into T_0 and V for training and out-of-sample holdout period. Next, we compute the MSFE of each predictor-based model j in validation sample. See section (3.3.2) for details.

It is obvious that this approach may select different predictors for each asset. In order to identify common predictors to pool, we use alternative criteria from asset pricing tests based on the cross-sectional performance.[31] More specifically, we use t-statistics of cross-sectional zero-beta rate of each univariate predictor-based model j, denoted as t_j. The subset variables, $X_t^* \subset X_t$, only include X_j if $\mid t_j \mid > 1.96$. Let us use P for selected variables. We estimate following models to generate conditional betas:

$$R_{i,t+1} = a_0 + \gamma_{0,j} R_{m,t+1} + \gamma_{1,j}(R_{m,t+1}.X_{jt}^*) + e_{i,t+1} \qquad (j = 1, \ . \ . \ . \ . \ , \ P) \tag{3.62}$$

$$\hat{\beta}_{i,j,t} = \gamma_{0,j,t} + \hat{\gamma}_{1,j,t}.X_{j,t}^* \qquad (j = 1, \ . \ . \ . \ . \ , \ P) \tag{3.63}$$

$$\hat{\beta}_{i,t}^{Pool-APC} = \sum_{j=1}^{P} \omega_t^*.\hat{\beta}_{i,j,t} \tag{3.64}$$

We use equal weights for individual forecasts.

3.4.3.1.2 Cluster combination Aiolfi & Timmermann (2006) proposed cluster combination approaches, which are conditional combination approaches that incorporate information from previous forecast results of individual models as well as forecast persistence. Following Rapach & Strauss (2010), we use the Previous Best Conditional Combination algorithm, defined as $C(D, PB)$, where C indicates a cluster combination, D refers to the number of clusters, and PB is used for the Previous Best conditional combination strategy. With this combination approach, the first combining forecast is calculated using the MSFE by clustering the individual model forecasts over the initial holdout out-of-sample period into D clusters of the same size. The individual models with the next lowest MSFE values are in the second cluster, and so on. The cluster CF is then simply an average of the cluster's forecasts. In this study, we follow Aiolfi & Timmermann (2006) and consider $D = 2$ and $D = 3$.

3.4.3.1.3 Pooling Forecast Island The most important downside of using top quantiles discussed in the previous section is that there is no performance metric used in selecting the best quantile; instead, the cut-off level is arbitrarily determined. As a consequence, the results must be sensitive to the cut-off level chosen by the modeller. Because of this, Kourentzes et al. (2019) introduced a heuristic for forming forecast pools, named "forecast island." Let C indicates some appropriate performance criteria. Following them, we use $C = MSFE^{CV}$ given in equation (3.61) (MSFE in validation sample) and order the forecasts from best to worst.[32] We then create $C' = (0, \Delta C)$ from the sorted metric, where Δ is used for differencing and a 0 indicates the first prediction. Here C' represents the rate of change of the metric allocated to each prediction. As a consequence, the pool of selected forecasts consists of all predictions before the first steep rise, which can be obtained by using the formula generally used to find outliers in the boxplot. This can be given as: $\mathbb{T} = Q3 + 1.5IQR$, here $Q3$ represents the 3^{rd} quartile, while IQR denotes the inter-quartile range. As each additional forecast is considered, $\mathbb{T}$ is progressively determined. All forecasts are included in the pool before $C' \geq \mathbb{T}$ and then are combined using a simple average.

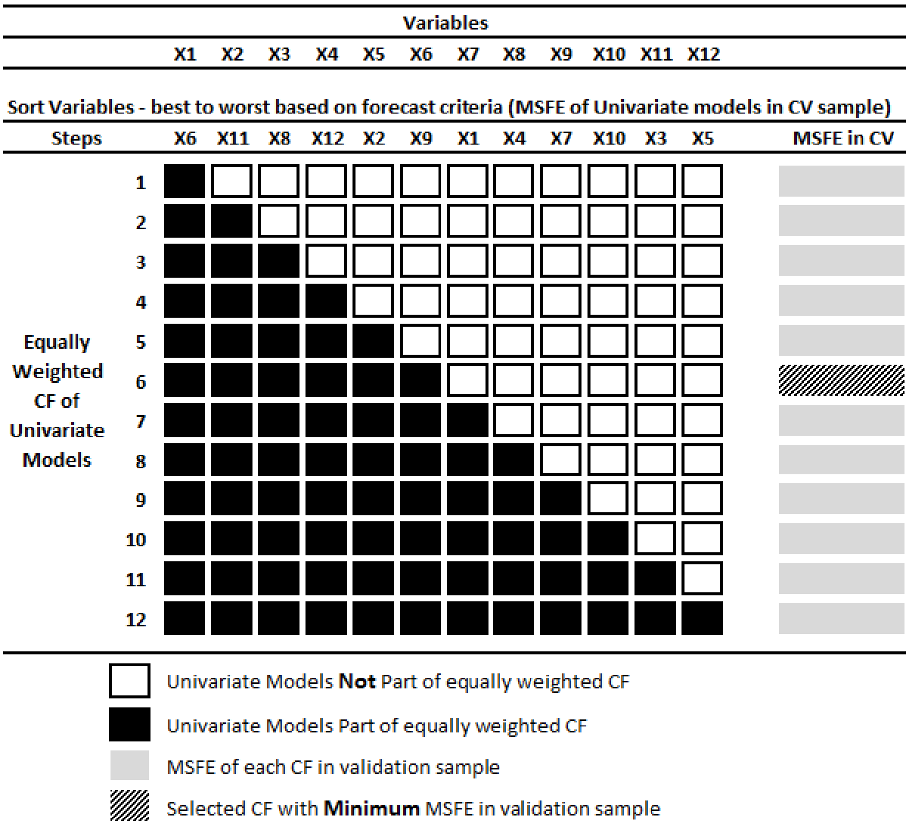
3.4.3.1.4 Auto-pooling (Optimal pooling) with asset pricing criteria Under this approach, we introduce a novel approach to automatically choose the optimal forecasts to combine. More specifically, we use out-of-sample holdout validation sample to evaluate the performance

[31] These asset pricing criteria are discussed in section (3.4.1.1.2.3).

[32] We also use some other criteria such as AIC, AICc, BIC and adjusted R^2 but there is no significant difference in results.

of each CF with specific k. The implementation of this process is similar to stepwise forward selection, however our objective is to find the univariate predictor-based models that best perform in CF framework. Figure 3.2 shows the implementation of our approach.[33] To implement this, we follow the standard approaches of selecting top quantiles or forecast island based on MSFE in validation sample of individual forecasts. In the first step, we sort the forecasts (predictors) from best to worst based on $MSFE_j^{CV}$.[34]

Figure 3.2: Optimal Pooling



Next, we start adding on predictor at a time to existing predictors to form a combined forecast based on individual forecasts and compute the MSFE in validation sample.[35] The process continues till all predictors from univariate models are finished. Note that each step will provide a CF based on active individual forecasts, where the first model is based on the best univariate predictor-based model (standard best univariate model selection approach), and the last model combines all the available forecasts (standard equally-weighted combining forecasts approach). The MSFE of each of k can be given as:

$$MSE_k = \frac{1}{V}\sum_{t=1}^{V}(\hat{\alpha}_{k,t})^2 \tag{3.65}$$

where $k = 1, 2, \ldots, K$ and each k consists of $P = 1 : k$ forecasts in the model. $\alpha_{k,t}$, is one-period ahead forecast error based on equally-weighted beta forecasts for a specific k, at time

[33]Note that the figure provides an example of 12 predictors, however, this can be generalised to any given number of predictors or forecasts.

[34]We also use cross-sectional adjusted R^2, and t-statistics of cross-sectional zero-beta rate of each predictor j as alternative criteria which effectively helps to identify the predictors to pool which are common to all portfolios.

[35]Note that the MSFE is computed for CF, not the individual forecasts.

t in out-of-sample holdout validation period, which is given as:

$$\hat{\alpha}_{k,t+1} = R_{t+1} - (\overline{\hat{\beta}^{CF}_{k,t}} R_{m,t+1}) \tag{3.66}$$

where $\overline{\hat{\beta}^{CF}_{k,t}}$ is defined as equally-weighted average beta forecast in each k, given as:

$$\overline{\hat{\beta}^{CF}_{k,t}} = \frac{1}{k} \sum_{j=1}^{k} \hat{\beta}_{j,t} \tag{3.67}$$

As shown in Figure 3.2, with maximum K models of CF, we get K values of $MSFE^{CV}$ and we choose optimal CF with one producing minimum MSFE and that is used to get out-of-sample beta at t defined as $\hat{\beta}^{CF-optP^*}_{i,t}$. The process is repeated at each period, leading to a beta forecast of a given asset with time-varying optimal pooling.

3.4.3.2 Combining forecasts principal components (CFPC)

Combining forecasts principal components (CFPC) is a combination of CF and PC methods, as the name suggests. In this approach, we follow Huang & Lee (2010) and Tu & Lee (2019) to combine the dimension reduction approaches, in particular principal components (PCs), given in section (3.4.1.1.1) and CF method proposed by Granger & Ramanathan (1984) (GR henceforth). The main motivation for the CFPC method is to address issues with the traditional GR approach, which involves regressing the target variable y_t on forecasts $\hat{y}_{j,t}$ obtained from a specific model j.

$$y_t = \omega_0 + \sum_{j=1}^{M} \omega_j \hat{y}_{j,t} + e_t \tag{3.68}$$

The estimated regression coefficients ($\hat{\omega}_0$, and $\hat{\omega}_j$) indicate the combining weights for a particular model.[36] If we have M models, generating M forecasts of target variable y, we need to estimate a multivariate regression with M explanatory variables. If the M is very large, there are chances of overfitting. Therefore, to overcome this issue of the standard GR method, instead of regressing y on all the M forecasts, dimension reduction methods such as principal components (PCs) are applied on M forecasts to construct the P factors, where $P < M$. Now the model can be estimated by regressing y on P factors.

The implementation of this approach is not straightforward when applied to CCAPM because we are mainly interested in estimates of betas and the approach relies on direct forecasts of the target variable. To implement this approach, we first need to obtain the estimated betas, $\hat{\beta}_{i,j,t}$, in either testing period or out-of-sample holdout period. Next, we require the estimated returns, $\hat{R}_{j,t+1} = \hat{\beta}_{j,t} R_{m,t+1}$, for each asset i based on each univariate predictor-based model j. The principal components are then extracted from the collection of individual predictions. Let $\hat{\mathbf{R}}_{t+1} = (\hat{R}_{1,t+1}, \hat{R}_{2,t+1}, \ldots, \hat{R}_{M,t+1})'$. Now, by considering a factor model of $\hat{\mathbf{R}}_t$ (similar to the factor model of X_t in equation (3.24) for *CI-PCR*). Let $\widehat{\mathbf{F}}_{t+1} = (\widehat{F}_{1,t+1}, \widehat{F}_{2,t+1}, \ldots, \widehat{F}_{P,t+1})'$ denote the first P principal components of $\hat{\mathbf{R}}_{t+1} = (\hat{R}_{1,t+1}, \hat{R}_{2,t+1}, \ldots, \hat{R}_{M,t+1})'$. Then the forecasting equation is:

$$R_{t+1} = \omega_0 + \sum_{j=1}^{P} \omega_j \widehat{F}_{j,t+1} + u_{t+1} \tag{3.69}$$

[36] Note that the regression weights in equation (3.68) are unconstrained, there are also two other variates of Granger & Ramanathan (1984) methods, putting some restrictions on combining weights. The first requires estimating the equation (3.68) without an intercept. Under the second approach, there is no intercept, and the sum of regression weights is constrained to one.

Equation (3.69) suggests that the asset returns depend upon the constructed factors from fitted returns, which are based on conditional betas. To estimate the conditional beta forecasts made at time t, we first need to calculate the fitted returns by using the estimated parameters of $\hat{\omega}_0$ and $\hat{\omega}_j$ from (3.69).

$$\hat{R}_{t+1} = \hat{\omega}_0 + \sum_{j=1}^{P} \hat{\omega}_j \widehat{F}_{j,t+1} \tag{3.70}$$

We know that the estimated returns for a particular asset can be given as $\hat{R}_{t+1} = \hat{\beta}_t R_{m,t+1}$. Since we know the estimated returns $\hat{R}_{t+1}$ and realised market returns $R_{m,t+1}$, the beta estimate can be defined as:

$$\hat{\beta}_t^{CF\text{-}PC} = \hat{R}_{t+1}/R_{m,t+1} \tag{3.71}$$

This forecasting scheme is denoted as CF-PC.

3.4.3.3 Combining forecasts based on univariate PCR

Here we combine the univariate forecasts of principal components regression (PCR). The principal components of the explanatory variables are used as regressors in PCR instead of explicitly regressing the dependent variable on the explanatory variables. Assume that we have r principal components ($\hat{F}_t$), which we use as individual predictors ($\hat{X}_t$). Let us use P for selected principal components through some selection criteria discussed in (3.4.1.1.1). We estimate the following models to estimate conditional betas:

$$R_{i,t+1} = a_0 + \gamma_{0,j} R_{m,t+1} + \gamma_{1,j}(R_{m,t+1}.\hat{F}_{jt}) + e_{i,t+1} \qquad (j = 1, \ . \ . \ . \ . \ , \ P) \tag{3.72}$$

$$\hat{\beta}_{i,j,t}^{PC} = \gamma_{0,j,t} + \hat{\gamma}_{1,j,t}.\hat{F}_{j,t} \qquad (j = 1, \ . \ . \ . \ . \ , \ P) \tag{3.73}$$

$$\hat{\beta}_{i,t}^{CF-UPCR} = \sum_{j=1}^{P} \omega_{j,t}^{*}.\hat{\beta}_{i,j,t}^{PC} \tag{3.74}$$

where $\omega_{j,t}^{*}$ represents optimal weights formed at time t, which can be obtained through different methods discussed in section (3.4.2). However, we use equal weights for individual forecasts.

3.4.3.4 Combining forecasts using cross-sectional beta premium

In section (3.4.1.1.5), we discussed the process of constructing factors from cross-sectional beta premium from a panel of $(T \times M)$ cross-sectional premium. Here we use individual CSBP as a predictor to obtain individual forecast. Given that we have K predictors and we can obtain K CSBP, $\hat{\lambda}_{1,j}$ $(j = 1, \ 2, \ . \ . \ . \ . \ , \ K)$. Now by using CSBP $(\hat{\lambda}_{1,j})$ as a predictor we can estimate following models.

$$R_{i,t+1} = a_0 + (\gamma_{0,j} + \gamma_{1,j}\hat{\lambda}_{1,j,t})R_{m,t+1} + e_{i,t+1} \qquad (j = 1, \ . \ . \ . \ . \ , \ K) \tag{3.75}$$

$$\hat{\beta}_{i,j,t} = \hat{\gamma}_{0,j,t} + \hat{\gamma}_{1,j,t}.\hat{\lambda}_{j,t} \qquad (j = 1, \ . \ . \ . \ . \ , \ K) \tag{3.76}$$

$$\hat{\beta}_{i,t}^{CF-CSBP} = \sum_{j=1}^{K} \omega_{j,t}^{*}.\hat{\beta}_{i,j,t} \tag{3.77}$$

Optimal weights (ω^{*}) can be obtained through various schemes discussed in section (3.4.2). However, we only report the results based on equal weights.

3.4.3.5 Combining Machine Learning and Traditional methods

Under this section, we discuss some of the methods that combine machine learning methods in particular shrinkage and bagging with CI or CF approaches.

3.4.3.5.1 Combining Forecast using Shrinkage methods Most recently, Rapach & Zhou (2020) employ the elastic net for refining the simple combination forecast.[37] Instead of averaging all the individual univariate predictor-based forecasts, the combination ENet (C-ENet) forecast takes the simple average across individual forecasts selected by the ENet. Following them, we use the ENet method to preselect the variables and then using them for combining individual forecasts of preselected variables. Let ENet selects the subset of relevant predictors $X_t^* = \{X_{jt} \in X_t \mid \hat{\gamma}^{ENet} \neq 0\}$. Let us use P for selected variables. To obtain conditional betas, we estimate the following models:

$$R_{i,t+1} = a_0 + \gamma_{0,j} R_{m,t+1} + \gamma_{1,j}(R_{m,t+1}.X_{jt}) + e_{i,t+1} \quad (j = 1, \ldots, P) \tag{3.78}$$

$$\hat{\beta}_{i,j,t} = \gamma_{0,j,t} + \hat{\gamma}_{1,j,t}.X_{j,t} \quad (j = 1, \ldots, P) \tag{3.79}$$

$$\hat{\beta}_{i,t}^{FC-ENet} = \sum_{j=1}^{P} \omega_{j,t}^{*}.\hat{\beta}_{i,j,t} \tag{3.80}$$

where $\omega_{j,t}^*$ represent optimal weights can be obtained through formed at time t using methods discussed in section (3.4.2). However, here we use equal weights for individual forecasts.

3.4.3.5.2 Bootstrap Aggregation (Bagging) Our implementation of "Bootstrap Aggregation" (Bagging or BAGG) follows the lines of Inoue & Kilian (2008), Rapach & Strauss (2010), and Borup & Schütte (2020).[38] We first apply a hard threshold on the variables in X_t, such that only the statistically significant variables based t-statistics at 5% significance level remain.[39] The subset variables, $X_t^* \subset X_t$, only include X_j if $\mid t_j \mid > 1.96$. See section (3.4.1.1.2.1) for details on hardthresholding. Let us use P for selected variables. In the next step, we run the following multivariate regression using X_t^* using data up to time t as:

$$R_{i,t+1} = a_0 + \gamma_0 R_{m,t+1} + \sum_{j=1}^{P} \gamma_{1,j}(R_{m,t+1}.X_{j,t}^*) + e_{i,t+1} \quad (j = 1, \ldots, P) \tag{3.81}$$

Next, we get the out-of-sample beta forecast at time t as:

$$\hat{\beta}_{i,b,t} = \hat{\gamma}_{0,t} + \sum_{j=1}^{P} \hat{\gamma}_{1,j,t}.X_{j,t}^* \quad (j = 1, \ldots, P) \tag{3.82}$$

The procedure is then augmented by using a moving block bootstrap to reduce variance coming from model uncertainty. We produce B pseudo samples of the size of training sample for the $R_{i,t+1}$, $R_{m,t+1}$, and X_t with replacement.[40] For a given pseudo sample (indexed by b), we use

[37] Note that this method is similar to forecast pooling discussed in section (3.4.3.1). However, we place this method here because the variable selection under this method is made using a machine learning method (ENet)

[38] Note that Bagging method is normally placed in combining forecasts (CF) category, in particular combination across samples see section (3.2.2.3) for details. However, we place this method under combining CI and CF framework because targeting predictors for each pseudo sample is related to the subset variable selection that we categorise as CI approach and the average across samples is CF approach. So in its design, bagging with targeted variables is a combination of CI and CF.

[39] Following Rapach & Strauss (2010), we use Newey & West (1987) standard errors.

[40] Following Inoue & Kilian (2008) and Rapach & Strauss (2010), we use $B = 100$. More details on generating bootstrap sample can also be found in these studies.

the decision rule ($tval = 1.96$) to eliminate insignificant predictors and estimate equation (3.81) with data up to time t. Next, we use equation (3.82) to compute the hard-threshold bootstrap beta forecast, $\hat{\beta}_{i,b,t}$, using bootstrap coefficients and original data X^*. The bagging beta forecast is then given as the average of the B hard-threshold bootstrap forecasts:

$$\hat{\beta}_{i,t}^{BAGG} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{i,b,t} \tag{3.83}$$

3.4.3.5.3 Bagging with asset pricing criteria Under this method, we use two approaches, indicated as BAGG-APC1 and BAGG-APC2. The framework for both approaches remains the same discussed in the previous section (3.4.3.5.2). The only difference is the selection of significant predictors. In the first approach, we use asset pricing criteria discussed in section (3.4.1.1.2.3) to preselect the predictors. More specifically, we use t-statistics of cross-sectional zero-beta rate of each predictor j, denoted as t_j. The subset variables, $X_t^* \subset X_t$, only include X_j if $|t_j| > 1.96$.

In the second approach, we want to directly compare the standard approach where significant predictors in each pseudo sample are selected through t-statistics and the approach based on asset pricing criteria. Here we use cross-sectional adjusted R^2 as our asset pricing criteria where first we get the adjusted R_j^2 for each univariate predictor-based model j. Next, we sort the adjusted R_j^2 from high to low values and select the first P models as significant predictors used to estimate betas in each sample. Here P indicates the number of predictors in X_t^* selected by standard method, $X_t^* \subset X_t$, and $\{X_j \in X_t^* \mid |t_{X_j}| > 1.96\}$. This way, we can ensure that both methods, standard and asset pricing criteria use the same number of variables in each sample, but the only difference is how we select these predictors.

3.4.3.5.4 CI approaches using Bagging The original approach of Bagging is based on selecting a subset of variables by evaluating the bivariate relationship (hard-threshold based on t-statistics) between a given predictor and the target variable. However, as discussed earlier in CI approaches (see section (3.4.1)) that we can apply various techniques to identify the subset of variables. Hirano & Wright (2017) also find that subset selection through sample-split CV and bagging application can improve the forecasting accuracy. Considering this, for a given subset variable selection method, we take an average of out-of-sample betas, $\hat{\beta}_{i,t}^{CI}$, across pseudo samples. For example, in our implementation of shrinkage methods, the out-of-sample beta for LASSO is given as $\hat{\beta}^{CI-lasso}$. Assume that the same process is repeated with B pseudo samples, the bagging forecast – the average across samples – can be given as:

$$\hat{\beta}_{i,t}^{CI-lasso-BAGG} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{i,b,t}^{CI-lasso} \tag{3.84}$$

3.4.3.5.5 CF approaches using Bagging Liu & Xie (2019) and Xie et al. (2020) find that the combination of bagging and CF strategies can significantly improve forecasting accuracy. Following this, we apply bagging on CF approaches discussed in section (3.4.2). For example, the CF of beta using robust weighting method discussed in section (3.4.2.4) is given by $\hat{\beta}_{i,t}^{CF-RobR}$, Given that we repeat the same process on B samples, the bagging forecast can be given as:

$$\hat{\beta}_{i,t}^{CF-RobR-BAGG} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{i,b,t}^{CF-RobR} \tag{3.85}$$

3.4.3.5.6 Combining CI and CF approaches using Bagging

Under this framework, any given CI and CF framework can the estimated with bagging. For example, CF using shrinkage methods recently used by Rapach & Zhou (2020) discussed in section (3.4.3.5.1). Where we define out-of-sample beta with original sample as $\hat{\beta}_{i,t}^{CF-ENet}$. The bagging forecast of this approach with B sample can be given as:

$$\hat{\beta}_{i,t}^{CF-ENet-BAGG} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{i,b,t}^{CF-ENet} \tag{3.86}$$

3.4.3.6 Testing combination of CI and CF models

Let $\hat{\beta}_{i,t}^{CI-CF}$ be the beta estimates made at time t with a given hybrid of CI and CF (CI-CF) method. The cross-sectional regression at each time t can be given as:

$$R_{i,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}\hat{\beta}_{i,t}^{CI-CF} + \alpha_{i,t+1} \tag{3.87}$$

Finally, we evaluate the performance of each model using tests discussed in section (3.3.3.2) and (3.3.3.3).

3.5 Data and Benchmark Models

3.5.1 Data

Following a large body of empirical research on explaining cross-sectional variation in expected returns, we use the 25 size and book-to-market portfolios to perform the tests for a sample period from July 1926 to December 2018. In addition, for robustness tests, we also use 25 size and momentum portfolios, 30 industry portfolios, and 10 momentum portfolios. The returns of all these portfolios and market factor are calculated in excess of risk-free rate. The data on portfolio returns, risk-free rate, and market factor are taken from Kenneth French's website.[41]

The conditioning information variables are from Goyal & Welch (2008), who provide detailed descriptions of the data and their sources. The dataset includes 14 variables considered relevant in predicting equity premium in past empirical studies.[42] These variables include stock characteristics (the dividend yield (DY), the dividend-price ratio (DP), the dividend-payout ratio (DE), the earning-price ratio (EP), the book-to-market ratio (BM), the net equity expansion (NTIS), and the stock variance (SVAR)), interest rate related variables (the Treasury bill rate (TBL), the long-term return (LTR), the long-term yield (LTY), the term spread (TMS), the defaults-return spread (DFR), and the default-yield spread (DFY)), and inflation (INFL) to represent the macroeconomy. We use monthly data for all these variables spanning from July 1926 to December 2018.

3.5.2 Benchmark Models

To analyse the performance of the various CI, CF, and hybrid of CI and CF approaches relative to the standard asset pricing models, we consider three standard conditional CAPM and two multi-factor models. Standard CCAPM approaches include models where beta dynamics is captured through: i) 60-month rolling window (Fama & MacBeth 1973), ii) short window (Lewellen & Nagel 2006), and iii) predetermined conditioning variable (Ferson & Harvey 1999). Multi-factor models include: i) the Fama & French (1993) three-factor model, and ii) the four-factor model of Carhart (1997). For all these models, we use Fama & MacBeth (1973) two-stage regressions,

[41] We use updated version of dataset available at
https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[42] We use an updated version of the dataset available at http://www.hec.unil.ch/agoyal/

where time-series regressions are estimated using a 60-month rolling window. See Chapter 2 for details.

3.6 Empirical Results

In this section, we discuss the empirical results of the various CI, CF and hybrid of CI and CF approaches applied to the CCAPM. This section consists of two subsections: *(i)* cross-sectional results of approaches considered to address variable-selection uncertainty, *(ii)* comparison of results from *(i)* with benchmark models.

3.6.1 Cross-sectional Results for CI, CF, and Combining CI and CF approaches

In this section, we discuss the results of the cross-sectional tests of the CCAPM based on various approaches to address variable-selection uncertainty (VSU), applied to CCAPM-IV, which include: i) combining information (CI), ii) combining forecasts (CF), and iii) combining CI and CF.

Our CCAPM cross-sectional tests are based on mainstream literature that evaluates the pricing abilities of a given model by looking at the significance of Fama & MacBeth (1973) parameter estimates.[43] The second category of results accounts for the explanation of size, value, and momentum anomalies which specifically tests whether a given model has already explained the three anomalies in the first pass regressions. This means that the loadings of the size, value, and momentum anomalies on net returns from the two factors given in equation (3.11) should not be priced.[44] More than one model can produce insignificant loadings, so to compare the models, we use adjusted R^2. A high value of adjusted R^2 indicates that the three factors can explain the large portion of three anomalies, which means CCAPM in the first pass has not explained these anomalies. This means that a lower value of adjusted R^2 indicates better performance. In addition, we assess the performance of models through SSPE, RMSPE, cross-sectional adjusted R^2, a number of mispriced assets at 1% and 5%, joint alpha tests which include JA and CPE, the magnitude of pricing error tests consisting of PEM and HJ (see section (3.3.3.3) for details). The test assets are the 25 assets sorted by size and book-to-market ratio. The conditional betas are estimated with a 60-month rolling window, and the out-of-sample period is August 1936 to December 2018.[45]

3.6.1.1 Results from CI and CF approaches

In this section, we discuss and compare the results of the CI, and CF approaches. Panel A of Table 3.1 presents the results from CI approaches, including subset variable selection (A.1) and dimension reduction approaches (A.2). In Chapter 2, we present a detailed analysis of subset variable selection approaches, including the best subset selection, stepwise selection, and shrinkage methods. Under the best subset variable selection, we use adjusted R^2, Akaike information criterion (Akaike 1973), Bayesian information criterion (Schwarz 1978), Mallows's C_P (Mallows 1973) to choose the best model. Our stepwise selection approaches include forward selection, backward elimination, and stepwise regression. The shrinkage methods include the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996), Adaptive LASSO

[43] We use t-statistics to test the significance using Newey & West (1987) heteroskedasticity and autocorrelation consistent standard errors.

[44] Net represents indicate the returns not explained by the conditional beta given as $R^*_{i,t+1} = R_{i,t+1} - (R_{m,t+1}\beta^C_{i,t})$, where $R_{i,t}$ is excess return on portfolio i at time t, $\beta^C_{i,t}$ is obtained either through CI, CF, or hybrid of CI and CF model and $R_{m,t+1}$ is realised risk premium at time $t+1$.

[45] Note that our sample starts from July 1926 to December 2018. The in-sample period is ten years because some methods use five years of data as a validation sample. To compare all the models, our out-of-sample analysis is based on August 1936 to December 2018.

(Zou 2006), and Elastic Net (ENet) (Zou & Hastie 2005). In addition, we also consider a new approach that selects the beta models that perform the best based on standard asset pricing criteria on past data at each point in time. We call this approach a dynamically selected beta model (DSBM). Results show that the DSBM approach, where each model is selected based on MSFE in a cross-validation sample, outperforms all the DMS approaches. Considering that we use the same data in our analysis, more specifically, the sample, test assets, and conditioning variables are the same used in Chapter 2. We only report the results of the best performing model DSM-III ($MSFE - CV$), which achieves an adjusted R^2 of 39.84%.

Panel A.2 in Table 3.1 presents the results for dimension reduction approaches. We find that similar to variable selection approaches in Chapter 2, all the dimension reduction approaches fail to produce an insignificant excess zero-beta rate. Moreover, none of the methods can explain the value premium as the loadings on the HML factors are priced in the cross-section of unexplained asset returns – significant at 5% level. If we compare the two standard DR approaches of principal component regression (PCR) and Kelly & Pruitt (2013) three-pass filter (KP3F), which is based on partial least squares (PLS), we find that KP3F outperforms the PCR by achieving an adjusted R^2 of 33.81%. PCR, on the other hand, achieves an adjusted R^2 of 30.86%. KP3F also outperforms PCR in other metrics such as SSPE and RMSPE by achieving lower values, implying lower pricing errors for KP3F. However, the PCR of cross-sectional beta premium and the supervised PCR – targeted PCR with asset pricing criteria – perform better than other approaches by achieving an adjusted R^2 of 46.46% and 41.52%, respectively. If we compare these results with those obtained through the best model selection approach DSM-III ($MSFE - CV$) given in Panel A.1 of Table 3.1, we find that the dimension reduction methods outperform. These results are consistent with studies such as Bai & Ng (2008), Ajana et al. (2019), and others who find that predictors have a natural grouping structure and dimension reduction techniques are more effective compared to variable selection approaches, including shrinkage approaches for predicting a given target variable accurately. Moreover, consistent with Tu & Lee (2019), we find the importance of supervised factor models where the relationship between the target variable and predictor variables is considered. However, the benefits of supervised factors when the relationship is viewed from the perspective of asset pricing are important, as the best dimension reduction approaches consider the target-predictor relationship from the perspective of conditional betas, which is consistent with our previous findings in Chapter 2.

[Insert Table 3.1 about here]

Panel B in Table 3.1 shows the cross-sectional results of all the combining forecasts (CF) approaches. Consistent with dynamic variable selection and dimension reduction approaches, all the CF approaches produce significant excess zero-beta rate and fail to price value premium. Moreover, joint alpha tests based on JA and CPE suggest all the CF models reject the null hypothesis that all the pricing errors are jointly zero. However, we can compare the various CF approaches based on multiple performance metrics. The simplest of all the CF approaches is the EW (mean) forecast given in Panel B.1 of Table 3.1. The results show that it achieves an adjusted R^2 of 42.66%, which can be used to compare it to other models. This implies that even a simple average forecast outperforms combining information (CI) approaches, such as the variable-selection approaches discussed in Chapter 2 and the dimension reduction methods presented in Panel A of Table 3.1. However, results suggest that the value-weighted (VW) methods, in which weights are based on individual model performance in a cross-validation (CV) sample, perform marginally better than the equally-weighted (EW) forecasts. The best performing models include the discounted MSFE (B.6 DMSFE-CV) with a discount factor of 0.9 ($\theta = 0.9$) and Robust Rank based on MSFE-CV (B.7) models, which achieve an adjusted R^2 of 47.08%

and 44.65%, respectively, which is slightly better than EW which achieves an adjusted R^2 of 42.66%.[46] The important finding is that the weights based on CV performance and the consideration of recent forecast performance (discounted forecast errors) improves the forecast accuracy. These two findings are in line with many forecasting studies such as Rapach et al. (2010) and others.

3.6.1.2 Results from hybrid of CI and CF approaches

Next, we discuss the cross-sectional results of approaches that are based on combining CI and CF methods. The first six models (A.1 to A.6) of Table 3.2 represent the results from pooling approaches. Pooling is the strategy where, rather than combining all the forecasts, a subset of forecasts is first selected then combined based on a simple average. Note that our pooling methods (A.1 to A.6) are based on $MSFE^{CV} - APC$ criteria.[47] The results show that our method of auto-combining (A.6), where we optimally select the best equally weighted forecast from available K clusters, perform the best among all the pooling strategies by achieving an adjusted R^2 of 48.90%. Moreover, the model not only explains the size and value anomalies but also produces the smallest joint pricing errors, which are insignificant at 5% level. These results are fascinating because our model (A.6) not only significantly outperforms the existing traditional pooling approaches but also the standard CF approaches given in Table 3.1.[48]

[Insert Table 3.2 about here]

There are two main takeaways from these results. First, we provide evidence that supports the idea of combining a subset of the available predictions rather than all of them (see Timmermann 2006, Aiolfi & Timmermann 2006, Kourentzes et al. 2019). Second, the optimal pooling based on $MSFE^{CV}$ APC performs better than the arbitrary choice for the number of forecasts. The method works better than the forecast island of Kourentzes et al. (2019) because our approach is based on evaluating both the individual forecast performance and all possible K clusters by adding one forecast at a time to existing forecasts.[49] In other words, it optimally chooses the best forecast to combine between the best univariate and the CF of all available forecasts. This flexibility enables the model to choose the forecasts that improves the forecast accuracy optimally.

The next two results (A.7 and A.8) given in Table 3.2 are based on CF using principal components. The first model, CF-PC (A.7) uses the PCs of forecasting returns ($\hat{R}_{t+1} = \hat{\beta}_t \,.\, R_{m,t+1}$) in Granger & Ramanathan (1984) regression analysis whereas the CF Univariate PCR (A.8) combines the individual forecasts of the P selected PCs.[50] Results show that CF-PC (A.7) and CF Univariate PCR (A.8) achieve an adjusted R^2 of 45.02% and 30.28%, respectively. The main comparison of these two models would be with their counterparts in CI and CF approaches. CF-PC (A.7) is basically used to improve the performance of standard CF approaches such as simple average. Our results are consistent with those of Stock & Watson (2004), Huang & Lee (2010), and Tu & Lee (2019) that the CF-PC (A.7) outperforms the standard CF approaches by achieving significantly higher adjusted R^2.[51] On the other hand, the comparison between

[46] The difference of adjusted R^2 is significant at 5% level using bootstrapped p-values.

[47] APC indicates the asset pricing criteria, conditional pricing errors, $\hat{\alpha}_{t+1} = R_{t+1} - \hat{\beta}_t . R_{m,t+1}$

[48] Based on the bootstrapped p-values for the difference between adjusted R^2.

[49] The forecast island of Kourentzes et al. (2019) is based on the outlier detection method (box-plot) and automatically identifies the pools of forecasts without relying on any performance metric. Instead, the performance metric is used only to sort the forecasts and not for identifying the optimal forecast to combine.

[50] Note that A.7 constructs the factors from returns ($\hat{R}$), whereas A.8 constructs the factors from predictors' data (X)

[51] At 5% bootstrapped p-values.

CF Univariate PCR (A.8 in Table 3.2) and PCR approach given in Table 3.1 show that there is no significant difference between the performance of these two models.[52]

The next approach (A.9 in Table 3.2) is combining forecasts (CF) based on cross-sectional beta premium (CSBP).[53] Under this approach, rather than using individual predictor to generate the beta forecast, we first use each predictor to generate the time-series of CSBP and then use lagged CSBP instead of the lagged predictor (see section 3.4.3.4). Results show that the model CF-CSBP outperforms all the models discussed so far by achieving an adjusted R^2 of 50.18%. Though the model produces a significant excess zero-beta rate but the estimated risk-premium on loadings on size and value factors anomalies are insignificant. This suggests that the model can explain the size and value anomalies for 25 size and book-to-market portfolios. These results are supported by the fact that the joint alpha tests show that the pricing errors produced by the $CF - CSBP$ model are jointly zero.

The final model in combining CI-CF approaches is the pooling through shrinkage method (A.10 in Table 3.2) recently used by Rapach & Zhou (2020) in forecasting equity premium. Under this method, we first identify the relevant predictors through ENet, which are then used in a univariate regression model to produce individual forecasts. Finally, we take the simple average to construct the final combined forecast. The results show that the CF-ENet model does not perform well as it only achieves an adjusted R^2 of 24.71% which is significantly lower than combining all individual forecasts. This supports our earlier findings that shrinkage methods perform poorly in variable selection approaches because they depend on residuals of time-series regression instead of conditional pricing errors. This evidence is further supported by the fact that the pooling strategies based on conditional pricing errors given in Table 3.2 (Models A.1, A.4, A.5, and A.6) outperform the CF-ENet model by producing significantly higher adjusted R^2.

3.6.1.3 Results from Bootstrap Aggregation (Bagging) approaches

Finally, we present the results with bootstrap aggregation (bagging) methods.[54] Bootstrap aggregating or bagging (Breiman 1996) is a common strategy to overcome the model uncertainty. The bagging estimate is calculated by the mean of the bootstrap samples (see section (3.4.3.5.2) for details). Panel A in Table 3.3 presents the results of model selection approaches using Bagging strategies where first three models A.1 to A.3 are based on the standard approach of bagging (see Inoue & Kilian 2008, Rapach & Strauss 2010, Borup & Schütte 2020) where we first apply a hard-threshold on the variables in X_t, such that only the statistically significant variables based t-statistics at 5% significance level remain.[55] The subset variables, $X_t^* \subset X_t$, only include X_j if $\mid t_{X_j} \mid > 1.96$. See section (3.4.1.1.2.1) for details on hard-thresholding.

The main difference between three approaches (A.1, A.2, and A.3) is that the A.1 is based on the direct relationship between R_i and X_j where the t-statistics on X_j of model, $R_{i,t+1} = a_0 + (\gamma_0 + \gamma_{1,j} X_{j,t}) R_{m,t+1} + e_{i,t+1}$, is considered. On the other hand, the models A.2 (BAGG-APC1) and A.3 (BAGG-APC2) are based on asset pricing criteria discussed in section (3.4.1.1.2.3) to preselect the predictors. In model A.2, we use t-statistics of cross-sectional zero-beta rate of

[52] Note that both these methods are unsupervised – does not consider the target variable. In unreported results, we compare the performance of supervised PCR with supervised CF univariate PCR where the factors are constructed from targeted predictors. Results were consistent with unsupervised PCR – no significant difference.

[53] To account for the criticism of Lewellen et al. (2010), who claim that the cross-sectional results can vary depending on the choice of test assets, we include diverse assets where in addition to 25 size-B/M portfolios, we include 10 momentum and 30 industry portfolios as test assets.

[54] We use bagging for all the methods, including subset variable selection, shrinkage, combining forecasts, and combining CI and CF. However, we only present the key results of bagging in Table 3.3.

[55] Following Rapach & Strauss (2010), we use Newey & West (1987) standard errors.

each predictor j, whereas in A.3 we use cross-sectional adjusted R^2 as our asset pricing criteria where first we get the adjusted R_j^2 for each predictor j. Next, we sort the adjusted R_j^2 from high to low values and select first P models as significant predictors used to estimate betas in each sample. Here P indicates the number of predictors in X_t^* selected by standard method (A.1), $X_t^* \subset X_t$, and $\{X_j \in X_t^* \mid |t_{X_j}| > 1.96\}$. This way we can ensure that both methods standard (A.1) and with asset pricing criteria (A.3) use the same number of variables in each pseudo sample but the only difference is the way we select these predictors.

The results show that the BAGG-APC2 performs best among these three methods by achieving an adjusted R^2 of 45.64%.[56] The improved performance over A.1 is consistent with previous findings that conditional beta-based performance measures outperform criteria based on a direct relationship between return and predictors. However, the significant difference between the performance of A.2 and A.3 is a surprising result. On exploring why A.3 outperforms A.2, we find that most univariate predictor-based models generate a significant excess zero-beta rate. In that case, only a single predictor with minimum t-statistics is selected. This makes the model univariate, i.e. based on a single best predictor at each period. More specifically, under A.2 on average, about 80% of times, the out-of-sample forecasts are based, univariate model. This makes other $K-1$ predictors irrelevant when they may have a significant role to play in the beta forecast. This also adds to the evidence that bagging is robust, as the average across pseudo sample can only increase forecasting performance if the underlying variable selection methods and criteria are chosen correctly.

The standard Bagging approaches are based on univariate model selection. Next, we apply Bagging to subset variable selection approaches.[57] Model A.4 (BAGG-MSFE-CV (APC)) in Panel A of Table 3.3 indicates the model where subset variable selection is based on MSFE in CV sample. We see a significant improvement in the performance as adjusted R^2 goes to 52.13% from 39.84% when estimated with original data only. Moreover, the estimated risk-premium on loadings on size and value factors anomalies are also insignificant, suggesting that after accounting for model uncertainty through bagging, the model can explain the size and value anomalies for 25 size and book-to-market portfolios. These results are supported by the fact that the joint alpha tests show that the pricing errors produced by $MSFE^{CV}$ APC model are jointly zero. We also find that the results with standard model selection approaches are not improved even after accounting for model uncertainty.[58] This supports our previous findings that standard methods are based on time-series regression residuals rather than pricing errors, and that a model chosen based on less appropriate criteria cannot improve even when estimating in multiple samples.

[Insert Table 3.3 about here]

Next, we apply the bagging strategy to dimension reduction (DR) models. We find that, with the exception of targeted PCR (APC), there is no significant difference in adjusted R^2 between standard dimension reduction methods and the same methods estimated using bagging. The results of Model B.1 (BAGG-PCR (APC)) in Table 3.3 show that the adjusted R^2 of the targeted PCR where factors are selected using asset pricing criteria goes to 45.51% from 41.52%. The main reason for this is that DR methods already account for model uncertainty by combining information from multiple predictors. Moreover, we test whether the average out-of-sample beta with original data is significantly different than the average beta estimated with bagging, and we do not find any difference between the two at 5% significance level. Panel C in Table 3.3 presents

[56] The difference between adjusted R^2 is significant at 5% level based on bootstrapped p-values.

[57] We only present results for MSFE-APC (CV) approaches as bagging improves this model's results significantly.

[58] For brevity, the results are not reported.

the CF results when CF betas under each method are estimated as an average across B bootstrap samples. This is surprising to see that bagging does not make significant improvements in the adjusted R^2 of standard CF methods. The reason behind this improvement is the fact that in simple average methods, in particular, EW, the model uncertainty is already addressed (e.g., Hendry & Clements 2004) Timmermann (2006). Panel D in Table 3.3 represent the results when bagging is applied to various combining CI and CF strategies given in Table 3.2. Results show that bagging improves the performance of most of the models. The top three models with significant improvement include the BAGG-Pool (Auto Combine) (D.1), BAGG-CF-CSBP (D.2) and BAGG-CF PCR (D.3), which achieve an adjusted R^2 of 52.94%, 52.10%, and 49.98%, respectively. This shows that bagging helps in overcoming the problem of model uncertainty related to subset variable selection for pooling, model D.1 in this case. On the other hand, the improvements in results of CF-PC are also consistent with the findings of (Yang et al. 2017), who find that the bagging applied to the principal component (PC) combination approaches helps in lowering the MSFE compared to the simple average.

To summarise our results on CI, CF, and combination of CI and CF approaches, we find that three top-performing models include BAGG–Pool (Auto Combine using $MSFE^{CV}$ APC), BAGG–$MSFE^{CV}$ APC, and BAGG–CF-CSBP with an adjusted R^2 of 52.94%, 52.13%, and 52.10%, respectively.[59] However, the difference between the adjusted R^2 of these models is insignificant, which means that these models' performance is almost identical. The choice of the model depends upon the complexity of the model. More complex models should be excluded. Given this, we choose the model BAGG–Pool (Auto Combine using $MSFE^{CV}$ APC) because it is based on the combination of various standard approaches. We name this method as BAPCAM, **B**agging of **A**uto combined **P**ooling based on **C**ross validation sample using **A**sset pricing criterion of **MSFE**. This is a summary of the different methods for accounting for variable-selection uncertainty (VSU). The best-performing strategy is one that combines a subset of forecasts rather than all, ii) uses automated quantile selection rather than arbitrarily selected by the researcher, iii) uses bagging to resolve variable (forecast) selection uncertainty, iv) uses a validation sample to test the performance of a given model, and v) uses the correct performance criterion.

3.6.2 Comparing Results with Benchmark Models

In this section, we provide a direct comparison of the best performing approach, BAPCAM, to the various standard CCAPM and multi-factor models. Standard CCAPM approaches include models where beta dynamics is captured through: i) 60-month rolling window (Fama & MacBeth 1973), ii) short window (Lewellen & Nagel 2006), iii) four standard variables (Ferson & Harvey 1999), and iv) kitchen sink model containing all variables. Multi-factor models include: i) Fama & French (1993) three-factor model, and ii) the four-factor model of Carhart (1997). See section (3.5.2) for more details on benchmark models.

The results of the Fama & MacBeth (1973) cross-sectional regressions presented in Table 3.4 show that consistent with our findings for CI, CF, and hybrid of CI and CF approaches, the constant (excess zero-beta rate) is significantly different from zero for all the models including BAPCAM and multi-factor models including Fama & French (1993) three-factor model and Carhart (1997) four-factor model, even though all the models are estimated on excess returns.[60] In this case, the theory behind all the models would predict the constant to be zero. This implies

[59] Note that before selecting these models as the best performing models, we first perform all the robustness tests discussed in section (3.7) and find that these methods consistently outperform other methods.

[60] Note that the models are estimated with a constant (excess zero-beta rate). In unreported results, we also estimate the model without constant assuming that $R_{zr} = R_f$ and find consistent results, i.e. same ranking of models.

that there are other factors not included in the model can explain the cross-section of returns.

However, results show that our BAPCAM outperforms the CCAPM benchmark models presented in Panel A of Table 3.4. The CCAPM benchmark models generally do a poor job. Starting with the excess zero-beta rate, all the benchmark models produce a significant zero-beta rate with a minimum of 0.820 achieved by CAPM (β_{RW}). However, BAPCAM produces a lower excess zero-beta rate compared to benchmark models with 0.412. The differences between the excess zero-beta rates of BAPCAM and CCAPM benchmark models are significant at 5% level. Next, we can see that the estimated risk premium for all the CCAPM benchmark models is negative. However, BAPCAM produces a positive risk premium of 0.472. In terms of asset pricing performance based on other metrics, our BAPCAM achieves an adjusted R^2 of 52.21%, which significantly higher than the best performing CCAPM benchmark, CAPM (β_{RW}), which achieves adjusted R^2 of 25.3%. Moreover, compared to our BAPCAM model, all the CCAPM models not only produce significantly larger errors measured by SSPE and RMSPE but also reject the null hypothesis of producing zero pricing errors ($H_0 : \bar{\hat{\alpha}} = 0$) for all the portfolios measured by measures of joint alpha (JA) and composite pricing errors (CPE). There are two main conclusions from these findings. First, conditioning information can play a significant role in capturing the beta dynamics compared to a model where beta is a function of time, i.e. using rolling window (β_{RW}) or short window (β_{SW}) to model time-variation in betas. Second, time-varying conditioning information based changing set of predictors is key to the better performance of CCAPM with IV framework.[61]

[Insert Table 3.4 about here]

We can also examine the pricing performance of each model by plotting the average monthly fitted excess return and the average monthly realised excess return for each asset. If a model perfectly fits the returns, then the test assets should lie on the 45^0 line in Figure 3.3. For a given test asset, the coordinates below (above) the 45^0 line correspond to the negative (positive) pricing errors, $\hat{\alpha}_i$. The Figure 3.3 provides a graphical illustration of the findings presented in Table 3.4. The CCAPM benchmark models fit the data poorly as there is a substantial deviation of assets returns from the 45^0 line. In contrast, the BAPCAM has asset returns relatively close to the 45^0 line. However, dispersion is still high, suggesting that even after optimally combining the beta forecasts, conditional CAPM does not fully explain the cross-section of asset returns for size and book-to-market sorted portfolios. On the other hand, if we compare our BAPCAM model's performance with FF3F and Carhart models, we see that asset returns are relatively close to the 45^0 line, which implies that factor models fit the data well compared to all given models.

[Insert Figure 3.3 about here]

Panel B in Table 3.4 reports the results of Fama & French (1993) three-factor model (FF3F) and Carhart (1997) four-factor model (Carhart). Results show that factor models produce a negative risk premium ($\bar{\hat{\lambda}}_{MKT}$) with -0.3113 and -0.2785 achieved by FF3F and Carhart, respectively. However, the HML factor is highly significant and positive, and the SMB factor is positive but significant at 10% level for both models. On comparison of the performance of the BAPCAM and factor models, we find that consistent with findings through graphical illustration

[61] Note that the most of CI, CF, and combining CI and CF (Table 3.2) approaches outperform these CCAPM benchmark models. This shows that after accounting for variable-selection uncertainty in CCAPM-IV approach, we can improve the performance.

in Figure 3.3 the FF3F (Carhart) performs better than our BAPCAM by achieving an adjusted R^2 of 72.10% (77.26%) which is about 20% (25%) higher than our BAPCAM model for FF3F (Carhart) model. This finding, however, should be interpreted with caution for at least three points. First, the high explanatory power of multi-factor models may simply be attributed to the extra degrees of freedom; the FF3F (Carhart) model has three (four) parameters that can be freely estimated (Campbell & Vuolteenaho 2004). Second, since the excess return on the zero-beta portfolio is substantial and high in magnitude (i.e. $R_{zb} - R_f \neq 0$), the multi-factor models do not handle the equity premium adequately. Third, the multi-factor models model, like the other models, yield significant pricing errors measured as JA and CPE.

In conclusion, our results are mixed, suggesting that none of the asset pricing models analysed perfectly describes the cross-section of realised returns. However, we consider the relative performance of our BAPCAM to be a success based on a single factor.

3.6.2.1 Sources of improvement

In this section, we identify the source of improvement for BAPCAM relative to benchmark models. Table 3.5 presents the pricing errors for each size and book-to-market quintile measured as the sum of squared pricing errors (SSPE). Consistent with past studies, results suggest that the large pricing errors for the CAPM (β_{RW}) are clustered in the small, big, growth and value quintiles. The performance of other benchmark models is almost similar. However, when we compare the performance of BAPCAM with the CCPAM benchmark models, we find an average reduction of 19.6%, 80.4%, 45.8%, 32.9% in SSPE of small, big, growth, and value quintiles, respectively. Therefore, the rationale behind our BAPCAM's empirical success compared to benchmark models in pricing size and B/M sorted portfolios is as follows. Take the value premium into consideration. Since value stocks have high average returns but low expected betas, CAPM fails to price value stocks in an OLS setting. However, according to our BAPCAM approach, the systematic risk tracks the business cycle, and on average, beta forecasts are higher than OLS betas. As a result, a high beta calculation is combined with high expected returns, lowering the expected pricing errors of "Value" stocks. "Small" stocks, which have experienced a reduction in beta, have a similar intuition. The opposite occurs for the "Big" and "Growth" stocks.

[Insert Table 3.5 about here]

3.6.2.2 Which predictors matter?

In this section, we investigate the relative importance of individual predictors in BAPCAM betas. We follow Gu et al. (2020) and calculate the reduction in out-of-sample cross-sectional adjusted R^2 by excluding a particular predictor. Given K predictors, the importance measure can be given as:

$$\Delta R^2{}_i = (R^2_K - R^2_{K-1}) \tag{3.88}$$

where $\Delta R^2{}_i$ is reduction in cross-sectional adjusted R^2 due to absence of ith predictor, R^2_K is adjusted R^2 using all predictors and R^2_{K-1} is adjusted R^2 without ith predictor. Now the importance factor can be calculated as:

$$\Phi_{1i} = \frac{\Delta R^2_{OOSi}}{\sum_{i=1}^{K} \Delta R^2_{OOSi}} \tag{3.89}$$

where Φ_i is the predictor importance of each predictor which is normalised to sum to one.

Figure 3.4 plots the R^2 based predictor importance where we can see that the variable EP is the most influential variable as the absence of this variable from the dataset contributes 19.2% in an overall reduction in cross-sectional adjusted R^2. The variables INFL and TMS stand second and third with 16.4% and 14.2%, respectively. This shows that the variable EP plays a significant part in explaining the value premium. Inflation (INFL) and term-spread (TMS) are two other essential variables. The importance of TMS and INFL indicate the connection to macroeconomic activity as (Stock & Watson 2003) and others show that TMS, a commonly used stock market return predictor, has strong predictive power for aggregate output. Since most of the predictors are highly correlated, so we classify them: i) equity - which includes DP, DY, EP, DE, BM and SVAR, and ii) non-equity - which includes DFR, DFY, LTY, LTR, TBL, TMS, INFL. If we sum individual contributions across each group, then we find that the non-equity group plays a more significant role with 60% contribution in the importance measure. Whereas four variables EP, INFL TMS and DP contribute more than 60%. This also provides evidence on how macroeconomic variables and financial variables complement each other.

We can link these predictor contributions with the findings on the improvement in the cross-sectional performance discussed in the previous section, which shows that the improvement compared to benchmark models is mainly due to improvements in beta estimates of value stocks. To do this, we compute the correlation between conditional betas of value premium and state variables and find a significant and negative correlation of the conditional beta of the value premium with INFL and PE and significant negative correlation with TMS and thus changes countercyclically across time.[62] By identifying EP as an important state variable, we confirm that the stock price indeed has essential effects on the dynamic of the value premium's conditional market beta. Moreover, because inflation is one of the most closely watched gauges of aggregate economic activity by the Federal Reserve, our results confirm a vital implication of the investment-based asset pricing models, (e.g., Zhang 2005), that the value premium's conditional market beta changes with business conditions.

[Insert Figure 3.4 about here]

3.7 Robustness Tests

We test the robustness of our key findings in a number of different ways. First, we look at different sample dates and estimation windows and find similar results in all alternative samples and different estimation windows. Second, we allow alphas to be time-varying following Christopherson et al. (1998), Ferson & Harvey (1999), among others. More specifically, we redo the asset pricing tests by using equation, $R_{i,t+1} = a_{0,i} + a'_{1,1}X_t + (\gamma_{0,i} + \gamma'_{1,i}X_t)R_{m,t+1} + \varepsilon_{i,t+1}$. Results show that the findings are very similar to the corresponding results in the benchmark tests. These results suggest that allowing alphas to be time-varying has a negligible impact on our main results.

Third, we examine the models' success in explaining cross-sectional variance in excess monthly returns using additional test portfolios, including 10 momentum sorted portfolios and 30 industry portfolios, as well as 25 size and momentum cross sorted portfolios. We find no significant change in our conclusions from the benchmark tests. Finally, following Boguth et al. (2011), we include 1-, 6-, and 36-month lagged realised beta as candidate conditioning variables. More specifically, we estimate each asset's realised betas either with 1 month or 6 months of past daily returns implementing Dimson (1979) correction for infrequent trading, or 36 months of

[62]Following Lewellen & Nagel (2006), we define value premium as the the difference between "Value" and "Growth" (VMG). We take the average of the five low-B/M (low-B/M) portfolios for "Growth" ("Value".).

past monthly returns. The results show that the inclusion of the realised betas as conditioning variables does not change our main findings in any qualitative manner. For brevity, we do not report the results of all these tests.

[Insert Table 3.6 about here]

3.8 Conclusion

In this chapter, we apply and compare various methods to address the issue of variable-selection uncertainty (VSU), applied to conditional CAPM (CCAPM). This chapter contributes to the existing literature in many ways. First, to our knowledge, this is the first study to include a detailed comparison of various well-known approaches to dealing with VSU from a CCAPM perspective. We find that CF approaches dominate the CI approaches in explaining the cross-section of assets returns measured as adjusted R^2. The better performance of CF is consistent with the past studies that claim that CF not only helps to reduce forecast errors but also helps to mitigate model uncertainty (Bates & Granger 1969, Timmermann 2006, Rapach et al. 2010). However, hybrid models that combine CF and CI methods improve the performance of CCAPM beyond CF. These results are consistent with Huang & Lee (2010), Kourentzes et al. (2019), Rapach & Zhou (2020), and others that show the superior performance of hybrid approaches compared to individual CI and CF approaches. Moreover, such studies claim that hybrid approaches help to reduce variable-selection uncertainty (VSU). Thus, the reduction in VSU, we believe, is the primary reason for the improved results of CCAPM-IV.

Moreover, we find that the best performing models include hybrid models where CF is combined with machine learning methods such as bagging. This is in line with studies that suggest that bagging helps reduce variance without raising forecast bias (Hastie et al. 2009) and handle the data, model and parameter uncertainty simultaneously (Petropoulos et al. 2018), which has been shown to increase forecasting accuracy (Rapach & Strauss 2010, Liu & Xie 2019, Borup & Schütte 2020).

References

Adrian, T. & Franzoni, F. (2009), 'Learning about beta: Time-varying factor loadings, expected returns, and the conditional capm', *Journal of Empirical Finance* **16**(4), 537–556.

Adrian, T. & Rosenberg, J. (2008), 'Stock returns and volatility: Pricing the short-run and long-run components of market risk', *The Journal of Finance* **63**(6), 2997–3030.

Aiolfi, M. & Favero, C. A. (2005), 'Model uncertainty, thick modelling and the predictability of stock returns', *Journal of Forecasting* **24**(4), 233–254.

Aiolfi, M. & Timmermann, A. (2006), 'Persistence in forecasting performance and conditional combination strategies', *Journal of Econometrics* **135**(1-2), 31–53.

Ajana, S., Acar, N., Bretillon, L., Hejblum, B. P., Jacqmin-Gadda, H. & Delcourt, C. (2019), 'Benefits of dimension reduction in penalized regression methods for high-dimensional grouped data: a case study in low sample size', *Bioinformatics* **35**(19), 3628–3634.

Akaike, H. (1973), Information theory and an extension of maximum likelihood principle, *in* 'Proceedings of Second International Symposium on Information Theory', pp. 267–281.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Andronoudis, D., Dargenidou, C., Konstantinidi, E. & Pope, P. F. (2019), 'Conservative accounting and risk: The case of research and development', *Available at SSRN 3006537* .

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. & Ng, S. (2008), 'Forecasting economic time series using targeted predictors', *Journal of Econometrics* **146**(2), 304–317.

Bair, E., Hastie, T., Paul, D. & Tibshirani, R. (2006), 'Prediction by supervised principal components', *Journal of the American Statistical Association* **101**(473), 119–137.

Bali, T. G. & Engle, R. F. (2010), 'Resurrecting the conditional capm with dynamic conditional correlations', *Available at SSRN 1298633* .

Barnard, G. A. (1963), 'New methods of quality control', *Journal of the Royal Statistical Society. Series A (General)* **126**(2), 255–258.

Barrow, D. K. & Kourentzes, N. (2016), 'Distributions of forecasting errors of forecast combinations: implications for inventory management', *International Journal of Production Economics* **177**, 24–33.

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Bühlmann, P. & Yu, B. (2002), 'Analyzing bagging', *The Annals of Statistics* **30**(4), 927–961.

Bjørnland, H. C., Gerdrup, K., Jore, A. S., Smith, C. & Thorsrud, L. A. (2012), 'Does forecast combination improve norges bank inflation forecasts?', *Oxford Bulletin of Economics and Statistics* **74**(2), 163–179.

Black, F. (1972), 'Capital market equilibrium with restricted borrowing', *The Journal of Business* **45**(3), 444–455.

Boguth, O., Carlson, M., Fisher, A. & Simutin, M. (2011), 'Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas', *Journal of Financial Economics* **102**(2), 363–389.

Boivin, J. & Ng, S. (2006), 'Are more data always better for factor analysis?', *Journal of Econometrics* **132**(1), 169–194.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A capital asset pricing model with time-varying covariances', *Journal of Political Economy* **96**(1), 116–131.

Bordignon, S., Bunn, D. W., Lisi, F. & Nan, F. (2013), 'Combining day-ahead forecasts for british electricity prices', *Energy Economics* **35**, 88–103.

Borup, D. & Schütte, E. C. M. (2020), 'In search of a job: Forecasting employment growth using google trends', *Journal of Business and Economic Statistics, Forthcomning* .

Breiman, L. (1996), 'Bagging predictors', *Machine Learning* **24**(2), 123–140.

Campbell, J. Y. & Vuolteenaho, T. (2004), 'Bad beta, good beta', *American Economic Review* **94**(5), 1249–1275.

Cang, S. & Yu, H. (2014), 'A combination selection algorithm on forecasting', *European Journal of Operational Research* **234**(1), 127–139.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of Finance* **52**(1), 57–82.

Cederburg, S. & O'Doherty, M. S. (2016), 'Does it pay to bet against beta? on the conditional performance of the beta anomaly', *The Journal of Finance* **71**(2), 737–774.

Chan, C. K., Kingsman, B. G. & Wong, H. (1999), 'The value of combining forecasts in inventory management–a case study in banking', *European Journal of Operational Research* **117**(2), 199–210.

Christopherson, J. A., Ferson, W. E. & Glassman, D. A. (1998), 'Conditioning manager alphas on economic information: Another look at the persistence of performance', *The Review of Financial Studies* **11**(1), 111–142.

Clark, T. E. & McCracken, M. W. (2010), 'Averaging forecasts from vars with uncertain instabilities', *Journal of Applied Econometrics* **25**(1), 5–29.

Clemen, R. T. (1989), 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting* **5**(4), 559–583.

Cochrane, J. (2005), Financial markets and the real economy, Report, National Bureau of Economic Research.

Cochrane, J. H. (2009), *Asset pricing: Revised edition*, Princeton University Press.

Cooper, M. J. & Gubellini, S. (2011), 'The critical role of conditioning information in determining if value is really riskier than growth', *Journal of Empirical Finance* **18**(2), 289–305.

Dimson, E. (1979), 'Risk measurement when shares are subject to infrequent trading', *Journal of Financial Economics* **7**(2), 197–226.

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004), 'Least angle regression', *The Annals of Statistics* **32**(2), 407–499.

Elliott, G. & Timmermann, A. (2005), ‘Optimal forecast combination under regime switching’, *International Economic Review* **46**(4), 1081–1102.

Elliott, G. & Timmermann, A. (2016), ‘Forecasting in economics and finance’, *Annual Review of Economics* **8**, 81–110.

Engle, R. (2002), ‘Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models’, *Journal of Business and Economic Statistics* **20**(3), 339–350.

Fama, E. F. & French, K. R. (1993), ‘Common risk factors in the returns on stocks and bonds’, *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. & MacBeth, J. D. (1973), ‘Risk, return, and equilibrium: Empirical tests’, *Journal of Political Economy* **81**(3), 607–636.

Ferson, W. E. & Harvey, C. R. (1999), ‘Conditioning variables and the cross section of stock returns’, *The Journal of Finance* **54**(4), 1325–1360.

Ghysels, E. (1998), ‘On stable factor structures in the pricing of risk: do time-varying betas help or hurt?’, *The Journal of Finance* **53**(2), 549–573.

Goyal, A. & Welch, I. (2008), ‘A comprehensive look at the empirical performance of equity premium prediction’, *The Review of Financial Studies* **21**(4), 1455–1508.

Graefe, A., Armstrong, J. S., Jones Jr, R. J. & Cuzán, A. G. (2014), ‘Combining forecasts: An application to elections’, *International Journal of Forecasting* **30**(1), 43–54.

Granger, C. W. & Ramanathan, R. (1984), ‘Improved methods of combining forecasts’, *Journal of Forecasting* **3**(2), 197–204.

Gu, S., Kelly, B. & Xiu, D. (2020), ‘Empirical asset pricing via machine learning’, *The Review of Financial Studies* **33**(5), 2223–2273.

Hansen, L. P. & Jagannathan, R. (1997), ‘Assessing specification errors in stochastic discount factor models’, *The Journal of Finance* **52**(2), 557–590.

Harvey, C. R. (2001), ‘The specification of conditional expectations’, *Journal of Empirical Finance* **8**(5), 573–637.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The elements of statistical learning: data mining, inference, and prediction*, Springer Science and Business Media.

Hendry, D. F. & Clements, M. P. (2004), ‘Pooling of forecasts’, *The Econometrics Journal* **7**(1), 1–31.

Hibon, M. & Evgeniou, T. (2005), ‘To combine or not to combine: selecting among forecasts and their combinations’, *International Journal of Forecasting* **21**(1), 15–24.

Hirano, K. & Wright, J. H. (2017), ‘Forecasting with model uncertainty: Representations and risk reduction’, *Econometrica* **85**(2), 617–643.

Huang, H. & Lee, T.-H. (2010), ‘To combine forecasts or to combine information?’, *Econometric Reviews* **29**(5-6), 534–570.

Inoue, A. & Kilian, L. (2008), ‘How useful is bagging in forecasting economic time series? a case study of us consumer price inflation’, *Journal of the American Statistical Association* **103**(482), 511–522.

Jagannathan, R. & Wang, Z. (1996), ‘The conditional capm and the cross-section of expected returns’, *The Journal of Finance* **51**(1), 3–53.

Jin, S., Su, L. & Ullah, A. (2014), ‘Robustify financial time series forecasting with bagging’, *Econometric Reviews* **33**(5-6), 575–605.

Jostova, G. & Philipov, A. (2005), ‘Bayesian analysis of stochastic betas’, *Journal of Financial and Quantitative Analysis* **40**(4), 747–778.

Kelly, B. & Pruitt, S. (2013), ‘Market expectations in the cross-section of present values’, *The Journal of Finance* **68**(5), 1721–1756.

Kelly, B. & Pruitt, S. (2015), ‘The three-pass regression filter: A new approach to forecasting using many predictors’, *Journal of Econometrics* **186**(2), 294–316.

Kolassa, S. (2011), ‘Combining exponential smoothing forecasts using akaike weights’, *International Journal of Forecasting* **27**(2), 238–251.

Kourentzes, N., Barrow, D. & Petropoulos, F. (2019), ‘Another look at forecast selection and combination: Evidence from forecast pooling’, *International Journal of Production Economics* **209**, 226–235.

Lettau, M. & Ludvigson, S. (2001), ‘Consumption, aggregate wealth, and expected stock returns’, *The Journal of Finance* **56**(3), 815–849.

Lewellen, J. & Nagel, S. (2006), ‘The conditional capm does not explain asset-pricing anomalies’, *Journal of Financial Economics* **82**(2), 289–314.

Lewellen, J., Nagel, S. & Shanken, J. (2010), ‘A skeptical appraisal of asset pricing tests’, *Journal of Financial Economics* **96**(2), 175–194.

Li, J. & Chen, W. (2014), ‘Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models’, *International Journal of Forecasting* **30**(4), 996–1015.

Liu, Y. & Xie, T. (2019), ‘Machine learning versus econometrics: prediction of box office’, *Applied Economics Letters* **26**(2), 124–130.

Ludvigson, S. C. & Ng, S. (2007), ‘The empirical risk–return relation: A factor analysis approach’, *Journal of Financial Economics* **83**(1), 171–222.

Mahmoud, E. (1989), ‘Combining forecasts: Some managerial issues’, *International Journal of Forecasting* **5**(4), 599–600.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018), ‘Statistical and machine learning forecasting methods: Concerns and ways forward’, *PLOS One* **13**(3), 1–26.

Makridakis, S. & Winkler, R. L. (1983), ‘Averages of forecasts: Some empirical results’, *Management Science* **29**(9), 987–996.

Mallows, C. (1973), ‘Some comments on cp’, *Technometrics* **15**(1), 661–675.

McCracken, M. W. & Ng, S. (2016), ‘Fred-md: A monthly database for macroeconomic research’, *Journal of Business and Economic Statistics* **34**(4), 574–589.

Neely, C. J., Rapach, D. E., Tu, J. & Zhou, G. (2014), ‘Forecasting the equity risk premium: the role of technical indicators’, *Management Science* **60**(7), 1772–1791.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Panopoulou, E. & Vrontos, S. (2015), 'Hedge fund return predictability; to combine forecasts or combine information?', *Journal of Banking and Finance* **56**, 103–122.

Pesaran, M. H., Pick, A. & Pranovich, M. (2013), 'Optimal forecasts in the presence of structural breaks', *Journal of Econometrics* **177**(2), 134–152.

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Petkova, R. & Zhang, L. (2005), 'Is value riskier than growth?', *Journal of Financial Economics* **78**(1), 187–202.

Petropoulos, F., Hyndman, R. J. & Bergmeir, C. (2018), 'Exploring the sources of uncertainty: Why does bagging for time series forecasting work?', *European Journal of Operational Research* **268**(2), 545–554.

Rapach, D. E. & Strauss, J. K. (2010), 'Bagging or combining (or both)? an analysis based on forecasting us employment growth', *Econometric Reviews* **29**(5-6), 511–533.

Rapach, D. E., Strauss, J. K. & 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.

Rapach, D. E. & Zhou, G. (2020), 'Time-series and cross-sectional stock return forecasting: New machine learning methods', *Machine Learning for Asset Management: New Developments and Financial Applications* **1**(1), 1–33.

Rapach, D. & Zhou, G. (2013), 'Forecasting stock returns', *Handbook of Economic Forecasting* **2**, 328–383.

Reeves, G. R. & Lawrence, K. D. (1982), 'Combining multiple forecasts given multiple objectives', *Journal of Forecasting* **1**(3), 271–279.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

Shanken, J. (1990), 'Intertemporal asset pricing: An empirical investigation', *Journal of Econometrics* **45**(1-2), 99–120.

Smith, J. & Wallis, K. F. (2009), 'A simple explanation of the forecast combination puzzle', *Oxford Bulletin of Economics and Statistics* **71**(3), 331–355.

Stock, J. H. & Watson, M. W. (2003), 'Forecasting output and inflation: The role of asset prices', *Journal of Economic Literature* **41**(3), 788–829.

Stock, J. H. & Watson, M. W. (2004), 'Combination forecasts of output growth in a seven-country data set', *Journal of Forecasting* **23**(6), 405–430.

Stock, J. H. & Watson, M. W. (2012), 'Generalized shrinkage methods for forecasting using many predictors', *Journal of Business and Economic Statistics* **30**(4), 481–493.

Tian, J. & Anderson, H. M. (2014), 'Forecast combinations under structural break uncertainty', *International Journal of Forecasting* **30**(1), 161–175.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58**(1), 267–288.

Timmermann, A. (2006), 'Forecast combinations', *Handbook of Economic Forecasting* **1**, 135–196.

Timmermann, A. (2018), 'Forecasting methods in finance', *Annual Review of Financial Economics* **10**, 449–479.

Tu, Y. & Lee, T.-H. (2019), 'Forecasting using supervised factor models', *Journal of Management Science and Engineering* **4**(1), 12–27.

Vendrame, V., Guermat, C. & Tucker, J. (2018), 'A conditional regime switching capm', *International Review of Financial Analysis* **56**, 1–11.

Winkler, R. L. (1989), 'Combining forecasts: A philosophical basis and some current issues', *International Journal of Forecasting* **5**(4), 605–609.

Winkler, R. L. & Makridakis, S. (1983), 'The combination of forecasts', *Journal of the Royal Statistical Society* **146**(2), 150–157.

Xie, T., Jun, Y. & Zeng, T. (2020), 'Econometric methods and data science techniques: A review of two strands of literature and an introduction to hybrid methods', *SMU Economics and Statistics Working Paper Series, Paper No. 16-2020* .

Yang, K., Tian, F., Chen, L. & Li, S. (2017), 'Realized volatility forecast of agricultural futures using the har models with bagging and combination approaches', *International Review of Economics and Finance* **49**, 276–291.

Yin, A. (2020), 'Equity premium prediction and optimal portfolio decision with bagging', *The North American Journal of Economics and Finance* **54**, 101274.

Zhang, L. (2005), 'The value premium', *The Journal of Finance* **60**(1), 67–103.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society* **67**(2), 301–320.

Table 3.1: Out-of-sample Cross-sectional Results for CI and CF Approaches

This table reports the results from the second step of the Fama & MacBeth (1973) methodology for Combining Information (CI) and Combining Forecasts (CF) approaches. CI given in Panel A consist of subset variable selection where we report results of DSM-III from Chapter 2 which selects the best model at each period based on MSFE in cross-validation sample. A.2 indicate dimension reduction approaches given in section (**??**). Panel B indicate the CF approaches given in section (3.4.2). Results are classified into two groups. First, the 'I. Cross-sectional Explanation', which consists of price of risk and performance evaluation. Under price of risk, $\bar{\bar{\lambda}}_0$ is average return on zero beta portfolio, $\bar{\bar{\lambda}}_1$ is estimated risk premium, to test the null hypothesis that the price of risk is equal to zero, the Newey & West (1987) t-statistics is reported below the coefficient estimates. [*], [**] and [***] asterisks denote the significance of coefficients at a 10%, 5% and 1% level, respectively. The performance evaluation criteria include SSPE, RMSFE, Adj. R^2, mispriced asset (MA), JA, and CPE are sum of square pricing errors, root mean square pricing error, Jagannathan & Wang (1996) adjusted R^2, joint alpha test, and composite pricing error. Under MA, at 1% and MA 5% indicate the number of portfolios for which pricing errors are significant at 1% and 5% levels respectively. JA and CPE statistics are used for testing the null hypothesis that pricing errors are jointly zero. [**] indicates that the null hypothesis is rejected at 5% critical value based on the normal distribution for JA, and for CPE, we use bootstrap distribution. The second category of results, 'II. Pricing anomalies' refers to the ability of respective model to explain the size, value, and momentum anomalies which is based on cross sectional regression analysis, $R^{RA}_{i,t+1} = \lambda_{0,t+1} + \lambda_{s,t+1}\hat{\beta}^{SMB}_{i,t} + \lambda_{h,t+1}\hat{\beta}^{HML}_{i,t} + \lambda_{m,t+1}\hat{\beta}^{MOM}_{i,t} + \alpha_{i\,t+1}$, with the SMB, HML, and MOM factors on net returns (Risk adjusted returns, R^{RA}) based on realised market returns, $R^{RA}_{i,t+1} = R_{i,t+1} - (R_{m,t+1}\beta_{i,t})$, and estimated risk premium, $R^{RA}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1})$. $\bar{\bar{\lambda}}_{SMB}$, $\bar{\bar{\lambda}}_{HML}$ and $\bar{\bar{\lambda}}_{MOM}$ indicate the estimated risk premium on size, value and momentum factors, respectively. t-statistic is presented in parentheses immediately below the coefficient estimates. All models are estimated using monthly observations. The test assets are the 25 assets sorted by size and book-to-market ratio. The out-of-sample period is August 1936 to December 2018.

Model	Cross-Sectional Explanation											Pricing Anomalies							
	Price of Risk		Performance Evaluation Criteria									Realised market premium ($R^{RA}_{i,t+1} = R_{i,t+1} - \hat{\beta}_{i,t}R_{m,t+1}$)				Estimated market Premium ($R^{RA}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1})$)			
						Mispriced Assets		Joint Test		Magnitude									
	$\bar{\bar{\lambda}}_0$	$\bar{\bar{\lambda}}_1$	SSPE	RMSPE	Adj. R^2	at 5%	at 1%	JA	CPE	PEM	HJ	$\bar{\bar{\lambda}}_{SMB}$	$\bar{\bar{\lambda}}_{HML}$	$\bar{\bar{\lambda}}_{MOM}$	Adj. R^2	$\bar{\bar{\lambda}}_{SMB}$	$\bar{\bar{\lambda}}_{HML}$	$\bar{\bar{\lambda}}_{MOM}$	Adj. R^2
A. Combining Information (CI)																			
A.1 Subset Variable Selection																			
A.1 MSFE-CV [DSM-III]	0.7125***	0.3919	0.5694	0.1509	0.3984	9	7	67.1012**	0.0185**	0.1284	0.1696	0.0435	0.2420**	0.1021	0.5150	0.0392	0.2178**	0.0919	0.4435
	2.9415	1.1381										0.4114	2.3755	0.5531		0.3702	2.1379	0.4978	
A.2 Dimension Reduction																			
A.2.1 PCR	0.8296***	0.2333	0.6219	0.1577	0.3086	11	7	68.7749**	0.0257**	0.1603	0.2061	0.0640	0.2586***	0.1469	0.6185	0.0602	0.2431**	0.1381	0.5814
	3.1016	0.9582										0.6609	2.5869	0.6041		0.6212	2.4317	0.5679	
A.2.2 KP's 3PF	0.7609***	0.3180	0.5889	0.1535	0.3381	10	7	78.6412**	0.0169**	0.1302	0.1797	0.0610	0.2431**	0.1410	0.5755	0.0574	0.2285**	0.1326	0.5410
	3.0413	1.0442										0.6192	2.4747	0.5939		0.5821	2.3262	0.5583	
A.2.3 PCR with Targeted Predictors																			
A.2.3.1 Hardthresholding	0.7835***	0.2736	0.6381	0.1598	0.3259	10	7	73.6558**	0.0198**	0.1409	0.1795	0.0648	0.2454**	0.1423	0.5965	0.0609	0.2306**	0.1338	0.5607
	3.0834	1.0192										0.6335	2.5448	0.6107		0.5955	2.3921	0.5741	
A.2.3.2 Softhresholding	0.7734***	0.2939	0.5912	0.1538	0.3355	10	7	84.1451**	0.0268**	0.1298	0.1800	0.0624	0.2547**	0.1477	0.5802	0.0586	0.2394**	0.1389	0.5454
	3.0501	1.0399										0.6262	2.5262	0.6063		0.5886	2.3746	0.5699	
B.2.3.3 Cross-sectional APC (HT)	0.6832***	0.4012	0.5501	0.1499	0.4152	9	6	66.3450**	0.0181**	0.1280	0.1667	0.0314	0.2378**	0.1380	0.5154	0.0295	0.2236**	0.1297	0.4845
	2.8834	1.2431										0.3013	2.3143	0.5554		0.2832	2.1754	0.5221	
A.2.4 Single Index with APC	0.7301***	0.3623	0.5826	0.1527	0.3842	9	7	67.4753**	0.0188**	0.1312	0.1712	0.0503	0.2423**	0.1406	0.5425	0.0473	0.2278**	0.1321	0.5099
	2.9823	1.0832										0.4827	2.4124	0.5790		0.4538	2.2677	0.5442	
A.2.5 Cross-sectional Beta Premium	0.5901***	0.4433	0.5245	0.1448	0.4646	8	4	64.7770**	0.0161**	0.1434	0.1548	0.0248	0.2188**	0.1212	0.3934	0.0233	0.2058**	0.1139	0.3763
	2.6571	1.5670										0.2533	2.1898	0.5255		0.2381	2.0584	0.4940	
B. Combining Forecasts (CF)																			
B.1 Mean	0.6822***	0.4016	0.5617	0.1499	0.4266	9	6	64.7828**	0.0164**	0.1467	0.1602	0.0288	0.2353**	0.1035	0.4998	0.0268	0.2188**	0.0963	0.4648
	2.8727	1.2574										0.2930	2.2950	0.4131		0.2725	2.1344	0.3842	
B.2 Trimmed Mean	0.7237***	0.3737	0.5786	0.1521	0.3887	9	7	68.7058**	0.0188**	0.1287	0.1709	0.0485	0.2445**	0.1076	0.5326	0.0451	0.2275**	0.1001	0.4953
	2.9704	1.1140										0.4720	2.4083	0.4335		0.4390	2.2397	0.4031	
B.3 Median	0.7571***	0.3257	0.5757	0.1517	0.3510	10	7	83.9262**	0.0213**	0.1275	0.1776	0.1189	0.2419**	0.1065	0.5664	0.1106	0.2250**	0.0990	0.5267
	3.0256	1.0552										0.5950	2.4607	0.4429		0.5533	2.2885	0.4119	
B.4 VW (CS adj. R^2)	0.6739***	0.4134	0.5596	0.1496	0.4287	9	6	67.8992**	0.0201**	0.1419	0.1599	0.1682	0.2316**	0.1019	0.4943	0.1565	0.2154**	0.0948	0.4597
	2.8378	1.3152										0.2752	2.2742	0.4093		0.2559	2.1150	0.3807	
B.5 VW (MSFE-CV)	0.6662***	0.4234	0.5487	0.1481	0.4399	8	5	69.2589**	0.0217**	0.1473	0.1584	0.0249	0.2294**	0.1009	0.4868	0.0232	0.2133**	0.0939	0.4527
	2.7893	1.3646										0.2545	2.2493	0.4049		0.2366	2.0918	0.3765	
B.6 DMSFE-CV (0.9)	0.5804***	0.4503	0.5184	0.1440	0.4708	8	4	62.2528**	0.0155**	0.1397	0.1539	0.0248	0.2061**	0.0907	0.3136	0.0230	0.1916*	0.0843	0.2916
	2.4030	1.5711										0.2533	1.9674	0.3541		0.2356	1.8297	0.3293	
B.7 Robust Rank (MSFE-CV)	0.6550***	0.4246	0.5422	0.1473	0.4465	8	5	63.2945**	0.0167**	0.1443	0.1574	0.0249	0.2173**	0.0956	0.4697	0.0232	0.2021**	0.0889	0.4368
	2.7765	1.4343										0.2545	2.1965	0.3954		0.2366	2.0427	0.3677	
Sample: August 1936 to December 2018																			

Table 3.2: Out-of-sample Cross-sectional Results for hybrid of CI and CF approaches

This table reports the results from the second step of the Fama & MacBeth (1973) framework for various approaches combining CI and CF approaches given in section (3.4.3). See Table (3.1) for details on tests and performance metrics.

	Cross-Sectional Explanation											Pricing Anomalies							
	Price of Risk		Performance Evaluation Criteria									Realised market premium $(R^{RA}_{i,t+1} = R_{i,t+1} - \hat{\beta}_{i,t} R_{m,t+1})$				Estimated market Premium $(R^{RA}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1}))$			
						Mispriced Assets		Joint Test		Magnitude									
	$\bar{\hat{\lambda}}_0$	$\bar{\hat{\lambda}}_1$	**SSPE**	**RMSPE**	**Adj. R^2**	**at 5%**	**at 1%**	**JA**	**CPE**	**PEM**	**HJ**	$\bar{\hat{\lambda}}_{SMB}$	$\bar{\hat{\lambda}}_{HML}$	$\bar{\hat{\lambda}}_{MOM}$	**Adj. R^2**	$\bar{\hat{\lambda}}_{SMB}$	$\bar{\hat{\lambda}}_{HML}$	$\bar{\hat{\lambda}}_{MOM}$	**Adj. R^2**
A.1 Pool (75%)	0.7337*** 2.9972	0.3620 1.0821	0.5832	0.1527	0.3838	9	7	67.5428**	0.0169**	0.1314	0.1714	0.0504 0.1807	0.2425** 2.4370	0.1334 0.5361	0.5425	0.0464 0.1663	0.2232** 2.2420	0.1227 0.4932	0.4991
A.2 Pool (50%)	1.0362*** 3.2935	-0.0948 -0.5132	1.0257	0.2026	0.1586	12	8	72.4003**	0.0221**	0.1485	0.1929	0.1343 0.3258	0.3093*** 3.3431	0.1701 0.7355	0.6537	0.1236 0.2998	0.2845*** 3.0756	0.1565 0.6766	0.6014
A.3 Pool (25%)	1.1243*** 3.4645	-0.1739 -0.8273	1.1474	0.2142	0.0589	13	9	90.6096**	0.0241**	0.1552	0.2040	0.1998 0.4724	0.4183*** 3.3953	0.2301 0.7470	0.7624	0.1838 0.4346	0.3848*** 3.1237	0.2117 0.6872	0.7014
A.4 Cluster Combination	0.7127*** 2.9433	0.3950 1.1878	0.5606	0.1498	0.4079	9	6	64.9559**	0.0170**	0.1306	0.1683	0.0318 0.4832	0.2308** 2.4148	0.1269 0.5313	0.5032	0.0293 0.4446	0.2123** 2.2217	0.1168 0.4888	0.4629
A.5 Pool (Forecast Island)	0.7249*** 2.9822	0.3739 1.1310	0.5904	0.1537	0.3975	9	7	**73.6089****	0.0219**	0.1481	0.1643	0.0496 0.6466	0.2684** 2.5573	0.1476 0.5626	0.5324	0.0456 0.5949	0.2470** 2.3527	0.1358 0.5176	0.4898
A.6 Pool (Auto Combine)	0.5557** 2.3354	0.4710 1.5923	0.5006	0.1415	0.4890	8	4	62.8257**	0.0189**	0.1378	0.1513	0.0256 0.3789	0.2011** 2.3589	0.1106 0.4773	0.3563	0.0235 0.3486	0.1849** 2.1702	0.1017 0.4391	0.3278
A.7 CF PCR	0.6216*** 2.7680	0.4288 1.4890	0.5386	0.1468	0.4502	8	5	62.2657**	0.0206**	0.1434	0.1569	0.0249 1.2134	0.2203*** 3.1236	0.1212 0.5272	0.4635	0.0229 1.1163	0.2027*** 2.8737	0.1115 0.4850	0.4264
A.8 CF Univariate PCR	0.8520*** 3.1119	0.2236 0.9500	0.6292	0.1586	0.3029	11	7	79.0755**	0.0183**	0.1354	0.1857	0.1060 1.4006	0.3037*** 3.2024	0.1670 0.7045	0.6218	0.0975 1.2885	0.2794*** 2.9462	0.1537 0.6482	0.5721
A.9 CF CSBP	0.5099** 2.1992	0.5172 1.6071	0.4720	0.1374	0.5018	8	4	61.0241**	0.0185**	0.1363	0.1493	0.0137 0.7419	0.1719** 2.2502	0.0946 0.8031	0.4131	0.0126 0.6825	0.1582** 2.0702	0.0870 0.7388	0.3800
A.10 CF Shrinkage (ENet)	0.8924*** 3.2124	0.2175 0.8824	0.6623	0.1628	0.2471	11	7	80.4724**	0.0274**	0.1682	0.2135	0.1183 1.2545	0.3014*** 3.1421	0.1657 0.7125	0.7143	0.1088 1.1541	0.2772*** 2.8907	0.1525 0.6555	0.6571

Sample: August 1936 to December 2018

Table 3.3: Out-of-sample Cross-sectional Results for Bagging Strategies

This table reports the results from the second step of the Fama & MacBeth (1973) framework for Bagging strategies discussed in section (3.4.3.5.2). See Table (3.1) for details on tests and performance metrics.

	I. Cross-Sectional Explanation											II. Pricing Anomalies							
	Price of Risk		Performance Evaluation Criteria									Realised market premium $(R^{RA}_{i,t+1} = R_{i,t+1} - \hat{\beta}_{i,t}R_{m,t+1})$				Estimated market Premium $(R^{RA}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1}))$			
						Mispriced Assets		Joint Test		Magnitude									
	$\bar{\lambda}_0$	$\bar{\lambda}_1$	SSPE	RMSPE	Adj. R^2	at 5%	at 1%	JA	CPE	PEM	HJ	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	Adj. R^2	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	Adj. R^2
A. Selection																			
A.1 BAGG-Hardthresholding	0.8912*** 3.2345	0.2175 0.8832	0.6633	0.1629	0.2463	11	7	60.5136**	0.0274**	0.1683	0.2143	0.1184 0.1036	0.3016** 2.3460	0.1689 0.5161	0.6349	0.1101 0.0963	0.2805** 2.1818	0.1571 0.4800	0.5904
A.2 BAGG-APC1	0.7274*** 2.9853	0.3734 1.1129	0.5791	0.1522	0.3884	9	7	68.7745**	0.0164**	0.1281	0.1711	0.0486 0.1167	0.2843** 2.3173	0.1592 0.5098	0.5326	0.0452 0.1086	0.2644** 2.1551	0.1481 0.4741	0.4953
A.3 BAGG-APC2	0.6143*** 2.7070	0.4432 1.4979	0.5325	0.1459	0.4564	8	5	63.2676**	0.0200**	0.1415	0.1560	0.0249 0.1168	0.2498** 2.1740	0.1399 0.4783	0.4636	0.0232 0.1087	0.2323** 2.0218	0.1301 0.4448	0.4312
A.4 BAGG-MSFE-CV (APC)	0.5056** 2.1435	0.5263 1.6337	0.4613	0.1358	0.5213	7	4	60.4413**	0.0181**	0.1335	0.1452	0.0119 0.1036	0.1984** 2.0405	0.1213 0.4489	0.2247	0.0111 0.0963	0.1845* 1.8977	0.1128 0.4175	0.2089
B. Dimesion Reduction																			
B.1 BAGG-PCR-APC	0.6158*** 2.7123	0.4441 1.5012	0.5336	0.1461	0.4551	8	5	63.2891**	0.0202**	0.1428	0.1556	0.0249 0.2545	0.2495** 2.1116	0.1397 0.4646	0.4432	0.0231 0.2366	0.2320** 1.9638	0.1299 0.4320	0.4122
C. CF																			
C.5 BAGG-Mean	0.6556*** 2.7793	0.4242 1.4329	0.5427	0.1473	0.4469	8	5	62.2232**	0.0208**	0.1442	0.1573	0.0249 0.4114	0.2271** 2.2155	0.1216 0.5226	0.4692	0.0231 0.3826	0.2112** 2.0604	0.1131 0.4860	0.4364
C.6 BAGG-DMSFE-CV (0.9)	0.5809** 2.4054	0.4499 1.5696	0.5189	0.1441	0.4713	8	4	61.1736**	0.0195**	0.1396	0.1538	0.0247 0.4146	0.2659*** 3.1267	0.1153 0.6879	0.3633	0.0230 0.3856	0.2472*** 2.9078	0.1072 0.6397	0.3379
D. CF and CI																			
D.1 BAGG-Pool (Auto Combine)	0.5018** 2.1251	0.5386* 1.6943	0.4609	0.1358	0.5294	7	4	60.4327**	0.0178	0.1334	0.1451	0.0107 0.4719	0.1809** 1.9929	0.0677 0.2404	0.2363	0.0099 0.4389	0.1683* 1.8534	0.0630 0.2236	0.2198
D.2 BAGG-CF CSBP	0.5052** 2.1610	0.5260 1.6123	0.4624	0.1360	0.5210	7	4	60.4843**	0.0182	0.1338	0.1467	0.0119 0.6529	0.1986** 2.0043	0.0944 0.2804	0.2449	0.0111 0.6072	0.1847* 1.8640	0.0878 0.2607	0.2277
D.3 BAGG-CF PCR	0.5189** 2.2392	0.5072 1.5011	0.4735	0.1376	0.4998	8	4	61.0919**	0.0190	0.1376	0.1499	0.0107 0.0523	0.2122** 2.1619	0.0964 0.3216	0.2919	0.0099 0.0487	0.1974** 2.0106	0.0897 0.2991	0.2715

Sample: August 1936 to December 2018

Table 3.4: Out-of-sample Cross-sectional Results of BAPCAM and Benchmark Models

This table reports the results from the second step of the Fama & MacBeth (1973) framework for our BAPCAM and benchmark asset pricing models. Panel A reports the results of various CCAPM benchmark models where models differ on the basis of capturing time-variation in betas. Model CAPM (β_{RW}), and CAPM (β_{SW}) report the results of CAPM where the time variation in beta is captured through 60 monthly rolling window and short window approach of Lewellen & Nagel (2006), respectively. CAPM (β_{FH}) reports the results of CAPM with beta defined as predetermined set of four conditioning variables used by Ferson & Harvey (1999). These variables include term spread, default yield, t-bill rate and spread between 3 months and one-month t-bill rate. CAPM (β_{KS}) reports the results of CAPM with beta defined as function of all the 12 predictors of Goyal & Welch (2008). Panel B reports the results of factor models which include Fama & French (1993) three-factor and Carhart (1997) four-factor models. See Table (3.1) for details on tests and performance metrics.

	Cross-sectional Explanation											Pricing Anomalies							
	Prices of Risk		Performance Evaluation Measures									Realised market premium ($R^{RA}_{i,t+1} = R_{i,t+1} - \hat{\beta}_{i,t}R_{m,t+1}$)				Estimated market Premium ($R^{RA}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1})$)			
						Mispriced Assets		Joint Test		Magnitude									
	$\bar{\lambda}_0$	$\bar{\lambda}_1$	SSPE	RMSPE	Adj. R^2	at 5%	at 1%	JA	CPE	PEM	HJ	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	Adj. R^2	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	Adj. R^2
BAPCAM	0.5018** 2.1251	0.5386* 1.6943	0.4609	0.1358	0.5294	7	4	60.4327**	0.0178	0.1334	0.1451	0.0107 0.4719	0.1809* 1.9929	0.0677 0.2404	0.2363	0.0099 0.4389	0.1682* 1.8534	0.0630 0.2236	0.2198
Panel A - CCAPM Benchmark Models																			
A.1 CAPM (RW)	0.9022*** 3.4158	-0.2494 -0.6508	0.5904	0.1537	0.3365	11	8	72.4341**	0.0209**	0.1303	0.1799	0.0621 0.7253	0.2945*** 3.4666	0.4303** 2.3631	0.6762	0.0565 0.6872	0.2063*** 3.1678	0.2814 1.5533	0.6113
A.2 CAPM(SW)	0.9499*** 4.0056	-0.2486 -1.2921	0.7203	0.1697	0.1828	12	7	80.9653**	0.0263**	0.1694	0.2194	0.0468 0.5094	0.3382*** 3.8770	0.4491** 2.3424	0.7213	0.0415 0.4810	0.2626*** 3.8094	0.2631 1.4900	0.6826
A.3 CAPM(FH)	0.8404*** 5.4505	-0.0586 -0.4341	0.6112	0.1564	0.3130	11	7	69.4144**	0.0236**	0.1481	0.1932	0.0155 0.1756	0.3498*** 4.0224	0.3038 1.5863	0.7049	0.0133 0.1635	0.2383*** 3.6018	0.3586* 1.9060	0.6518
A.4 CAPM(KS)	0.9250*** 5.6163	-0.0599 -1.1625	0.7149	0.1691	0.1965	12	10	81.2772**	0.0222**	0.1692	0.2108	0.1744* 1.6962	0.2467*** 3.8241	0.1993 1.4214	0.7217	0.1376 1.6073	0.2597*** 3.3468	0.2376 1.3044	0.6646

Panel B - Factor Models

	Prices of Risk					Performance Evaluation Measures								
									Mispriced Assets		Joint Test		Magnitude	
	$\bar{\lambda}_0$	$\bar{\lambda}_{MKT}$	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	SSPE	RMSPE	Adj. R^2	at 5%	at 1%	JA	CPE	PEM	HJ
B.1 FF3F	0.9917*** 5.8432	−0.3113* -1.6833	0.1070 1.1602	0.3338*** 3.5625		0.2482	0.0996	0.7210	7	3	65.3967**	0.0063**	0.080	0.117
B. Carhart	0.9509*** 5.3710	-0.2785 -1.4590	0.1288 1.4063	0.3281*** 3.5430	0.1487 0.7889	0.2023	0.0900	0.7726	6	4	71.1373**	0.0052**	0.073	0.105

Table 3.5: Pricing errors of individual assets and SSPE

This table reports the pricing errors of individual assets and sum of squared pricing errors (SSPE) for our BAPCAM and benchmark asset pricing models. Values in bold indicate that the pricing errors are significant at 5% level.

A. BAPCAM

A.1) CAPM with BAPCAM

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.3917**	-0.0935	**-0.1437**	**0.1658**	0.0175	0.2106
2nd Q	0.0351	0.0858	0.0538	-0.0522	**-0.1203**	0.0287
3rd Q	-0.0823	0.0604	-0.0095	-0.0310	-0.0314	0.0125
4th Q	**0.2092**	0.0975	0.0969	0.0419	**-0.2630**	0.1336
Big	**0.1830**	**0.1657**	0.0429	0.0717	-0.0825	0.0747
SSPE	0.2387	0.0567	0.0349	0.0381	0.0917	**0.4601**

B. CAPM - Time variation in beta is function of time

B.1) CAPM with 60 months rolling beta

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.3003**	**-0.2025**	**-0.1511**	-0.0753	**-0.1989**	0.1993
2nd Q	-0.0898	0.0618	0.0051	-0.0864	**-0.1735**	0.0495
3rd Q	-0.0247	0.0552	0.0026	-0.0206	-0.0817	0.0108
4th Q	**0.2161**	**0.1345**	**0.1048**	0.0430	**-0.1564**	0.1021
Big	**0.3244**	**0.2582**	**0.2107**	0.1041	0.0409	0.2288
SSPE	0.2508	0.1326	0.0782	0.0262	0.1025	**0.5904**

B.2) CAPM with short window

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.3182**	**-0.1722**	**-0.1407**	-0.0820	**-0.2809**	0.2363
2nd Q	-0.0569	0.0657	0.0159	-0.0930	**-0.2319**	0.0702
3rd Q	0.0092	0.0930	0.0146	0.0230	-0.1128	0.0222
4th Q	**0.2229**	**0.1669**	**0.1421**	-0.0126	**-0.2134**	0.1435
Big	**0.3545**	**0.2583**	**0.2191**	0.0834	0.0281	0.2481
SSPE	0.2799	0.1372	0.0885	0.0230	0.1917	**0.7203**

C. CAPM - Time variation in beta is function of Constant IVs

C.1) CAPM - Ferson and Harvey (1999)

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.3415**	**-0.1935**	**-0.1389**	-0.0586	**-0.2075**	0.2198
2nd Q	-0.0875	0.0483	0.0241	-0.0807	**-0.1819**	0.0502
3rd Q	0.0071	0.0686	0.0147	0.0109	-0.0866	0.0126
4th Q	**0.1934**	0.1519	**0.1237**	0.0352	**-0.1856**	0.1114
Big	**0.3206**	**0.2397**	**0.2078**	0.1174	-0.0009	0.2172
SSPE	0.2645	0.1250	0.0786	0.0251	0.1181	**0.6112**

C.2) CAPM with Kitchen Sink

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.2715**	**-0.2036**	-0.1247	-0.1003	**-0.2121**	0.1857
2nd Q	-0.0981	0.0660	0.0168	-0.1093	**-0.2197**	0.0745
3rd Q	0.0226	0.0914	0.0272	0.0022	-0.1322	0.0271
4th Q	**0.2048**	**0.1246**	0.0946	0.0323	**-0.2501**	0.1300
Big	**0.3928**	**0.2796**	**0.2153**	**0.1364**	0.0148	0.2976
SSPE	0.2801	0.1478	0.0718	0.0417	0.1735	**0.7149**

D. Factor Models

D.1) Fama French Three Factor

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.2890**	-0.0665	0.0329	**0.0912**	0.0157	0.0976
2nd Q	-0.1247	0.0627	**0.1002**	0.0050	-0.0926	0.0381
3rd Q	**-0.0854**	0.0304	0.0173	0.0268	-0.0141	0.0094
4th Q	**0.0763**	0.0744	0.0755	0.0223	**-0.2010**	0.0580
Big	**0.1468**	0.1104	0.0362	0.0407	-0.0916	0.0451
SSPE	0.1337	0.0270	0.0184	0.0112	0.0578	**0.2482**

D.2) Carhart Four Factor

	Growth	2nd Q	3rd Q	4th Q	Value	SSPE
Small	**-0.2588**	-0.0851	0.0372	**0.0794**	0.0293	0.0828
2nd Q	-0.0880	0.0640	**0.0868**	0.0217	-0.1128	0.0326
3rd Q	-0.0820	0.0220	0.0207	0.0076	0.0122	0.0078
4th Q	**0.0764**	0.0669	0.0745	0.0071	**-0.1783**	0.0477
Big	**0.1343**	0.0881	0.0123	0.0315	-0.0671	0.0314
SSPE	0.1053	0.0241	0.0151	0.0079	0.0500	**0.2023**

Sample: August 1936 to December 2018

Table 3.6: Out-of-sample Cross-sectional Results with other portfolios

This table reports the out-of-sample Fama & MacBeth (1973) coefficient estimates and performance measures for additional test portfolio which include 25 assets sorted by size and momentum, 25 assets sorted by size and book-to-market and 30 industry, and 25 assets sorted by size and book-to-market, 30 industry, and 10 momentum portfolios. See Table (3.4) for details on models and performance metrics.

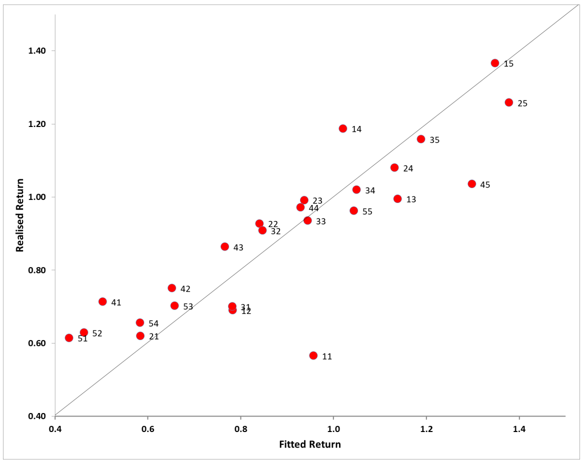
Model	Assets	Cross-sectional Explanation											Pricing Anomalies							
		Prices of Risk		Performance Evaluation Measures									Realised market premium ($R^{RA}_{i,t+1} = R_{i,t+1} - \hat{\beta}_{i,t} R_{m,t+1}$)				Estimated market Premium ($R^{RA}_{i,t+1} = R_{i,t+1} - (\hat{\lambda}_{0,t+1} + \hat{\beta}_{i,t}\hat{\lambda}_{1,t+1})$)			
							Mispriced Assets		Joint Test		Magnitude									
		$\bar{\hat{\lambda}}_0$	$\bar{\hat{\lambda}}_1$	SSPE	RMSPE	Adj. R^2	at 5%	at 1%	JA	CPE	PEM	HJ	$\bar{\hat{\lambda}}_{SMB}$	$\bar{\hat{\lambda}}_{HML}$	$\bar{\hat{\lambda}}_{MOM}$	Adj. R^2	$\bar{\hat{\lambda}}_{SMB}$	$\bar{\hat{\lambda}}_{HML}$	$\bar{\hat{\lambda}}_{MOM}$	Adj. R^2
BAPCAM	25 Size and Momentum	0.441^{**} 2.1724	0.3396^{*} 1.7312	0.4785	0.1384	0.5331	11	6	106.1681^{**}	0.0182^{**}	0.1491	0.1717	0.0110 0.4819	0.0809 0.5929	0.4903^{***} 3.0781	0.5205	0.0105 0.4627	0.0777 0.5692	0.4706^{**} 2.9550	0.4997
	25 Size and BM, 30 IND	0.5431^{***} 2.8642	0.2317 1.4453	1.0265	0.1366	0.2964	12	7	186.9551^{**}	0.0348^{**}	0.1945	0.1683	0.0733 0.6293	0.1691^{*} 1.9825	0.1370 1.1771	0.2654	0.0696 0.5978	0.1606^{*} 1.8834	0.1302 1.1182	0.2417
	25 Size and BM, 30 IND, 10 MOM	0.5834^{**} 2.5146	0.3143^{*} 1.6821	1.8136	0.1803	0.2241	13	8	173.1532^{**}	0.0554^{**}	0.2355	0.1946	0.0662 0.5930	0.1691^{*} 1.8253	0.4370^{***} 3.1771	0.5149	0.0609 0.5456	0.1555^{*} 1.6793	0.4021^{***} 2.9229	0.4737
Panel A - CCAPM Benchmark Models																				
A.1 CAPM (RW)	25 Size and Momentum	0.5059^{***} 2.7713	0.2840 1.2364	2.3874	0.3090	0.1391	14	12	118.5957^{**}	0.0651^{**}	0.2551	0.3420	0.2184^{***} 2.9606	0.2822^{**} 2.4364	0.7782^{***} 6.0732	0.8823	0.1376 1.5098	0.0706 0.5799	0.6145^{***} 5.9083	0.8069
	25 Size and BM, 30 IND	0.6606^{***} 3.8362	0.1024 0.4913	1.1588	0.1452	0.1223	14	8	167.9551^{**}	0.0409^{**}	0.2021	0.1780	0.1332^{**} 2.1225	0.2633^{***} 3.2321	0.2193^{**} 2.0139	0.4413	0.0632 0.7779	0.1915^{***} 2.9383	0.1916 1.6012	0.4021
	25 Size and BM, 30 IND, 10 MOM	0.7070^{***} 3.8856	-0.0301 -0.3423	2.5803	0.1992	0.0468	20	14	194.4556^{**}	0.0809^{**}	0.2845	0.2499	0.1566^{**} 2.1861	0.2704^{***} 3.3655	0.5845^{***} 5.0463	0.6685	0.0596 0.6517	0.1953^{***} 3.0680	0.5128^{***} 4.9534	0.6296
A.2 CAPM(SW)	25 Size and Momentum	0.6123^{***} 3.6257	0.0397 1.2031	2.4686	0.3142	0.0777	16	14	165.5445^{**}	0.0675^{**}	0.2599	0.3609	0.2383^{***} 3.1160	0.2473^{*} 1.9400	0.7767^{***} 6.3410	0.8674	0.1586 1.5554	0.1474 1.2039	0.5980^{***} 6.1713	0.8228
	25 Size and BM, 30 IND	0.7001^{***} 4.8370	0.0699 0.4915	1.4042	0.1598	0.0502	14	10	214.4788^{**}	0.0501^{**}	0.2238	0.2025	0.1903^{***} 2.8318	0.2920^{***} 3.4283	0.2261^{**} 2.2121	0.4758	0.0608 0.6979	0.2234^{***} 3.1122	0.1770^{*} 1.6710	0.4688
	25 Size and BM, 30 IND, 10 MOM	0.6999^{***} 4.9341	-0.0083 -0.0804	2.4658	0.1948	0.0329	19	12	198.9120^{**}	0.1030^{**}	0.3210	0.2635	0.1365^{**} 2.1090	0.3105^{***} 3.7280	0.5130^{***} 4.9515	0.6750	0.0069 0.0826	0.2024^{***} 3.1266	0.4762^{***} 3.9943	0.6016
A.3 CAPM(FH)	25 Size and Momentum	0.5398^{***} 3.1663	0.1284 1.3896	2.3364	0.3057	0.1471	16	12	112.3635^{**}	0.0622^{**}	0.2494	0.3343	0.2134^{***} 2.6897	0.3249^{***} 2.5881	0.5955^{***} 6.0176	0.8052	0.1032 1.1164	0.1518 1.2582	0.6725^{***} 5.2992	0.7250
	25 Size and BM, 30 IND	0.6434^{***} 4.4068	0.1140 0.6918	1.2065	0.1481	0.1264	13	9	130.4471^{**}	0.0420^{**}	0.2050	0.1877	0.1441^{**} 2.2130	0.3043^{***} 3.6240	0.2797 1.6182	0.4718	0.0196 0.2323	0.2044^{***} 3.1081	0.1502 1.1522	0.4206
	25 Size and BM, 30 IND, 10 MOM	0.7372^{***} 3.5946	-0.0411 -0.1945	2.6107	0.2004	0.0287	20	14	201.3725^{**}	0.1046^{**}	0.3232	0.2710	0.1737^{***} 2.6225	0.2239^{***} 3.3464	0.4772^{***} 5.0381	0.6653	0.0462 0.5330	0.3023^{***} 3.5739	0.5607^{***} 4.6258	0.6394
A.4 CAPM(KS)	25 Size and Momentum	0.6653^{***} 3.7725	0.0705 1.2199	2.4182	0.3110	0.1286	15	13	127.2059^{**}	0.0642^{**}	0.2533	0.3511	0.1608^{*} 1.9027	0.2279^{*} 1.8190	0.6272^{***} 6.1330	0.9199	0.1475^{*} 1.7456	0.2091^{*} 1.6688	0.5753^{***} 5.6266	0.8439
	25 Size and BM, 30 IND	0.7841^{***} 4.9017	0.0055 0.1309	1.2755	0.1523	0.1171	12	8	170.0615^{**}	0.0474^{**}	0.2177	0.1930	0.1762^{**} 2.2440	0.2119^{***} 2.6318	0.1826 1.1725	0.5028	0.1616^{**} 2.0588	0.1944^{**} 2.4145	0.1675 1.0757	0.4613
	25 Size and BM, 30 IND, 10 MOM	0.7604^{***} 5.1499	0.0016 0.0395	2.1104	0.1802	0.0984	17	11	184.8403^{**}	$0.0774^{**}64$	0.2782	0.2320	0.1473^{*} 1.9080	0.2019^{**} 2.5444	0.3789^{***} 3.6714	0.6164	0.1340^{*} 1.7363	0.1838^{**} 2.3154	0.3448^{***} 3.3410	0.5609

Table (3.6) Continue

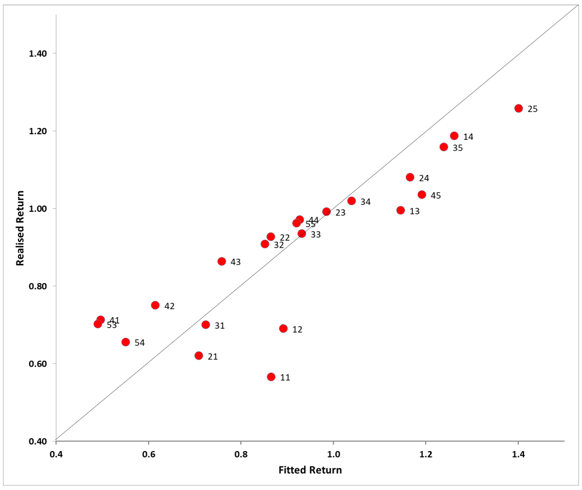
Panel B - Factor Models																
		Prices of Risk					Performance Evaluation Measures									
										Mispriced Assets		Joint Test		Magnitude		
		$\bar{\lambda}_0$	$\bar{\lambda}_{MKT}$	$\bar{\lambda}_{SMB}$	$\bar{\lambda}_{HML}$	$\bar{\lambda}_{MOM}$	SSPE	RMSPE	Adj. R^2	at 5%	at 1%	JA	CPE	PEM	HJ
B.1 FF3F	25 Size and Momentum	1.3447*** 5.6497	−0.6091** -2.3580	0.2379** 2.4273	-0.1483 -0.8686		0.6258	0.1582	0.7662	10	5	108.1681**	0.0142**	0.1191	0.1817
	25 Size and BM, 30 IND	0.8214*** 5.9002	-0.1401 -0.8373	0.1095 1.2603	0.2662*** 2.9163		0.8569	0.1248	0.3867	14	5	185.3015**	0.0297**	0.1724	0.1582
	25 Size and BM, 30 IND, 10 MOM	0.8859*** 6.4137	-0.2060 -1.2393	0.0976 1.1391	0.2418*** 2.6940		1.7579	0.1645	0.2491	13	8	167.8736**	0.0566**	0.2379	0.2117
B.2 Carhart	25 Size and Momentum	1.0577*** 5.3883	-0.3450 -1.6390	0.2639*** 2.7505	0.0026 0.0199	0.6468*** 4.9909	0.2647	0.1029	0.9011	8	6	87.1427**	0.0062**	0.0787	0.1182
	25 Size and BM, 30 IND	0.7675*** 5.6894	-0.0867 -0.5184	0.1440* 1.6432	0.2629*** 2.9265	0.1915 1.1958	0.7325	0.1154	0.4757	10	4	162.4989**	0.0259**	0.1609	0.1463
	25 Size and BM, 30 IND, 10 MOM	0.7935*** 5.8655	-0.1093 -0.6605	0.1334 1.5353	0.2696*** 3.0394	0.4136*** 3.6178	0.8140	0.1119	0.6523	12	5	193.2309**	0.0274**	0.1654	0.1441

Figure 3.3: Actual (Realised) vs Fitted Returns

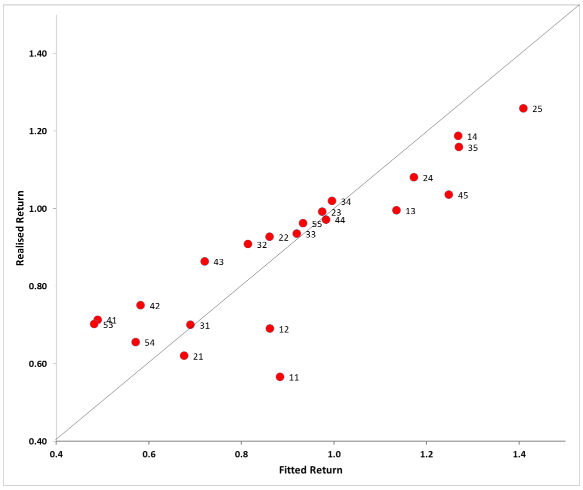
(a) BAPCAM



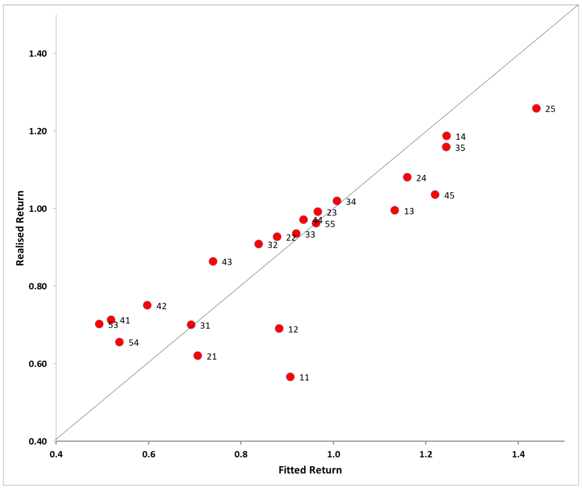
(b) CAPM (β_{RW})



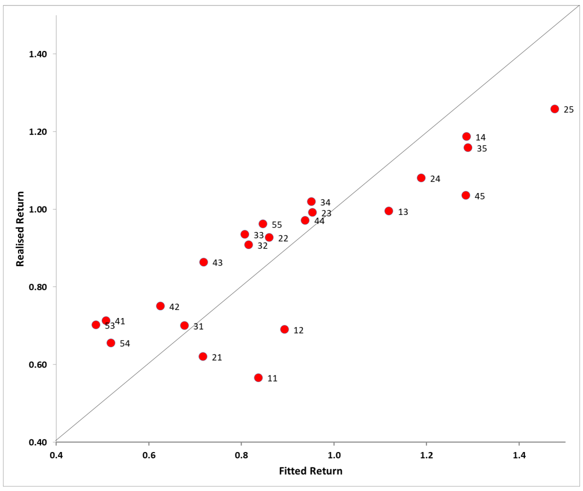
(c) CAPM (β_{SW})



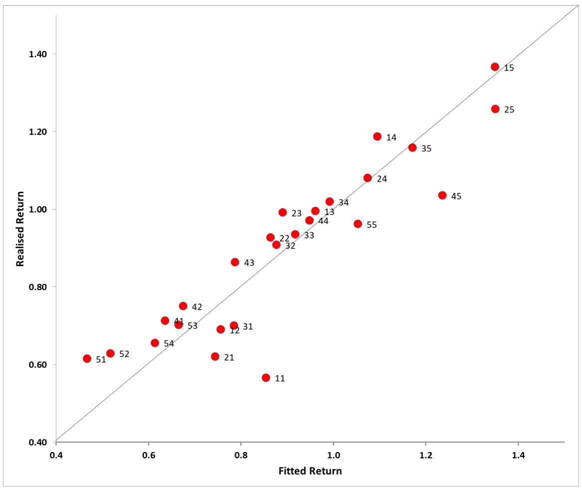
(d) CAPM (β_{FH})



(e) CAPM (β_{KS})

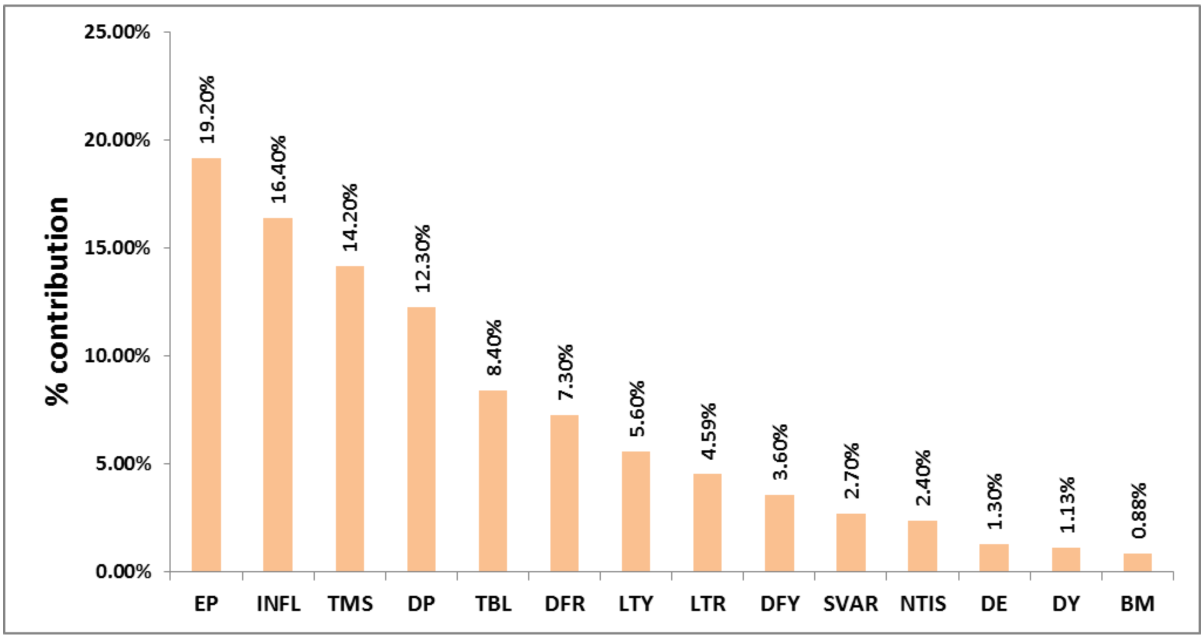


(f) FF3F



Note: This figure compares the average fitted excess returns of 25 size and book-to-market sorted portfolios and their realised (actual) excess returns for various models. The two numbers indicate the individual portfolios where the first, and second digit indicate the size and the book-to-market quantile, respectively.

Figure 3.4: Significance of Variables



Note: This figure shows the relative importance of each predictor for our BAPCAM model. Predictor importance of a specific variable j is defined as the reduction in cross-sectional adjusted R^2 for the model estimated using all the variables compared to the same model reestimated using all the variables except the variable j.

Chapter 4

Equity Premium Prediction under Variable-selection Uncertainty (VSU) and Parameter Instability (PI)

4.1 Introduction

4.1.1 Background

Scholars and practitioners alike are interested in forecasting future stock returns because return predictability has applications in a number of fields, including asset pricing, risk management, asset allocation, and the evaluation of investment managers' performance.[1] However, the seminal work by Goyal & Welch (2008) suggests that it is challenging to identify an economic variable that generates consistently superior out-of-sample forecasts of aggregate stock returns compared to the historical average. This failure in establishing the existence of reliable out-of-sample predictor variables has been attributed to various reasons. Among them, scholars appear to accept that structural breaks, or model parameter instability, might have contributed in many linear predictive models' weak out-of-sample forecasting performance (e.g, Rapach & Wohar 2006, Paye & Timmermann 2006, Dangl & Halling 2012). The other explanation for weak out-of-sample prediction results could be linked to the model-selection uncertainty. For example, the poor performance of simple univariate linear prediction models does not rule out the possibility of substantial predictive improvements by moving to more complex models that introduce new predictor variables that have not been extensively examined in the past (Phan et al. 2015, Huang et al. 2015, Chen 2017, Rapach et al. 2016, Chava et al. 2015). Overall, the weak out-of-sample forecasts observed in Goyal & Welch (2008) has been linked to the issue of *parameter instability* (PI) surrounding the data generating process for stock returns and *variable-selection uncertainty* (VSU) arising from the lack of theoretical guidance on which subset of predictors should be used in a forecasting model (Timmermann 2018, Yin 2019).

To address these two issues, several researchers try to establish return predictability employing advanced econometric methods.[2] This study contributes to the literature by introducing a new combining forecasts (CF) approach that relies on a variance-covariance method that simultaneously addresses parameter instability and variable-selection uncertainty to improve out-of-sample

[1]More details on the implications of return predictability in various field of finance can be found in Fama & French (1989), Barberis (2000), Avramov & Wermers (2006), and others.

[2]Econometric methods include economically motivated restrictions (Campbell & Thompson 2008, Pettenuzzo et al. 2014, Zhang et al. 2019), variable selection based on shrinkage methods (Buncic & Tischhauser 2017, Li & Tsiakas 2017), combining forecasts Rapach et al. (2010), combining information through dimension reduction methods (Ludvigson & Ng 2007, Kelly & Pruitt 2013, Neely et al. 2014, Kelly & Pruitt 2015), regime shifts (Guidolin & Timmermann 2007, Dangl & Halling 2012), machine learning (Gu et al. 2020, Rapach & Zhou 2020), and others.

equity premium forecasts.

4.1.2 Motivation

Our motivation stems from research in macroeconomics and finance that has linked parameter instability to forecast failure (e.g., Stock & Watson 1996, Pesaran et al. 2006, Giacomini & Rossi 2009, Inoue & Rossi 2011, Inoue et al. 2017). The main argument of these studies is that the issue of structural breaks in predictive models changes the underlying relationship between the variables in the model. Since standard forecasting models assume that the relationship between these variables remains constant over time, any parameter instability can lead to weak out-of-sample performance.

Structural breaks caused by a number of factors trigger parameter instability. These factors include extreme events, significant changes in financial market conditions, presidential elections, regime switches in monetary policies, business changes, new technology, and significant changes in government regulations. In particular, in the predictive relationship between different economic variables and stock returns, Rapach & Wohar (2006) find clear evidence of structural breaks. As shown by Pesaran & Timmermann (2007) the performance of a forecasting model when structural breaks are present depends on the number of observations (window length) used to estimate the out-of-sample forecast. However, there is no clear consensus in the literature on the number of observations to be used in estimation, which is usually referred as *estimation window uncertainty* (EWU) (Pesaran & Timmermann 2007). Due to this issue, it is recommended rather than including all available observations for estimating the parameters, only the most recent observations be used (the so-called "rolling estimation" method). However, most of the existing forecasting strategies in *equity premium prediction* (EPP) use an expanding window method, for example, Rapach et al. (2010) use combining forecast (CF) technique, whereas Neely et al. (2014) use combining information (CI) approach. Both of these strategies use a recursive expanding window, which uses all the observations available and, as a result, will be non-optimal in the presence of structural breaks. There are few studies which use rolling window approach, for example, most recently Li & Tsiakas (2017) and Yin (2020) use rolling window in implementing the shrinkage approaches such as LASSO and Elastic Net (ENet) to predict equity premium.

In most studies that use the rolling window technique, the size of the rolling window is arbitrarily selected or supported by the results of past studies. However, the forecasting accuracy of the rolling scheme is found to be sensitive to the choice of window size (e.g., Pesaran & Timmermann 2007, Inoue et al. 2017). This implies that though the existing forecasting strategies (such as combining forecasts, combining information, and shrinkage methods) using either expanding window or rolling window approaches account for VSU but do not choose the window optimally, as a result, fail to address the issue of EWU. In this study, we address this by using techniques to choose both the optimal variables and optimal window. Therefore, our new forecasting approach addresses both VSU (choice of variable) and EWU (choice of window) simultaneously to improve the out-of-sample forecasts of the equity premium.

4.1.3 Research Gaps and Objectives

Estimating breaking dates and including post-break observations for estimating the parameters, then generating forecasts, is a standard technique for dealing with parameter instability in the presence of structural breaks (see Bai & Perron 1998). However, this approach has been criticised in the literature for not improving forecast accuracy measured as mean squared forecasting error (MSFE). For example, in a study on the usefulness of pre-break data for estimating parameters, Pesaran & Timmermann (2007) show that the exclusion of pre-break data leads to high estimation uncertainty resulting in high MSFE.

Alternatively, studies such as Pesaran & Timmermann (2007), Pesaran & Pick (2011), Pesaran et al. (2013), Tian & Anderson (2014), Wang et al. (2020) and others show that combining forecasts (CF) obtained from the same model but over different estimation windows can be beneficial. However, combining weights under these methods are either based on the simple average or MSFE of an individual window. This implies that the existing CF approaches across windows methods fail to consider the correlation among forecasting errors of models estimated through multiple windows. However, from CF literature, we learn that one of the widely used CF techniques is the variance-covariance (VC) method (see Bates & Granger 1969, Newbold & Granger 1974, Figlewski 1983, Wong et al. 2007, Cang & Yu 2014, Croce 2016, Chan & Pauwels 2018, and others). The VC approach emphasises the consideration of correlation among forecasting errors, and the optimal weights are obtained as a solution to minimising the error variance-covariance matrix. It is shown that VC can provide a diversification effect and improve forecast accuracy (Bates & Granger 1969). Therefore, we aim to contribute to the literature by implementing the VC approach to obtain the optimal out-of-sample forecast for a particular economic predictor-based model based on different windows. More specifically, our primary question in this study is: "Can a model addressing *estimation window uncertainty* (EWU) through variance-covariance approach fully explain the weak predictive performance of univariate economic predictor-based models examined by Goyal & Welch (2008)?"

Next, Pesaran et al. (2013) show that *estimation window uncertainty* (EWU) and *variable-selection uncertainty* (VSU) are relevant problems for predicting macroeconomic and financial variables and introduced a new approach called average-average (AveAve). They argue that the two differently used approaches based on a simple average for accounting VSU (forecasts from various models, all estimated on a single window, are averaged, AveM) and EWU (calculated as the averages of forecasts generated from the same model over multiple windows, AveW) can be combined into one approach (AveAve) to generate a single forecast. They show that out-of-sample "AveAve" forecasts outperform the AveM, as well as the AveW forecasts. However, most of the equity premium literature considers VSU and EWU as two different and independent issues. For example, Rapach et al. (2010) account for VSU by taking a simple average across individual predictive models (AveM), however all individual models are estimated with an expanding window. On the other hand, Tian & Zhou (2018) apply three alternative methods for directly dealing with EWU for various univariate and multivariate models based on Goyal & Welch (2008) predictors to forecast out-of-sample equity premium; however, they do not consider VSU. This provides an opportunity to contribute to the existing literature by implementing the VC approach to simultaneously address EWU and VSU issues. Thus, our second research question is: "Can a model addressing EWU and VSU simultaneously based on VC panel approach improve the out-of-sample forecasts of equity premium and outperform simple average approaches such as AveM, AveW, and AveAve?"

Finally, most recently, Wang et al. (2020) use a different approach to address EWU and VSU at the same time. They show that a two-step procedure, in which the first step generates forecasts of individual models using a "time-dependent weighted least squares (TWLS)" approach, and the second step takes a simple average across forecasts of univariate models produced in the first step, performs better than the AveAve method. Following them, we can contribute to the literature by introducing a two-step strategy based on the variance-covariance method, in which the first addresses the EWU by combining forecasts across estimation windows of a single model, and the second step involves combining forecasts across multiple models obtained from the first step. Thus, our final question of this study is: "Can a model addressing EWU and VSU simultaneously through VC approach under a two-step process outperform the panel approach that combines forecasts from different models and windows in one step?"

4.1.4 Summary of Methodology

Following the literature on equity premium prediction, our primary analysis is based on a simple linear regression, $r_{t+1} = \beta X_t + \epsilon_{t+1}$, where r_{t+1} indicates aggregate market returns in excess of the risk-free rate, and X_t represents a vector consisting of predictor variables and an intercept. In this paper, our focus is on univariate models. More specifically, X_t only incorporates a variable of interest and the intercept. We can estimate N different models with N predictors. Since each univariate predictor-based model can be estimated with M different windows, this essentially results in $N \times M$ models, each generating one forecast at a time. Since our primary focus in this paper is to apply VC approach for combining i) forecasts from different univariate predictive models all estimated on a single-window (across N forecasts when $M = 1$, $N \times 1$ forecasts), ii) forecasts from the same univariate predictor-based model estimated across different windows (across M forecasts when $N = 1$, $1 \times M$ forecasts), and iii) forecasts from both different univariate predictor-based models and windows (both across N and M, $N \times M$ forecasts). Thus, we derive the optimal weights to be applied to the individual forecasts as a function of the covariance matrix of forecast errors, which provides optimal weights in the sense that the variance of the combined forecast error is minimised. Following mainstream literature, we use a sample covariance matrix. We also use a "single-index" model (SIM) introduced by Figlewski (1983), which is similar to the "market model" of finance. Following Pesaran & Timmermann (2007), we use cross-validation method, which reserves the last E observations of the data (T observations) for an out-of-sample estimation exercise. More specifically, we estimate forecasting errors, error-covariance matrix, and optimal weights over out-of-sample holdout (cross-validation, CV) observations.

4.1.5 Data and Principal Results

Our main results are based on updated data from Goyal & Welch (2008), spanning from January 1927 to December 2018. We include 14 variables for which monthly data is available. Our out-of-sample evidence suggests that the introduction of the variance-covariance approach based on single-index model (VC-SIM) applied to combine forecast across estimation windows for individual predictor-based models helps to produce a smaller MSFE for 12 (14) out of 14 individual predictive models when estimated with traditional expanding (rolling) window. Our VC-SIM approach applied to a panel of forecasts (Panel (SIM)) generated from different univariate models and windows also improves forecasting accuracy. It achieves an out-of-sample R^2 (R^2_{OS}) of 0.922%, whereas the AveAve (Pesaran et al. 2013) and equally weighted based on expanding window (AveM) (Rapach et al. 2010) achieves an R^2_{OS} of 0.808% and 0.689%, respectively. The results are even better for the VC-SIM approach, which uses a two-step method to combine forecasts across estimation windows first, and then individual model forecasts are combined COMCOM (SIM), which achieves an R^2_{OS} of 1.426% which remains the highest among all the approaches. A heuristic calculation suggested by Cochrane (1999) shows that the Sharpe ratio (s^*) earned by an active investor exploiting predictive information (summarised by the regression R^2) and the Sharpe ratio (s_o) earned by a buy-and-hold investor are related by $s^* = \sqrt{\frac{s_0^2+R^2}{1-R^2}}$. Using data back to 1871, Campbell & Thompson (2008) calculated a monthly equity buy-and-hold Sharpe ratio of 0.108. As a result, an out-of-sample predictive R^2 of 0.922% (1.426%) for Panel-SIM (COMCOM-SIM), suggests that an active investor using our approach would achieve a Sharpe ratio improvement of approximately 34% (50%), over a buy-and-hold investor, using real-time information in Goyal & Welch (2008) predictors.

We also analyse portfolio performance for a mean-variance investor allocating the wealth into risk-free assets (Treasury bill) and equity. Return forecasts for the next period are used to calculate the weights of the stock index in the portfolio. Certainty equivalent returns (CER) is a popular utility-based metric to analyse the equity premium forecasts (e.g., Goyal & Welch

2008, Campbell & Thompson 2008, Rapach et al. 2010, Neely et al. 2014, Li & Tsiakas 2017). The CER can be viewed as the performance fees that a risk-averse investor with a specific risk aversion level should pay to switch from a risk-free asset to a risky portfolio. To evaluate the portfolio performance, we use the gains in CER (ΔCER), which is defined as the difference between CER of portfolio formed on the basis of forecasts obtained from a given model (VC approaches in our case), and CER generated by portfolio based on the forecast of the historical average. Results suggest that using the VC-SIM method, we improve CER gains in 12 of the 14 univariate predictor-based models in the full sample. Our Panel (SIM) and COMCOM (SIM) achieve CERs of 207.1 bps and 252.4 bps, respectively, implying that our methods improve portfolio performance. On the other hand, AveM and AveAve achieve CER of 107.6 bps and 120.9 bps, respectively. This confirms that our findings are both statistically and economically significant. We also compare our results with some of the standard approaches including dynamic factor models based on principal components (Neely et al. 2014), three pass filter (Kelly & Pruitt 2013); and shrinkage methods including LASSO (Tibshirani 1996), Adaptive LASSO (Zou 2006), Ridge (Hoerl & Kennard 1970), and Elastic Net (Zou & Hastie 2005). Our both approaches, Panel (SIM) and COMCOM (SIM), outperform these methods.

Moreover, we show that our equity premium forecasts generated by Panel (SIM) and COMCOM (SIM) are linked to the real economy. Increased risk aversion, according to Fama & French (1989) and Cochrane (1999), generally requires a higher risk premium during economic downturns, resulting in equity premium predictability. Considering this, we analyse variations in equity premium forecasts generated by our VC-SIM approaches over the business cycle. More precisely, we analyse that changes in combined forecasts obtained with our VC approaches are closely linked to business-cycle phases as measured by the National Bureau of Economic Research (NBER). We find a clear pattern in our equity premium forecasts, with a sharp increase in equity premium forecasts in periods of recession and a decline during expansions. The six [five] highest points achieved by our forecasts generated by COMCOM (SIM) [Panel (SIM)] during recessions, and if we compare our forecasts with the historical average, then we see that the historical average is smooth and does not respond to business cycles. Overall, we demonstrate that the NBER business-cycle phases are closely tracked by the equity premium forecasts obtained through our VC-SIM approach. This forecasts behaviour is consistent with the findings of Fama & French (1989) and Cochrane (1999).

4.1.6 Contribution

This study contributes to the existing literature in multiple ways. First, this study complements the growing literature on methods for directly dealing with estimation window uncertainty (EWU) in forecasting.[3] This is the first study to the best of our knowledge, to apply the VC approach for combining forecasts from estimation windows of individual models. All past studies either use a simple average (e.g., Pesaran et al. 2013) or inverse of MSFE (e.g., Pesaran & Timmermann 2007) to combine forecasts across different windows. We show that considering the correlation among forecast errors across estimation windows can significantly improve the forecasting accuracy for individual models.

Second, this is the first study, to our knowledge, to present a comprehensive evaluation of models addressing parameter instability in the perspective of equity premium prediction. For univariate and multivariate models, Tian & Zhou (2018) compare three alternative parameter instability approaches to conventional rolling and expanding window in equity premium prediction. The three methods to address parameter instability under structural breaks include the Bai–Perron method of using post-break observations (Bai & Perron 1998), optimally weighted observations

[3] A partial list include Pesaran & Timmermann (2007), Pesaran & Pick (2011), Rossi & Inoue (2012), Pesaran et al. (2013) and others.

(Pesaran et al. 2013) and combining forecasts across different estimation windows (Pesaran & Timmermann 2007, Tian & Anderson 2014). The authors' focus was limited to two simple combining approaches not requiring the estimation of weights. However, in this study, we compare our VC approach's performance with ten alternative methods to deal with parameter instability. These methods are summarised in Table 4.2 and discussed in section (4.2.3). Consistent with the findings of Tian & Zhou (2018), we find that none of the univariate predictor-based models based on all parameter instability approaches, including our variance-covariance approach, can outperform the historical average forecasts. These findings are also consistent with Pesaran et al. (2013), who argue that the ultimate out-of-sample forecast for any given target variable should account for both variable-selection uncertainty and parameter instability at the same time.

Third, for the first time in the forecasting literature, we introduce a panel combination approach based on VC approach to address the issues of EWU and VSU simultaneously. Based on out-of-sample equity premium prediction, we show that our new model not only outperforms the existing AveAve approach of Pesaran & Timmermann (2007) but also existing approaches in equity premium such as combining forecasts (CF) methods (Rapach et al. 2010), dimension reduction (DR) methods (Kelly & Pruitt 2013, Neely et al. 2014) and shrinkage methods (Zhang et al. 2020). Moreover, we also complement the findings of Wang et al. (2020) by showing that a two-step procedure where first addressing EWU and then combining forecasts of univariate models generated from the first step (COMOM (SIM)) performs better than panel approach (Panel (SIM)).

Finally, in recessions, the superior performance of our VC-SIM approach based on Panel (SIM) and COMCOM (SIM) is important because predictive awareness of economic fundamentals over recessions is more valuable to an investor. This is because investors are more risk-averse during recessions requiring a higher risk premium, and there is also high volatility, making the historical mean a weak forecast (see Li & Tsiakas 2017, for more details). Kacperczyk et al. (2016) also show that the economic outlook affects how investors process information. In recessions, fund managers are more concerned with aggregate shocks because stocks have a higher aggregate risk. In summary, in recessions, economic fundamentals information is most important for predicting the equity premium, and we find that this is when our predictive process works best.

The remainder of this chapter is organised as follows. Section (4.2) provides an overview of literature related to equity premium prediction and forecasting methods to deal with variable-selection uncertainty and parameter instability. Section (4.3) outlines the econometric methodology. Section (4.4) discusses the implementation of variance-covariance approach in addressing the issues of EWU and VSU. Data and benchmark models are provided in section (4.5). Section (4.6) discusses the empirical results. Results from various robustness tests are provided in section (4.7). Section (4.8) concludes.

4.2 Literature review

In this section, we provide a brief overview of the literature that would serve as a basis for implementing our variance-covariance (VC) approach to address the parameter instability and variable-selection uncertainty simultaneously. We start our discussion with a brief overview of equity premium literature highlighting earlier success and challenges in establishing out-of-sample predictability. Next, we discuss the recent developments in out-of-sample equity premium prediction literature that would not only guide to identify the research gaps but also help to determine the benchmark models for performance comparison with our models. Since our motivation for this study is parameter instability, we then discuss some of the well-known approaches used in forecasting literature to address the parameter instability issues. These approaches would also be the benchmark models to compare the performance of VC approach. Considering that our VC approach lies in the body of combining forecasts (CF) literature, next, we summarise standard methods of combining forecasts which would also serve as benchmark models.

4.2.1 Overview of equity premium forecasting literature

Until the 1970s, asset returns were commonly thought to be unpredictable, and asset prices were assumed to follow a random walk (Cochrane 2009). Nonetheless, a vast literature compiles evidence starting in the late 1970s that various economic variables can forecast aggregate stock returns using the predictive regression provided in equation (4.1).

$$r_{t+1} = \alpha + \beta x_t + u_{t+1} \tag{4.1}$$

where x_t represent a set of predictors known at time t. The most widespread in-sample predictor variables of stock returns identified in the literature include the dividend–price ratio, dividend-payout ratio, book-to-market ratio, the earnings–price ratio, volatility of stock market, interest rate spreads, nominal interest rates, inflation rate, corporate issuing activity, and many other.[4] However, in an important review of all the commonly used predictor variables, Goyal & Welch (2008) argue that, although many conditioning variables appear to strongly predict returns based on in-sample regressions, it is challenging to identify a predictor variable that consistently outperforms the historical average based on out-of-sample tests.

Goyal & Welch (2008) also use a multivariate regression model ("Kitchen sink model"), incorporating all K variables:

$$r_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k x_{k,t} + \varepsilon_{t+1} \tag{4.2}$$

In addition, they also use the "model selection" (MS) approach where they first estimate all possible combinations of predictor variables.[5] Then, they choose the forecast that has made the biggest impact by achieving the smallest value of cumulative mean squared error (CMSE) up to that point at each time period. They find that both "Kitchen sink model" and "model selection" approaches also fail to outperform historical average. Overall, Goyal & Welch (2008) argue that univariate predictor-based and multivariate predictive regression models are unstable and that the historical average is not outperformed by traditional forecasting models.

This failure in establishing out-of-sample predictability may be attributed to various reasons. However, the equity premium forecasting literature generally agrees that the existence of permanent shift in the underlying distribution of the premium, known as structural breaks, is one of the key explanations for weak out-of-sample prediction performance (e.g., Paye & Timmermann 2006, Rapach & Wohar 2006, Dangl & Halling 2012, Rapach & Zhou 2013).[6]

There have been several advances in equity premium prediction literature to respond to the critique of Goyal & Welch (2008) (See Figure 4.1). Recent academic studies show that the out-of-sample predictability of stock returns can be enhanced by *certain new predictors* and *new econometric methods*. Many new and effective predictors have been developed in recent studies to outperform the historical average benchmark and improve stock return predictability. Some of these predictors include technical indicators (Neely et al. 2014, Lin 2018), investor sentiment and attention (Huang et al. 2015, Ni et al. 2015, Sun et al. 2016, Chen 2017, Coqueret 2020, Zhang et al. 2021), manager sentiment (Jiang et al. 2019), the short interest index (Rapach

[4]See Rapach & Zhou (2013) for a comprehensive review of equity premium prediction literature.

[5]The implementation of the MS approach involves estimating 2^N models and selecting one of the best models.

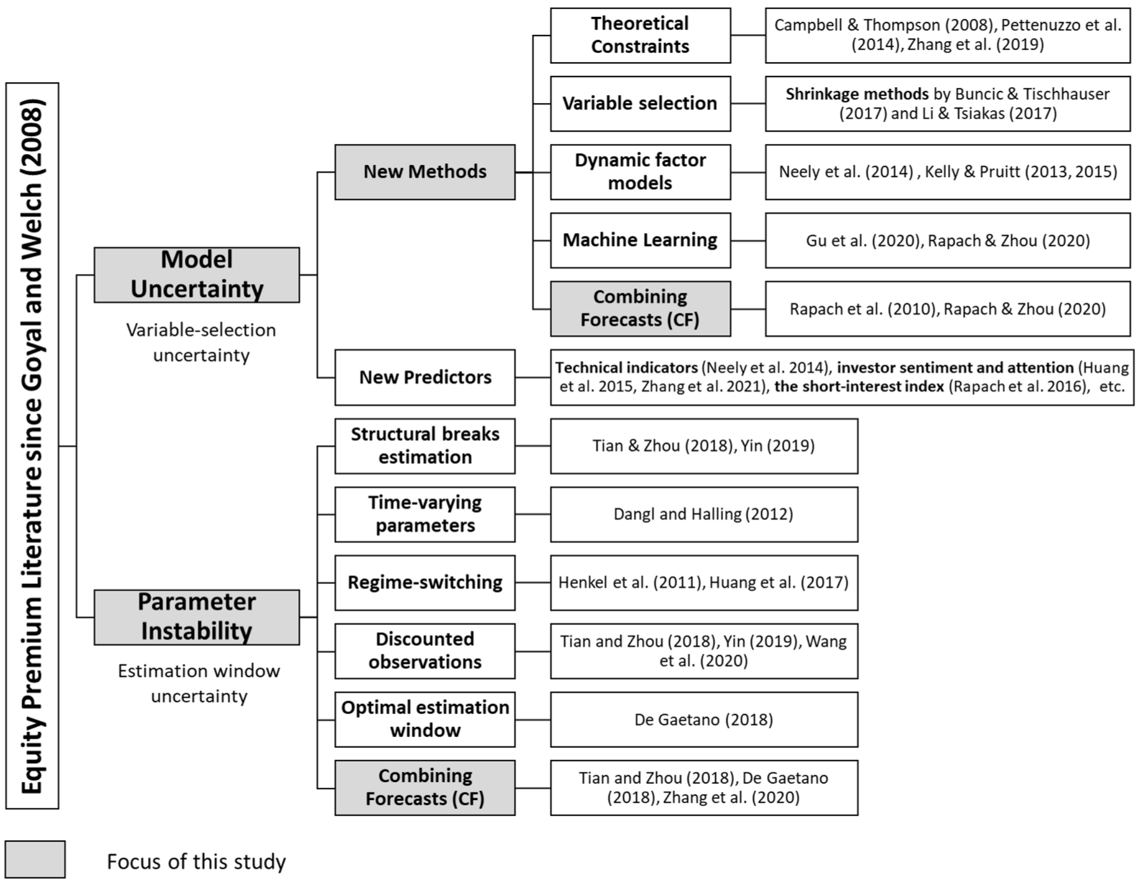
[6]See Timmermann (2018) who discusses some of the challenges in establishing equity premium predictability.

et al. 2016), the variance risk premium (Bollerslev et al. 2009), news-implied volatility (Manela & Moreira 2017), financial news (Narayan & Bannigidadmath 2017), bitcoin prices (Salisu et al. 2019), and credit quality (Chava et al. 2015), among others.

Other research studies aim to improve predictability by using the same set of traditional predictors but applying different econometric methodologies. Econometric methods include economically motivated restrictions (Campbell & Thompson 2008, Pettenuzzo et al. 2014, Zhang et al. 2019), variable selection based on shrinkage methods (Buncic & Tischhauser 2017, Li & Tsiakas 2017), combining forecasts (Rapach et al. 2010), combining information through dimension reduction methods (Ludvigson & Ng 2007, Kelly & Pruitt 2013, Neely et al. 2014, Kelly & Pruitt 2015), regime shifts (Guidolin & Timmermann 2007, Dangl & Halling 2012), machine learning (Gu et al. 2020, Rapach & Zhou 2020), and among others.

This study fits in the second group of studies that address parameter instability and variable-selection uncertainty issues using standard Goyal & Welch (2008) predictors. In the next subsection, we discuss various econometric approaches mentioned above in more detail.

Figure 4.1: Advances in Equity premium literature since Goyal & Welch (2008)



4.2.2 Strategies to establish equity premium predictability

The approaches for dealing with the variable-selection uncertainty (VSU) and parameter instability (PI) are summarised in Figure 4.1.

4.2.2.1 Theoretical Constraints

Campbell & Thompson (2008) (CT) noted that one of the drawbacks of the analysis conducted by Goyal & Welch (2008) was that out-of-sample analyses are based on unrestricted predictive regressions. Consequently, they propose imposing economically motivated restrictions on stock return predictive regression forecasts. Specifically, they suggest that a rational investor would disregard the negative forecasts, so any negative forecast produced by the model should be set to zero. Statistically, the CT forecasts at month $t+1$ with predictor i, $\hat{r}^{CT}_{t+1}$, can be given as

$$\hat{r}^{CT}_{t+1} = \max(0, \hat{\alpha}_{i,t} + \hat{\beta}_{i,t} x_{i,t}) \tag{4.3}$$

The other theoretical constraint is given by Pettenuzzo et al. (2014) (PTV) who propose to limit the in-sample annualized Sharpe ratios between zero and one. To implement this approach we need to estimate equation (4.1) with constrained least squares,

$$0 \leq \frac{\sqrt{12}(\hat{\alpha}^{PTV}_{i,\tau} + \hat{\beta}^{PTV}_{i,\tau} x_{i,\tau})}{\hat{\sigma}_\tau} \leq 1, \qquad \tau = 1, \ldots, t \tag{4.4}$$

where $\hat{\alpha}^{PTV}_{i,\tau}$, and $\hat{\beta}^{PTV}_{i,\tau}$ indicate the constrained coefficient estimates, and $\hat{\sigma}_\tau$ represents the the empirical estimate of stock volatility at month τ. The PTV forecasts at month $t+1$ with predictor i, $\hat{r}^{PTV}_{i,t+1}$, can be given as

$$\hat{r}^{PTV}_{i,t+1} = \hat{\alpha}^{PTV}_{i,t} + \hat{\beta}^{PTV}_{i,t} x_{i,t} \tag{4.5}$$

Most recently, Zhang et al. (2019) (ZWMY) propose a new constraint which is based on the well-known three-sigma rule where extreme positive and negative values of return forecasts are truncated. The ZWMY constrained forecasts at month $t+1$ with predictor i, $\hat{r}^{ZWMY}_{i,t+1}$, can be given as

$$\hat{r}^{ZWMY}_{i,t+1} = \begin{cases} r_{it} + 3\sigma_t, & \text{if } \hat{r}_{i,t+1} > r_{it} + 3\sigma_t \\ r_{it} - 3\sigma_t, & \text{if } \hat{r}_{i,t+1} < r_{it} - 3\sigma_t \\ \hat{r}_{i,t+1} & \text{otherwise,} \end{cases} \tag{4.6}$$

where $\hat{r}_{i,t}$ is the unconstrained forecast generated by equation (4.1), and σ_t represents the standard deviation of excess stock returns for month t.

4.2.2.2 Combining Information (Dynamic Factor Models)

Combining or pooling information provides a way to track the key co-movements in many predictors conveniently. One of the most prominent approaches is the dynamic factor model (DFM) or the diffusion index (DI) approach that assumes a latent factor model structure for the (demeaned) potential predictors:

$$x_{it} = \lambda_i' f_t + e_{i,t}, \qquad (i = 1, \ \ldots . N) \tag{4.7}$$

where f_t and λ_i represent the q-vector of latent factors and factor loadings, respectively and $e_{i,t}$ indicates zero-mean disturbance term. Under (4.7), relatively small number of factors ($q \ll N$) mainly represent the major co-movements in the predictors. Next, the estimated latent factors f_t serve as predictors in predictive regression given in equation (4.1):

$$r_{t+1} = \alpha_{DI} + \beta'_{DI} f_t + \varepsilon_{t+1} \tag{4.8}$$

Many studies show improvement in forecasting accuracy using DFM. For example, Ludvigson & Ng (2007) show the better out-of-sample performance of quarterly equity premium forecasts based on dynamic factors extracted from 172 financial and 209 macroeconomic predictors. Moreover, Neely et al. (2014) also demonstrate the forecasting gains for equity premium based on dynamic factors extracted from a set of popular technical indicators and Goyal & Welch (2008) predictors. Along the same lines, Kelly & Pruitt (2013) propose a new method of dimension reduction that considers the relationship between the target variable and predictors. Their approach consists of three steps; they call it a three-pass filter (3PF) and show the improvement in out-of-sample equity premium prediction based on factors derived from a collection of disaggregated valuation ratios. The first pass of three passes runs N time-series regressions separately, one for each predictor.[7] The predictor is the dependent variable in these first pass regressions given in equation (4.9), the proxies are the regressors, and the estimated coefficients define the predictor's sensitivity to factors represented by the proxies.

$$x_{i,t} = \phi_{0,i} + z_t'\phi_i + e_{i,t} \tag{4.9}$$

Next, in the second pass regressions, the estimated first-pass coefficients $\hat{\phi}_i'$ are used in T independent cross-sectional regressions given in equation (4.10). The individual predictor is once again the dependent variable in these second pass regressions, while the first-pass coefficients $\hat{\phi}_i'$ are the regressors.

$$x_{i,t} = \phi_{0,t} + \hat{\phi}_i'\mathbf{F}_t + \varepsilon_{i,t} \tag{4.10}$$

This gives time series of estimated factors, $\hat{\mathbf{F}}_t$, which are then used in equation (4.8) to estimate equity premium forecasts.

4.2.2.3 Combining Forecasts

Many studies, such as Bates & Granger (1969), Clemen (1989), and others, argue that combining individual forecasts is advantageous since it provides diversification benefits over depending on forecasts from a single forecasting approach. Moreover, it is also shown that combining forecasts (CF) guards against misspecification of the model (Timmermann 2006). Considering this, Rapach et al. (2010) demonstrate the statistical and economic significance of the CF approach in generating out-of-sample forecasts for the equity premium. They use various versions of CF, which include equally weighted, trimmed mean, median and discount mean square forecast error (DMSFE). We provide the details on CF strategies in section (4.2.4).

4.2.2.4 Shrinkage Methods

Buncic & Tischhauser (2017) and Li & Tsiakas (2017) show that the shrinkage methods that impose the statistical constraints on regression coefficients can significantly improve the out-of-sample equity premium forecasts. The shrinkage methods improve the performance of a full model given in equation (4.11), containing all the predictors by shrinking the coefficients of irrelevant predictors to zero.

$$r_{t+1} = \alpha + \sum_{i=1}^{N} \beta_i x_{i,t} + \varepsilon_{t+1}, \tag{4.11}$$

The shrinkage of the regression coefficients from equation (4.11) is based on solving the following system:

$$\begin{aligned} &\min_{\beta} \quad \frac{1}{2}\sum_{t=1}^{T-1}\left(r_{t+1} - \alpha - \sum_{i=1}^{N}\beta_i x_{i,t})\right)^2 \\ &\text{subject to} \quad \sum_{i=1}^{N}|\beta_i| < s_1 \quad \text{and} \quad \sum_{i=1}^{N}\beta_i^2 < s_2 \end{aligned} \tag{4.12}$$

[7]Predictors need to be standardised to have unit variance.

where s_1 and s_2 represent two positive constants. To estimate these constraints, one needs to minimise the mean squared forecast errors (MSFE). The type of shrinkage method depends upon the constraint. For example, the Elastic Net (ENet) originally proposed by Zou & Hastie (2005) given in equation (4.12), is a general estimator containing two well-known special cases. For example, if the first constraint, s_1, in unbound i.e., $s_1 = \infty$, then equation (4.12) becomes the ridge regression introduced by (Hoerl & Kennard 1970). On the other hand, if the second constraint, s_2, in unbound i.e., $s_2 = \infty$, then equation (4.12) reduces to the LASSO (Least Absolute Shrinkage and Selection Operator) regression (Tibshirani 1996).

4.2.2.5 Machine Learning

The availability of high dimensional data has motivated many researchers in finance to rely on machine learning (ML) methods. The initial application of ML was limited to shrinkage methods discussed in the previous section. However, there is a growing trend of ML methods in finance and predicting stock returns in particular. For example, Gu et al. (2020) perform a comparative analysis of ML techniques to measure asset risk premia. Their study is based on analysing about 30,000 individual stocks over 60 years from 1957 to 2016. For each stock, the dataset includes 94 characteristics, interactions of each characteristic with eight aggregate time-series predictors, and 74 industry sector dummies. In total, their analysis is based on more than 900 baseline signals. They find that ML methods, in particular trees and neural networks, can predict asset returns and provide large economic gains to investors. Complete detail of these methods is beyond the scope of this paper. However, the applications of ML methods in predicting stock returns in both time-series and cross-sectional can be found in Rapach & Zhou (2020) where the authors extend the ML techniques introduced in Han et al. (2019) to forecast cross-sectional stock returns to a time-series context.

4.2.2.6 Strategies to address Parameter Instability

Most of the above methods either use an expanding window or a rolling window approach, which make an assumption about the presence of structural breaks in data (see section 4.2.3 for details). To address parameter instability, several forecasting strategies have been proposed in equity premium prediction literature. For example, following Hamilton (1989), Guidolin & Timmermann (2007) report forecasting gains from using a multivariate Markov switching model. In their application, they define four regimes as 'crash', 'slow growth', 'bull', and 'recovery'. Moreover, by using a regime-switching vector auto-regression model based on two states closely resembling the NBER-dated business cycles, Henkel et al. (2011) find the out-of-sample forecasting gains for stock returns. The important findings include that fundamental variables such as the dividend yield only provide valuable information during recessions. However, the historical average forecast remains the best out-of-sample indicator during expansion.

Lettau & Van Nieuwerburgh (2008), however, criticise the regime-shifting models for problems associated with correctly specifying the timing and the size of regime shifts. To address this criticism, Huang et al. (2017) introduce a state-dependent predictive regression model given in equation (4.13).

$$r_{t+1} = \alpha_i + \beta_i^{good} x_{i,t} I_t^{good} + \beta_i^{bad} x_{i,t}(1 - I_t^{good}) + \varepsilon_{i,t+1}, \tag{4.13}$$

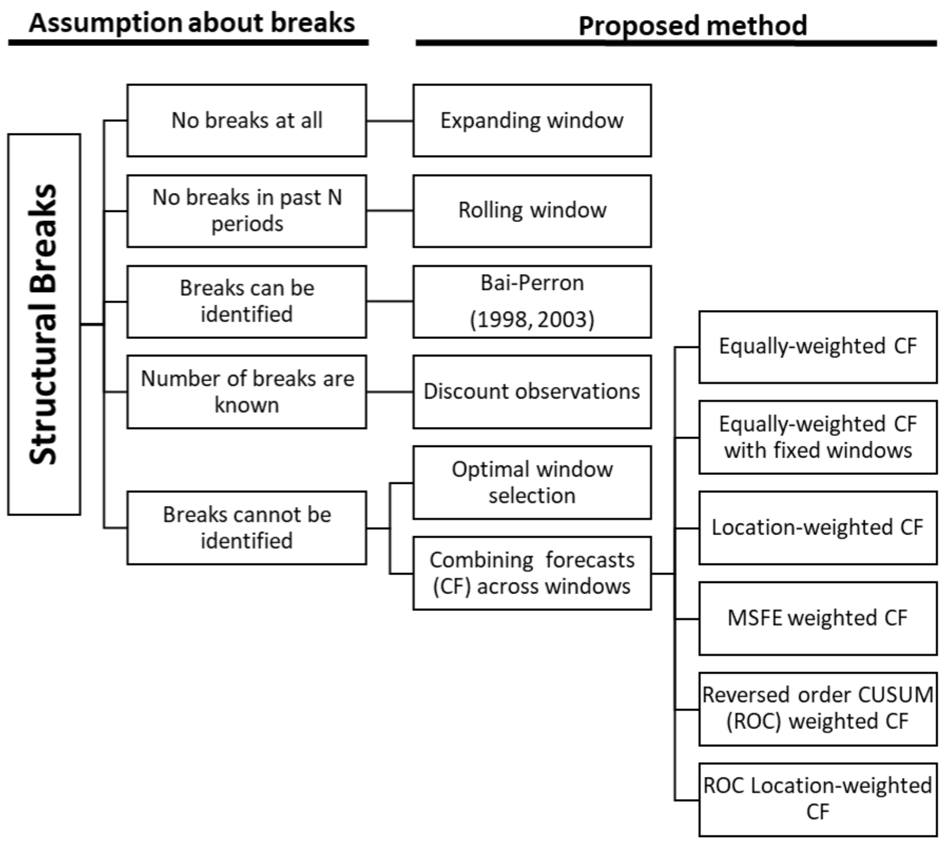
the indicator I_t is the proxy for market state which depends upon the past return information and takes the value of one if the past six-month (log) returns are positive, zero otherwise.

There are also few other studies such as Tian & Zhou (2018), Zhang et al. (2020), and Wang et al. (2020) that implement various strategies that directly address the parameter instability (see section 4.2.3 for details) from the equity premium perspective. Tian & Zhou (2018) compare three approaches to dealing with parameter instability associated with structural break uncertainty (SBU) with traditional rolling and expanding window approaches to forecasting out-of-sample equity premium. On the other hand, Zhang et al. (2020) apply window combination approach to various variable selection and model averaging approaches to forecast stock returns. The main difference between these two studies is that Tian & Zhou (2018) focus on parameter instability. In contrast, Zhang et al. (2020) compare the methods dealing model uncertainty and parameter instability simultaneously. Moreover, Wang et al. (2020) use a different approach to address parameter instability and model uncertainty at the same time from equity premium perspective. They show that a two-step procedure where first addressing parameter instability based on a "time-dependent weighted least squares (TWLS)" method to generate forecasts of univariate predictor-based models and then in the second step taking a simple average across forecasts of univariate models from the first step can improve the forecast accuracy.

4.2.3 Forecasting methods to deal with Parameter Instability

In this section, we present an overview of alternative methods that address parameter instability due to structural break uncertainty (SBU). Figure 4.2 summarises these methods, demonstrating how different methods can be classified based on assumptions regarding structural breaks. It is worth noting that our aim is not to assess the merits and drawbacks of each approach. Instead, we present the key intuition behind each method to better understand these approaches, which we also use as benchmark models.

Figure 4.2: Forecasting methods under assumptions about structural breaks



4.2.3.1 Expanding window

The expanding window (EXP) approach ignores the possible structural breaks in data. Consequently, it uses all the observations up to time T $(1:T)$ to make a forecast at time $T+1$. So as T increases by one, the observations used for estimating parameters β_T^{EXP} also increase by one. Many studies in equity premium prediction literature such as Rapach et al. (2010), Neely et al. (2014) and others use expanding window approach. The forecast under this approach $\hat{y}_{T+1}$ can be computed as:

$$\hat{y}_{T+1} = X_T' \hat{\beta}_T^{EXP} \tag{4.14}$$

where

$$\hat{\beta}_T^{EXP} = (X_{1:T-1}' X_{1:T-1})^{-1} X_{1:T-1}' y_{2:T} \tag{4.15}$$

with the observation matrices, $X_{1:T}' = [X_1,\ X_2,\ X_3,\ .\ .\ .\ ,\ X_T]$ and $y_{1:T}' = [y_1,\ y_2,\ y_3,\ .\ .\ .\ ,\ y_T]$, respectively.

4.2.3.2 Rolling window

The evidence of structural break in long samples (e.g., Paye & Timmermann 2006, Rapach & Wohar 2006) motivate researchers to use most recent observations. This leads to rolling window approach which assumes that there are no structural breaks in data in recent past. Consequently, it only uses last w observations to make a forecast at time $T+1$. Unlike expanding window approach, this approach use a fixed number of observations for estimating parameters β_T^{ROLL}. Given that we choose w as window size and our data on target variable (y) and predictor variables (X) include $X_{T-w+1:T}' = [X_{T-w+1},\ X_{T-w+2},\ X_{T-w+3},\ .\ .\ .\ ,\ X_T]$ and $y_{T-w+1:T}' = [y_{T-w+1},\ y_{T-w+2},\ y_{T-w+3},\ .\ .\ .\ ,\ y_T]$. The forecast under this approach $\hat{y}_{T+1}$ can be computed as:

$$\hat{\beta}_T^{ROLL} = (X_{T-w+1:T-1}' X_{T-w+1:T-1})^{-1} X_{T-w+1:T-1}' y_{T-w+2:T} \tag{4.16}$$

$$\hat{y}_{T+1} = X_T' \hat{\beta}_T^{ROLL} \tag{4.17}$$

Studies as such Fama & MacBeth (1973), Li & Tsiakas (2017), and others use rolling window approach. This approach has been criticised for disposing of all the data prior to the recent w observations, on the other hand, the window size (w) is arbitrarily selected. Studies such as Pesaran & Timmermann (2007), Inoue & Rossi (2011), and others show that the forecasting accuracy of the rolling scheme is sensitive to window size choice.

4.2.3.3 The Bai–Perron method

When forecasters are unaware of details about the structural breaks in the sample, the method of Bai & Perron (1998, 2003) is one of the most commonly used approaches for estimating break dates as well as the number of breaks. The presence of several breaks is accommodated by this approach and is thus most often used in studies involving a long period of observations. Since by identifying break dates, the Bai-Perron method chooses its estimation window (the post-break observations) for generating forecasts, its predictive accuracy is dependent on the accurate identification of the breakpoints. To implement this approach for identifying the optimal number of breaks, one needs to estimate the model with various numbers of breaks, l, ranging from zero to a maximum number of breaks assumed by the researcher. One can estimate the break points, $T_1,\ .\ .\ .\ ,\ T_l$, for each specified l by minimising the sum of the squared residuals or information criteria such as the Bayesian information criterion (BIC). If the optimal number of breaks, $l*$, is zero then it suggests that there are no breaks in the past, and then one needs to use all the past observations to estimate the parameters. However, if we detect structural breaks, then our estimated parameters would be based on the observations after the last estimated breakpoint $\hat{T}_{l*}$ and the forecast can be given as:

$$\hat{\beta}_T^{BP} = (X'_{\hat{T}_{I^*}+1:T-1} X_{\hat{T}_{I^*}+1:T-1})^{-1} X'_{\hat{T}_{I^*}+1:T-1} y_{\hat{T}_{I^*}+2:T} \tag{4.18}$$

$$\hat{y}_{T+1} = X'_T \hat{\beta}_T^{BP} \tag{4.19}$$

This method is widely used in forecasting (e.g., Choi & Jung 2009, Cró & Martins 2017). However, studies such as Pesaran & Timmermann (2007) criticise this approach and show that due to limited post-break data, this approach introduces high estimation uncertainty which adversely affects the forecast accuracy measured as MSFE.

4.2.3.4 Cross validation: Minimum MSFE

Pesaran & Timmermann (2007) propose a cross-validation method for selecting a single estimation window from estimation windows of different sizes. The authors are not concerned with detecting the exact location of the break, but instead with the optimum sample size to be used to estimate the model parameters to predict out-of-sample, assuming that there has been a structural break. The cross-validation strategy retains the last E observations of overall T for an out-of-sample estimation exercise and selects the estimation window on this sample that produces the smallest MSFE value. It is further assumed that to estimate the parameters of the forecasting model, a minimum of w observations are needed; this implies that $w + E$ data points are needed for implementing this approach. The recursive pseudo out-of-sample MSFE value for each possible starting point of the estimation window, m, can be computed as:

$$MSFE(m|T, E) = E^{-1} \sum_{\tau=T-E}^{T-1} (y_{\tau+1} - X'_\tau \hat{\beta}_{m:\tau})^2, \quad m = 1,\ 2,\ \ldots,\ T - w - E \tag{4.20}$$

Now suppose forecaster has already estimated the breakpoint, $\hat{T}_1$ and define $m^*(T, \hat{T}_1, w, E)$ as that value of $m \in\ 1,\ 2,\ \ldots,\ \hat{T}_1 + 1$ or $m \in\ 1,\ 2,\ \ldots,\ T - w - E$, whichever is smallest, since it would only be necessary to look for windows that start before $\hat{T}1 + 1$ that minimize the out-of-sample MSFEE for efficiency reasons:

$$m^*(T, \hat{T}_1, w, E) = \arg \underset{m=\ 1,\ \ldots,\ \min(\hat{T}_1+1, T-w-E)}{\text{minimise}} \left\{ E^{-1} \sum_{\tau=T-E}^{T-1} (\hat{y}_{\tau+1} - X'_\tau \hat{\beta}_{m:\tau})^2 \right\} \tag{4.21}$$

Now we can compute the corresponding forecast as:

$$\hat{y}_{T+1} = X'_T \hat{\beta}_{m^*:T} \tag{4.22}$$

However, one can assume that the break date is unknown, which means there is no need to estimate $\hat{T}_1$. In that case, the optimal window for estimation is obtained from

$$m^*(T, w, E) = \arg \underset{m=\ 1,\ \ldots,\ \min(\hat{T}_1+1, T-w-E)}{\text{minimise}} \left\{ E^{-1} \sum_{\tau=T-E}^{T-1} (\hat{y}_{\tau+1} - X'_\tau \hat{\beta}_{m:\tau})^2 \right\} \tag{4.23}$$

this effectively searches for m^* along the points $m = 1,\ 2,\ \ldots,\ T - w - E$ without relaying on the break date.

4.2.3.5 Robust optimal weights on observations

Pesaran et al. (2013) propose a new method called robust optimal weights (ROW), a type of generalised least squares estimator. These weights are called as optimal because these are based on minimising the one-step ahead MSFE. Moreover, by incorporating the optimal weights with regard to uniformly distributed break dates, they are robust to the uncertainty concerning parameter instability timing. The important feature of this approach is that there is no need to estimate break dates and size to produce an out-of-sample forecast. Since the weights are independent of the models and data and are determined solely by the sample size T and the assumed number of past breaks, the ROW method is simple to implement in practice. For a large sample and a single structural break, the weights can be given as:[8]

$$w_t^* = \frac{-\log(1 - t/T)}{T-1}, \quad \text{for } t = 1,\ 2,\ ,\ T-1, \text{ and} \tag{4.24}$$

$$w_t^* = \frac{\log(T)}{T-1} \tag{4.25}$$

Therefore, the standardised robust optimal weights of w_t^R, which add up to unity, are determined as

$$w_t^R = \frac{w_t^*}{\sum\limits_{s=1}^{T} w_s^*} \tag{4.26}$$

To obtain a weighted observation matrix, we need to multiply the robust optimal weights w_t^* by the matrix of regressors, $wx'_{1:T} = [x_1 \mathrm{x} w_1^R,\ x_2 \mathrm{x} w_2^R,\ \ldots\ ,x_T \mathrm{x} w_T^R]$. The (weighted) least square estimator is

$$\hat{\beta}_T^{WLS} = (wx'_{1:T-1}X_{1:T-1})^{-1} wx'_{1:T-1} y_{2:T} \tag{4.27}$$

and now we can obtain the one-step-ahead forecast as

$$\hat{y}_{T+1} = X'_T \hat{\beta}_T^{WLS} \tag{4.28}$$

4.2.3.6 Combining forecasts across estimation windows

Under this approach, the forecasts obtained through different window sizes are combined to generate a final out-of-sample forecast. As opposed to the Bai & Perron (1998) method, this method has the advantage of not relying on calculating break sizes and break dates, which can be difficult to quantify due to the noisy time series. In the literature, the following methods have been proposed.

4.2.3.6.1 Equally weighted forecast combination Pesaran & Timmermann (2007) propose an equally-weighted combining forecasts approach, estimated with different sample (window) sizes. Assume a minimum of w observations are required for estimation, the one-step-ahead CF can be given as

$$\hat{y}_{T+1} = \frac{1}{T-w} \sum_{\tau=1}^{T-w} (X'_T \hat{\beta}_{\tau:T}) \tag{4.29}$$

where $\hat{\beta}_{\tau:T} = (X'_{\tau:T-1} X_{\tau:T-1})^{-1} X'_{\tau:T-1} y_{\tau+1:T}$ for $\tau = 1,\ 2,\ \ldots\ ,\ T-w$.

[8] See Pesaran et al. (2013), and Tian & Zhou (2018) for optimal weights in case of more than one breaks.

4.2.3.6.2 Equally weighted with fixed windows Pesaran & Pick (2011) propose a simple method where instead of combining forecasts across all possible windows, one only requires a fixed number of windows that automatically update between a minimum given window and full sample. Let us consider the observation window $\mathbb{W} = \{y_{t+1}, X_t\}_{t=0}^{T-1}$, and divide it into m estimation windows of size

$$\mathbb{W}_i = \{y_{t+1}, X_t\}_{t=w_i}^{T-1}, \qquad i = 1, \ldots\ldots, m \tag{4.30}$$

where w_i is the size of the i^{th} estimation window, defining w_{min} as the minimum size of estimation window, w_i can be given as;

$$w_i = w_{min} + (\frac{i-1}{m-1})(T - w_{min}) \tag{4.31}$$

Now, the CF across m estimation windows, $AveW$, can be given as

$$\hat{y}_{T+1}^{AveW} = \frac{1}{m} \sum_{i=1}^{m} \hat{y}_{T+1}(\mathbb{W}_i) \tag{4.32}$$

where $\hat{y}_{T+1}(\mathbb{W}_i)$ indicates the forecast of any given model with particular estimation window $\mathbb{W}_i$.

4.2.3.6.3 Location weighted forecast combination Tian & Anderson (2014) suggest a different CF method that combines forecasts across estimation windows by putting more weight on recent observations. The weights under this approach are proportional to the position of each estimation window's start date (i.e. τ); CF are basically location weighted forecasts and can be given as:

$$\hat{y}_{T+1} = \sum_{\tau=1}^{T-w} \left[\frac{\tau}{\sum_{\tau=1}^{T-w} \tau} (X_T' \hat{\beta}_{\tau:T}) \right] \tag{4.33}$$

4.2.3.6.4 MSFE weighted forecast combination Pesaran & Timmermann (2007) propose a CF method that combines forecasts obtained through different estimation windows. The combining weights under this approach are proportional to the inverse of the corresponding out-of-sample MSFE values computed over cross-validation period. This method is similar to cross-validation approach discussed in section (4.2.3.4) but instead of selecting the best window, this approach recommends to combine these forecasts based on their forecasting performance under each estimation window. Given that E and w represent the cross-validation sample and minimum observations required to estimate a model, respectively, the recursive pseudo out-of-sample MSFE value for each possible starting point of the estimation window, m, can be computed as:

$$MSFE(m|T,E) = E^{-1} \sum_{\tau=T-E}^{T-1} (\hat{y}_{\tau+1} - X_\tau' \hat{\beta}_{m:\tau})^2, \qquad m = 1,\ 2,\ \ldots,\ T - w - E \tag{4.34}$$

The CF across estimation windows based on MSFE weights is then given as:

$$\hat{y}_{T+1} = \frac{\sum_{m=1}^{T-w-E} (X_T' \hat{\beta}_{m:T}) MSFE(m|T,E)^{-1}}{\sum_{m=1}^{T-w-E} MSFE(m|T,E)^{-1}} \tag{4.35}$$

4.2.3.6.5 ROC weighted forecast combination Tian & Anderson (2014) propose a new approach to combine forecasts over various estimation windows which relies on the Reverse Ordered CUSUM (ROC) structural break test. The ROC is a two-stage technique for prediction. A sequence of ROC test statistics is determined in the first step, starting from the most recent observations and going back through time. Each location in the sample is regarded as the most recent possible breakpoint. This test is similar to the Brown et al. (1975) classical CUSUM test, but differs from the standard CUSUM test in that the test sequence considers potential breakpoints in reverse chronological order, which is done by first putting all observations in reverse order, then performing the conventional CUSUM test on the rearranged set of data.

In the first stage of the approach, for $\tau = T - w + 1,\ T - w,\ \ldots 2,\ 1$, let :

$$y'_{T:\tau} = (y_T,\ y_{T-1},\ \ldots,\ y_{\tau+1},\ y_\tau) \tag{4.36}$$

$$X'_{T:\tau} = (X_T,\ X_{T-1},\ \ldots,\ X_{\tau+1},\ X_\tau) \tag{4.37}$$

be the observations matrices, and let:

$$\hat{\beta}^{ROC}_{T:\tau} = (X'_{T:\tau} X_{T:\tau})^{-1} X'_{T:\tau} y_{T:\tau} \tag{4.38}$$

be a series of β estimates obtained through least squares linked to the reverse-ordered datasets. The ROC test statistics s_τ is given as:

$$s_\tau = \frac{\sum_{t=\tau}^{T-w} \xi_t^2}{\sum_{t=1}^{T-w} \xi_t^2}, \quad \text{for} \quad \tau = T - w,\ T - w - 1,\ \ldots 2,\ 1 \tag{4.39}$$

where ξ_t are the standardised one-step-ahead recursive residuals given as:

$$\xi_t = \frac{y_t - X'_t \hat{\beta}^{ROC}_{T:T+1}}{\sqrt{(1 + X'_t (X'_{T:t+1} X_{T:t+1})^{-1} X_t)}} \tag{4.40}$$

All dates t are viewed as potential options for the final breakpoint in the second stage of the procedure. The combining weight on each t is then given as:

$$cw_\tau = \frac{\left| s_\tau - \left(\frac{T-w-\tau+1}{T-w}\right) \right|}{\sum_{\tau=1}^{T-w} \left| s_\tau - \left(\frac{T-w-\tau+1}{T-w}\right) \right|} \quad , \qquad \tau = 1,\ 2,\ \ldots,\ T - w \tag{4.41}$$

Since, under the null hypothesis of no structural break in τ, it is:

$$E(s_\tau) = \frac{T - w - \tau + 1}{T - w} \tag{4.42}$$

the combining weights differ from $s\mathcal{T}$ to its predicted value depending on the absolute distances. As a result, if this distance is high, cw_τ is larger. This suggests that the evidence of a structural break is strong. On the other hand, if no proof of a significant breakpoint exists in τ, the associated weight, cw_τ, is negligible.

In addition, the weights are not determined by locating and dating a structural break. However, if, under the null hypothesis, the absolute values of the gap between the ROC statistics and their expectations start to rise (indicating a possible structural break), the higher weights to the

observations on data, following τ, will be assigned.
The one-step-ahead forecast, $\hat{y}_{T+1}$, based on ROC statistics can be given as:

$$\hat{y}_{T+1} = \sum_{\tau=1}^{T-w} \left(cw_{\tau}(X'_T \hat{\beta}_{\tau:T}) \right) \tag{4.43}$$

4.2.3.6.6 ROC Location weighted forecast combination Tian & Anderson (2014) show that it is possible to take into account an additional weight function l_t in the definition of ROC weights by considering a prior belief on the likelihood that a time t may be the latest breakpoint. The combining weight with ROC location can be given as:

$$cw_{\tau} = \frac{\left| s_{\tau} - \left(\frac{T-w-\tau+1}{T-w} \right) \right| l_{\tau}}{\sum_{\tau=1}^{T-w} \left| s_{\tau} - \left(\frac{T-w-\tau+1}{T-w} \right) \right| l_{\tau}} \quad , \qquad \tau = 1,\ 2,\ \ldots,\ T-w \tag{4.44}$$

For example, if a single breakpoint seems to be equally likely at each time point, the obvious choice is $l_{\tau} = 1$ for $\tau = 1,\ 2,\ \ldots,\ T-w$. In this situation, the combining weights are only related to the magnitude of the ROC statistics, and the weights are obtained through (4.43). However, if the identification of the latest break is important, the prior weight l_t could be chosen in the full sample as an increasing function of the location of time τ. The most natural alternative is $l_{\tau} = \tau$ in the sense of CF with location weights.

4.2.4 Overview of Combining Forecasts (CF) Literature

The concept of combining forecasts (CF) was first proposed by Bates & Granger (1969) to improve forecast accuracy, and since then, the CF techniques have been used in many fields. A growing consensus suggests that CF improves forecasting accuracy and minimises forecast error variance. The following four points summarise the motivation for using the forecast combination method to achieve superior performance. (I) In explaining the variation in the target variable, an individual model may not be appropriate because the best in-sample model may perform poorly out-of-sample. (II) Usually, individual forecasts characterise the time series data-generating process from numerous and somewhat complementary perspectives. (III) Adaptive CF strategies may provide a full image of a variety of partial solutions, i.e., CF allows individual forecasts to "cover considerable grounds." (IV) CF may help mitigate structural breaks, model uncertainties, and model misspecifications, thereby improving the forecast accuracy. In short, CF can compensate for the disadvantages of individual forecasts, take advantage of interactions between individual forecasts, and minimise the risks of relying on a single forecast.[9]

4.2.4.1 Approaches for combining forecasts

In this section, we consider some of the important approaches of CF that have been used in various field. Following Newbold & Granger (1974), all CF methods can be described as a linear combination such that:

$$\hat{y}_t^c = \sum_{i=1}^{N} \omega_{it} \hat{y}_{it} = \mathbf{w}'_t \hat{\mathbf{y}}_t \tag{4.45}$$

where $\hat{\mathbf{y}}_t$ indicates the column vector of one-step-ahead forecasts $(\hat{y}_{1t},\ \hat{y}_{2t},\ \ldots,\hat{y}_{Nt})$ at time t generated by the i^{th} forecasting model, and $\mathbf{w}_t$ represents the column vector of combining weights for the set of N forecasting methods $(\omega_{1t},\ \omega_{2t},\ \ldots,\omega_{Nt})$. Generally, the combining weights, ω_{it}, would depend on the historical precision of the base forecasts. Therefore, in order to forecast at time $t+1$, all the observations up to time t are used to estimate all base forecast

[9] See Chapter 3 for more details on CF approach.

model parameters and combining weights. The well-known CF approaches that vary primarily in the way the combining weights are obtained are discussed below.

4.2.4.1.1 Simple average Of all the combining forecast approaches, this is the simplest. It is famous because of its ease of implementation, robustness, and strong economic and business forecasting record (Jose & Winkler 2008, Timmermann 2006, Rapach et al. 2010). The weights under this approach are obtained as:

$$\omega_i = \frac{1}{N} \tag{4.46}$$

The mean (simple average) forecast is susceptible to outliers and implies that distributions are symmetrical. It is possible to use alternative combination operators such as the trimmed mean combined forecast that sets $\omega_{it} = 0$ forecasts having the lowest and highest values, and $\omega_{it} = 1/(N-2)$ for the rest. Alternatively the median and model of $\hat{\mathbf{y}}_t$ can also be used (see Rapach et al. (2010) for more details).

4.2.4.1.2 Discounted mean square forecast error (DMSFE) method Under DMSFE method, recent forecasts are more heavily weighted than distant ones. Following Winkler & Makridakis (1983), the combining weights under DMSFE can be given as

$$\omega_i = \frac{1/\sum_{t=1}^{T} \theta^{T-t-1} \upsilon_{it}^2}{1/\sum_{i=1}^{N}\sum_{t=1}^{T} \theta^{T-t-1} \upsilon_{it}^2} \tag{4.47}$$

where θ indicates the discounting factor with $0 < \theta \leq 1$, υ_{it} is the i^{th} forecast error, whereas T and N represent the total observation and the number of individual forecasts, respectively. If $\theta = 1$, then ω is proportional to the inverse of the models' MSFE-values:

$$\omega_i = \frac{MSFE_i^{-1}}{\sum_{i=1}^{N} MSFE_i^{-1}} \tag{4.48}$$

4.2.4.1.3 Robust weighting scheme Aiolfi & Timmermann (2006) propose a robust weighting method that inversely weighs forecast models to their rank based on some performance criteria such as MSFE.

$$\omega_i = \frac{Rank_i^{-1}}{\sum_{i=1}^{N} Rank_i^{-1}} \tag{4.49}$$

where best model gets a rank of 1, second best model a rank of 2, etc.

4.2.4.1.4 Regression Method Granger & Ramanathan (1984) introduced a regression method to combine forecasts. This method involves regressing the target variable y_t on the forecasts f_t obtained through a particular model. The estimated regression coefficients indicate the combining weights for a particular model. Granger & Ramanathan (1984) consider the following three regression models:

$$y_t = \alpha + \mathbf{w}' f_t + e_t \tag{4.50}$$

$$y_t = \mathbf{w}' f_t + e_t \tag{4.51}$$

$$y_t = \mathbf{w}' f_t + e_t, \quad \text{subject to} \sum_{i=1}^{N} \omega_i = 1 \tag{4.52}$$

In equation (4.50) regression weights are unconstrained, while in equation (4.51) there is no intercept term and in equation (4.52) there is no intercept and sum of regression weights is constrained to one. If some of the predictive models f_{it} are biased predictors, then the weighting schemes generated by (4.51) and (4.52) could be biased, but (4.50) still generates unbiased predictor as the bias in f_{it} is picked up by the intercept. Therefore, if sample size is large, we should expect the mean squared prediction error of (4.50) $\leq$ mean squared prediction error of (4.51) $\leq$ mean squared prediction error of (4.52) unless the restriction is correct.

4.2.4.1.5 The optimal or variance-covariance method This method is one of the widely used CF techniques in forecasting literature (Bates & Granger 1969, Newbold & Granger 1974, Figlewski 1983, Cang & Yu 2014, Wong et al. 2007, Croce 2016). The VC approach emphasises the consideration of correlation among forecasting errors, and the optimal weights are obtained as a solution to minimising the combined forecast variance based on error variance-covariance matrix. It is shown that VC can provide diversification effect and improve forecast accuracy Bates & Granger (1969).

To obtain the optimal weights, $\mathbf{w}_t$, under VC approach, following formula is used

$$\mathbf{w}_t = \frac{\Omega^{-1} e}{e' \Omega^{-1} e} \tag{4.53}$$

where e is the $(N \times 1)$ unit vector and Ω is the $(N \times N)$ covariance matrix of one-step-ahead forecast errors.

To illustrate the VC approach, we assume that there two individual forecasts. If $\alpha_{it} = y_t - f_{it}$, $i = 1, 2$ and y_t represent the actual value of the corresponding forecast series, and we define ε_t as forecast error on forecast combination, then

$$\varepsilon_t = y_t - f_{ct} = \omega_1 \alpha_{1t} + \omega_2 \alpha_{2t} \tag{4.54}$$

considering $\sum_{i=1}^{N} \omega_i = 1$, equation (4.54) can be rewritten as;

$$\varepsilon_t = \omega_1 \alpha_{1t} + (1 - \omega_1) \alpha_{2t} \tag{4.55}$$

which has mean zero. Now we can obtain the variance as

$$\sigma_c^2 = E(\varepsilon_t^2) = \omega_1^2 \sigma_1^2 + (1 - \omega_1)^2 \sigma_2^2 + 2\omega_1 (1 - \omega_1) \sigma_{12}, \tag{4.56}$$

where σ_c^2 is the variance of the combined forecast, σ_i^2 represents the variance of the i^{th} individual forecast, and σ_{ij} is the covariance between the i^{th} and the j^{th} individual forecasts.
Now we can obtain the weight vector ω_i by minimizing σ_c^2:

$$\omega_1 = \frac{(\sigma_2^2 - \sigma_{12})}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \tag{4.57}$$

$$\omega_2 = \frac{(\sigma_1^2 - \sigma_{12})}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \tag{4.58}$$

The error-covariance method primarily depends upon the estimation of the error covariance matrix. The first choice is a sample covariance matrix which requires estimation of $N(N+1)/2$

distinct elements. To overcome issues of the sample covariance matrix, Figlewski (1983) provides a 'single-index' model. The single index model restricts the nature of the covariance matrix and is similar to the market model in finance. Another way is to assume that forecasts are independent, which means that the diagonal of Ω is sufficient. This helps in mitigating the estimation issues in the presence of short time series (Bates & Granger 1969, Granger & Ramanathan 1984). There is also another variant of the VC approach that assumes that weights cannot be negative (i.e., must belong to the interval $[0, 1]$) and sum to one.

4.3 Econometric Methodology

This section discusses the econometric framework to estimate the out-of-sample equity premium forecasts. Before discussing the models and tests, we first introduce notation, then explain the process of sample splitting. Next, we discuss the predictive regression framework for our variance-covariance approach and out-of-sample forecast evaluation methods.

4.3.1 Notation

There are N predictors, which we index using the subscript i, from T dates (e.g., months), which we index using the subscript t. Each model can be estimated with M estimation windows indexed by j. W and E denote the in-sample and the out-of-sample holdout period, respectively. S denotes the total number of out-of-sample observations ($S = T - W - E$) to evaluate the performance of a given model. We let $\upsilon_{i,t}$ denote the forecast errors for model i at date t for $1 \leq i \leq N$ and $1 \leq t \leq T$. The vector $\upsilon_t := (\upsilon_{1,t}, \ . \ . \ . \ . \ , \upsilon_{N,t})'$ consists of the forecast errors of each model on date t. Finally, $\mu_t := E[\upsilon_t]$ and $\Sigma_t := \mathrm{Cov}(\upsilon_t)$ denote the expected forecast errors and the error covariance matrix for date t, respectively.

4.3.2 Sample Split

All the methods used in this study are based on out-of-sample analysis to prevent look-ahead bias. To obtain optimal combining weights, we follow the mainstream literature and estimate forecast errors, error-covariance matrices and optimal weights from the data in a validation sample. In particular, we divide our total sample of T observations into three different periods that maintain the temporal ordering of the data. The first subsample (training sample) consists of W observations to obtain initial parameters. The second subsample (validation sample) consist of E observations, used to obtain optimal weights. The concept of validation is to simulate an out-of-sample test of a given model. Note that the validation sample fits are not really out of the sample since they are used to achieve an optimum combination of weights, which, in turn, is an input to the calculation of the final forecast. Thus, the third subsample (testing sample) consisting of $S = T - W - E$ observations, which is not used for estimation or validation, is simply out of the sample and is instead used to assess the predictive performance of a given model.

4.3.3 Out-of-sample optimal forecast of equity premium

Following mainstream literature (e.g., Goyal & Welch 2008, Campbell & Thompson 2008, Rapach et al. 2010, Neely et al. 2014), we use the standard predictive regression model, which can be represented as:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1}, \tag{4.59}$$

where r_{t+1} is excess return on stock market index (equity premium) at time $t + 1$, and $x_{i,t}$ is the predictive variable at time t, indexed by i, and $\varepsilon_{i,t+1}$ is corresponding disturbance term of equity premium.

In this study, our focus is on univariate models. More specifically, equation (4.59) only incorporates one variable of interest i. We can estimate N different models with N predictors. Since each univariate predictor-based model can be estimated with M different windows, this essentially results in $N \times M$ models, each generating one forecast at a time. Our primary focus in this paper is to apply combining forecast (CF) approach based on combining i) forecasts from different univariate predictive models all estimated on a single-window (across N forecasts when $M = 1$, $N \times 1$ forecasts), ii) forecasts from the same univariate predictor-based model estimated across different windows (across M forecasts when $N = 1$, $1 \times M$ forecasts), and iii) forecasts from both different univariate predictor-based models and windows (both across N and M, $N \times M$ forecasts). The procedure for generating CF using these three methods is described below.

4.3.3.1 CF across models

Rapach et al. (2010) propose the CF approach to combine forecasts obtained from various univariate predictor-based models. Given that, we have N forecasts (indexed by i). The CF across models can be given as:

$$\hat{r}_{t+1}^{CF-Mod} = \sum_{i=1}^{N} \hat{\omega}_{i,t}(\hat{r}_{i,t+1}) \tag{4.60}$$

where $\hat{r}_{t+1}^{CF-Mod}$ is a combined forecast across models at time $t+1$. $\hat{\omega}_{i,t}$ indicates the optimal weight for forecast i, obtained through variance-covariance approach discussed in section (4.4). $\hat{r}_{i,t+1}$ indicates the forecast obtained from model i, based on an expanding window.

4.3.3.2 CF across estimation windows

Most of the studies in equity premium forecasting (e.g., Rapach et al. 2010, Neely et al. 2014, and others) use a recursive expanding window to estimate out-of-sample forecasts. However, many studies such as Pesaran & Timmermann (2007), Pesaran & Pick (2011), Rossi & Inoue (2012), Tian & Anderson (2014), Wang et al. (2020), and others show the benefit of combining the forecasts of a single model estimated with different estimation windows. Suppose we have M forecasts (indexed by j) from different window lengths and combining weights formed at time t are given as $\hat{\omega}_{i,j,t}$, the combined forecast across different estimation windows for model i can be defined as:

$$\hat{r}_{i,t+1}^{CF-Win} = \sum_{j=1}^{M} \hat{\omega}_{i,j,t}(\hat{r}_{i,j,t+1}) \tag{4.61}$$

The above-mentioned studies combine forecasts based on different estimation windows, using equal weights or weights proportional to the inverse of the loss function of the out-of-sample, such as mean square forecasting error (MSFE). However, in this study, we use the variance-covariance approach discussed in section (4.4.2).

4.3.3.3 CF across models and estimation windows

The CF approach given in equation (4.60) represents the CF across N models, all estimated with a single window (expanding). On the other hand, model in equation (4.61) indicates the CF forecasts across M estimation windows for any given model i. Now we consider combining forecasts based on various models (N) and different estimation windows for individual model (M). We consider three variants of this method discussed below.

4.3.3.3.1 Panel approach The panel approach implies that we have a panel of $(P = M \times N)$ forecasts indexed by g. The combined forecasts can be given as:

$$\hat{r}_{t+1}^{CF-Panel} = \sum_{g=1}^{P} \hat{\omega}_{g,t}(\hat{r}_{g,t+1}) \tag{4.62}$$

where $\hat{r}_{t+1}^{CF-Panel}$ is combined forecast across windows and models. $\hat{\omega}_{g,t}$ indicate the optimal weights obtained through variance-covariance approach discussed in section (4.4.2).

4.3.3.3.2 SELCOM approach Next, we are interested in a hybrid of model selection (SEL) and combination (COM) approach. We call this approach as "SELCOM" and the combined forecast with this approach can be given as:

$$\hat{r}_{t+1}^{SELCOM} = \sum_{i=1}^{N} \hat{\omega}_{i,t}(\hat{r}_{i,t+1}^{*}) \tag{4.63}$$

where $\hat{r}_{i,t+1}^{*}$ is equity premium forecast at time $t+1$ based on the optimal estimation window approach proposed by Pesaran & Timmermann (2007), discussed in section (4.2.3.4). Specifically, this approach chooses the optimal window size by minimising the MSFE in validation sample. $\hat{\omega}_{i,t}$ indicates the optimal combining weights obtained through variance-covariance approach discussed in section (4.4.2).

4.3.3.3.3 COMCOM approach Finally, we consider a double combining method, which first combines the forecasts of individual predictor-based models across estimation windows given in equation (4.61), and then uses the variance-covariance approach to combine the optimally weighted forecasts across models. Therefore, we name this approach as "COMCOM", and the combined forecast using this method is as follows:

$$\hat{r}_{t+1}^{COMCOM} = \sum_{i=1}^{N} \hat{\omega}_{i,t}(\hat{r}_{i,t+1}^{CF-Win}) \tag{4.64}$$

where $\hat{r}_{i,t+1}^{CF-Win}$ is the combined forecast of individual model across M estimation windows obtained from equation (4.61), and $\hat{\omega}_{i,t}$ indicates the optimal combining weights obtained through variance-covariance approach discussed in section (4.4.2).

4.3.4 Forecast Evaluation

In this section, we discuss different tests to evaluate the out-of-sample performance of our forecasting approaches and benchmark models. Each model's performance is evaluated over S out-of-sample observations spanning from January 1947 to December 2018. The out-of-sample evaluation period is indexed by e $(e = 1,\ 2 \ldots S)$.[10] This section consists of two parts. First, we discuss the measures to evaluate the statistical significance of a given model and the measures for evaluating the economic significance are discussed in the final part of this section.

4.3.4.1 Measures to evaluate statistical significance

The most common metric for measuring forecast accuracy is mean squared forecasting error (MSFE), and it is not surprising that MSFE is regularly reported in stock return predictability

[10] Note that each time e represents the one-step-ahead forecast. For example, a forecast at time $e = 1$ would indicate the first out-of-sample forecast, $\hat{r}_{t+1}$.

studies. The MSFE for approach for a given approach $\mathbf{m}$ over evaluation period can be given as:

$$MSFE_{\mathbf{m}} = \frac{1}{S}\sum_{e=1}^{S}(r_e - \hat{r}_{\mathbf{m},e})^2 \tag{4.65}$$

We compare the performance of a given forecasting model $\mathbf{m}$, with the historical average (HA) benchmark model which assumes that the expected excess returns remain constant over time and can be given as:

$$\overline{r}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} r_s \tag{4.66}$$

and its MSFE over evaluation period can be given as:

$$MSFE_{HA} = \frac{1}{S}\sum_{e=1}^{S}(r_e - \overline{r}_e)^2 \tag{4.67}$$

To compare MSFEs of a given model and historical average, the out-of-sample R^2 (Campbell & Thompson 2008) is a convenient statistic. It is similar to the traditional in-sample R^2 and calculates the proportional decrease in MSFE relative to the historical average for a given forecasting approach $\mathbf{m}$. It can be given as:

$$R^2_{OOS,\mathbf{m}} = 1 - \frac{MSFE_{\mathbf{m}}}{MSFE_{HA}} \tag{4.68}$$

The out-of-sample R^2 lies in the range $(-\infty, 1)$, where a negative value indicates that the forecasting performance of a given model is poor and fails to outperform the historical average.

Next, we test whether the CF models produce significantly lower MSFE than the benchmark models i.e., historical averages and others. We test the null hypothesis of $R^2_{OOS} \leq 0$ against the alternative hypothesis of $R^2_{OOS} > 0$. We use MSFE-adjusted statistic given by Clark & West (2007). In order to calculate MSFE-adjusted statistic, we first need to define:

$$f_e = (r_e - \overline{r}_e)^2 - [(r_e - \hat{r}_{\mathbf{m},e})^2 - (\overline{r}_e - \hat{r}_{\mathbf{m},e})^2] \tag{4.69}$$

Next, we regress $\{f_e\}_{\mathrm{e=1}}^{\mathrm{S}}$ on a constant, and calculate the t-statistic. Finally, using the standard normal distribution, we calculate a p-value for a one-sided (upper-tail) test based on 1%, 5%, and 10% significance levels.

In addition to the Clark & West (2007) statistic and the R^2_{OOS}, we also compute and plot the cumulative difference of the squared forecast errors (CumSFE) of the historical average (HA) and a given $\mathbf{m}$ model over the out-of-sample forecast evaluation period. The CumSFE is frequently used as a visual method in the forecasting literature to illustrate the predictive performance over time of a proposed model relative to the benchmark (e.g., Goyal & Welch 2008, Rapach et al. 2010). The CumSFE can be given as:

$$CumSFE_e = \sum_{s=1}^{e}\left([r_s - \overline{r}_s]^2 - [r_s - \hat{r}_{\mathbf{m},s}]^2\right), \qquad \forall e = 1, \ldots\ldots, S \tag{4.70}$$

A value above zero of $CumSFE_e$ suggests that the proposed $\mathbf{m}$ model produces more accurate forecasts relative to historical average.

4.3.4.2 Measures to evaluate economic significance

The performance metrics discussed in the previous section only indicate the statistical significance; they do not specifically account for the risk that an investor bears during the out-of-sample period. Considering this, we also evaluate the economic significance of the proposed models. More specifically, following Campbell & Thompson (2008), Goyal & Welch (2008), Rapach et al. (2010), Neely et al. (2014), Li & Tsiakas (2017), and others, we use a dynamic asset allocation strategy to assess the economic significance of equity risk premium predictions for an investor with a certain degree of risk-averse level. The strategy assumes that an active portfolio consisting of investments in risky (equities) and risk-free (Treasury bill) assets requires monthly rebalancing to allocate the resources based on different equity risk premium predictions.

The average utility gain for a mean-variance investor is a popular utility-based metric to analyse the equity premium forecasts. At the end of period t, the risk-averse investor optimally distributes the following proportion of the portfolio to equities during period $t+1$:

$$\omega_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right) \tag{4.71}$$

where γ is the degree of risk-aversion, $\hat{r}_{t+1}$ indicates the equity premium forecast with any given model and $\hat{\sigma}^2_{t+1}$ represents the variance of the forecast. The weights assigned to risk-free asset can be given as $1-\omega_t$. The return on portfolio at period $t+1$, $r_{p,t+1}$, based on optimal weights can be given as:[11]

$$r_{p,t+1} = \omega_t \hat{r}_{t+1} + r_{f,t+1} \tag{4.72}$$

Following Campbell & Thompson (2008), our estimates of $\hat{\sigma}^2_{t+1}$ are based on a five-year rolling average of the variance of past monthly returns and do not allow short-selling and leverage is also limited to no more than 50%. We do this by limiting the weights of risky assets to be $w_t \in [0, 1.5]$.

The certainty equivalent rate (CER) for the portfolio can now be given as:

$$CER = \left(\bar{\hat{r}}_p \;-\; \frac{1}{2}\gamma\hat{\sigma}^2_p\right) \tag{4.73}$$

where $\bar{\hat{r}}_p$ and $\hat{\sigma}^2_p$ represent the mean portfolio return and portfolio variance, respectively over the forecast evaluation period for an investor with a risk-aversion level of γ. The CER can be viewed as the performance fees that a risk-averse investor with a specific risk aversion level should pay to switch from a risk-free asset to a risky portfolio. Our focus in this study is the relative CER also known as CER gain (ΔCER), which is defined as the difference between the CER of any given model and the CER of the historical mean benchmark. To present this difference (ΔCER) into average annualised percentage return, we multiply it by 1200, which can be understood as the annual percentage portfolio management fee that an investor is ready to pay in order to access a given model's forecasts in comparison to the historical average forecast. We also take the impact of transaction costs into account to provide a practical measure of the profitability of dynamic trading strategies. Specifically, we follow forecasting literature (e.g., Neely et al. 2014) and calculate the ΔCER net of proportional transaction costs of 50 bps per month.

In addition to the CER, we also consider two other measures including Sharpe ratio (SR) and portfolio turnover (TO). The SR is by far the most widely used measure of performance and is

[11] Remember that the equity premium is measured as the return on S&P 500 index in excess of risk-free rate (Treasury bill), $r^{S\&P500} - r_f$. The weighted average portfolio based on a risky and a risk-free asset, therefore, be given as $\omega_t \hat{r}^{S\&P500}_{t+1} + (1-\omega_t) r_{f,t+1}$, which can be redefined as $\omega_t(\hat{r}^{S\&P500}_{t+1} - r_{f,t+1}) + r_{f,t+1}$, where the former term represents the equity premium by definition.

characterised as a portfolio's average excess return divided by the standard deviation of portfolio returns.

$$SR = \frac{\bar{\hat{r}}_p}{\hat{\sigma}_p} \tag{4.74}$$

where $\bar{\hat{r}}_p$ and $\hat{\sigma}_p$ represents the mean and standard deviation, respectively, of the investor's portfolio over the forecast evaluation period. To assess the statistical significance, we use the Ledoit & Wolf (2008) bootstrap two-sided test, which checks if the alternative model's SR is distinct from the benchmark.

Finally, we consider the average portfolio turnover measure TO, which can be viewed as the average fraction of portfolio value is being traded each period. We can define TO as the average of the sum of absolute change in portfolio weight across the assets for any given model over all available rebalancing periods, $T-1$:

$$TO = \frac{1}{T-1}\sum_{t=1}^{T-1}(|w_{t+1} - w_{t+1}^{-}|) \tag{4.75}$$

where $T-1$ indicates the number of trading periods, w_{t+1} represents the weight on the risky asset at time $t+1$, and $w_{t+1}^{-} = w_t \frac{1+r_{t+1}}{1+r_{p,t+1}}$ indicates the risky asset's weight right before rebalancing at time $t+1$. Note that the TO indicates the average monthly trading volume but in results our TO measure is in relative sense, which is the ratio of the alternative model's average turnover divided by the benchmark model's average turnover.

4.4 Optimal CF with variance-covariance (VC) approach

This section discusses the optimal combining forecast (CF) approach using the variance-covariance method. We first discuss the optimal CF problem and explain how it relates to covariance matrix estimation. In the last subsection, we discuss the implementation of the variance-covariance approach to generate out-of-sample forecasts for equity premium.

4.4.1 Optimal forecast combination problem

The idea of combining forecasts (CF) was first proposed by Bates & Granger (1969) as a way of improving forecast accuracy by minimising the variance of composite forecast. The optimisation problem is similar to Markowitz portfolio theory, which deals with the problem of assigning weights ω to a universe of M possible assets in order to minimise the variance of the portfolio. More precisely, the optimal portfolio weights (forecast weight in our application) ω on time t are found by solving

$$\underline{\omega} = \frac{\Sigma^{-1}\underline{e}}{\underline{e}'\,\Sigma\,\underline{e}} \tag{4.76}$$

where $\underline{\omega}$ is M vector of weights, $\underline{e}$ is the $(M \times 1)$ unit vector, Σ is the $(M \times M)$ covariance matrix of asset returns. The solution can also be found through optimisation which is identical to (4.76).

$$\begin{aligned} \underset{\underline{\omega}}{\text{minimise}} \quad & \underline{\omega}'\,\Sigma\,\underline{\omega} \\ \text{subject to} \quad & \underline{\omega}'\underline{e} = 1 \end{aligned} \tag{4.77}$$

We can see that this problem is similar to portfolio optimisation (global minimum variance) to obtain optimal portfolio weights for M assets where optimal weights $\underline{\omega}$ are obtained by minimising the portfolio variance.

Many studies (e.g., Newbold & Granger 1974, Figlewski 1983, Timmermann 2006) find that allowing negative weights can produce poor forecasts as the change in weights from one period to another can be large and consequently many studies use the additional constraint of no-short sales in equation (4.77) to obtain optimal weights which results in:

$$\begin{aligned} \underset{\underline{\omega}}{\text{minimise}} \quad & \underline{\omega}' \, \Sigma \, \underline{\omega} \\ \text{subject to} \quad & \underline{\omega}' \underline{e} = 1 \\ & \omega_j \geq 0 \end{aligned} \tag{4.78}$$

To implement the variance-covariance method, we first need to estimate the error-covariance matrix. The standard choice is the sample covariance matrix which requires estimation of $M(M+1)/2$ distinct elements. To overcome issues of sample covariance matrix, Figlewski (1983) provides a 'single-index' model. The single index model restricts the nature of the covariance matrix, and is similar to the market model in finance. Another way is to assume that forecasts are independent which means that the diagonal of Σ is sufficient. This also mitigates estimation issues when only short time series are available (Bates & Granger 1969, Granger & Ramanathan 1984). To estimate the error variance-covariance matrix, we use all three methods: diagonal, sample, and single-index model.

4.4.1.1 Sample Covariance Matrix

The sample covariance matrix is the standard measure for estimating the pair-wise covariances of asset classes. Given E observations of forecasting errors, we can define the sample covariance matrix as:

$$\Sigma_{\mathbb{S}} = \frac{1}{E-1} \sum_{t=1}^{E} (\upsilon_t - \overline{\upsilon})(\upsilon_t - \overline{\upsilon})' \tag{4.79}$$

where $\upsilon_t = [\alpha_{1,t}, \ \ldots \ , \alpha_{M,t}]$, $\overline{\upsilon} = [\overline{\alpha}_1, \ \ldots \ , \overline{\alpha}_M]$ and $\overline{\upsilon}_j = \frac{1}{E} \sum\limits_{t=1}^{E} \alpha_{j,t}$.

4.4.1.2 Diagonal Covariance Matrix

Diagonal covariance treats the forecast errors as independent by setting all off-diagonal elements to zero. This approach ignores correlations that help prevent complications caused by dependency when assigning optimal weights to forecasts.

$$\Sigma_{DIAG} = \begin{bmatrix} \sigma_j^2 & & \\ & \ddots & \\ & & \sigma_M^2 \end{bmatrix}$$

$$\Sigma_{DIAG} = I_M \Sigma_{\mathbb{S}} = \mathrm{Diag}(\Sigma_{\mathbb{S}}) \tag{4.80}$$

where I_M is identity matrix and $\Sigma_{\mathbb{S}}$ denotes sample covariance matrix discussed earlier.

4.4.1.3 Single Index Model Covariance Matrix

Figlewski (1983) employs a "single-index model," which is analogous to the market model of finance, in which he models the information structure in such a way that certain individual model differences are allowed, but the error covariance matrix (Σ) is a function of a small number of parameters. He uses a cross-sectional average of forecast errors to measure "systematic" error forecast, which is common to all models.

$$\overline{\upsilon}_t = \frac{1}{M}\sum_{j=1}^{M} \alpha_{j,t}, \tag{4.81}$$

We can now describe our single-index model as follows, using E time-series of M individual forecast errors ($E \times M$) and $E \times 1$ time-series of average errors:

$$\begin{aligned} \alpha_{j,t} = A_j + C_j\overline{\upsilon}_t + u_{j,t}, \\ \overline{\upsilon} \sim \mathcal{N}(0,\, \Theta^2), \\ u_j \sim \mathcal{N}(0,\, \sigma^2_{u_j}), \\ cov(\overline{\upsilon}, u_j) = 0, \\ cov(u_i, u_j) = 0 \qquad \forall\ i \neq j \end{aligned} \tag{4.82}$$

This can also be given in matrix form. Assume Ψ is the error vector.

$$\Psi_t = \ A + C \quad \overline{\Psi}_t \ + U_t \tag{4.83}$$

$$\Psi_t = \begin{bmatrix} \hat{\alpha}_{1,t} \\ \hat{\alpha}_{2,t} \\ \hat{\alpha}_{3,t} \\ \bullet \\ \bullet \\ \bullet \\ \hat{\alpha}_{M,t} \end{bmatrix}, C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \bullet \\ \bullet \\ \bullet \\ C_M \end{bmatrix}, U_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ \bullet \\ \bullet \\ \bullet \\ u_{M,t} \end{bmatrix}$$

then

$$\Sigma_{SIM} = \text{var}(\Psi) = C\text{var}\overline{\Psi}_t C' + \text{var}(U_t) \tag{4.84}$$

$$\Sigma_{SIM} = \Theta^2.CC' + D \tag{4.85}$$

where Θ^2 . CC' is shared information deficiency and $D = \text{diag}(\sigma^2_{u_1},\ \sigma^2_{u_2},\ \sigma^2_{u_3},\ .\ .\ .\ .\ .\ ,\ \sigma^2_{u_M})$, model specific error variance.

4.4.2 Implementing Variance-Covariance approach

In this section, we discuss the implementation of the optimal forecast method with the variance-covariance approach in forecasting out-of-sample equity premium. This section is divided into three subsections based on the form of uncertainty that the VC approach addresses: i) estimation window uncertainty (EWU), ii) variable-selection uncertainty (VSU), and iii) both EWU and VSU.

4.4.2.1 Applying VC approach to address EWU

Our first application is to combine forecasts from different estimation windows for univariate predictor-based models. To obtain optimal weights, we need to estimate the error-covariance matrix in the validation sample, which requires the estimates of forecast errors generated by individual models across all possible windows. Since the last E data observations are reserved for an out-of-sample estimation exercise by the cross-validation approach. And we further assume that a minimum of w_{min} observations are needed to estimate the forecasting model parameters; this means that this method needs $w_{min} + E$ data points. In our application, we set the minimum estimation window to 180 observations and the CV sample to 60 observations.[12] All possible estimation windows can be defined as $M = T - w_{min} - E$. Here M indicates the total possible models with different estimation windows which is indexed by j ($j = 1, 2, \ldots\ldots M$).

The recursive pseudo out-of-sample forecasting errors for each possible starting point of the estimation window, j, can be computed as:

$$A_\tau(j|T,E) = r_{\tau+1} - x'_\tau \hat{\beta}_{j:\tau}, \quad j = 1,\ 2,\ \ldots,\ T - w_{min} - E \ \text{ and } \tau = T - E,\ \ldots,\ T-1 \tag{4.86}$$

where A_τ represents $E \times M$ error matrix at time τ, $\hat{\beta}_{j:\tau}$ represents the OLS estimate from equation (4.59) using $[j, \tau]$ observation window. Since the number of estimation windows, M, increases by 1 with an increase in time τ, so error matrix also increases with each out-of-sample period. This means that the first out-of-sample evaluation period will contain error-matrix with $E \times 1$ and last period will have $E \times (T - w_{min} - E)$ errors.

Since we have $E \times M$ cross-validation forecast errors and now we can obtain $M \times M$ error covariance matrix using diagonal, sample covariance and the single-index model discussed in the previous section.[13] After estimating the error-covariance matrix now we can find optimal weights under our error-covariance (EC) ω_j as:

$$\begin{aligned} &\underset{\underline{\omega}}{\text{minimise}} \quad \underline{\omega}' \, \Sigma \, \underline{\omega} \\ &\text{subject to} \quad \underline{\omega}'\underline{e} = 1, \qquad \underline{e} \text{ is vector of 1's} \\ &\qquad\qquad\quad \omega_j \geq 0 \end{aligned} \tag{4.87}$$

this will generate $1 \times M$ optimal weights, which can be applied to the out-of-sample forecast. The one-period ahead forecast based on the VC method is given as:

$$\hat{r}_{i,t+1}^{CFWin-VC} = \sum_{j=1}^{M} \hat{\omega}_{i,j,t}(\hat{r}_{i,j,t+1}) \tag{4.88}$$

where $\hat{r}_{i,t+1}^{CFWin-VC}$ is combined out-of-sample forecast of equity premium across M estimation windows with predictor i using variance-covariance approach.

4.4.2.2 Applying VC approach to address VSU

We use the variance-covariance (VC) method to combine forecasts from various univariate predictor-based models, each of which is estimated with a single-window. More precisely, a CF approach that tackles variable-selection uncertainty but uses a non-optimal estimation window. Given that we have N univariate predictor-based models indexed by i ($i = 1,\ 2,\ \ldots,\ N$), the combining forecast based on VC approach, $\hat{r}^{CFMod-VC}$, can be given as:

$$\hat{r}_{t+1}^{CFMod-VC} = \sum_{i=1}^{N} \hat{\omega}_{i,t}(\hat{r}_{i,t+1}^{EXP}) \tag{4.89}$$

where $\hat{r}_{i,t+1}^{EXP}$ is a forecast of the univariate predictor-based model estimated with expanding window (e.g., Rapach et al. 2010) and $\hat{\omega}_{i,t}$ indicates the optimal combining weights obtained through VC approach.

[12] Many studies (e.g., Rapach et al. 2010) use 15 years data for estimation of initial parameters, and 5 year data for holdout validation sample.

[13] Since cross-validation forecasting errors are not zero mean, so to make them unbiased, we follow Figlewski (1983) and remove the mean of individual errors.

Forecasting literature suggests four potential ways to estimate the optimal weights based on the error covariance matrix. i) cross-validation approach; ii) using all past forecasting errors, i.e. estimating covariance matrix with expanding window, iii) using recent forecasting errors, i.e. estimating error covariance matrix with rolling window, and iv) optimal covariance matrix. The main difference between approach (i) and other three approaches (ii to iv) is that (i) considers the forecasting performance in the CV sample; however, approaches ii, iii, and iv consider the past performance.

Since the evidence for these methods is mixed, we use all four. A review of the literature on portfolio optimisation shows that several studies, including Fleming et al. (2003), Hautsch & Voigt (2019), and others, suggest the use of rolling covariance, in which estimates are re-estimated at each time. Considering this, our rolling window covariance consists of 15 years rolling window by using equation (4.59). On the other hand, the optimal covariance matrix is based on the minimisation of combined forecast variance. More specifically, at each period, we get the covariance matrix $\hat{\Sigma}$, optimal weights $\underline{\omega}^*$ and composite forecast variance $\underline{\omega}^{*\,\prime}\ \hat{\Sigma}\ \underline{\omega}^*$ with all possible window lengths with minimum of 15 years. Next, we choose the optimal length one that produces the minimum composite variance. Next, the combining weights under optimal window length are applied to the out-of-sample forecasts of the individual model obtained through expanding window.

4.4.2.3 Applying VC approach to address both EWU and VSU

This section introduces some of the methods for simultaneously addressing EWU and VSU problems.

4.4.2.3.1 A panel approach of Combining Forecasts In this section, we discuss the panel approach to combine various forecasts from both univariate predictor-based models and estimation windows. As shown in section (4.4.2.1) that a validation sample approach for each predictor-based model would generate $E \times M$ forecast errors which are used as input for the error-covariance matrix. Instead of estimating N separate error-covariance matrices at each period, we use a panel of forecast errors ($E \times N \times M$) that requires estimation of a single error-covariance matrix. After estimating error covariance matrix, we can obtain the optimal weights through equation (4.87) and define the combined forecast as:

$$\hat{r}_{t+1}^{CFPanel-VC} = \sum_{g=1}^{P} \hat{\omega}_{g,t}(\hat{r}_{g,t+1}) \tag{4.90}$$

where $\hat{r}_{t+1}^{CFPanel-VC}$ is combined forecast across estimation windows and models. $\hat{\omega}_{g,t}$ indicate the optimal weights obtained through variance-covariance approach.

4.4.2.3.2 Selection-Combination (SELCOM) approach Under this approach, we first select the optimal window for each individual predictor-based model and then combine them using the variance-covariance approach. Specifically, at each period, we get the individual forecasts with all possible estimation windows with minimum w_{min} observations, and by following Pesaran & Timmermann (2007), we compute MSFE for each estimation windows in the CV sample. We then choose the optimal window length as one with minimum MSFE. In the next step, we obtain the optimal weights for the individual model based on the variance-covariance approach discussed earlier and then form the combined forecast as:

$$\hat{r}_{t+1}^{SELCOM} = \sum_{i=1}^{N} \hat{\omega}_{i,t}(\hat{r}_{i,t+1}^{*}) \tag{4.91}$$

where $\hat{r}^*_{i,t+1}$ indicates the forecasts of the univariate predictor-based model with optimal estimation window based on MSFE in the validation sample. And $\hat{\omega}_{i,t}$ indicates the optimal combining weights for each model i are obtained through the variance-covariance approach. We consider all four methods for estimating optimal weights discussed in section (4.4.2.2).

4.4.2.3.3 Combination-Combination (COMCOM) approach This strategy is similar to that of SELCOM. However, in the first step, we replace the selection with combination, and the CF across windows for univariate predictor-based models is based on the VC method. The second combining step is the same as the SELCOM method described in the previous subsection.

$$\hat{r}^{COMCOM}_{t+1} = \sum_{i=1}^{N} \hat{\omega}_{i,t}(\hat{r}^{CW}_{i,t+1}) \tag{4.92}$$

4.5 Data and Benchmark Models

4.5.1 Data

We follow the mainstream equity premium predictability literature (e.g., Goyal & Welch 2008, Rapach et al. 2010, and others) to construct the equity premium as well as the economic and financial predictor variables. The target return to be predicted is the market return, which is described as the S&P 500 index log return (including dividends) minus the log return on one-month T-bill rate, available from Amit Goyal's website. The predictor variables are from Goyal & Welch (2008), who also provide comprehensive descriptions of the data, including their origins. The dataset includes 14 variables considered relevant in predicting equity premium in past empirical studies.[14] These variables include stock characteristics (the dividend yield (DY), the dividend-price ratio (DP), the dividend-payout ratio (DE), the earning-price ratio (EP), the book-to-market ratio (BM), the net equity expansion (NTIS), and the stock variance (SVAR)), interest rate related variables (the Treasury bill rate (TBL), the long-term return (LTR), the long-term yield (LTY), the term spread (TMS), the defaults-return spread (DFR), and the default-yield spread (DFY)), and inflation (INFL) to represent the macroeconomy. We use monthly data for all these variables spanning from July 1926 to December 2018.[15]

4.5.2 Benchmark Models

To analyse the performance of our various optimal forecast approaches (see section 4.4) relative to the standard equity premium prediction models, we consider various benchmark models from existing forecasting literature. To provide a direct comparison to our variance-covariance approach, we divide our benchmark models into three groups based on the type of uncertainty they address: i) estimation window uncertainty (EWU), ii) variable-selection uncertainty (VSU), and iii) both EWU and VSU.

4.5.2.1 Standard approaches to address EWU only

The various well-known methods used in literature to address the issue of EWU for a single model are included in our first group of benchmark models. These benchmarks allow us to directly compare forecast performance with our variance-covariance approach that optimally combines the forecasts of a single model based on different estimation windows. We denote the forecast for any given individual predictor-based model i under one of the benchmark models b as $\hat{r}^{EWU^*}_{i,t+1,b}$. The details on these model are given in section (4.2.3).

[14] We use an updated version of the dataset available at http://www.hec.unil.ch/agoyal/

[15] The transformations that we use to build the predictor variables are given in Table 1 of Chapter 2.

4.5.2.2 Standard approaches to address VSU only

In the next group of benchmark models, we consider well-known models that are thought to solve the issue of VSU but are not based on an optimum estimation window. The detail on these models can be found in section (4.2.2) and Table 4.2. The primary reason for including these strategies is that they typically use an expanding (recursive) window and presume that either no structural breaks exist or that the applied strategy often addresses parameter instability issues. However, studies such as Pesaran & Timmermann (2007), Pesaran & Pick (2011), Rossi & Inoue (2012), Pesaran et al. (2013), and Zhang et al. (2020) show that EWU should be tackled separately. In particular, when compared to some of the benchmark models given in section (4.2.2) that only address VSU, Pesaran et al. (2013), Zhang et al. (2020), and Wang et al. (2020) show that considering VSU and EWU simultaneously improves forecast accuracy.

4.5.2.3 Standard approaches to address both EWU and VSU

Our final set of benchmark models includes models that first address the problem of EWU and then take a simple average across all models to address the issue of VSU. The main benchmark under this approach is the average-average (AveAve) of Pesaran et al. (2013), where the first step involves taking a simple average across estimation windows for individual predictor-based models (AveW). Then a simple average across models is taken to reach the final forecast (AveAve). We extend our benchmark set in this category to all the approaches discussed in the previous section (4.5.2.1) which address the EWU issues. Given that $\hat{r}^{EWU^*}_{i,t+1,b}$ is the optimal or final forecast for i predictor-based model using approach b, the average forecast across model then be given as:

$$\hat{r}^{Ave-PI^*}_{b,t+1} = \frac{1}{N}\sum_{i=1}^{N}\hat{r}^{PI^*}_{i,t+1,b} \tag{4.93}$$

The inclusion of these models as benchmarks allows us to compare our SELCOM and COMCOM models more effectively. For example, a direct benchmark for our SELCOM model will be models that choose the best window for each individual predictor-based model based on MSFE in the CV sample and then take a simple average across all models. The only distinction between these two methods is how forecasts across models are combined, since we use a variance-covariance approach to combine forecasts across models based on forecasting results rather than a simple average as in benchmark models. Similarly, any method that combines forecasts from different estimation windows first and then takes a simple average across models will be an excellent benchmark for our COMCOM approach.

4.6 Empirical Results

The out-of-sample forecasting results of the equity risk premium from January 1947 to December 2018 are discussed in this section. There are six subsections in this section. First, we discuss the results of the variance-covariance (VC) approach in addressing estimation window uncertainty (EWU), variable-selection uncertainty (VSU), and both EWU and VSU. The VC approach's findings are then compared to the various benchmark models discussed in section (4.5.2). The third subsection compares the results of the VC approaches and benchmark models over recession and expansion periods. In subsection four, the findings on economic significance based on the mean-variance strategy are discussed. The importance of individual predictors in combined equity premium forecasts are discussed in subsection 5. The last subsection discusses the evidence on the linkage between equity premium forecasts generated by our VC approaches and the real economy.

4.6.1 Results from VC approach

This section discusses the results of variance-covariance approach. Following the literature, we use Campbell & Thompson (2008) out-of-sample R^2_{OOS} metric to evaluate the forecasting performance of all the models considered. The R^2_{OOS} measures the reduction in mean squared forecasting error (MSFE) relative to an alternative model, which is a historical average in our case. A positive R^2_{OOS} suggests the model of interest outperforms the benchmark model. In order to test whether our model produces less MSFE than the benchmark model, we use Clark & West (2007) statistic.

4.6.1.1 Results from VC approach addressing EWU only

Table 4.3 reports the results from univariate predictor-based models based on our VC approach. These models are summarised in Table 4.1 and details are given in section (4.4.2.1). The results from VC approach are based on diagonal, sample, and single index model (SIM) covariance matrices. All of these covariance matrices are calculated using the Cross-validation (CV) method, which uses an out-of-sample holdout sample (cross-validation sample) to estimate forecasting errors, error-covariance matrix, and optimal weights.

[Insert Table 4.3 about here]

We report the out-of-sample R^2 (R^2_{OOS}) statistics in percentage form to compare the forecasts provided by prediction models with the historical average forecast. The results show that all univariate predictor-based models under VC-DIAG and VC-SAMPLE perform poorly as R^2_{OOS} is negative for all models. However, for VC-SIM, three individual models, DE, SVAR, and TMS, yield positive R^2_{OOS}. This indicates that the VC-SIM approach outperforms the VC-SAMPLE and VC-DIAG approaches by achieving the highest R^2_{OOS} for all individual models. This is in line with studies such as Figlewski (1983) that performance of sample covariance is poor due to high estimation error, and diagonal covariance matrix does not consider the correlation among forecasting errors. This indicates that the optimal weights under the single-index model help to improve forecasting accuracy. However, despite accounting for estimation window uncertainty (EWU) using a variance-covariance method, univariate predictor-based models fail to outperform the historical average in general. The best performing model, DE, has an R^2_{OOS} of 0.23%, which is significant at 5% based on Clark & West (2007) statistic. However, according to Campbell & Thompson (2008) argue that R^2_{OOS} in excess of 0.5% can be considered a significant improvement over the benchmark model (historical average). However, all of the models using VC approaches have an R^2_{OOS} of less than 0.5%.

4.6.1.2 Results from VC approach addressing VSU only

Table 4.4 reports the forecasting results of our VC approach addressing variable-selection uncertainty only by combining forecasts across univariate predictor-based models. These models are summarised in Table 4.1 and details are given in section (4.4.2.2). Following mainstream equity premium prediction (e.g., Rapach et al. 2010, Neely et al. 2014), each univariate model is estimated with an expanding window. The variance-covariance method produces results using three types of covariance matrices: diagonal, sample, and single-index model. Moreover, each covariance matrix is estimated using four different ways, including cross-validation (CV), rolling-window (ROLL), expanding-window (EXP), and optimal-window (Optimal). CV denotes a cross-validation approach similar to the one used in combining forecasts across different estimation windows discussed in the previous section. ROLL (EXP) indicates that the historical forecasting errors for individual models are estimated with a rolling window of 15 years (expanding window including all the observations). The optimal covariance matrix at a given time effectively determines the optimal past observations for estimation of the error-covariance matrix using minimal variance criteria (see section (4.4.2.2) for details).

[Insert Table 4.4 about here]

Results in Table 4.4 suggest that CF across univariate predictor-based models based on VC approach improves the forecasting accuracy for most of the models. All the models except those which use expanding window for estimating the covariance matrix achieves an R^2 in excess of 0.50%, which is really a big improvement compared to various methods considered in Table 4.3 for addressing EWU issues for univariate predictor-based models. Moreover, consistent with findings on univariate models given in Table 4.3, we find that VC-SIM outperforms VC-DIAG and VC-SAMPLE by achieving an R^2_{OOS} of 0.754% and 0.804% for covariance matrix based on 15 years rolling window and optimal covariance, respectively. These R^2_{OOS} are significant at 1% based on Clark & West (2007) statistic.

4.6.1.3 Results from VC approach addressing EWU and VSU

Given that addressing variable-selection uncertainty using a VC approach improves equity premium forecasting accuracy, we are now focusing on addressing EWU and VSU at the same time using our VC approach. These models are summarised in Table 4.1 and details are given in section (4.4.2.3). Table 4.5 reports the forecasting results of our VC approach addressing EWU and VSU simultaneously by combining forecasts across estimation windows and individual predictor-based models. Panel A presents the results of a panel approach in which a single error-covariance matrix is estimated to obtain the optimal weights for individual predictor-based models and estimation windows. We begin by comparing the results of the three covariance matrices that were used in this analysis. Results show that among diagonal, sample and single-index model, the Panel (SIM) obtains the highest monthly out-of-sample R^2 with 0.922%. Panel(DIAG) and Panel (SAMPLE) achieves R^2 of 0.838% and 0.792%, respectively. All these R^2 statistics are significant at 5% level based on Clark & West (2007) test.

[Insert Table 4.5 about here]

Next, we analyse the optimal weights generated by each approach using original and bootstrapped samples. We find that the optimal weights under the panel approach are more volatile for Panel (SAMPLE) than Panel (SIM). On the other hand, the diagonal matrix generates more stable weights. Despite the better performance of Panel (SIM), in some periods, the model is biased towards a single univariate model where optimal weights are spread over different estimation windows of the same predictor-based model. This implies that a panel approach can be biased toward accounting for EWU while neglecting to account for VSU. This motivates us to use a two-step approach of Wang et al. (2020), in which they demonstrate that it is best to address EWU for univariate predictor-based models first and then take a simple average across forecasts of univariate models obtained from the first step. This way, all univariate predictor-based models have the opportunity to contribute towards the ultimate forecast. Following Wang et al. (2020), we first apply VC approach to EWU by combining forecasts of individual predictor-based models across windows and in the next step, we take a simple average across individual models. We call this approach Mean (VC-DIAG), Mean (VC-SAMPLE), and Mean (VC-SIM) for diagonal, sample, and SIM covariance matrix, respectively. Results in Panel B of Table 4.5 suggests that there is significant improvement in R^2 for SIM as Mean (VC-SIM) achieves an R^2 of 1.124%, significant at 5% level based on Clark & West (2007) test.[16]

[16]The difference between R^2 of Mean (CV-SIM) and Panel (SIM) is significant at 5% level using bootstrapped p-values.

Next, we replace the simple average in the second step of the two-step process with optimal weights for univariate predictor-based models obtained through VC approach discussed in the previous section. Panel C of Table 4.5 reports the results of our SELCOM approach where by following Pesaran & Timmermann (2007), we first compute MSFE for each estimation windows in CV sample. We then choose the optimal window length as one with minimum MSFE. In the next step, we obtain the optimal weights for the individual model based on VC approach discussed earlier and then form the combined forecast (see section (4.4.2.3.2) for details). Results show that SELCOM (SIM) outperforms SELCOM (DIAG) and SELCOM (SAMPLE) in all specifications of covariance matrix estimation. In addition, we find that Roll covariance performs better than Exp, which implies considering recent observations and forecasting errors improves forecasting accuracy. Moreover, the best performing model is SELCOM (SIM), estimated with optimal covariance matrix, which achieves an R^2 of 1.215% which is significant at 5% level based on Clark & West (2007) test.[17]

Panel D of Table 4.5 reports the results of our COMCOM approach, which applies VC approach to both CF across windows for univariate predictor-based models and in the second step, it also applies the VC approach across models to reach the final forecast. Results are consistent with SELCOM approach that COMCOM (SIM) outperforms COMCOM (DIAG) and COMCOM (SAMPLE) and the best performing model is COMCOM (SIM-Optimal) which achieves an R^2 of 1.426% which is significant at 5% level based on Clark & West (2007) test.[18]

To support these results, following Goyal & Welch (2008), Rapach et al. (2010) and many others, we present time-series plots of the cumulative square forecasting error (CumSFE) for our VC-SIM approaches relative to historical average (HA) in Figure 4.3 for the full sample. This provides an informative graphical representation about the consistency of a given model in outperforming the benchmark model over the entire out-of-sample period. When the curve for a given model in Figure 4.3 increases, the model outperforms the historical average, whereas when the curve decreases, the model underperforms the historical average. The horizontal zero line to correspond to the start of the out-of-sample period; the plot easily indicates whether our VC-SIM models have a lower MSFE than the historical average for any specific out-of-sample period. The plot clearly shows if our VC-SIM models have a lower MSFE than the historical average for any given out-of-sample time, with the horizontal zero line corresponding to the start of the out-of-sample period. Figure 4.3 shows that all of our four VC-SIM approaches always have a positive curve, implying that our models outperform the historical average for every out-of-sample period.

To summarise our results on our VC approach, the SIM covariance outperforms diagonal and sample covariance in all the specifications. Forecasting results for univariate predictor-based models estimated with VC approaches fail to outperform the historical average. These findings are consistent with Tian & Zhou (2018), who report that addressing EWU for univariate predictor-based models alone does not outperform the historical average. The Panel (SIM) that addresses the EWU and VSU at the same time, on the other hand, produces significantly lower MSFE than the historical average. Moreover, we find that a two-step approach addressing EWU and VSU separately strengthens the panel approach, which is consistent with Wang et al. (2020). Our three approaches Mean(VC), SELCOM (VC), COMOM (VC) under SIM framework achieves an R^2 of 1.124%, 1.215%, and 1.426%, respectively.

[17] The difference between R^2 of SELCOM (SIM-Optimal) and all other approaches under SELCOM is significant at 5% level using bootstrapped p-values.

[18] The difference between R^2 of COMCOM (SIM-Optimal) and all other approaches under COMCOM is significant at 5% level using bootstrapped p-values.

4.6.2 VC-SIM approach compared with benchmark models

In this section, we analyse the performance of our VC approaches relative to the standard benchmark models discussed in section (4.5.2).

4.6.2.1 Comparing with benchmarks addressing EWU

Table 4.6 reports the results from univariate predictor-based models based on our VC approach and various benchmark models. There are two panels in this: Panel A displays results from VC approach based on diagonal, sample, and single index model (SIM) covariance matrices. All of these covariance matrices are calculated using the Cross-validation (CV) method, which uses an out-of-sample holdout sample (cross-validation sample) to estimate forecasting errors, error-covariance matrix, and optimal weights. Panel B presents forecasting results of the benchmark models listed in Table 4.2.

[Insert Table 4.6 about here]

To compare the forecast produced by prediction models with the historical average forecast, we report the out-of-sample R^2 statistics in percentage form. Values in bold indicate the best forecasting approach for a given univariate predictor-based model. Results show that VC based on SIM (VC-SIM) performs better than the expanding window (EXP) approach for 12 of 14 univariate economic predictor-based forecasting models. The EXP performs better than VC-SIM for TBL and DFR only. Results also confirm the poor performance of the rolling window approach (ROLL), where each forecasting model is estimated with the last 15 years data. Our VC-SIM outperforms all individual models when estimated with the ROLL approach. In addition, ROLL also performs poorly compared to EXP and other approaches that do not completely discard historical observations. This is also consistent with Pesaran & Timmermann (2007), who argue that past data may also be informative and produce accurate forecasts.

If we compare the results of our VC-SIM with other approaches used to address the parameter instability, then we find that VC-SIM performs better than all the approaches for 6 of 14 univariate economic predictor-based forecasting models. Consistent with Tian & Zhou (2018), we find that the robust optimal weighting method (RobW1 and RobW2) can also improve the forecasting performance of individual models as it outperforms the traditional expanding (rolling) for 10 (14) of 14 models. However, RobW1 and RobW2 together can outperform all the approaches for 5 of 14 models. However, consistent with Tian & Zhou (2018), although most forecasting strategies addressing parameter instability considered in this study can enhance the forecasting efficiency of economic predictor-based models, while still typically failing to outperform historical average benchmark forecasts based on R^2_{OS}.[19]

4.6.2.2 Comparing with benchmarks addressing EWU and VSU

The predictive accuracy of our VC-SIM approaches is compared in Table 4.7 with a comprehensive array of alternative approaches considered in the literature to address the issues of VSU, and both EWU and VSU. Panel A displays results from VC-SIM approaches whereas Panel B and C display forecasting results of the benchmark models listed in Table 4.2.

[19] In unreported results, we also perform the analysis to evaluate the economic significance of all the models based on the mean-variance trading strategy discussed in section (4.3.4.2). The results are consistent with our main findings based on out-of-sample R^2 that almost all the univariate predictor-based models fail to outperform the historical average.

[Insert Table 4.7 about here]

Panel B reports the results of benchmark models, which are developed to address EWU. However, following Wang et al. (2020), we apply a two-step process to enable all these models to address the issues of EWU and VSU simultaneously. More specifically, for each approach, we first estimate the forecasts for univariate predictor-based models, and in the second step, we take a simple average across individual forecasts from the first step. Results show that this approach works well as all the models considered produce lower MSFE relative to the historical average. The lowest R^2 is 0.503%, achieved by Mean (ROLL), whereas Mean (ROC) achieves an R^2 of 0.838% which remain the highest among benchmark models. It is interesting to see that the AveAve approach of Pesaran et al. (2013), which takes simple average across univariate predictor-based models and windows, achieves an R^2 of 0.808% which is the second-highest among benchmark models. When we compare these findings to the results of our VC-SIM approaches in Panel A, we can see that all the four approaches outperform the benchmark models. The best performing model, COMCOM-SIM, achieves an R^2 of 1.426%, which is significantly higher than the best performing benchmark model, Mean (ROC), which achieves an R^2 of 0.838%.[20]

Next, we compare the results with standard approaches given in Panel C of Table 4.7, which have been widely used in forecasting equity premium. The forecasts under all these approaches are based on expanding window, which assumes no structural breaks. These approaches include the Kitchen Sink model, dimension reduction, shrinkage, and combining forecasts. Results suggest that CF approaches outperform the other methods among benchmark models given in Panel C. The simple average across forecasts from univariate models achieves an R^2 of 0.689%. However, DMSFE with a discount factor of 0.9 ($\theta = 0.9$) model performs slightly better by achieving an R^2 of 0.731%. This shows that considering recent forecasts improves forecasting accuracy. But surprisingly, none of the models from CF outperforms our VC-SIM (Optimal) approach given in Table 4.4, which combines forecasts of univariate models using VC-SIM approach where each individual models are estimated with expanding window. This suggests that the implementation of VC approach improves not only the forecasting accuracy for univariate predictor-based models given in Table 4.6 but also CF based on a simple average. If we compare these results with our VC-SIM approaches addressing EWU and VSU simultaneously, we see that our all four approaches Panel (SIM), Mean (VC-SIM), SELCOM-SIM, and COMCOM-SIM outperform all the approaches by achieving R^2 of 0.922%, 1.124%, 1.215%, and 1.426%, respectively. This suggests that accounting for EWU and VSU together can improve the forecasting accuracy for equity premium.

4.6.3 Out-of-sample results over recession and expansion

To see how various models considered in this study perform during the business cycle, we compute R^2_{OS} for NBER-dated business cycle expansions (EXP) and recessions (REC):

$$R^2_{c,OOS} = 1 - \frac{\sum_{t=m}^{T} I_t^c \left(r_{t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{t=m}^{T} I_t^c \left(r_{t+1} - \overline{r}_{t+1}\right)^2} \qquad \text{for } c = \text{EXP}, \text{REC} \tag{4.94}$$

where I_t^{REC} (I_t^{EXP}) is an indicator function that takes a value of one when t is a recession (expansion) and zero otherwise.

Table 4.8 shows the results of VC-SIM approaches and benchmark models over recession and

[20]The difference between R^2s is significant at 5% using bootstrapped p-values.

expansion periods for 1-month returns. Results show that our VC-SIM approaches can predict equity premium over both recession and expansion periods. Our COMCOM-SIM achieves an R^2 of 0.934% in expansion and 2.811% during recession. Results are also consistent with the other three VC-SIM approaches. These results provide strong evidence that our VC-SIM approaches perform better in recession. If we see the performance of benchmark models, the results show that none of the models outperforms our approach by producing a higher R^2 across recession and expansion periods.

Panel B reports the results of a two-step process for addressing parameter instability first and then taking a simple average across models. Surprisingly, all these approaches perform good in both recession and expansion, but performance is better in the recession period. This is an important finding because this suggests that accounting for parameter instability and variable-selection uncertainty at the same time significantly improves the forecasting accuracy in both expansion and recession. This evidence is further supported by the findings of models given in Panel C that only accounts for variable uncertainty. Results suggest that none of the models consistently produce higher R^2 across recession and expansion periods. For example, PCR3 achieves an R^2 of 2.954% in recession, but it performs poorly in expansion as it achieves R^2 of -0.354%, which means it even fails to outperform the simple historical average.

An essential finding is the superior performance of our VC-SIM approaches in recessions because predictive awareness of economic fundamentals over recessions is more important to an investor. Because of the high volatility experienced during recessions, the historical mean is a poor indicator (see Li & Tsiakas 2017, for more details). Moreover, Kacperczyk et al. (2016) demonstrates that the economic outlook influences how investors interpret information. Since stocks carry more aggregate risk during recessions, fund managers are more concerned with aggregate shocks. In summary, in recessions, economic fundamentals information is most important for predicting the equity premium, and we find that this is when our predictive models (VC-SIM) work best.

4.6.4 Equity premium predictability and asset allocation

Table 4.9 presents the results of portfolio performance for a mean-variance investor with a degree of risk aversion of five. Consistent with our statistical findings, portfolio performance results show that our VC-SIM approaches perform better than benchmark models. For example, our COMCOM-SIM delivers an annualised utility gain (ΔCER) of 2.524%, which are higher than any of the benchmark models. The AveAve model performs the best among all benchmark models with (ΔCER) of 1.421% whereas simple average across model (Mean) delivers (ΔCER) of 1.076%. The net transaction costs results remain the same as our COMCOM-SIM model outperforms all the benchmarks even after considering 50bp transaction cost can generate (ΔCER) of 2.019%. Moreover, our COMCOM-SIM achieves an annualised Sharpe ratio of 0.497, which remains the highest among all the models.

Overall, we provide the evidence on out-of-sample predictability of the equity premium. We also show a high economic benefit in using our VC-SIM approaches in the context of a mean-variance strategy.

[Insert Table 4.9 about here]

4.6.5 Which predictors matter?

In this section, we investigate the relative importance of individual predictors in combined forecast over time for our COMCOM-SIM approach. We use three different approaches to account for the importance of each predictor in the equity premium forecast. Table 4.10 reports the importance of each predictor, which is discussed below.

[Insert Table 4.10 about here]

4.6.5.1 Importance as means of reduction in R^2_{OOS}

In the first method, we follow Gu et al. (2020) and calculate the reduction in out-of-sample R^2_{OOS} by excluding a particular predictor in each training sample, and then we get an average of these to obtain importance measure.

$$\Delta R^2_{OOSi} = \frac{1}{S}\sum_{e=1}^{S}(R^2_{N,e} - R^2_{N-1,e}) \tag{4.95}$$

where ΔR^2_{OOSi} is reduction in out-of-sample R^2 due to absence of ith predictor, $R^2_{N,e}$ is R^2 at time e of evaluation period using all predictors and $R^2_{N-1,e}$ is R^2 at time e without ith predictor. Now the importance factor can be calculated as:

$$\Phi_{1i} = \frac{\Delta R^2_{OOSi}}{\sum\limits_{i=1}^{N} \Delta R^2_{OOSi}} \tag{4.96}$$

where Φ_i is the variable importance of each predictor, which is normalised to sum to one.

Figure 4.4 plots the R^2 based predictor importance. We can see that NTIS is the most influential variable, as the absence of this variable from the dataset contributes 19.8% in an overall reduction in R^2. This is obvious as NTIS is one of the proxies to measure sentiment index, which influences the equity market (Baker & Wurgler 2006), and since this variable is not highly correlated with other variables so other variables may not take the effect of NTIS. DFR and DY stand second (16.6%) and third (13.4%), respectively. Since most of the predictors are highly correlated, we classify them: i) equity - which includes DP, DY, EP, DE, BM and SVAR, and ii) non-equity - which includes DFR, DFY, LTY, LTR, TBL, TMS, INFL. If we sum individual contributions across each group, then we find that the equity group plays a more significant role with 54.1% contribution in the composite forecast of the equity premium.

[Insert Figure 4.4 about here]

4.6.5.2 Importance as means of contribution in optimal forecast

In our next approach, we take the average of optimal weights of individual predictors obtained at each of the evaluation period e.

$$\overline{\omega}^*_i = \frac{1}{S}\sum_{e=1}^{S}\omega^*_{i,e} \tag{4.97}$$

where $\overline{\omega}^*_i$ is an average weight for predictor i, and $\omega^*_{i,e}$ is optimal weight for predictor i at time e.

To normalise the importance of predictor to one we use:

$$\Phi_{2i} = \frac{\overline{\omega}_i^*}{\sum_{i=1}^{N} \overline{\omega}_i^*} \tag{4.98}$$

where Φ_{2i} indicates the importance of ith predictor in composite forecast.
Figure 4.5 plots the optimal-weight based importance of each predictor over the full sample and NBER-dated recession periods. The main difference between R^2-based importance and importance represented in Figure 4.5 is that the R^2-based measures are averaged over all sample splits. In contrast, the optimal-weight based importance is based on the full sample (1947:01 - 2018:12) only. Results show that BM is the most influential variable in the full sample, expansion, and recession periods with a contribution of 12.61% in composite weight in the full sample. However, BM's contribution in recession increases to 16.03% and decreases to 12.05% in expansion, but it is still highest in all predictors. The predictors such as DFR, TBL, NTIS, and EP together contribute 42.4% with an individual contribution of 10.73%, 10.71%, 10.49% and 10.47%, respectively. SVAR seems to be the least important, with a contribution of only 1.28%. In expansion periods, the top three most influential predictors are BM, DFR, and TBL, with a contribution of 12.05%, 11.68%, and 11.15%, respectively. Whereas BM, EP, and LTR are highly contributing predictors at 16.03%, 12.51%, and 10.64%.

[Insert Figure 4.5 about here]

Next, we investigate the predictor importance over the periods of optimism and pessimism. Following Baker & Wurgler (2006), we define the optimism period with a sentiment index over zero and a negative value as pessimism. Since the data for sentiment index is available from 1963:01, therefore this analysis also provides the importance over sub-sample. Figure 4.6 plots the optimal weight-based importance for sample 1963:01-2018:12. Results show that DFR, NTIS, and EP are the most significant variables over the overall sub-sample with an individual contribution of 15.45%, 12.56%, and 11.76%, respectively. When we look at the optimism and pessimism periods, the optimism periods are dominated by DFR and DFY together, and they contribute 32.33% with an individual contribution of 16.51% and 15.82%, respectively. The most important thing is the role of NTIS as it seems to be the most important variable in pessimism periods with a contribution of 17.04% which is highest among all the predictors, TBL also gains significant importance in pessimism by standing second with 15.65% which was only 3.99% in optimism. The contribution of DFY falls from 15.82% (optimism) to 2.23% (pessimism), whereas DFR remains the most consistent variable through optimism (16.51%) and pessimism (14.53%) periods.

[Insert Figure 4.6 about here]

Next, we identify the optimism and pessimism periods in NBER-dated recession periods, which leads to four groups: i) expansion-optimism; ii) expansion-pessimism; iii) recession-optimism, and iv) recession-pessimism. 4.7 plots the optimal weight-based importance for sample 1963:01-2018:12. Results show that EP (22.38%) and DFR (19.08%) are the most important predictor in expansion-optimism and recession-optimism periods, respectively. In contrast, NTIS contributes the most in both expansion-pessimism and recession-pessimism periods with 20% and 16.74%, respectively.

To summarise the discussion on the overall importance of individual predictors, we can conclude that the NTIS is the most influential variable throughout all the samples, but it has less importance in optimism and optimism-recession periods, and the most significant influence is in expansion-pessimism periods. In financial ratios, EP is a consistent predictor with an average contribution of more than 10% in all periods except expansion-recession, where it only contributes 4.09%. The predictor, DFR, seems to be the most influential and consistent in macro variables in all periods, but it only contributes 4.95% and 4.27% in recession and expansion-optimism, respectively.

[Insert Figure 4.7 about here]

4.6.5.3 Contribution in optimal forecast overtime

The first two measures give an overall picture of predictor importance, but we cannot get an idea about the importance of predictors over time. Considering this, we plot the contribution (optimal weight) of each predictor at each time. Since the weight of composite forecast sums to one so, we can easily get an idea about the contribution of each predictor. Figure 4.8 shows that predictors DP and DY frequently appear in the 1970s, with DY achieving the highest weight of 54%, and most of the time, they contribute near recession points. DY has significantly appeared after the global crisis of 2008. Variable EP appears at the beginning (first five years) of out-of-sample with an average weight of about 40%, and then it disappears for a long period. Then it contributes heavily with an average 65% weight between the recession of 2001-2008 and disappears entirely after the global crisis of 2008. The exciting results are on SVAR as reported earlier that it contributes less in the composite forecast, and we can see that it appears between 1980-1985; otherwise, it appears occasionally. The next variable is BM, which dominates the other predictors between 1947 to 1955 with an average contribution of 70% and then disappears till the beginning of the 1970 recession, but since then, it never appears. The previous section shows the importance of the variable NTIS, and the time series plot confirms this by showing that NTIS variable is an essential variable in the expansion period, and it frequently appears in the period 1985 to 1995 and then two years post-global crisis 2008. In macro variables, TBL appears more regularly between 1960 to 1980 with an average contribution of 40%, whereas LTY only appears in the last five years of the sample (2013-18). The variables LTR, TMS, DFY, and DFR, seem to contribute more in periods between 1955 to 1990 by appearing more frequently in periods 1972-1985, 1960-1985, 1980-1990 and 1990-2000, respectively. The variable INFL shows its importance in periods 1955-1960 and post-global crisis 2008.

[Insert Figure 4.8 about here]

Next, we classify the predictors into equity (DP, DY, EP, DE, SVAR, BM, and NTIS) and non-equity (TBL, LTY, LTR, TMS, DFY, DFR, and INFL). Figure 4.9 shows the importance of each category as we can see that the financial variables dominate in the early sample (1947-1955), where variables like BM, NTIS, and EP contribute heavily. In contrast, variable LTR is the only variable that contributes from the non-equity group. Non-equity variables show absolute dominance in periods 1956-1968 and 1975-1980, where equity variables hardly contribute. The period between 1980 and 2001 shows the shared contribution of both groups, but the financial group dominates between the recession of 2001 and global crisis 2008, and it is interesting to see that the variables DP, DY, EP contribute heavily with an average 98% weight. Post global crisis, there is a shared contribution of each group where variables DY, NTIS, LTY, and INFL seem to play a vital role in predicting equity premium.

[Insert Figure 4.9 about here]

4.6.6 CF with VC-SIM and link to real economy

In this section, we analyse whether the equity premium estimates generated by our VC-SIM approaches are related to the real economy. These ties can provide additional support for our prediction methods as well as an economic justification for the out-of-sample forecasting improvements.

4.6.6.1 Equity premium forecast and NBER business cycles

Increased risk aversion, according to Fama & French (1989) and Cochrane (1999), generally requires a higher risk premium during economic downturns, resulting in equity premium predictability. Considering this, we analyse variations in equity premium forecasts generated by our VC-SIM approaches over the business cycle. More precisely, we analyse whether forecast fluctuations are strongly related to NBER-dated recession and expansion phases. Figure 4.10 plots the equity premium forecasts along with NBER-dated recession bars shows that there are well-defined patterns in our equity premium forecast as we can see a sharp increase in equity premium forecasts in periods of recession and decline in expansion. The six highest points achieved by our forecasts are during recessions, and if we compare our forecasts with the historical average, then we see that the historical average is smooth and does not respond to business cycles. Overall, our VC-SIM approaches generate equity premium forecasts that tend to follow NBER business-cycle phases, as shown in Figure 4.10. This forecasts' behaviour is consistent with the findings of Fama & French (1989) and Cochrane (1999).

[Insert Figure 4.10 about here]

4.6.6.2 Forecasting gains during "Good" and "Bad" growth periods

Following Rapach et al. (2010), we next analyse CF generated through our VC-SIM approaches over different regimes based on the economic growth rate. Specifically, we use the top, middle, and bottom thirds of sorted rates of growth to characterise "good", "normal," and "bad" regimes. Consistent with Rapach et al. (2010), our analysis uses growth rates of three variables: i) real profit, ii) real GDP, and iii) real net cash flow. We calculate R^2_{OS} for these three regimes. Results in Table 4.11 indicate that out-of-sample gains are frequently clustered in lower growth periods for our VC-SIM approaches. When comparing periods of low-growth periods to normal-growth, the R^2_{OS} is higher for all three variables. For real GDP growth, the R^2_{OS} is around four times greater in low-growth periods than the periods of normal-growth. Such differences are even larger when regimes are sorted on real profit growth, but the difference is about three times when we sort by real net cash flow growth.

[Insert Table 4.11 about here]

4.6.6.3 Correlation between equity premium forecasts and macro variables

Following Kelly & Pruitt (2013), we test the correlation between CF generated by VC-SIM approaches and indicators such as macroeconomic activity (GDP growth, industrial production growth and unemployment), macroeconomic uncertainty (volatilities in GDP growth, consumer growth, industrial production, real profit and uncertainty in financial, macro and real activity), sentiment (sentiment index variables of Baker & Wurgler (2006)) and credit (term spread). Table 4.12 reports the correlation results.

[Insert Table 4.12 about here]

First, the countercyclical nature of equity premium estimates is found to be substantially lower in periods of high GDP growth, industrial growth, real profit growth, low book-to-market aggregate ratios, and low unemployment. In periods of high real personal income growth and high Chicago Fed National Activity Index (CFNAI) levels, they are also lower but insignificant. We consider two indicators of recession probability. The first is a real-time probability of recession determined by Chauvet & Potter (2002), Chauvet & Piger (2008). The second is the "Anxious Index" from the Survey of Professional Forecasters, which asks survey panellists "to estimate the probability that real GDP will decline in the quarter in which the survey is taken and in each of the following four quarters." With both measures, we find a strong positive and significant association between expected equity returns and the likelihood of economic contraction.

Next, our findings show a strong association between our equity premium forecasts and macroeconomic uncertainty. We consider several uncertainty measures, which include: i) Bloom et al. (2018) measures - volatility in GDP estimated with a GARCH model, the cross-section standard deviation in the forecasts of industrial production growth made by professionals, and an uncertainty index which is a combination of seven variables of macroeconomic uncertainty and ii) Jurado et al. (2015) - financial, macro and real uncertainty. We find a strong positive association between our equity premium forecasts and these uncertainty measures. Considering the importance of consumption growth volatility from the theory of the long-run risk of Bansal & Yaron (2004), we also consider consumption uncertainty estimated with a GARCH model. Our results show a positive and significant correlation between our estimates of equity premium with COMCOM-SIM and consumption growth volatility with a correlation of 24.86%.

Next, we find a strong association between investor sentiment and equity premium forecasts generated by our VC-SIM approaches, which is consistent with findings of Baker & Wurgler (2006) that in times of high investor sentiment, discount rates decline. We use the investor sentiment index of Baker & Wurgler (2006) and four individual sentiment proxies and find that the sentiment is significantly negatively correlated with our forecasts of the equity premium. Finally, we find that forecasts are significantly positively correlated with default yield spread. However, the correlation between forecasts and the term spread is insignificant.

4.6.6.4 Forecasting Macroeconomic variables

Rapach et al. (2010) show that combining forecast approach with Goyal & Welch (2008) predictors can predict some of the macro variables including real profit growth, real GDP growth, and real net cash flow growth. On this note, we also use the same variables used in predicting equity premium to predict macroeconomic variables, including industrial production growth, real GDP growth, real profit growth, Unemployment rate, and GDP growth, and Chicago Fed National Activity Index, with our VC-SIM approaches. Data is obtained from the FRED dataset. Considering the availability of all variables, our sample consists of 1961:03 to 2018:12, where the first 10 years are used as training sample and the next 5 years for holdout validation sample, which leaves an out-of-sample period 1976:04 to 2018:12. Table 4.13 shows that our approach can predict above mentioned macroeconomic variables by achieving R^2 of 3.315%, 7.924%, 5.734%, 9.723% and 5.891% for real profit growth, industrial production growth, real GDP growth, CFNAI and unemployment rate, respectively.

[Insert Table 4.13 about here]

4.7 Robustness Tests

We establish the robustness of our main equity premium prediction results in many ways. First, we evaluate various sample dates and equity measures and find similar results in all alternative samples and equity premium measures such as unlogged equity premium and Fama & French (1993) excess market factor. Second, we use technical indicators given by Neely et al. (2014) as an alternative dataset and show that our approach produces consistent results with these indicators as well. Moreover, when these technical indicators are taken as the first principal component and included as an additional predictor to Goyal & Welch (2008) dataset, then results are improved. Third, we find strong evidence on good out-of-sample predictability of our approach for predicting excess returns for characteristic portfolios sorted on book-to-market, size, momentum, and industry. Finally, we test whether economic restrictions can improve the results. We use three economic constraints, CT, PTV, and ZWMY (see section 4.2.2.1 for details). Results show that these constraints have a positive impact on most of the models considered in this study. Consistent with Zhang et al. (2019), we find that ZWMY performs better than CT and PTV. Most importantly, consistent with our main results, after imposing these various constraints on all the models, our VC-SIM approaches still outperform the benchmark models. For brevity, we do not report the results of all these tests.

4.8 Conclusion

Combining forecasts (CF) approaches are widely used in economic forecasting, but they are less common in financial forecasting, specifically in the prediction of equity premium Timmermann (2018). This study investigates whether it is worth combining the equity premium forecasts from the same model but with different windows using the variance-covariance approach. Using Goyal & Welch (2008) 14 predictors, this study documents that the optimal CF based on a variance-covariance approach where the error-covariance matrix is estimated with the single-index model can be useful for predicting monthly US excess stock returns over the out-of-sample period of 1947–2018. We also introduce a panel CF approach that combines the forecasts across univariate predictor-based models and windows, which effectively addresses the estimation window uncertainty (EWU) and variable-selection uncertainty (VSU) simultaneously. Two other approaches SELCOM and COMCOM are also introduced. Results show that our approaches improve upon benchmark forecast combination models, as well as the historical average excess return by achieving an out-of-sample R^2 of 1.426% with our COMCOM-SIM model. The improvements in out-of-sample forecast accuracy are significant, both statistically and economically. Our COMCOM-SIM approach has superior market timing abilities, such that a mean-variance investor with a risk-aversion level of five would be willing to pay an annual performance fee of 2.54% to switch from the predictions offered by the historical average benchmark model to those of the COMCOM-SIM for monthly returns.

Moreover, our VC-SIM models can also predict equity premium over both recession and expansion periods. An essential finding is the superior performance of our VC-SIM approaches in recessions because predictive awareness of economic fundamentals over recessions is more important to an investor. This is because investors are more risk-averse during recessions requiring a higher risk premium, and there is also high volatility, making the historical mean a weak forecast (Li & Tsiakas 2017). In addition, Kacperczyk et al. (2016) also shows that the economic outlook affects how investors process information. In recessions, fund managers are more concerned with aggregate shocks because stocks have a higher aggregate risk. In summary, in recessions, economic fundamentals information is most important for predicting the equity premium, and we find that this is when our predictive framework works best.

References

Aiolfi, M. & Timmermann, A. (2006), 'Persistence in forecasting performance and conditional combination strategies', *Journal of Econometrics* **135**(1-2), 31–53.

Avramov, D. & Wermers, R. (2006), 'Investing in mutual funds when returns are predictable', *Journal of Financial Economics* **81**(2), 339–377.

Bai, J. & Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Bai, J. & Perron, P. (2003), 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics* **18**(1), 1–22.

Baker, M. & Wurgler, J. (2006), 'Investor sentiment and the cross-section of stock returns', *The Journal of Finance* **61**(4), 1645–1680.

Bansal, R. & Yaron, A. (2004), 'Risks for the long run: A potential resolution of asset pricing puzzles', *The Journal of Finance* **59**(4), 1481–1509.

Barberis, N. (2000), 'Investing for the long run when returns are predictable', *The Journal of Finance* **55**(1), 225–264.

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. & Terry, S. J. (2018), 'Really uncertain business cycles', *Econometrica* **86**(3), 1031–1065.

Bollerslev, T., Tauchen, G. & Zhou, H. (2009), 'Expected stock returns and variance risk premia', *The Review of Financial Studies* **22**(11), 4463–4492.

Boyd, J. H., Hu, J. & Jagannathan, R. (2005), 'The stock market's reaction to unemployment news: Why bad news is usually good for stocks', *The Journal of Finance* **60**(2), 649–672.

Brown, R. L., Durbin, J. & Evans, J. M. (1975), 'Techniques for testing the constancy of regression relationships over time', *Journal of the Royal Statistical Society* **37**(2), 149–163.

Buncic, D. & Tischhauser, M. (2017), 'Macroeconomic factors and equity premium predictability', *International Review of Economics and Finance* **51**, 621–644.

Campbell, J. Y. & Thompson, S. B. (2008), 'Predicting excess stock returns out of sample: Can anything beat the historical average?', *The Review of Financial Studies* **21**(4), 1509–1531.

Cang, S. & Yu, H. (2014), 'A combination selection algorithm on forecasting', *European Journal of Operational Research* **234**(1), 127–139.

Chan, F. & Pauwels, L. L. (2018), 'Some theoretical results on forecast combinations', *International Journal of Forecasting* **34**(1), 64–74.

Chauvet, M. & Piger, J. (2008), 'A comparison of the real-time performance of business cycle dating methods', *Journal of Business and Economic Statistics* **26**(1), 42–49.

Chauvet, M. & Potter, S. (2002), 'Predicting a recession: evidence from the yield curve in the presence of structural breaks', *Economics Letters* **77**(2), 245–253.

Chava, S., Gallmeyer, M. & Park, H. (2015), 'Credit conditions and stock return predictability', *Journal of Monetary Economics* **74**, 117–132.

Chen, T. (2017), 'Investor attention and global stock returns', *Journal of Behavioral Finance* **18**(3), 358–372.

Choi, K. & Jung, C. (2009), 'Structural changes and the us money demand function', *Applied Economics* **41**(10), 1251–1257.

Clark, T. E. & West, K. D. (2007), 'Approximately normal tests for equal predictive accuracy in nested models', *Journal of Econometrics* **138**(1), 291–311.

Clemen, R. T. (1989), 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting* **5**(4), 559–583.

Cochrane, J. H. (1999), 'New facts in finance', *Economic Perspectives Federal Reserve Bank of Chicago* **23**(3), 36–58.

Cochrane, J. H. (2009), *Asset pricing: Revised edition*, Princeton University Press.

Coqueret, G. (2020), 'Stock-specific sentiment and return predictability', *Quantitative Finance* **20**(9), 1531–1551.

Cró, S. & Martins, A. M. (2017), 'Structural breaks in international tourism demand: Are they caused by crises or disasters?', *Tourism Management* **63**, 3–9.

Croce, V. (2016), 'Can tourism confidence index improve tourism demand forecasts?', *Journal of Tourism Futures* **2**(1), 6–20.

Dangl, T. & Halling, M. (2012), 'Predictive regressions with time-varying coefficients', *Journal of Financial Economics* **106**(1), 157–181.

Fama, E. F. & French, K. R. (1989), 'Business conditions and expected returns on stocks and bonds', *Journal of Financial Economics* **25**(1), 23–49.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. & MacBeth, J. D. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**(3), 607–636.

Figlewski, S. (1983), 'Optimal price forecasting using survey data', *The Review of Economics and Statistics* **65**(1), 13–21.

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The economic value of volatility timing using "realized" volatility', *Journal of Financial Economics* **67**(3), 473–509.

Giacomini, R. & Rossi, B. (2009), 'Detecting and predicting forecast breakdowns', *The Review of Economic Studies* **76**(2), 669–705.

Goyal, A. & Welch, I. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *The Review of Financial Studies* **21**(4), 1455–1508.

Granger, C. W. & Ramanathan, R. (1984), 'Improved methods of combining forecasts', *Journal of Forecasting* **3**(2), 197–204.

Gu, S., Kelly, B. & Xiu, D. (2020), 'Empirical asset pricing via machine learning', *The Review of Financial Studies* **33**(5), 2223–2273.

Guidolin, M. & Timmermann, A. (2007), 'Asset allocation under multivariate regime switching', *Journal of Economic Dynamics and Control* **31**(11), 3503–3544.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* **57**(2), 357–384.

Han, Y., He, A., Rapach, D. & Zhou, G. (2019), 'Firm characteristics and expected stock returns', *Available at SSRN 3185335* .

Hautsch, N. & Voigt, S. (2019), 'Large-scale portfolio allocation under transaction costs and model uncertainty', *Journal of Econometrics* **212**(1), 221–240.

Henkel, S. J., Martin, J. S. & Nardari, F. (2011), 'Time-varying short-horizon predictability', *Journal of Financial Economics* **99**(3), 560–580.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Huang, D., Jiang, F., Jun, T. & Zhou, G. (2017), 'Forecasting stock returns in good and bad times: The role of market states', *27th Australasian Finance and Banking Conference* .

Huang, D., Jiang, F., Tu, J. & Zhou, G. (2015), 'Investor sentiment aligned: A powerful predictor of stock returns', *The Review of Financial Studies* **28**(3), 791–837.

Inoue, A., Jin, L. & Rossi, B. (2017), 'Rolling window selection for out-of-sample forecasting with time-varying parameters', *Journal of Econometrics* **196**(1), 55–67.

Inoue, A. & Rossi, B. (2011), 'Identifying the sources of instabilities in macroeconomic fluctuations', *Review of Economics and Statistics* **93**(4), 1186–1204.

Jiang, F., Lee, J., Martin, X. & Zhou, G. (2019), 'Manager sentiment and stock returns', *Journal of Financial Economics* **132**(1), 126–149.

Jose, V. R. R. & Winkler, R. L. (2008), 'Simple robust averages of forecasts: Some empirical results', *International Journal of Forecasting* **24**(1), 163–169.

Jurado, K., Ludvigson, S. C. & Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

Kacperczyk, M., Van Nieuwerburgh, S. & Veldkamp, L. (2016), 'A rational theory of mutual funds' attention allocation', *Econometrica* **84**(2), 571–626.

Kelly, B. & Pruitt, S. (2013), 'Market expectations in the cross-section of present values', *The Journal of Finance* **68**(5), 1721–1756.

Kelly, B. & Pruitt, S. (2015), 'The three-pass regression filter: A new approach to forecasting using many predictors', *Journal of Econometrics* **186**(2), 294–316.

Ledoit, O. & Wolf, M. (2008), 'Robust performance hypothesis testing with the sharpe ratio', *Journal of Empirical Finance* **15**(5), 850–859.

Lettau, M. & Van Nieuwerburgh, S. (2008), 'Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence', *The Review of Financial Studies* **21**(4), 1607–1652.

Li, J. & Tsiakas, I. (2017), 'Equity premium prediction: The role of economic and statistical constraints', *Journal of Financial Markets* **36**, 56–75.

Lin, Q. (2018), 'Technical analysis and stock return predictability: An aligned approach', *Journal of Financial Markets* **38**, 103–123.

Ludvigson, S. C. & Ng, S. (2007), 'The empirical risk–return relation: A factor analysis approach', *Journal of Financial Economics* **83**(1), 171–222.

Manela, A. & Moreira, A. (2017), 'News implied volatility and disaster concerns', *Journal of Financial Economics* **123**(1), 137–162.

Narayan, P. K. & Bannigidadmath, D. (2017), 'Does financial news predict stock returns? new evidence from islamic and non-islamic stocks', *Pacific-Basin Finance Journal* **42**, 24–45.

Neely, C. J., Rapach, D. E., Tu, J. & Zhou, G. (2014), 'Forecasting the equity risk premium: the role of technical indicators', *Management Science* **60**(7), 1772–1791.

Newbold, P. & Granger, C. W. (1974), 'Experience with forecasting univariate time series and the combination of forecasts', *Journal of the Royal Statistical Society* **137**(2), 131–146.

Ni, Z.-X., Wang, D.-Z. & Xue, W.-J. (2015), 'Investor sentiment and its nonlinear effect on stock returns—new evidence from the chinese stock market based on panel quantile regression model', *Economic Modelling* **50**, 266–274.

Paye, B. S. & Timmermann, A. (2006), 'Instability of return prediction models', *Journal of Empirical Finance* **13**(3), 274–315.

Pesaran, M. H., Pettenuzzo, D. & Timmermann, A. (2006), 'Forecasting time series subject to multiple structural breaks', *The Review of Economic Studies* **73**(4), 1057–1084.

Pesaran, M. H. & Pick, A. (2011), 'Forecast combination across estimation windows', *Journal of Business and Economic Statistics* **29**(2), 307–318.

Pesaran, M. H., Pick, A. & Pranovich, M. (2013), 'Optimal forecasts in the presence of structural breaks', *Journal of Econometrics* **177**(2), 134–152.

Pesaran, M. H. & Timmermann, A. (2002), 'Market timing and return prediction under model instability', *Journal of Empirical Finance* **9**(5), 495–510.

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Pettenuzzo, D., Timmermann, A. & Valkanov, R. (2014), 'Forecasting stock returns under economic constraints', *Journal of Financial Economics* **114**(3), 517–553.

Phan, D. H. B., Sharma, S. S. & Narayan, P. K. (2015), 'Stock return forecasting: some new evidence', *International Review of Financial Analysis* **40**, 38–51.

Rapach, D. E., Ringgenberg, M. C. & Zhou, G. (2016), 'Short interest and aggregate stock returns', *Journal of Financial Economics* **121**(1), 46–65.

Rapach, D. E., Strauss, J. K. & 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.

Rapach, D. E. & Wohar, M. E. (2006), 'Structural breaks and predictive regression models of aggregate us stock returns', *Journal of Financial Econometrics* **4**(2), 238–274.

Rapach, D. E. & Zhou, G. (2020), 'Time-series and cross-sectional stock return forecasting: New machine learning methods', *Machine Learning for Asset Management: New Developments and Financial Applications* **1**(1), 1–33.

Rapach, D. & Zhou, G. (2013), 'Forecasting stock returns', *Handbook of Economic Forecasting* **2**, 328–383.

Rossi, B. & Inoue, A. (2012), 'Out-of-sample forecast tests robust to the choice of window size', *Journal of Business and Economic Statistics* **30**(3), 432–453.

Salisu, A. A., Isah, K. & Akanni, L. O. (2019), 'Improving the predictability of stock returns with bitcoin prices', *The North American Journal of Economics and Finance* **48**, 857–867.

Stock, J. H. & Watson, M. W. (1996), 'Evidence on structural instability in macroeconomic time series relations', *Journal of Business and Economic Statistics* **14**(1), 11–30.

Sun, L., Najand, M. & Shen, J. (2016), 'Stock return predictability and investor sentiment: A high-frequency perspective', *Journal of Banking and Finance* **73**, 147–164.

Tian, J. & Anderson, H. M. (2014), 'Forecast combinations under structural break uncertainty', *International Journal of Forecasting* **30**(1), 161–175.

Tian, J. & Zhou, Q. (2018), 'Improving equity premium forecasts by incorporating structural break uncertainty', *Accounting and Finance* **58**(1), 619–656.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58**(1), 267–288.

Timmermann, A. (2006), 'Forecast combinations', *Handbook of Economic Forecasting* **1**, 135–196.

Timmermann, A. (2018), 'Forecasting methods in finance', *Annual Review of Financial Economics* **10**, 449–479.

Wang, Y., Hao, X. & Wu, C. (2020), 'Forecasting stock returns: A time-dependent weighted least squares approach', *Journal of Financial Markets* **53**, 100568.

Winkler, R. L. & Makridakis, S. (1983), 'The combination of forecasts', *Journal of the Royal Statistical Society* **146**(2), 150–157.

Wong, K. K., Song, H., Witt, S. F. & Wu, D. C. (2007), 'Tourism forecasting: to combine or not to combine?', *Tourism Management* **28**(4), 1068–1078.

Yin, A. (2019), 'Out-of-sample equity premium prediction in the presence of structural breaks', *International Review of Financial Analysis* **65**, 101385.

Yin, A. (2020), 'Equity premium prediction and optimal portfolio decision with bagging', *The North American Journal of Economics and Finance* **54**, 101274.

Zhang, H., He, Q., Jacobsen, B. & Jiang, F. (2020), 'Forecasting stock returns with model uncertainty and parameter instability', *Journal of Applied Econometrics* **35**(5), 629–644.

Zhang, Y., Chu, G. & Shen, D. (2021), 'The role of investor attention in predicting stock prices: The long short-term memory networks perspective', *Finance Research Letters* **38**, 101484.

Zhang, Y., Wei, Y., Ma, F. & Yi, Y. (2019), 'Economic constraints and stock return predictability: A new approach', *International Review of Financial Analysis* **63**, 1–9.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society* **67**(2), 301–320.

Table 4.1: List of Variance-Covariance (VC) approaches

Symbol	Model	Reference
A. Variance-Covariance (VC) Approach		See section (4.4)
VC-DIAG	Diagonal Covariance Matrix	Equation (4.79)
VC-SAMPLE	Sample Covariance Matrix	Equation (4.80)
VC-SIM	Single Index Model Covariance Matrix	Equation (4.85)
B. VC approach combining across estimation windows		See section (4.4.2.1)
VC-W-DIAG	Using diagonal error-covariance matrix based on CV sample	
VC-W-SAMPLE	Using sample error-covariance matrix based on CV sample	
VC-W-SIM	Using SIM error-covariance matrix based on CV sample	
C. VC approach combining across univariate predictor-based models		See section (4.4.2.2)
VC-M-DIAG	Using diagonal error-covariance matrix estimated with CV, ROLL, EXP, Optimal	
VC-M-SAMPLE	Using sample error-covariance matrix estimated with CV, ROLL, EXP, Optimal	
VC-M-SIM	Using sample error-covariance matrix estimated with CV, ROLL, EXP, Optimal	
CV	Out-of-sample Holdout period to estimate forecasting errors, error-covariance, and optimal weights	
EXP	Realised error covariance-matrix based on expanding window	
ROLL	Realised error covariance-matrix based on 15 year rolling window	
Optimal	Realised error covariance-matrix based on optimal window	
D. VC panel approach combining across univariate predictor-based models and windows		See section (4.4.2.3.1)
Panel (DIAG)	A combined forecast where optimal weights based on Diagonal Covariance Matrix	
Panel (SAMPLE)	A combined forecast where optimal weights based on Sample Covariance Matrix	
Panel (SIM)	A combined forecast where optimal weights based on Single Index Model Covariance Matrix	
E. VC two-step approach: i) VC to combine windows; ii) Simple average across models		See section (4.5.2.3)
Mean (VC-DIAG)	Combining windows of univariate model with diagonal then simple average across models	
Mean (VC-SAMPLE)	Combining windows of univariate model with sample then simple average across models	
Mean (VC-SIM)	Combining windows of univariate model with SIM then simple average across models	
F. VC two-step approach: i) Select optimal window; ii) VC across models		See section (4.4.2.3.2)
SELCOM-DIAG	Select optimal window based on minimum MSFE in CV sample, next CF based across models with diagonal	
SELCOM-SAMPLE	Select optimal window based on minimum MSFE in CV sample, next CF based across models with sample	
SELCOM-SIM	Select optimal window based on minimum MSFE in CV sample, next CF based across models with SIM	
G. VC two-step approach: i) Select optimal window; ii) VC across models		See section (4.4.2.3.3)
COMCOM-DIAG	CF across windows with diagonal next CF across univariate models with diagonal	
COMCOM-SAMPLE	CF across windows with sample next CF across univariate models with sample	
COMCOM-SIM	CF across windows with SIM next CF across univariate models with SIM	

Table 4.2: List of Benchmark Models

Symbol	Model	Key Literature	Reference
A. Benchmark Models in the presence of estimation window uncertainty (EWU)			See section (4.2.3)
EXP	Expanding Window	Goyal & Welch (2008); Rapach et al. (2010)	Equation (4.14)
ROLL	Rolling Window	Fama & MacBeth (1973); Pesaran & Timmermann (2002))	Equation (4.17)
BP	Bai-Perron Method - Post-break observations	Bai & Perron (1998)	Equation (4.19)
SEL-CV	Selection of optimal windows based on MSFE in CV	Pesaran & Timmermann (2007)	Equation (4.22)
RobW1	Robust optimal weight on observation with 1 break	Pesaran et al. (2013); Tian & Zhou (2018)	Equation (4.28)
RobW2	Robust optimal weight on observation with 2 breaks		Equation (4.28)
EW	Equally-weighted across all possible estimation windows	Pesaran & Timmermann (2007)	Equation (4.29)
EW10	Equally-weighted across 10 windows	Pesaran & Pick (2011)	Equation (4.32)
LW	Location Weighted across estimation windows	Tian & Anderson (2014)	Equation (4.33)
VW-MSFE	Value-weighted based on MSFE in CV	Pesaran & Timmermann (2007)	Equation (4.35)
ROC	Reverse Ordered CUSUM (ROC)	Tian & Anderson (2014)	Equation (4.43)
ROC-L	Location weighted Reverse Ordered CUSUM (ROC)	Tian & Anderson (2014)	Equation (4.44)
RS	Regime Shift	Boyd et al. (2005); Henkel et al. (2011); Huang et al. (2017)	Equation (4.13)
B. Standard Benchmark Models in Equity Premium Prediction to deal variable-selection uncertainty (VSU)			See section (4.2.2)
Kitchen Sink			
KS	Kitchen Sink Model	Goyal & Welch (2008)	Equation (4.13)
Dimesion Reduction			See section (4.2.2.2)
PCR1	Principal Component Regression (PCR) with 1 PC	Neely et al. (2014)	Equation (4.8)
PCR3	Principal Component Regression (PCR) with 3 PCs	Neely et al. (2014)	Equation (4.8)
KP-3PF	Kelly & Pruitt (2013) (KP) three pass filter (3PF)	Kelly & Pruitt (2013)	Equation (4.9 and 4.10)
Shrinkage			See section (4.2.2.4)
LASSO	Least Absolute Shrinkage and Selection Operator	Tibshirani (1996); Buncic & Tischhauser (2017)	Equation (4.12)
Ad-LASSO	Adaptive Least Absolute Shrinkage and Selection Operator	Zou (2006);	Equation (4.12)
Ridge	Ridge	Hoerl & Kennard (1970)	Equation (4.12)
ENet	Elastic Net	Zou & Hastie (2005)	Equation (4.12)
Combining Forecasts (CF) across univariate predictive models			See section (4.2.4.1)
Mean	Equally-weighted forecast	Rapach et al. (2010)	Equation (4.46)
T-Mean	Trimmed mean forecast	Rapach et al. (2010)	Equation (4.46)
Median	Median forecast	Rapach et al. (2010)	Equation (4.46)
DMSFE	Discount Mean Squared Error	Rapach et al. (2010)	Equation (4.47)
RA	Regression analysis (RA)	Rapach & Zhou (2020)	
	RA1, RA2, RA3		Equation (4.50), (4.51), and (4.52)
RR	Robust Rank (RA)	Aiolfi & Timmermann (2006)	Equation (4.49)

Table 4.3: Out-of-sample results: VC approach to address EWU

This table presents the out-of-sample R^2_{OOS} for 1-month returns of equity premium based on univariate predictor-based forecasting models for the out-of-sample period January 1947 to December 2018. All the results are from our variance-covariance approach, where the combined forecast of each model is taken as optimally weighted across estimation windows based on the error covariance matrix. The three covariance matrices include diagonal (DIAG), sample (SAMPLE) and single-index model (SIM), respectively. The details of the models is given in Table (4.2). Equity premium is estimated with univariate predictor-based model, $r_{t+1} = \alpha_i + \beta_i X_t + \epsilon_{i,t+1}$, where Xi, t is one of the 14 Goyal & Welch (2008) predictors. R^2_{OOS} measures the reduction in MSFE for the competing forecast relative to the historical average forecast. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The significance is based on Clark & West (2007) statistic given in Equation 4.69.

	VC-DIAG	**VC-SAMPLE**	**VC-SIM**
DP	-0.2492	-1.1693	-0.0693
DY	-0.1257	-1.3420	0.1091
EP	-0.6429	-2.2166	-0.9219
DE	-0.8088	-0.2375	0.2288**
SVAR	-0.2552	0.0008	0.1688**
BM	-1.4586	-0.7946	-0.7568
NTIS	-0.8513	-1.1227	-0.4754
TBL	-0.8571	-2.5962	-0.8893
LTY	-0.5762	-2.3006	-0.1438
LTR	-0.5158	-0.2582	-0.2900
TMS	-0.5030	-1.2967	0.1508*
DFY	-0.9706	-2.2067	-0.1474
DFR	-0.7954	-0.9247	-0.9187
INFL	-0.3942	-0.4490	-0.0349

Table 4.4: Out-of-sample results: VC approach to address VSU

This table reports the out-of-sample results for 1-month return for equity premium approach with VC approach addressing variable uncertainty only by combining forecasts across univariate predictor-based models. Univariate predictive models are estimated with expanding window where are CV, Roll, Exp, and optimal represent four different approaches to estimate covariance matrix. CV indicates the cross-validation approach. Roll and Exp indicate that the historical forecasting errors for individual models are estimated with a rolling window of 15 years and expanding window including all the observations, respectively. The optimal indicates the optimal covariance matrix at a period that effectively identifies the optimal past observations based on minimum variance criteria to estimate the error-covariance matrix (see section (4.4.2.2) for details). MSFE represents the mean squared forecast error. R^2_{OOS} indicates the reduction in MSFE for the competing forecast relative to the historical average forecast. MSFE-adjusted is the Clark & West (2007) statistic given in Equation 4.69. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. $\bar{\hat{e}}^2$ represents the squared forecast bias, and var($\hat{e}$) indicates the forecast error variance.

Forecasting model	**MSFE**	R^2_{OS} **(%)**	**MSFE-adj.**	**p-value**	$\bar{\hat{e}}^2$	**var($\hat{e}$)**
Historical Average (HA)	17.224				0.035	17.189
A. VC-DIAG						
CV	17.115	0.636***	3.018	0.001	0.004	17.111
EXP	17.127	0.566**	2.336	0.010	0.024	17.103
ROLL	17.110	0.664***	3.019	0.001	0.024	17.086
Optimal	17.106	0.691***	2.481	0.006	0.009	17.097
B. VC-SAMPLE						
CV	17.133	0.531***	2.922	0.002	0.005	17.128
EXP	17.320	−0.553	0.424	0.336	0.001	17.319
ROLL	17.119	0.609***	3.009	0.001	0.030	17.089
Optimal	17.117	0.624***	3.215	0.001	0.033	17.084
C. VC-SIM						
CV	17.100	0.726***	2.968	0.001	0.005	17.095
EXP	17.341	−0.678	0.641	0.261	0.003	17.338
ROLL	17.095	0.754***	3.215	0.001	0.024	17.071
Optimal	17.086	0.804***	3.063	0.001	0.029	17.057

Table 4.5: Out-of-sample results: VC approach to address EWU and VSU

This table reports the out-of-sample results for 1-month return for equity premium approach with VC approach addressing EWU and VSU simultaneously. Details of models models is given in Table (4.1). MSFE is the mean squared forecast error. R^2_{OOS} measures the reduction in MSFE for the competing forecast relative to the historical average forecast. MSFE represents the mean squared forecast error. R^2_{OOS} indicates the reduction in MSFE for the competing forecast relative to the historical average forecast. MSFE-adjusted is the Clark & West (2007) statistic given in Equation 4.69. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. $\bar{e}^2$ represents the squared forecast bias, and var($\hat{e}$) indicates the forecast error variance.

Forecasting model	MSFE	R^2_{OS} (%)	MSFE-adj	p-value	$\bar{e}^2$	var($\hat{e}$)
Historical Average (HA)	17.224				0.035	17.189
A. Panel Approach						
Panel (DIAG)	17.081	0.838***	3.199	0.001	0.001	17.079
Panel (SAMPLE)	17.088	0.792**	1.711	0.044	0.001	17.088
Panel (SIM)	17.032	0.922***	2.473	0.007	0.004	17.028
B. MeanCOM Approach						
Mean (DIAG)	17.087	0.802***	2.571	0.005	0.004	17.083
Mean (SAMPLE)	17.091	0.779**	2.192	0.014	0.014	17.076
Mean (SIM)	17.066	1.124***	2.808	0.002	0.003	17.064
C. SELCOM Approach						
C.1 SELCOM (DIAG)						
CV	17.105	0.692**	2.339	0.010	0.004	17.102
EXP	17.118	0.616***	3.021	0.001	0.023	17.095
ROLL	17.100	0.721***	2.484	0.006	0.022	17.078
Optimal	17.095	0.751***	3.022	0.001	0.008	17.087
C.2 SELCOM (SAMPLE)						
CV	17.125	0.578***	2.925	0.002	0.004	17.120
EXP	17.328	−0.602	0.424	0.336	0.001	17.367
ROLL	17.110	0.663***	3.012	0.001	0.029	17.083
Optimal	17.108	0.678***	3.218	0.001	0.031	17.079
C.3 SELCOM (SIM)						
CV	17.090	0.781***	2.971	0.001	0.004	17.086
EXP	17.352	−0.738	0.642	0.261	0.003	17.388
ROLL	17.010	1.256***	2.93	0.00	0.023	17.024
Optimal	17.000	1.315***	3.066	0.001	0.028	16.972
D. COMCOM Approach						
D.1 COMCOM (DIAG)						
CV	17.099	0.732***	2.563	0.006	0.004	17.110
EXP	17.102	0.713***	2.192	0.014	0.023	17.131
ROLL	17.081	0.832***	2.798	0.003	0.014	17.067
Optimal	17.072	0.889***	3.066	0.001	0.014	17.058
D.2 COMCOM (SAMPLE)						
CV	17.124	0.583**	2.175	0.015	0.004	17.120
EXP	17.316	−0.531	0.425	0.337	0.001	17.367
ROLL	17.102	0.709**	2.339	0.010	0.014	17.088
Optimal	17.097	0.744***	2.925	0.002	0.014	17.082
D.3 COMCOM (SIM)						
CV	17.014	1.233***	3.110	0.001	0.004	17.010
EXP	17.344	−0.693	0.642	0.261	0.003	17.341
ROLL	17.006	1.282***	2.971	0.001	0.021	16.984
Optimal	16.998	1.325***	3.417	0.000	0.028	16.971

Table 4.6: VC approach vs. Benchmark models to address EWU

This table reports the out-of-sample R^2_{OOS} for 1-month returns of equity premium based on univariate economic predictor-based forecasting models. Panel A reports the results from our variance-covariance approach where combined forecast of each model is taken as optimally weighted across estimation windows based on error covariance matrix. The three covariance matrices include diagonal (DIAG), sample (SAMPLE) and single-index model (SIM), respectively. Panel B reports the results from various approaches used in literature to deal with parameter instability (PI) in presence of structural breaks (SBU). The details of models is given in Table (4.2). Equity premium is estimated with univariate predictor-based model, $r_{t+1} = \alpha_i + \beta_i X_t + \epsilon_{i,t+1}$, where Xi, t is one of the 14 Goyal & Welch (2008) predictors. R^2_{OOS} measures the reduction in MSFE for the competing forecast relative to the historical average forecast. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The significance is based on Clark & West (2007) statistic given in Equation 4.69.

	A. Variance-Covariance			**B. Benchmark Models**												
	VC-DIAG	**VC-SAMPLE**	**VC-SIM**	**EXP**	**ROLL**	**BP**	**SEL-CV**	**RobW1**	**RobW2**	**EW**	**EW10**	**LW**	**VW-MSFE**	**ROC**	**ROC-L**	**RS**
DP	-0.2492	-1.1693	-0.0693	-0.0781	-1.0343	-0.4683	-0.2611	**0.1648**	0.1599	-0.1739	-0.1792	-0.1896	-0.2175	-0.1868	-0.1870	-0.4496
DY	-0.1257	-1.3420	0.1091	-0.4032	-0.7782	-0.8065	-0.1490	**0.2142**	0.1928	-0.0764	-0.0629	-0.1560	-0.1127	-0.0036	-0.0345	-0.7742
EP	-0.6429	-2.2166	-0.9219	-1.4505	-2.5401	-1.7406	-0.6399	-0.6536	-0.6863	-0.5212	-0.5580	-0.5387	-0.5805	-0.4398	**-0.4372**	-1.6710
DE	-0.8088	-0.2375	**0.2289**	-1.4258	-2.6367	-2.8516	-0.7640	-1.0807	-0.9727	-0.6527	-0.6164	-0.6680	-0.7083	-0.6504	-0.6620	-2.7375
SVAR	-0.2552	0.0008	**0.1688**	-0.0660	-0.8448	-0.0640	-0.2517	0.1576	0.1592	-0.1909	-0.4179	-0.0998	-0.2213	-0.3869	-0.2823	0.0456
BM	-1.4586	-0.7946	**-0.7568**	-1.4896	-1.9856	-1.4913	-1.4719	-0.7676	-0.8750	-1.2053	-1.0859	-1.4560	-1.3386	-1.3256	-1.4204	-1.4316
NTIS	-0.8513	-1.1227	-0.4754	-0.6780	-1.5888	-1.2204	-0.8710	**-0.2420**	-0.3630	-0.9670	-0.8686	-0.9336	-0.9190	-0.9837	-0.9403	-1.1716
TBL	-0.8571	-2.5962	-0.8893	**0.0717**	-3.1802	-0.0954	-0.8568	-0.2298	-0.2597	-0.3590	-0.3207	-0.5055	-0.6079	-1.0082	-1.0568	-0.0916
LTY	-0.5762	-2.3006	**-0.1438**	-0.6548	-1.8265	-1.3096	-0.5396	-0.5362	-0.3754	-0.5595	-0.5667	-0.4833	-0.5496	-0.5248	-0.4216	-0.9953
LTR	-0.5158	-0.2582	-0.2900	-0.7280	-1.3362	-0.7287	-0.5026	-0.1084	**-0.0846**	-0.2911	-0.2769	-0.3631	-0.3969	-0.2147	-0.2857	-0.6995
TMS	-0.5030	-1.2967	**0.1509**	0.1070	-0.1838	0.1321	-0.5040	0.0379	0.0303	-0.2182	-0.1171	-0.3505	-0.3611	-0.7613	-0.8050	0.1268
DFY	-0.9706	-2.2067	-0.1474	-0.1698	-1.5490	-0.4244	-0.9811	-0.1589	**-0.1271**	-0.6854	-0.4611	-0.7446	-0.8332	-1.0940	-0.9929	-0.4074
DFR	-0.7954	-0.9247	-0.9187	**-0.1906**	-1.9540	-0.3312	-0.7861	-0.7734	-0.2320	-0.6481	-0.5953	-0.5236	-0.7171	-0.5285	-0.5150	-0.3180
INFL	-0.3942	-0.4490	**-0.0349**	-0.0405	-0.6467	-0.0406	-0.3964	-0.0583	-0.0933	-0.2235	-0.2308	-0.2897	-0.3099	-0.4053	-0.4202	-0.0390

Table 4.7: VC approach vs. Benchmark models to address EWU and VSU

This table compares the the out-of-sample results for 1-month return for equity premium with various benchmark models. The details of VC-SIM and benchmark models are given Table (4.1) and (4.2), respectively. MSFE represents the mean squared forecast error. R^2_{OOS} indicates the reduction in MSFE for the competing forecast relative to the historical average forecast. MSFE-adjusted is the Clark & West (2007) statistic given in Equation 4.69. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. $\bar{\hat{e}}^2$ represents the squared forecast bias, and var($\hat{e}$) indicates the forecast error variance.

Forecasting model	**MSFE**	R^2_{OS} **(%)**	**MSFE-adj.**	**p-value**	$\bar{\hat{e}}^2$	**var($\hat{e}$)**
Historical Average (HA)	17.224				0.035	17.189
A. Variance-covariance approach						
Panel (SIM)	17.066	0.922***	2.473	0.007	0.004	17.062
Mean (SIM)	17.066	1.124***	2.808	0.002	0.003	17.064
SELCOM (SIM^*)	17.000	1.215***	3.066	0.001	0.028	16.972
COMCOM (SIM^*)	16.998	1.426***	3.417	0.000	0.028	16.971
B. Benchmark Models to deal with SBU						
Mean (EXP)	17.106	0.689***	3.050	0.001	0.029	17.077
Mean (ROLL)	17.122	0.503*	1.336	0.091	0.003	17.120
Mean (BP)	17.118	0.618**	1.646	0.050	0.009	17.109
Mean (SEL-CV)	17.096	0.747***	2.563	0.006	0.005	17.092
Mean (RobW1)	17.108	0.675***	2.584	0.005	0.004	17.104
Mean (RobW2)	17.109	0.673***	2.571	0.005	0.004	17.104
Mean (EW) [AveAve]	17.086	0.808***	2.798	0.003	0.003	17.083
Mean (EW10)	17.093	0.766***	2.339	0.010	0.004	17.089
Mean (LW)	17.090	0.781***	3.218	0.001	0.004	17.086
Mean (VW-MSFE)	17.091	0.777**	1.899	0.029	0.001	17.090
Mean (ROC)	17.081	0.838***	3.199	0.001	0.001	17.079
Mean (ROC-L)	17.085	0.814***	3.417	0.000	0.002	17.083
Mean (RS)	17.112	0.656***	2.563	0.006	0.005	17.107
C. Standard Benchmark Models in Equity Premium Prediction						
C.1 Kitchen Sink						
KS	19.613	−12.184	1.899	0.029	0.320	19.294
C.2 Dimension Reduction						
PCR1	17.173	0.294**	1.951	0.026	0.163	17.010
PCR3	17.140	0.491**	2.173	0.015	0.188	16.951
KP-3PF	17.348	−0.714	0.91	0.18	0.197	17.151
C.3 Shrinkage						
LASSO	17.288	−0.374	0.024	0.509	0.000	17.288
Ad-LASSO	17.325	−0.584	0.463	0.322	0.002	17.323
Ridge	17.339	−0.662	0.977	0.164	0.199	17.139
ENet	17.286	−0.362	0.872	0.192	0.000	17.286
C.4 Forecast Combination						
Mean	17.106	0.689***	3.050	0.001	0.029	17.077
T-Mean	17.114	0.641***	3.020	0.001	0.029	17.085
Median	17.125	0.576***	3.176	0.001	0.015	17.110
DMSFE ($\theta = 0.9$)	17.099	0.731***	3.116	0.001	0.028	17.071
DMSFE ($\theta = 1$)	17.106	0.688***	3.047	0.001	0.028	17.077
RA1	17.654	−2.439	0.702	0.241	0.000	17.654
RA2	17.435	−1.213	0.038	0.485	0.014	17.421
RA2	17.107	0.684***	4.204	0.000	0.018	17.088
RR	17.165	0.342**	1.719	0.043	0.024	17.141
VC-SIM (Optimal)	17.086	0.804***	3.063	0.001	0.029	17.057

Table 4.8: Out-of-sample results over recession and expansion

This tables reports out-of-sample R^2_{OOS} for NBER-dated expansions and recessions. The details of VC-SIM and benchmark models are given Table (4.1) and (4.2), respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Forecasting model	Full	Expansion	Recession
A. Variance-covariance approach (SIM)			
Panel (SIM)	0.922^{***}	0.702^{**}	1.163^{***}
Mean (SIM)	1.124^{***}	0.859^{**}	1.273^{***}
SELCOM (SIM^*)	1.215^{***}	0.892^{***}	2.546^{***}
COMCOM (SIM^*)	1.426^{***}	0.934^{***}	2.811^{***}
B. Benchmark Models to deal with SBU			
Mean (EXP)	0.689^{***}	0.600^{***}	0.949^{***}
Mean (ROLL)	0.593^{*}	0.098	1.474
Mean (BP)	0.618^{**}	0.480^{*}	1.139^{*}
Mean (SEL-CV)	0.747^{***}	0.203^{*}	1.243^{***}
Mean (RobW1)	0.675^{***}	0.225^{*}	1.274^{**}
Mean (RobW2)	0.673^{***}	0.225^{*}	1.272
Mean (EW)	0.808^{***}	0.369^{**}	1.399^{**}
Mean (EW10)	0.766^{***}	0.314^{**}	1.565^{*}
Mean (LW)	0.781^{***}	0.388^{**}	1.147^{***}
Mean (VW-MSFE)	0.777^{**}	0.126	1.694
Mean (ROC)	0.838^{***}	0.336^{**}	1.468^{***}
Mean (ROC-L)	0.814^{***}	0.363^{**}	1.321^{***}
Mean (RS)	0.656^{***}	0.212^{*}	1.264^{***}
C. Standard Benchmark Models in Equity Premium Prediction			
C.1 Kitchen Sink			
KS	−12.184	−13.511	-8.316
C.2 Dimension Reduction			
PCR1	0.294^{**}	−0.456	1.222^{**}
PCR3	0.491^{**}	−0.354	2.954^{***}
KP-3PF	−0.714	−1.499	1.573^{**}
C.3 Shrinkage			
LASSO	−0.374	−0.216	−1.511
Ad-LASSO	−0.584	−0.323	−1.228
Ridge	−0.662	−0.282	−1.522
ENet	−0.362	−0.180	−1.815
C.4 Forecast Combination			
Mean	0.689^{***}	0.600^{***}	0.949^{**}
T-Mean	0.641^{***}	0.526^{**}	0.975^{**}
Median	0.576^{***}	0.506^{**}	0.780^{**}
DMSFE ($\theta = 0.9$)	0.731^{***}	0.605^{***}	1.098^{**}
DMSFE ($\theta = 1$)	0.688^{***}	0.542^{**}	1.028^{***}
RA1	−2.439	−2.018	−3.730
RA2	−1.213	−0.332	−3.778
RA3	0.684^{***}	0.571^{**}	0.997^{**}
RR	0.342^{**}	0.475^{**}	−0.046
VC-SIM (Optimal)	0.804^{***}	0.613^{***}	1.565^{***}

Table 4.9: Results from asset allocation

The table represents the results of portfolio performance for a mean-variance investor having a relative risk-aversion level of five allocating the resources between risk-free assets (Treasury bills) and equity. The weights are determined through the forecasts of equity premium generated either through a given model or historical average (HA). ΔCER is the annualised certainty equivalent return (CER) gain for an investor who uses underlying model (VC-SIM or benchmark) instead of the historical average forecast. $Turnover$ refers to relative average turnover based on a given model given in equation (4.75). $\Delta CER(50bp)$ refers to the net CER gain where an investor pays a is a transaction cost of 50 basis points for each transaction. SR is the Sharpe ratio of portfolio. The details of VC-SIM and benchmark models are given Table (4.1) and (4.2), respectively.

Forecasting model	**Δ (CER) (ann %)**	**SR**	**Turnover**	**Δ (CER) (ann %, 50bp)**
A. Variance-covariance approach (SIM)				
Panel (SIM)	2.071	0.461	3.357	1.657
Mean (SIM)	2.239	0.483	3.362	1.791
SELCOM (SIM^*)	2.378	0.489	4.405	1.922
COMCOM (SIM^*)	2.524	0.497	4.535	2.019
B. Benchmark Models to deal with SBU				
Mean (EXP)	1.076	0.416	3.337	0.828
Mean (ROLL)	0.930	0.442	3.243	0.711
Mean (BP)	0.939	0.447	3.983	1.082
Mean (SEL-CV)	0.925	0.376	3.543	0.648
Mean (RobW1)	0.926	0.376	3.546	0.648
Mean (RobW2)	0.925	0.376	3.539	0.647
Mean (EW) [AveAve]	1.421	0.442	3.638	0.846
Mean (EW10)	1.412	0.420	3.599	0.988
Mean (LW)	0.975	0.380	3.546	0.682
Mean (VW-MSFE)	1.404	0.429	3.894	0.982
Mean (ROC)	1.064	0.392	3.780	0.745
Mean (ROC-L)	0.964	0.383	3.735	0.675
Mean (RS)	0.907	0.369	3.472	0.635
C. Standard Benchmark Models in Equity Premium Prediction				
C.1 Kitchen Sink				
KS	0.181	0.346	2.943	0.096
C.2 Dimension Reduction				
PCR1	0.674	0.380	3.807	0.425
PCR3	1.045	0.429	3.435	0.658
KP-3PF	1.041	0.453	2.529	0.956
C.3 Shrinkage				
LASSO	0.226	0.169	3.205	0.129
Ad-LASSO	0.424	0.299	3.406	0.242
Ridge	0.469	0.340	2.355	0.267
ENet	0.249	0.254	3.878	0.142
C.4 Forecast Combination				
Mean	1.076	0.416	3.337	0.828
T-Mean	1.052	0.414	3.339	0.810
Median	0.798	0.391	3.754	0.615
DMSFE ($\theta = 0.9$)	1.225	0.431	3.315	0.943
DMSFE ($\theta = 1$)	1.040	0.411	3.606	0.801
RA1	-2.989	0.183	4.953	-3.416
RA2	-0.538	0.335	4.371	-0.948
RA3	1.148	0.423	3.352	0.884
RR	0.789	0.387	3.407	0.608
VC-SIM (Optimal)	1.102	0.430	3.414	0.851

Table 4.10: Contribution of predictors in equity premium forecasts

This table shows the importance of each predictor in forecasting equity premium. Panel A shows importance on the basis of average reduction in out-of-sample R^2 due to absence of a particular predictor from information set. The remaining results (Panel B to L) are based on average contribution of each predictor in combines forecast overtime using COMCOM-SIM model. Panel B takes average of individuals weights over all sample splits where as panel C to E report the average weight contribution based on full out-of-sample (1947:01 - 2018:12) for overall, expansion and recession periods respectively. Panel F to H for the sample period 1974:01 - 2018:12 (availability of sentiment data) for overall, optimism and pessimism periods respectively. Optimism and pessimism periods are identified on the basis of Baker & Wurgler (2006) sentiment index (SI) where following Baker & Wurgler (2006) periods with positive index are classified as optimism while other periods as classified as pessimism.

	Full Sample 1947:01 - 2018:12					**Sample 1974:01 - 2018:12**						
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
	Contribution in Reduction in R^2 Avg. over sample splits	**Contribution in optimal weight Avg. over sample splits**	**Average Contribution in optimal weight - Overall sample**									
			Full Sample	**Expansion**	**Recession**	**Overall**	**Optimism**	**Pessimism**	**Exp - Opt**	**Exp - Pess.**	**Rec - Opt.**	**Rec - Pess.**
DP	1.10	3.17	2.94	2.66	4.66	2.85	3.31	2.44	4.61	1.48	3.04	2.54
DY	13.40	9.91	5.54	5.69	4.61	7.86	6.87	8.72	5.30	8.22	7.20	8.77
EP	11.76	13.95	10.47	10.14	12.51	11.76	14.51	9.35	22.38	4.09	12.86	9.87
DE	3.20	1.39	1.58	1.35	2.98	1.85	1.66	2.01	0.00	11.07	2.00	1.11
SVAR	3.89	1.13	1.28	0.94	3.37	1.83	3.71	0.19	8.32	0.00	2.74	0.21
BM	0.89	1.59	12.61	12.05	16.03	1.26	2.62	0.07	0.13	0.00	3.15	0.07
NTIS	19.82	16.68	10.49	10.77	8.80	12.56	7.44	17.04	2.84	20.00	8.41	16.74
TBL	9.16	4.83	10.71	11.15	8.03	10.21	3.99	15.65	9.72	17.36	2.78	15.47
LTY	5.74	12.10	3.89	4.20	1.99	5.60	4.28	6.76	2.86	3.53	4.58	7.08
LTR	4.87	4.49	8.02	7.59	10.64	8.30	7.58	8.93	13.45	11.01	6.34	8.72
TMS	2.80	4.12	8.07	8.51	5.43	8.09	8.67	7.58	10.58	0.00	8.27	8.34
DFY	4.77	5.20	5.98	5.63	8.14	8.57	15.82	2.23	15.53	8.07	15.88	1.65
DFR	16.49	13.72	10.73	11.68	4.95	15.45	16.51	14.53	4.27	13.61	19.09	14.62
INFL	2.12	7.72	7.68	7.64	7.87	3.81	3.02	4.49	0.00	1.56	3.66	4.79

Table 4.11: R^2_{OOS} for during good, normal, and bad growth periods

This table reports R^2_{OOS} for equity premium forecasts generated through our VC-SIM approaches over different regimes based on the economic growth rate. Specifically, we use the top, middle, and bottom thirds of sorted rates of growth to characterise "good", "normal," and "bad" regimes. Following Rapach et al. (2010), our analysis uses growth rates of three variables: i) real profit, ii) real GDP, and iii) real net cash flow, as shown in panels A, B, and C. The R^2 statistics are calculated for the whole forecast testing period (Overall) as well as three subperiods (regimes). *, **, and *** represent the significance at the 10%, 5%, and 1% levels, respectively.

	Overall	**Good**	**Normal**	**Bad**
A. Real GDP growth				
Panel (SIM)	0.922^{***}	0.214	0.319	1.482^{***}
Mean (SIM)	1.124^{***}	0.286	0.492^{*}	1.617^{***}
SELCOM (SIM^*)	1.215^{***}	0.372	0.562^{**}	1.729^{***}
COMCOM (SIM^*)	1.426^{***}	0.426	0.618^{**}	1.922^{***}
B. Real profit growth				
Panel (SIM)	0.922^{***}	0.098	0.514^{*}	1.882^{***}
Mean (SIM)	1.124^{***}	0.128	0.668^{**}	2.109^{***}
SELCOM (SIM^*)	1.315^{***}	0.149	0.799^{**}	2.363^{***}
COMCOM (SIM^*)	1.426^{***}	0.158	0.843^{**}	2.418^{***}
C. Real net cash flow growth				
Panel (SIM)	0.922^{***}	0.330	1.105^{***}	1.417^{***}
Mean (SIM)	1.124^{***}	0.413	1.287^{***}	1.604^{***}
SELCOM (SIM^*)	1.315^{***}	0.528^{**}	1.413^{***}	1.833^{***}
COMCOM (SIM^*)	1.426^{***}	0.539^{**}	1.421^{***}	1.861^{***}

Table 4.12: Correlation between expected equity premium and macro variables

This table reports percentage correlations between our estimated 1-month expected return series under VC-SIM approaches and various macroeconomic time series consisting of four categories. Panel A consist of Macroeconomic activity including industrial production growth, GDP growth, real profit growth, real personal income, Chicago Fed National Activity Index, recession probability of Chauvet & Piger (2008), Anxious index of Survey of Professional Forecasters (SPF), unemployment rate, and aggregate book-to-market ratio of aggregate U.S. stock market. Panel B consists of volatility in macro variables reported in Panel A. In addition, we also consider the macroeconomic uncertainty index from Bloom et al. (2018). Panel C consists of sentiment variables from Baker & Wurgler (2006). Panel D consists of credit variables, i.e. term-spread and yield spread. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Variable	Panel (SIM)	Mean (SIM)	SELCOM (SIM^*)	COMCOM (SIM^*)
A. Macroeconomic activity				
Industrial Production growth	-4.12^{*}	-5.33^{**}	-6.88^{**}	-8.16^{***}
GDP growth	-3.95	-7.34^{*}	-8.68^{*}	-9.14^{*}
Real profit growth	-2.99	-3.29	-10.67^{***}	-11.23^{***}
Real personal income	-3.64	-5.01	-7.34^{**}	-7.72^{**}
CFNAI	-4.04^{*}	-6.45^{***}	-7.99^{***}	-8.42^{***}
Unemployment rate (FRED)	22.54^{***}	25.79^{***}	30.19^{***}	31.78^{***}
CP recession	12.61^{**}	11.87^{**}	13.20^{***}	13.89^{***}
SPF recession	13.16^{**}	15.62^{**}	20.81^{***}	21.90^{***}
Agg. Book-to-market	32.16^{***}	36.08^{***}	42.82^{***}	47.71^{***}
Surplus consumer ratio	-5.14^{*}	-6.94^{*}	-8.43^{**}	-8.92^{**}
B. Macroeconomic uncertainty				
GDP growth volatility	12.12^{*}	16.25^{***}	20.79^{***}	21.65^{***}
Consumer growth volatility	12.89^{*}	16.50^{**}	23.87^{***}	24.86^{***}
Ind. Pr. growth Volatility	14.05^{*}	17.17^{**}	22.92^{***}	23.87^{***}
Real Profit growth Volatility	15.87^{**}	19.77^{***}	27.54^{***}	28.68^{***}
SPF uncertainty	18.40^{**}	24.77^{***}	28.32^{***}	29.50^{***}
Financial Uncertainty	22.05^{***}	26.69^{***}	30.25^{***}	31.51^{***}
Macro Uncertainty	21.01^{***}	25.74^{***}	28.53^{***}	29.72^{***}
Real Uncertainty	15.113^{**}	17.55^{**}	20.96^{***}	21.83^{***}
Uncertainty Index	22.51^{***}	25.89^{***}	30.39^{***}	31.65^{***}
C. Sentiment (See Baker & Wurgler (2006) for more details on variables)				
Index (Baker & Wurgler (2006))	-12.63^{*}	-18.76^{***}	-21.55^{***}	-21.87^{***}
IPO first day return	-13.40^{*}	-17.79^{***}	-20.63^{***}	-21.05^{***}
IPO volume	-17.45^{**}	-19.62^{***}	-22.67^{***}	-23.14^{***}
Closed-end discount	-19.40^{***}	-21.33^{***}	23.99^{***}	24.08^{***}
Equity new issues	-12.49^{*}	-16.84^{***}	-18.32^{***}	-18.69^{***}
4. Credit				
Term spread	-0.95	-1.07	-2.45	-2.91
Baa-Aaa spread	23.95^{***}	25.43^{***}	30.46^{***}	31.80^{***}

Table 4.13: Macroeconomic Variable Prediction using VC-SIM approach

This table reports out-of-sample results for 1 month time series for industrial production growth, real GDP growth, real profit growth, Unemployment rate, and GDP growth, Chicago Fed National Activity Index with VC-SIM approaches. MSFE-adjusted is the Clark & West (2007) given in Equation 4.69. *, **, and *** represent the significance at the 10%, 5%, and 1% levels, respectively.

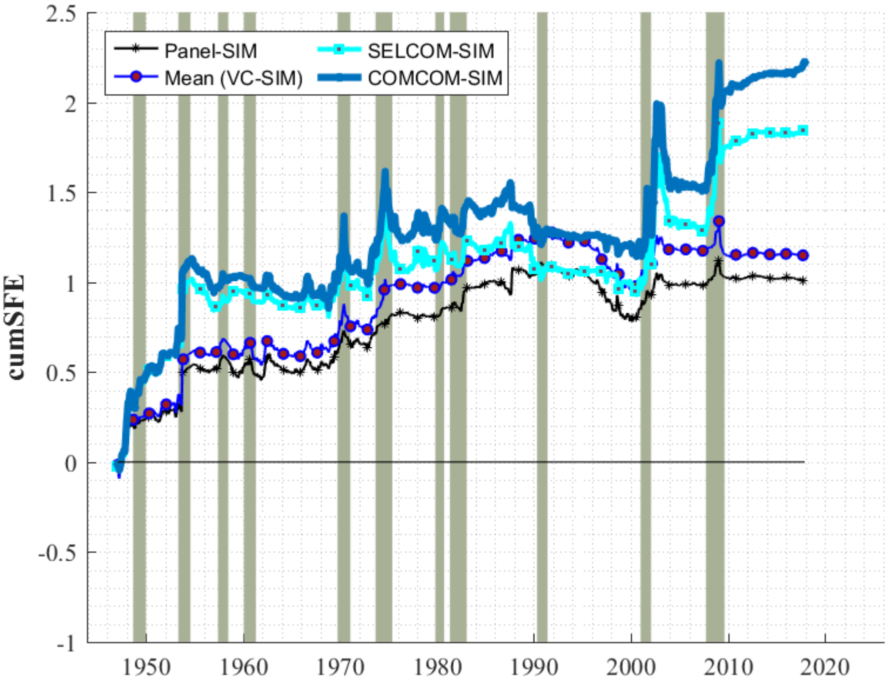
Variable	Panel (SIM)	Mean (SIM)	SELCOM (SIM^*)	COMCOM (SIM^*)
A. Real profit growth				
R^2_{OS} (%)	2.291***	2.621***	3.182***	3.315***
MSFE-adjusted	3.020	3.050	3.1162	3.1762
p-value	0.001	0.001	0.001	0.001
B. Industrial Production growth				
R^2_{OS} (%)	4.153***	5.327***	7.242***	7.924***
MSFE-adjusted	3.921	3.964	4.012	4.123
p-value	0.000	0.000	0.000	0.000
C. GDP growth				
R^2_{OS} (%)	3.548***	4.195***	5.492***	5.734***
MSFE-adjusted	3.882	3.925	3.972	4.082
p-value	0.000	0.000	0.000	0.000
D. CFNAI				
R^2_{OS} (%)	6.772***	7.211***	8.987***	9.723***
MSFE-adjusted	3.980	4.024	4.072	4.185
p-value	0.000	0.000	0.000	0.000
E. Unemployment rate (FRED)				
R^2_{OS} (%)	3.824***	4.273***	5.482***	5.891***
MSFE-adjusted	3.085	3.175	3.932	4.041
p-value	0.001	0.001	0.000	0.000

Table 4.14: Out-of-sample results: Alternative Sample

This table reports the out-of-sample results for sub-sample, 1974:01 to 2018:12. See Table (4.7) for details.

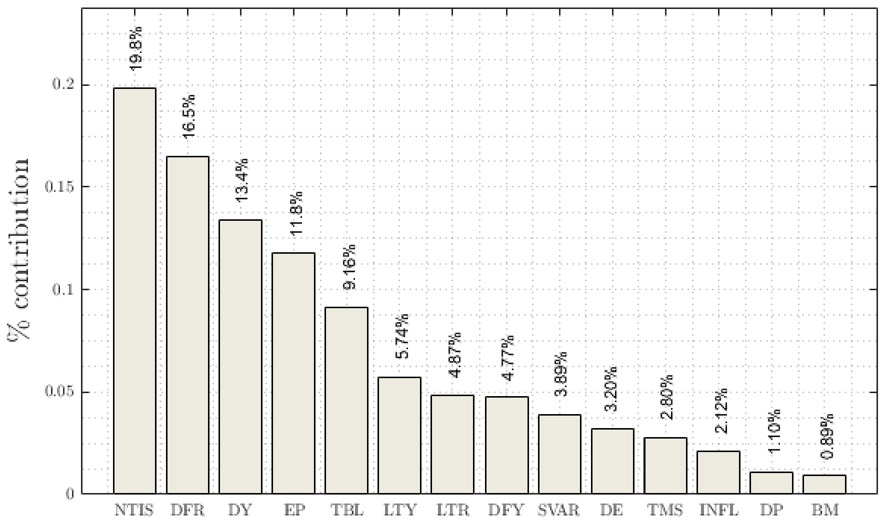
Forecasting model	**MSFE**	R^2_{OS} **(%)**	**MSFE-adjusted**	**p-value**	$\bar{\hat{e}}^2$	**var**($\hat{e}$)
Historical Average (HA)	16.150				0.037	16.113
A. Variance-covariance approach (SIM)						
Panel (SIM)	15.939	1.226***	3.012	0.001	0.004	15.934
Mean (SIM)	15.954	1.324***	3.218	0.001	0.003	15.951
SELCOM (SIM^*)	15.894	1.607***	3.361	0.000	0.026	15.868
COMCOM (SIM^*)	15.882	1.686***	3.447	0.000	0.028	15.854
B. Benchmark Models to deal with SBU						
Mean (EXP)	15.973	1.103***	3.052	0.001	0.029	15.945
Mean (ROLL)	15.998	0.948***	2.986	0.001	0.003	15.995
Mean (BP)	15.992	0.988***	3.046	0.001	0.009	15.983
Mean (SEL-CV)	15.959	1.195***	3.056	0.001	0.005	15.954
Mean (RobW1)	15.977	1.080***	2.984	0.001	0.004	15.973
Mean (RobW2)	15.977	1.077***	2.971	0.001	0.004	15.973
Mean (EW)	15.943	1.293***	3.398	0.000	0.003	15.940
Mean (EW10)	15.954	1.226***	3.339	0.000	0.004	15.950
Mean (LW)	15.950	1.249***	3.218	0.001	0.004	15.946
Mean (VW-MSFE)	15.951	1.243***	3.05	0.001	0.001	15.951
Mean (ROC)	15.936	1.341***	3.499	0.000	0.001	15.934
Mean (ROC-L)	15.942	1.302***	3.417	0.000	0.002	15.940
Mean (RS)	15.982	1.049***	3.013	0.001	0.005	15.977
C. Standard Benchmark Models in Equity Premium Prediction						
C.1 Kitchen Sink						
KS	18.015	-10.356	0.981	0.162	0.032	17.983
C.2 Dimension Reduction						
PCR1	16.064	0.529**	1.951	0.026	0.016	16.048
PCR3	16.008	0.884**	2.173	0.015	0.019	15.989
KP-3PF	16.238	-0.542	0.038	0.486	0.020	16.218
C.3 Shrinkage						
LASSO	16.192	-0.262	0.501	0.308	0.000	16.192
Ad-LASSO	16.216	-0.409	0.463	0.322	0.002	16.214
Ridge	16.225	-0.464	0.912	0.181	0.003	16.222
ENet	16.191	-0.253	0.872	0.192	0.000	16.191
C.4 Forecast Combination						
Mean	15.973	1.104***	3.052	0.001	0.029	15.945
T-Mean	15.986	1.025***	3.017	0.001	0.029	15.956
Median	16.002	0.922***	2.808	0.002	0.015	15.987
DMSFE ($\theta = 0.9$)	15.963	1.169***	3.121	0.001	0.028	15.935
DMSFE ($\theta = 1$)	15.974	1.101***	3.007	0.001	0.028	15.945
RA1	16.481	-2.012	0.976	0.163	0.000	16.481
RA2	16.303	-0.943	0.039	0.485	0.014	16.289
RA3	15.975	1.093***	3.440	0.000	0.018	15.957
RR	16.062	0.547**	1.721	0.042	0.024	16.038
VC-SIM (Optimal)	15.960	1.196***	3.102	0.001	0.004	15.956

Figure 4.3: CumSFE - CV-SIM approaches



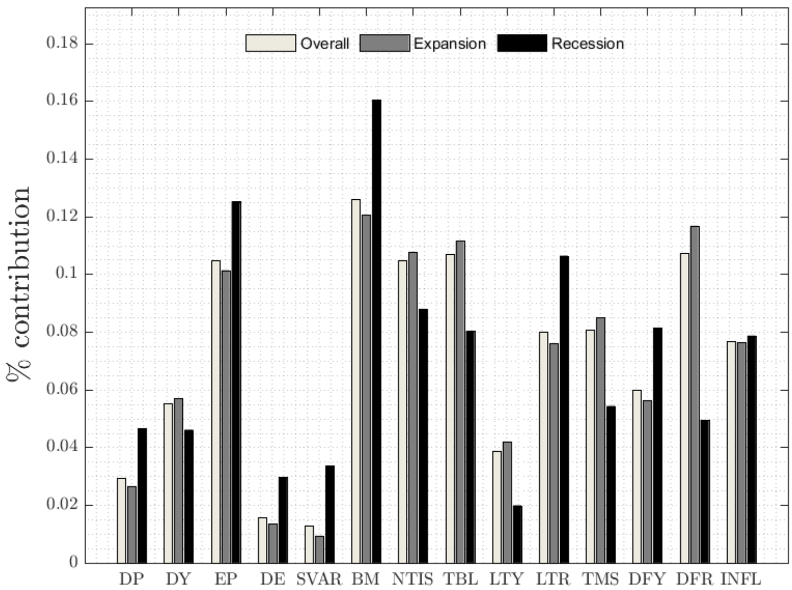
Note: This figures plots the Cumulative squared forecast error (CumSFE) for the historical average (HA) benchmark model minus the CumSFE for our VC-SIM approaches from 1947:1 to 2018:12. When the curve rises, the given VC-SIM model outperforms the benchmark, while the opposite holds when the curve falls.

Figure 4.4: R^2-based Importance of each predictor in OOS equity premium forecast



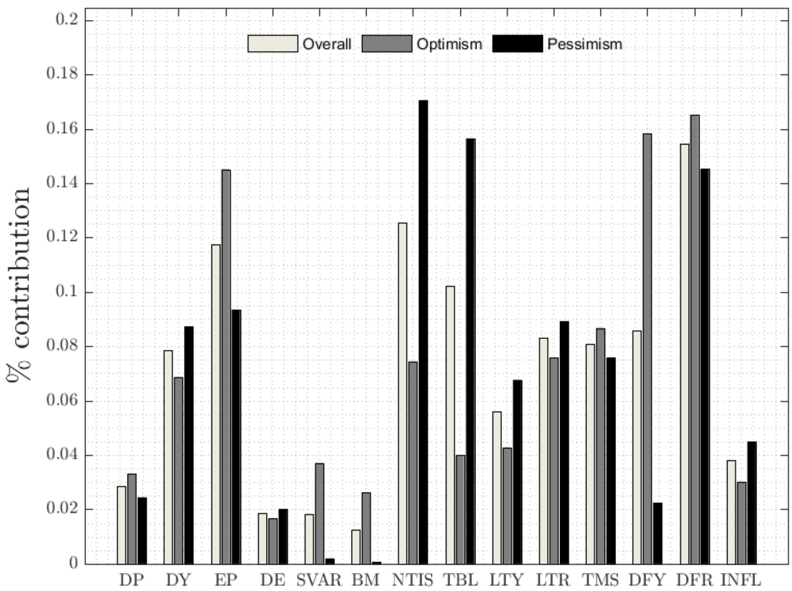
Note: This figure plots R^2-based importance of each predictor which is normalized to sum to one. Importance is defined as reduction in out-of-sample R^2 due to absence of a given predictor. Importance factor of individual is taken as average over sample-splits (1947-2000).

Figure 4.5: Importance of each predictor in EP predictability



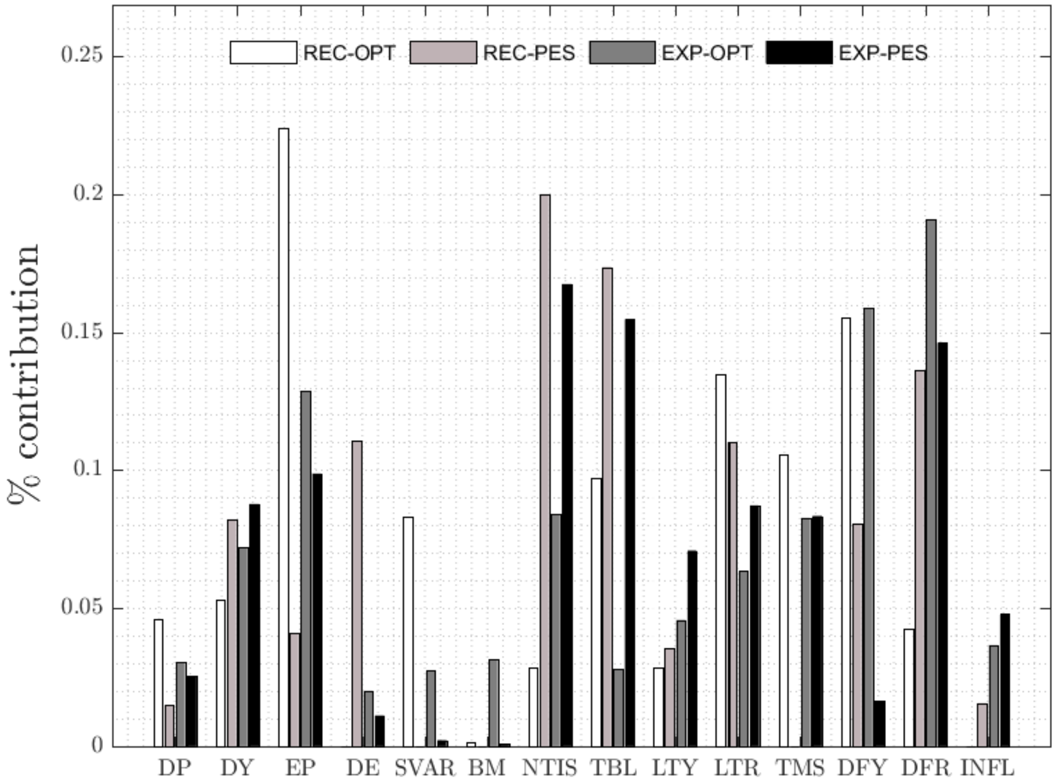
Note: This figure plots the average optimal weight of individual predictor over full sample which is normalized to sum to one. Recession and expansion periods are based on NBER dates.

Figure 4.6: Importance of each predictor over optimism and pessimism periods



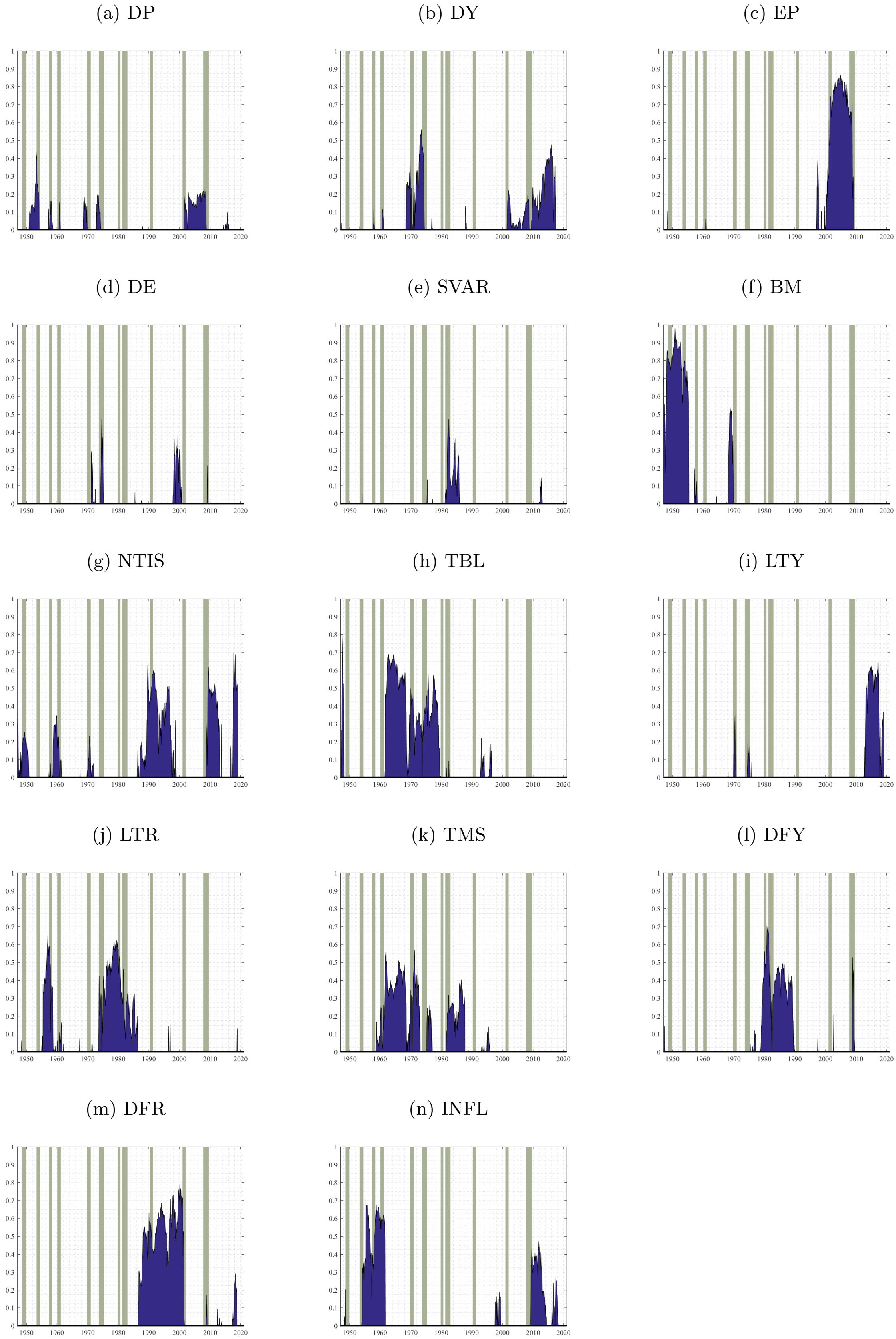
Note: This figure plots the average optimal weight of individual predictors over sub-sample (1974:01-2018:12) which is normalized to sum to one. Optimism and pessimism periods are defined following Baker & Wurgler (2006) where any period with positive (negative) sentiment index is defined as optimism (pessimism).

Figure 4.7: Importance of each predictor for NBER-dated recession



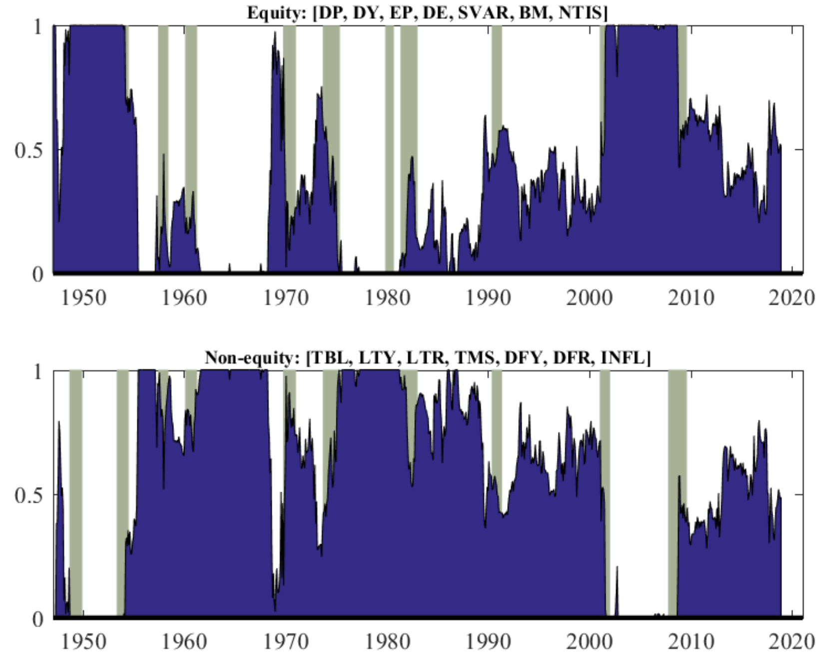
Note: This figure plots the average optimal weight of individual predictors over sub-sample (1974:01-2018:12) which is normalized to sum to one. REC, OPT, PES, and EXP indicate recession, optimism, pessimism and expansions respectively. Recessions and expansions are indicated by NBER dates whereas optimism and pessimism periods are defined following Baker & Wurgler (2006) where any period with positive (negative) sentiment index is defined as optimism (pessimism).

Figure 4.8: Contribution of each predictor in optimal OOS equity premium forecast



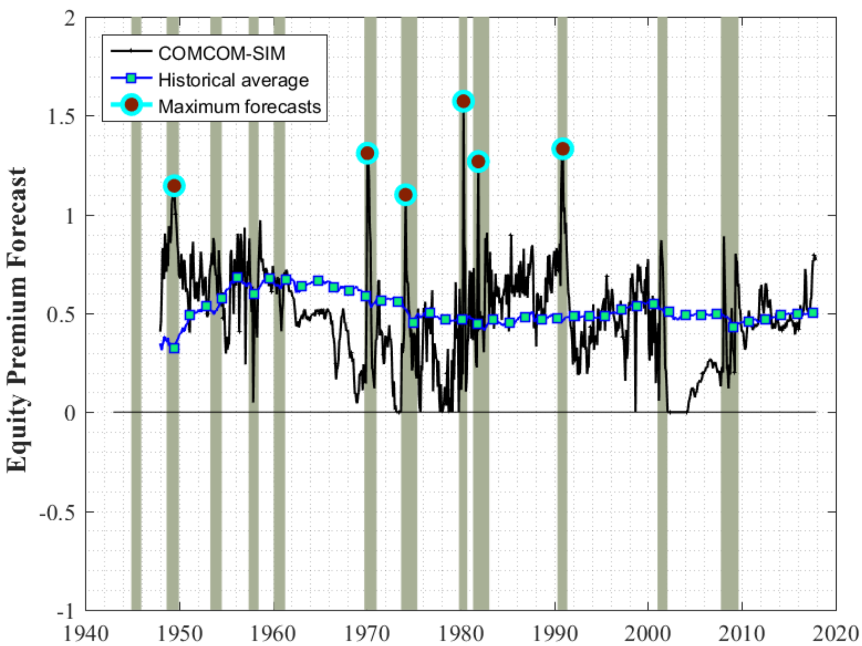
Note: This figure shows the contribution of each predictor in composite equity premium forecast at each time t. The sample period is 1947:01-2018-12 and shaded area represent NBER recession periods.

Figure 4.9: Contribution of equity and non-equity predictors



Note: This figure shows the contribution of predictor groups (financial ratios and macro) in composite equity premium forecast at each time t. Predictors are grouped into equity (DP, DY, EP, DP, BM, and SVAR) and non-equity (TBL, LTY, LTR, TMS, DFY, DFR and INFL). The sample period is 1947:01-2018-12 and shaded area represent NBER recession periods.

Figure 4.10: Equity Premium Forecasts - Monthly returns



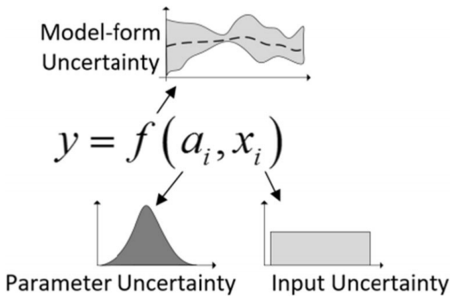
Note: This figure plots out-of-sample equity premium forecasts with COMCOM-SIM and historical average over 1947:01 to 2018:12. Following Campbell & Thompson (2008), negative forecasts are set to zero for better comparison with historical average.

Chapter 5

Conclusion and Future directions

In this thesis, we analysed two important subjects in Empirical Finance: the Conditional Capital Asset Pricing Model (CCAPM) and out-of-sample equity premium prediction. Our focus in three different essays remained on the fact that economic theory does not always define the functional relationship between the dependent (target) and explanatory (predictor) variables, nor does it always define a specific set of covariates. This implies that model uncertainty is widespread in empirical economics and finance. Uncertainty arises when there is a lack of general agreement in a particular field of study. In such instances, there are often many different models and methods attempting to describe the variable of interest. According to Ziyadi & Al-Qadi (2019), there are three key sources of model uncertainty.[1] Figure 5.1 shows these sources, where the dependent variable (y) is defined as a function of explanatory (independent) variables (x) and model parameters (α). The three sources of model uncertainty include: *i) input uncertainty* (e.g., relevance of explanatory variables, etc.); *ii) parameter uncertainty* (e.g., uncertainty of model parameters due to data quality and structural breaks, etc.); and *iii) model form uncertainty* (e.g., assumption on relationship between x and y, i.e., linear vs non-linear, etc.).

Figure 5.1: Three sources of model uncertainty



Source: *Ziyadi & Al-Qadi (2019)*

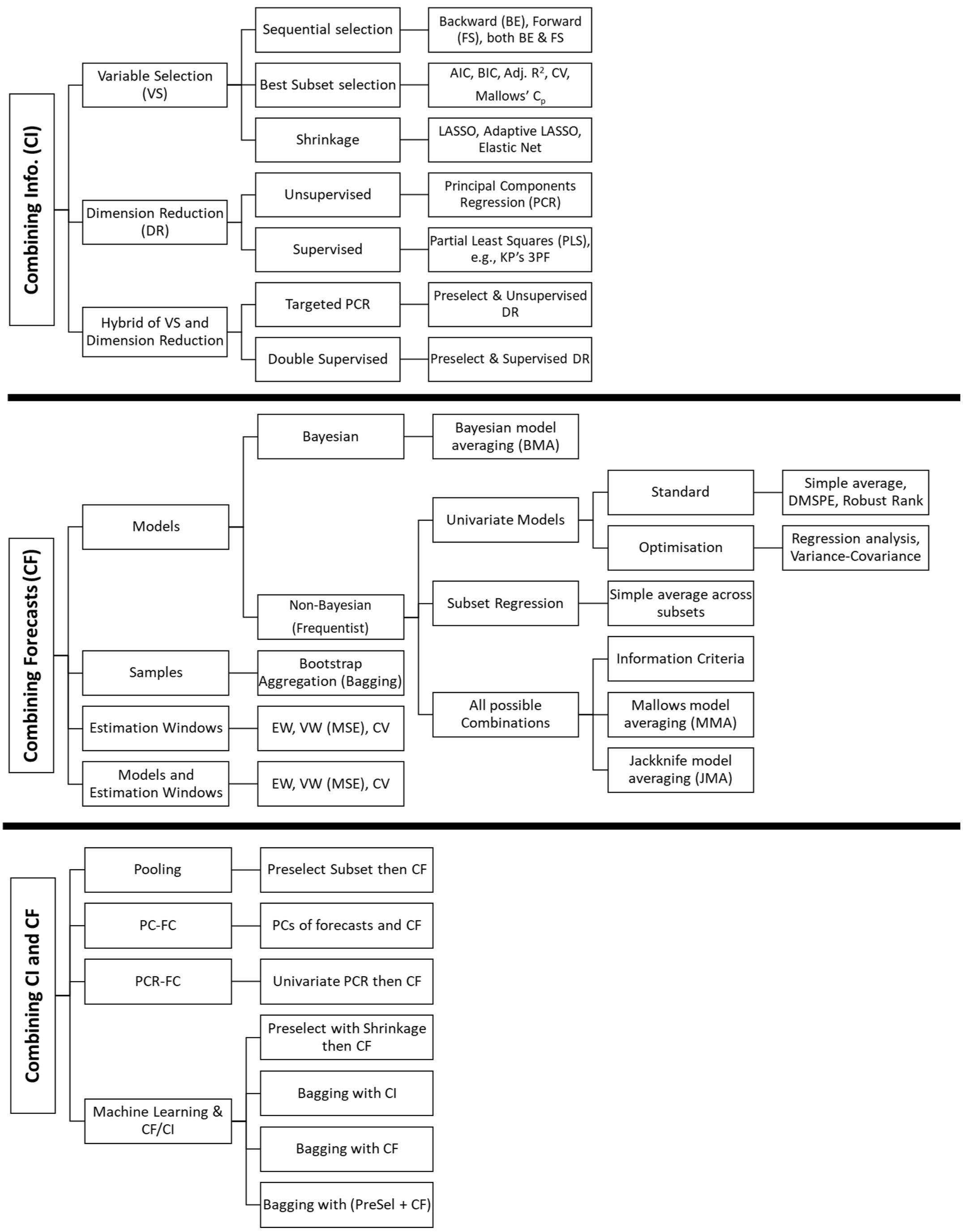
This thesis only focuses on the first two sources of model uncertainty, namely input and parameter uncertainties, and assumes that there exists a linear relationship between y and x. Therefore, this should be considered as a limitation of this study. The *input uncertainty* (IU) mainly arises when there is no guidance on the choice of independent variables. The *parameter uncertainty* (PU), on the other hand, mainly arises due to structural breaks affecting the underlying relationship between the variables in the model. Pesaran & Timmermann (2007) show that the performance of a forecasting model in the presence of structural breaks depends on the number

[1] For more details, see Clyde & George (2004), Loucks et al. (2005), Young (2018), and others.

of observations (window length) used to estimate the out-of-sample forecast. However, there is no clear consensus in the literature on the number of observations to be used in estimation, which is usually referred to as *estimation window uncertainty* (EWU) (Pesaran & Timmermann 2007).

Since our objective was to address these two issues from CCAPM and equity premium prediction perspective, therefore, a review of forecasting literature helped us to identify three main strategies to deal with these issues, which are summarised in Figure 5.2.

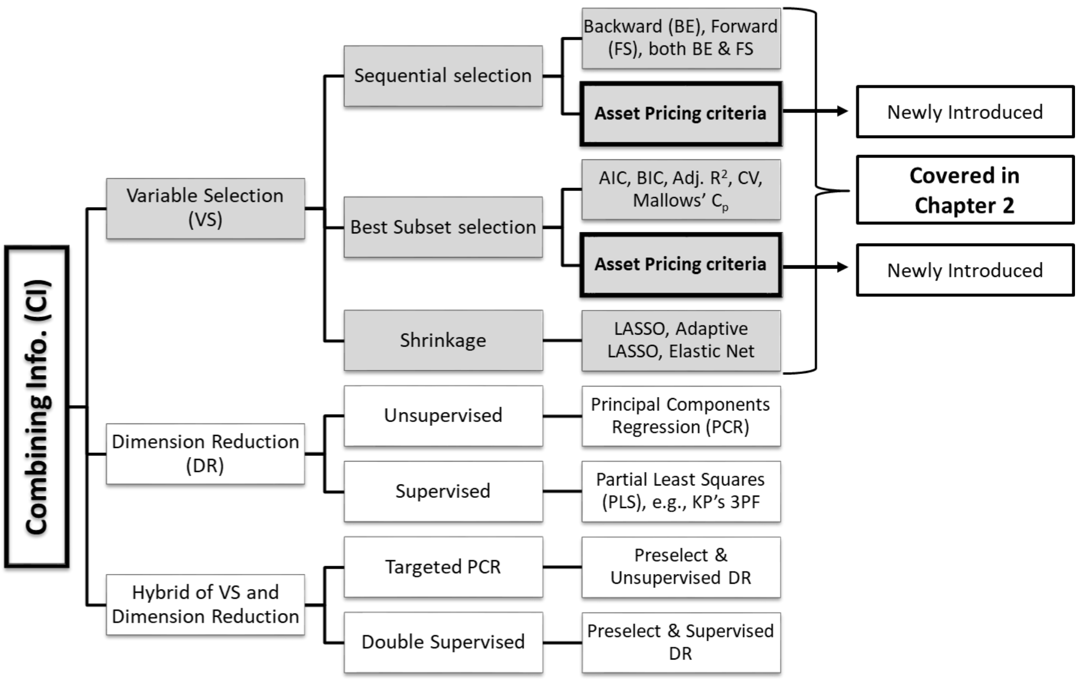
Figure 5.2: Strategies to overcome Model Uncertainty



5.1 Contribution to asset pricing literature

The first two essays were related to the tests of conditional CAPM. The main objective of the first essay was to address the *variable-selection uncertainty* (VSU) by applying the dynamic model selection (DMS) approaches to CCAPM-IV. Specifically, we introduced a CCAPM model where the choice of conditioning variables, used to capture the variation in conditional betas, is allowed to vary through time and is selected from a large pool of potential state variables. The main research question was whether the cross-section of expected returns on 25 Size-B/M portfolios could be explained by conditional CAPM with *variable selection* (VS) approaches. Figure 5.3 suggests that the focus of essay one has been on *combining information* (CI) approaches only where three different methods of VS were considered. The important contribution of our first essay is the introduction of DMS approach that considers each model by its performance in pricing the assets. We call this approach the *dynamically selected beta model* (DSBM). Specifically, this approach selects the beta models that perform the best based on standard asset pricing tests on past data at each point in time. Results suggest that DSBM performs better than traditional approaches such as sequential selection (e.g., forward selection, backward elimination, and stepwise regression), best subset selection (e.g., adjusted R^2, AIC (Akaike 1973), BIC (Schwarz 1978), Mallows's C_P (Mallows 1973) and shrinkage methods (e.g., LASSO ((Tibshirani 1996), Adaptive LASSO (Zou 2006), and Elastic Net (ENet) (Zou & Hastie 2005) by achieving higher R^2. One potential reason for the traditional methods' poor performance is their reliance on residuals from CCAPM-IV regression as their primary objective function. However, according to the CAPM theory, when the returns are measured in excess of the risk-free rate, the intercept term a_i^{IV} indicates the expected abnormal return, which should be zero. Therefore, the DSBM ensures that a model that minimises the pricing errors is selected to capture beta dynamics.

Figure 5.3: Strategies to overcome VSU used in Chapter 2 (Essay 1)



However, consistent with Lewellen & Nagel (2006), we find that all the DMS approaches, including DSBM, cannot explain the value and momentum anomalies. Using bootstrap methods to quantify the model uncertainty and instability, we find that DMS approaches do not fully account for variable-selection uncertainty (VSU). These findings are in line with recent criticism of DMS approaches regarding their failure to account for variable-selection uncertainty fully and to achieve model stability (e.g., Smith 2018, Petropoulos et al. 2018, Makridakis et al. 2020). These findings motivated us to consider alternative strategies from forecasting literature in the second essay to address variable-selection uncertainty from the CCAPM-IV perspective.

In the second essay, we applied *dimension reduction* (DR) (e.g., Ludvigson & Ng 2007, Neely et al. 2014, Kelly & Pruitt 2013), *combining forecasts* (CF) (e.g., Bates & Granger 1969, Timmermann 2006, Rapach et al. 2010) and a hybrid of *combining information* (CI) and CF (e.g., Huang & Lee 2010, Hirano & Wright 2017, Rapach & Zhou 2020) approaches to CCAPM-IV.[2] These approaches are summarised in Figure 5.4. Our second essay contributes to the literature in the following ways: to the best of our knowledge, this is the first study that provides a comprehensive comparison of various well-known approaches identified by literature to deal with model uncertainty related to variable selection from a CCAPM perspective. Our out-of-sample results suggest that CF approaches dominate the CI approaches in explaining the cross-section of assets returns. Finally, consistent with studies as Hirano & Wright (2017), Liu & Xie (2019), and Rapach & Zhou (2020), we show that a combination of conventional econometric methods and machine learning methods can outperform the individual methods. For example, we find the evidence on improved performance of CCAPM-IV with bagging (BAGG) method where, in each pseudo sample, we first select the subset of variables based on the mean squared forecasting error (MSFE) in cross-validation sample, and then take a simple average of beta estimates across all pseudo samples. This method performs as well as the Fama & French (1993) three-factor model in explaining the cross-sectional returns of 25 Size-B/M, 30 industry and ten momentum portfolios.

5.2 Future directions for asset pricing literature

Following points summarise the limitations and future directions that can be generalised to both essay 1 and 2.

5.2.1 Applying (non-linear) machine learning models

One of the limitations of essay 1 and 2 is that they are mainly related to the variable-selection uncertainty and make assumptions about functional form and parameter uncertainty. Following mainstream asset pricing literature (e.g., Jagannathan & Wang 1996, Ferson & Harvey 1999, Lettau & Ludvigson 2001), both essays assume that the relationship between asset returns and predictor variables is linear. However, some studies evaluate this assumption, Cai et al. (2015), for example, found strong evidence against the linearity assumption, arguing that such a strong assumption could lead to model misspecification.[3] Inference and estimation based on misspecification can become very deceptive, as Ghysels (1998) indicates. Moreover, Ghysels (1998) demonstrated that among several well-known time-varying beta models, severe misspecification could result in highly volatile time variance in the beta, which can lead to significant pricing errors. On the other hand, both the studies, essay 1 and 2, use a rolling window scheme criticised in the literature for not choosing the optimal estimation window. As a result, failing to address the issue of *estimation window uncertainty* (EWU) (e.g., Pesaran & Timmermann 2007, Inoue & Rossi 2011).

[2] Note that the data (sample, test portfolios, and predictor variables) are identical to essay 1.

[3] Also see Wang (2002) and Wang (2003) on the relationship between return and predictors.

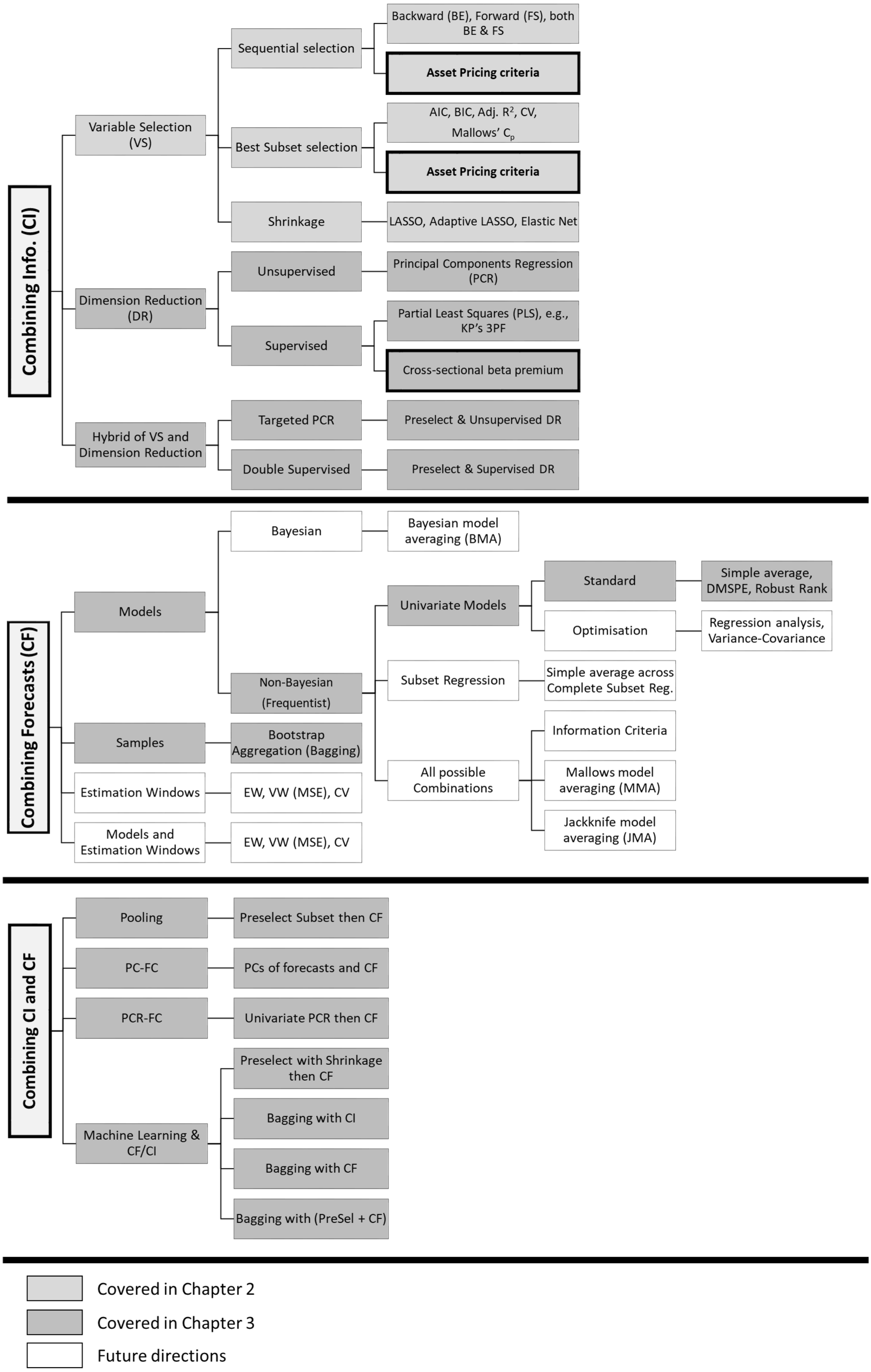
In future, these two sources of uncertainties (parameter and functional form) can be addressed from CCAPM-IV perspective. Specifically, recent research shows that non-linear regression strategies like regression trees and neural networks outperform linear regression techniques in forecasting (Gu et al. 2020). Moritz & Zimmermann (2016) and Gu et al. (2020) both point out the limitations of OLS regression when it comes to variable selection in high-dimensional datasets. From a CCAPM-IV standpoint, it is worthwhile to use these non-linear approaches. Moreover, the techniques used in Chapter 4 to address the issue of EWU in forecasting equity premium can be applied in future studies to address the EWU in estimating conditional betas from a CCAPM-IV perspective.

5.2.2 Applying alternative approaches of CF

A review of the literature summarised in Figure 5.2 suggests that the CF approaches can be classified into three groups based on combining forecasts: i) across models, ii) across samples, and iii) across estimation windows. In the second essay, we only considered CF across models and samples and does not focus on CF across estimation windows. Moreover, CF across models can be grouped into Bayesian and Non-Bayesian. Under the Bayesian approach, the Bayesian model averaging (BMA) is well-known in forecasting literature that assumes that prior knowledge is available on a set of possible models which contains the true model (Leamer 1978). The Non-Bayesian or Frequentist approach, on the other hand, requires an estimate of the parameter of interest as well as a related standard error by assigning appropriate weights to each model. Unlike the Bayesian approach, the Frequentist approach does not involve priors on parameters and models, and the corresponding estimators are solely based on data. In comparison to the Bayesian approach, the frequentist approach is thought to be less complicated and easy to implement. In addition, there is also some criticism on the Bayesian approach, in particular, for specifying the prior model probabilities. According to Brock et al. (2003) and Ley & Steel (2011), the chosen prior distributions may significantly impact the outcome of BMA analysis.[4] As a result, we only focused on Non-Bayesian approaches in this thesis, where combining weights are based on a simple average or other performance criteria like mean squared forecast error (MSFE). However, BMA might have been used as a benchmark model from a CCAPM perspective to provide a direct comparison with Frequentist approaches, but we did not consider this because we used CF approaches in essay 2, which required us to estimate all possible combinations of predictor variables, which was computationally intensive. Moreover, many studies such as Rapach et al. (2010) use simple average across univariate models in their applications. For these reasons, we restrict our CF analysis to combining forecasts from univariate predictor-based models, requiring only $M = K$ models instead of $M = 2^K$ models to be estimated.

[4]See Moral-Benito (2015), who provides an excellent survey on Bayesian and Frequentist approaches.

Figure 5.4: Strategies to overcome VSU used in Chapter 3



This also opens up the possibility of considering other approaches to combining beta forecasts from all possible models and the complete subset regression (CSR) methodology proposed by Elliott et al. (2013).[5] CSR is a direct and straightforward method for dealing with model uncertainty, model instability, and estimation error. The CSR approach involves using k out of K variables ($k \leq K$) to fit linear regressions for all possible combinations of the k variables. K is the total number of predictors. The final forecast is the equally-weighted average forecast computed from all regressions. In addition, instead of using a simple average, CSR can be applied to CCAPM-IV using asset pricing criteria to see whether we can address the criticism of Boot & Nibbering (2020) that there are not enough empirical studies using alternative weighting schemes and addressing the issues related with optimal model subspace (k). Specifically, there is a possibility to use value weights based on mean squared forecasting error based on cross-validation (MSFE-CV) approach and using the same approach MSFE-CV to identify the optimal k. This should result in developing various strategies to evaluate including: i) value-weighted forecast for specific k, ii) equally-weighted forecast for optimal k, and iii) value-weighted forecast for optimal k. Where the standard approach of the equally-weighted forecast for specific k can serve as a benchmark model.

Next, as shown in Figure 5.4, our essay 2 only covers the CF approach based on simple average (e.g., Rapach et al. 2010). However, as shown in Chapter 4 that the *variance-covariance* (VC) approach that emphasises the consideration of correlation among forecasting errors can improve the forecast accuracy. Considering this, the VC approach can also be applied in future studies to CCAPM-IV.

5.2.3 Combining betas across various approaches

The findings of essay 2 suggests that CF approaches applied to CCAPM-IV can outperform the individual models and CI approaches in explaining the cross-section of asset returns. There are also some other studies reporting better performance of CF approach in pricing assets. Specifically, these studies model the beta dynamics by combining betas estimated with different frequencies without relying on conditioning information variables. For example, Cenesizoglu & Reeves (2018) model an asset's conditional beta as a weighted average of short-, medium- and long-run betas calculated over different periods based on different frequency data, while González et al. (2012) employ MIDAS (mixed data sampling) regressions to estimate a portfolio's conditional beta as a weighted average of a high and low-frequency components. Our approach in essay 2 differs from these studies in the way that our focus was to apply CF approaches to estimate asset betas with the same frequency (monthly) based on different predictor variables. Overall, we have evidence of CF approaches' effectiveness in combining asset betas across different frequencies without relying on conditioning information (e.g., Cenesizoglu & Reeves 2018, González et al. 2012) and combining betas obtained using the same frequency but different variables (e.g., essay 2).

The application of CF can also be extended to other areas of asset pricing, in particular CCAPM. For example, we learn from the literature that there are several other methods for capturing beta-dynamics besides CCAPM-IV. Some of the famous approaches include those using data-driven filters such as beta calculated from a 60-month rolling window as in Fama & MacBeth (1973), or a short window approach (Lewellen & Nagel 2006) and high-frequency data (Andersen et al. 2003), multivariate GARCH (Bollerslev et al. 1988), dynamic conditional correlation (DCC) (Engle 2002, Bali & Engle 2010), regime-switching model (Vendrame et al.

[5] CSR has been applied to various research areas, including equity premium (Meligkotsidou et al. 2021), bond betas Aslanidis et al. (2019), inflation (Garcia et al. 2017), exchange rates (Beckmann & Schüssler 2016), employment growth (Borup & Schütte 2020), commodity prices (Gargano & Timmermann 2014), and many others.

2018), mean-reverting stochastic process (Jostova & Philipov 2005), Kalman filter (Adrian & Franzoni 2009), and others. Given that betas obtained through CF approaches can outperform the betas from a single model, the CF approach applied to combine betas obtained through various above-mentioned approaches may explain the cross-section of asset returns more than a single approach.

5.2.4 Extending the information set

One of the limitations of essay 1 and 2 is sticking with standard 14 variables of Goyal & Welch (2008) because among various approaches that we compared is the subset variable selection that requires estimation of all possible models (2^N). Therefore, a large dataset makes it computationally infeasible to evaluate all possible models. For example, the inclusion of 30 variables means that we need to estimate more than 1.07 billion models at each period for model selection or averaging. However, findings from essay 2 suggest that combining univariate beta forecasts, requiring N estimates only, performs better than dynamic model selection (DMS) approaches that require 2^N models. This opens up the possibility to test the CCAPM-IV approach with many predictors based on CF across univariate predictor-based models and dimension reduction approaches only. Many asset pricing researchers, in particular, are looking for novel predictive firm characteristics for explaining anomalies that traditional capital asset pricing and factor models fail to capture. Several independent analytical studies based on a data science methodology have recently demonstrated the value of employing a greater number of economically interpretable predictors relevant to firm characteristics and other common factors (e.g., Moritz & Zimmermann 2016, Gu et al. 2020, Feng et al. 2018). Gu et al. (2020), for example, examine a dataset of more than 30,000 individual stocks from 1957 to 2016 and identify over 900 baseline signals. Moreover, the dataset can also be extended to the Federal Reserve Economic Data (FRED) database (e.g., McCracken & Ng 2016) and new predictor variables that have proved essential in establishing equity premium predictability. These predictors include technical indicators (Neely et al. 2014, Lin 2018), investor sentiment and attention (Huang et al. 2015, Ni et al. 2015, Chen 2017, Coqueret 2020, Zhang et al. 2021), manager sentiment (Jiang et al. 2019), the short interest index (Rapach et al. 2016), bitcoin prices (Salisu et al. 2019), and credit quality (Chava et al. 2015, Narayan et al. 2017), among others.

5.2.5 Introducing Learning

Adrian & Franzoni (2009) suggest that the key explanation for the unconditional CAPM's failure is that the model does not mimic the learning process of investors. As a result, they propose a learning-based CAPM to estimate conditional betas. Their model is based on the assumption that investors conduct systematic learning activities on observed asset returns, infer risk loadings from available data, and update them optimally as new data becomes available. According to their findings, the learning-based CAPM outperforms the unconditional CAPM. In addition, they show that the introduction of conditioning information in their learning-CAPM can further improve the performance. However, they only use standard predictors such as lagged market index, the term spread, the value spread, and the consumption-to-wealth ratio (CAY) as conditioning variables. Therefore, it can be interesting to evaluate the CCAPM-IV models introduced in essay 1 and 2 with learning. The analysis will not only based on the extended set of conditioning information variables but also let the information be optimally selected over time. In addition, various approaches such as CF and *principal component forecast combination* (PCFC) can be considered to estimate the time-varying betas.

5.2.6 Linking optimal conditioning to equity premium forecasts

Our approaches, in particular, DMS and CF are used to estimate time-varying betas. However, it is worth analysing that the optimally selected information set at a particular period explaining the cross-section of asset returns can also forecast the equity premium out-of-sample. This can be done by using DSBM to select the best subset of variables and using those variables to generate out-of-sample forecasts of the equity premium. And in the case of CF, the optimal weights obtained through methods pricing asset at a given time can be applied to equity premium forecasts to generate combined forecasts.

5.3 Contribution to the equity premium literature

In the third essay, we proposed a new combining forecasts (CF) approach based on a variance-covariance (VC) method that addresses estimation window uncertainty (EWU) and variable-selection uncertainty (VSU) simultaneously to improve the out-of-sample forecasts of the equity premium. Our third essay contributes to the existing literature by complementing the current literature on methods for directly dealing with EWU in forecasting, such as Pesaran & Timmermann (2007), Pesaran & Pick (2011), Rossi & Inoue (2012), Pesaran et al. (2013), Tian & Anderson (2014), Wang et al. (2020) and others. This is the first study, to the best of our knowledge, to apply the VC approach for combining estimation windows of individual models. We show that considering the correlation among forecast errors across estimation windows can significantly improve the forecasting accuracy of individual models. Secondly, for the first time in the forecasting literature, we introduce a panel combination approach based on VC approach to address the model uncertainty and parameter instability simultaneously. Based on out-of-sample forecasting results of the equity premium, we show that our new model not only outperforms the existing AveAve approach of Pesaran et al. (2013) but also existing approaches in equity premium such as CF (Rapach et al. 2010), dimension reduction methods (Kelly & Pruitt 2013, Neely et al. 2014) and shrinkage methods (Li & Tsiakas 2017, Zhang et al. 2020).

5.4 Future directions for the equity premium literature

The limitations and future directions of our third essay are discussed below.

5.4.1 Extending EWU and VSU to non-linear models

We introduced the variance-covariance approach to address the issues of VSU and EWU. However, as discussed earlier, the functional form is the third source of model uncertainty that we do not consider in our third essay. However, given the success of non-linear machine learning methods (e.g., Gu et al. 2020), it is worth considering these methods from the perspective of addressing EWU by applying our proposed methods.

5.4.2 Extending the information set

To respond the critique of Goyal & Welch (2008), recent academic research shows that certain new *predictors* and *econometric methods* can improve the out-of-sample predictability of stock returns. Our third essay belongs to the second category of studies that use new econometric methods to addresses parameter instability and variable-selection uncertainty issues using standard Goyal & Welch (2008) predictors. In future, by expanding the dataset mentioned in section (5.2.4), the methods introduced in essay 3 can be applied to both linear (already considered in the third essay) and non-linear models.

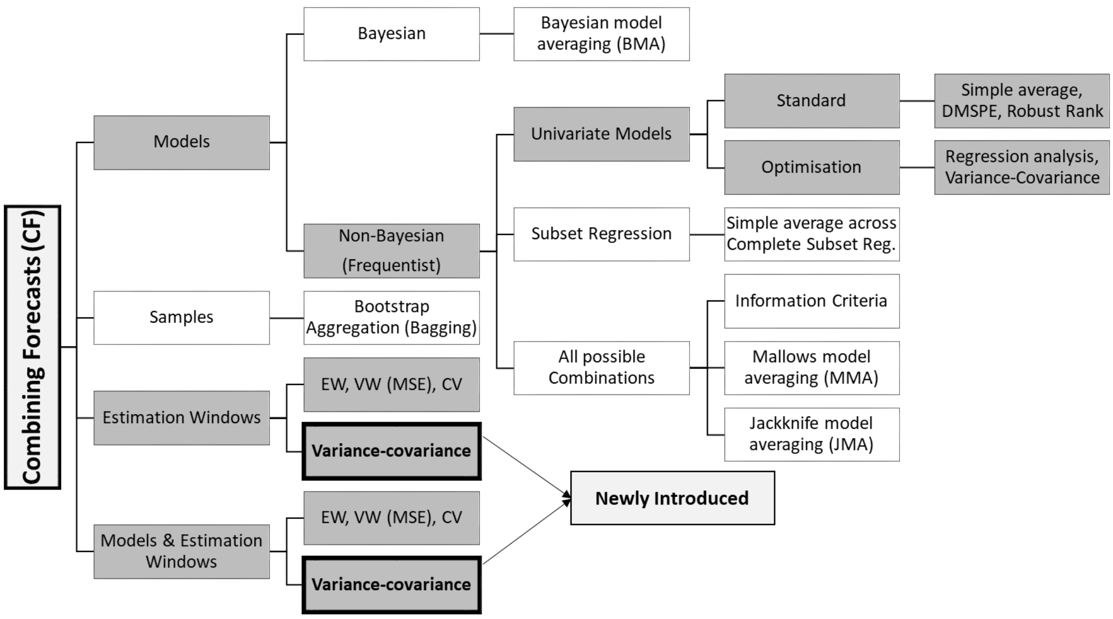
5.4.3 Applying hybrid of CI and CF approaches

Figure 5.5 shows that our focus in essay 3 was on the combining forecasts (CF) method, which is commonly used to address both the estimation window and variable-selection uncertainties. However, as shown in essay 2 that Bagging and hybrid of CI and CF approaches can improve the forecasting performance, there is potential to consider these techniques in addressing VSU and EWU.

5.4.4 Addressing EWU for other approaches

We demonstrated in essay 3 that our methods based on CF using variance-covariance approach outperform established well-known methods such as dimension reduction (DR) and shrinkage methods. However, the comparison is incomplete since we do not consider the EWU for these models; instead, they are all estimated using an expanding window. According to studies such as Zhang et al. (2020), addressing EWU for methods like shrinkage improves forecasting accuracy as compared to estimates made with a single-window, such as rolling or expanding. Given this, EWU for DR and shrinkage methods in forecasting out-of-sample equity premium forecasts could be addressed. This is also interesting because it would eventually result in a hybrid model; for example, the DR method would generate a forecast for each window, which would then be combined to generate a final forecast. As a result, the strategy effectively combines DR and CF.

Figure 5.5: Strategies to overcome VSU and EWU used in Chapter 4



5.4.5 Applying theoretical constraints

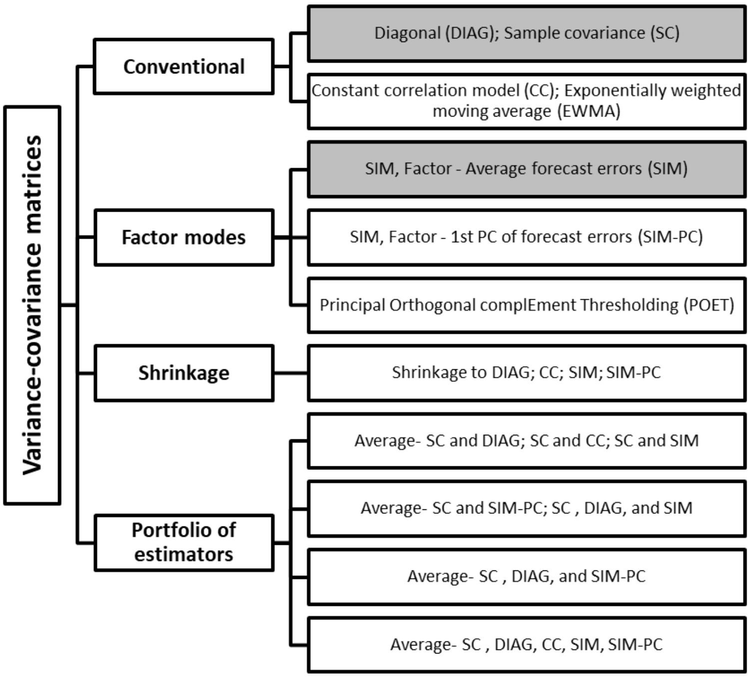
The optimal combining forecast (CF) method uses the variance-covariance approach that is based on the objective function of minimising the variance of the portfolio. Our study is limited to the standard approach. However, some studies use some theoretical constraints in generating the optimal forecast. For example, Hsiao & Wan (2014) implement different geometric methods in large samples and provide a simple eigenvector approach for combining forecasts. The authors empirically show that this approach can perform better than the existing optimal CF and simple average. Most recently, Chan & Pauwels (2018) present a method for analysing the issue of combining forecasts using various forecast metrics, such as Mean Squared Error (MSE), and demonstrate that CF outperforms simple average and optimal CF based on standard approach. Following these studies, theoretical constraints can be applied to the optimal CF methods discussed in the third essay. In addition, there is also a possibility to apply some of the other objective functions other than minimum portfolio variance that have been considered in the application of portfolio optimisation. These methods include: (i) inverse volatility (e.g., De Carvalho et al. 2012), ii) equal-risk-contribution (Maillard et al. 2010, Mausser & Romanko

2014), and iii) maximum diversification (e.g., Choueifaty & Coignard 2008). See Neffelli (2018), who provide comprehensive details on these methods.

5.4.6 Applying alternative covariance matrices

The variance-covariance approach of optimal CF relies on the estimation of the covariance matrix. Our analysis suggests that the results are sensitive to the selection of covariance-matrix. For example, *single index model* (SIM) of Figlewski (1983) outperforms diagonal and sample covariance matrices. We only considered these three approaches(diagonal, sample, and SIM) to estimate the error-covariance matrix. This can be considered as a limitation of our study because Figure 5.6 suggests that there are various methods of estimating covariance-matrix. We classify them into conventional, factor models, shrinkage, and portfolio of estimators.

Figure 5.6: Approaches to estimate covariance-matrix



The conventional methods include diagonal, sample, constant correlation model (CC) of Elton & Gruber (1973), Constant covariance model (CCov) of Pantaleo et al. (2011), and exponentially weighted moving average (EWMA) of Morgan (1996). However, in essay 3, we only considered two conventional methods, namely diagonal and sample. The number of estimated parameters to construct a sample covariance matrix grows with the square of the number of predictors, resulting in high estimation error. Ledoit & Wolf (2003) interpret the sample covariance matrix consisting of N assets as an N factor model (forecasting models in our case). Imposing a factor structure on the covariations among assets is a common way to reduce estimation error in the covariance matrix. Since the factor structure decreases the number of parameters to be calculated, the estimation error is reduced. However, with just a few factors, the estimated covariance matrix cannot capture all asset relationships, resulting in specification error. Figlewski (1983) propose a single index model (SIM) that requires a small number of estimates to construct the covariance matrix. Following Figlewski (1983), in essay 3, we used average forecast errors as a single factor. However, instead of using average asset returns as a factor, some studies use a principal component as the factor (e.g., Fan et al. 2013).

Ledoit & Wolf (2003) argue that though the single factor covariance model reduces the estimation error, it also increases bias due to its dependence on a single source of risk. According to the authors, the best trade-off is to find the optimum combination of bias and prediction error. This resulted in the development of covariance matrix estimators such as linear shrinkage (Ledoit & Wolf 2004) and extensions thereof (Ledoit & Wolf 2012, Engle et al. 2019, Ledoit & Wolf 2020). Instead of the optimally weighted shrinkage suggested by Ledoit & Wolf (2003, 2004), Jagannathan & Ma (2003) suggest weighted covariance estimators. The authors argue that using equal portfolio weights is the best method since little is known about the covariance structure of the estimation errors of various estimators. For example, taking a simple average across different covariance matrices such as the diagonal, sample, and single-index model. Thus, it may be worth applying these alternative covariance methods to obtain combining forecasts across models and estimation windows.

References

Adrian, T. & Franzoni, F. (2009), 'Learning about beta: Time-varying factor loadings, expected returns, and the conditional capm', *Journal of Empirical Finance* **16**(4), 537–556.

Akaike, H. (1973), Information theory and an extension of maximum likelihood principle, *in* 'Proceedings of Second International Symposium on Information Theory', pp. 267–281.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Aslanidis, N., Christiansen, C. & Cipollini, A. (2019), 'Predicting bond betas using macro-finance variables', *Finance Research Letters* **29**, 193–199.

Bali, T. G. & Engle, R. F. (2010), 'Resurrecting the conditional capm with dynamic conditional correlations', *Available at SSRN 1298633* .

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Beckmann, J. & Schüssler, R. (2016), 'Forecasting exchange rates under parameter and model uncertainty', *Journal of International Money and Finance* **60**, 267–288.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A capital asset pricing model with time-varying covariances', *Journal of Political Economy* **96**(1), 116–131.

Boot, T. & Nibbering, D. (2020), *Subspace methods*, Springer, pp. 267–291.

Borup, D. & Schütte, E. C. M. (2020), 'In search of a job: Forecasting employment growth using google trends', *Journal of Business and Economic Statistics, Forthcomning* .

Brock, W. A., Durlauf, S. N. & West, K. D. (2003), Policy evaluation in uncertain economic environments, Report, National Bureau of Economic Research.

Cai, Z., Ren, Y. & Yang, B. (2015), 'A semiparametric conditional capital asset pricing model', *Journal of Banking and Finance* **61**, 117–126.

Cenesizoglu, T. & Reeves, J. J. (2018), 'Capm, components of beta and the cross section of expected returns', *Journal of Empirical Finance* **49**, 223–246.

Chan, F. & Pauwels, L. L. (2018), 'Some theoretical results on forecast combinations', *International Journal of Forecasting* **34**(1), 64–74.

Chava, S., Gallmeyer, M. & Park, H. (2015), 'Credit conditions and stock return predictability', *Journal of Monetary Economics* **74**, 117–132.

Chen, T. (2017), 'Investor attention and global stock returns', *Journal of Behavioral Finance* **18**(3), 358–372.

Choueifaty, Y. & Coignard, Y. (2008), 'Toward maximum diversification', *The Journal of Portfolio Management* **35**(1), 40–51.

Clyde, M. & George, E. I. (2004), 'Model uncertainty', *Statistical Science* **19**(1), 81–94.

Coqueret, G. (2020), 'Stock-specific sentiment and return predictability', *Quantitative Finance* **20**(9), 1531–1551.

De Carvalho, R. L., Lu, X. & Moulin, P. (2012), 'Demystifying equity risk–based strategies: A simple alpha plus beta description', *The Journal of Portfolio Management* **38**(3), 56–70.

Elliott, G., Gargano, A. & Timmermann, A. (2013), 'Complete subset regressions', *Journal of Econometrics* **177**(2), 357–373.

Elton, E. J. & Gruber, M. J. (1973), 'Estimating the dependence structure of share prices–implications for portfolio selection', *The Journal of Finance* **28**(5), 1203–1232.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business and Economic Statistics* **20**(3), 339–350.

Engle, R. F., Ledoit, O. & Wolf, M. (2019), 'Large dynamic covariance matrices', *Journal of Business and Economic Statistics* **37**(2), 363–375.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. & MacBeth, J. D. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**(3), 607–636.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society* **75**(4).

Feng, G., Polson, N. G. & Xu, J. (2018), 'Deep learning in characteristics-sorted factor models', *Available at SSRN 3243683* .

Ferson, W. E. & Harvey, C. R. (1999), 'Conditioning variables and the cross section of stock returns', *The Journal of Finance* **54**(4), 1325–1360.

Figlewski, S. (1983), 'Optimal price forecasting using survey data', *The Review of Economics and Statistics* **65**(1), 13–21.

Garcia, M. G., Medeiros, M. C. & Vasconcelos, G. F. (2017), 'Real-time inflation forecasting with high-dimensional models: The case of brazil', *International Journal of Forecasting* **33**(3), 679–693.

Gargano, A. & Timmermann, A. (2014), 'Forecasting commodity price indexes using macroeconomic and financial predictors', *International Journal of Forecasting* **30**(3), 825–843.

Ghysels, E. (1998), 'On stable factor structures in the pricing of risk: do time-varying betas help or hurt?', *The Journal of Finance* **53**(2), 549–573.

González, M., Nave, J. & Rubio, G. (2012), 'The cross section of expected returns with midas betas', *Journal of Financial and Quantitative Analysis* **47**(1), 115–135.

Goyal, A. & Welch, I. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *The Review of Financial Studies* **21**(4), 1455–1508.

Gu, S., Kelly, B. & Xiu, D. (2020), 'Empirical asset pricing via machine learning', *The Review of Financial Studies* **33**(5), 2223–2273.

Hirano, K. & Wright, J. H. (2017), 'Forecasting with model uncertainty: Representations and risk reduction', *Econometrica* **85**(2), 617–643.

Hsiao, C. & Wan, S. K. (2014), 'Is there an optimal forecast combination?', *Journal of Econometrics* **178**, 294–309.

Huang, D., Jiang, F., Tu, J. & Zhou, G. (2015), 'Investor sentiment aligned: A powerful predictor of stock returns', *The Review of Financial Studies* **28**(3), 791–837.

Huang, H. & Lee, T.-H. (2010), 'To combine forecasts or to combine information?', *Econometric Reviews* **29**(5-6), 534–570.

Inoue, A. & Rossi, B. (2011), 'Identifying the sources of instabilities in macroeconomic fluctuations', *Review of Economics and Statistics* **93**(4), 1186–1204.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *The Journal of Finance* **58**(4), 1651–1683.

Jagannathan, R. & Wang, Z. (1996), 'The conditional capm and the cross-section of expected returns', *The Journal of Finance* **51**(1), 3–53.

Jiang, F., Lee, J., Martin, X. & Zhou, G. (2019), 'Manager sentiment and stock returns', *Journal of Financial Economics* **132**(1), 126–149.

Jostova, G. & Philipov, A. (2005), 'Bayesian analysis of stochastic betas', *Journal of Financial and Quantitative Analysis* **40**(4), 747–778.

Kelly, B. & Pruitt, S. (2013), 'Market expectations in the cross-section of present values', *The Journal of Finance* **68**(5), 1721–1756.

Leamer, E. E. (1978), *Specification searches: Ad hoc inference with nonexperimental data*, Vol. 53, Wiley New York.

Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit, O. & Wolf, M. (2004), 'Honey, i shrunk the sample covariance matrix', *The Journal of Portfolio Management* **30**(4), 110–119.

Ledoit, O. & Wolf, M. (2012), 'Nonlinear shrinkage estimation of large-dimensional covariance matrices', *The Annals of Statistics* **40**(2), 1024–1060.

Ledoit, O. & Wolf, M. (2020), 'Analytical nonlinear shrinkage of large-dimensional covariance matrices', *Annals of Statistics* **48**(5), 3043–3065.

Lettau, M. & Ludvigson, S. (2001), 'Consumption, aggregate wealth, and expected stock returns', *The Journal of Finance* **56**(3), 815–849.

Lewellen, J. & Nagel, S. (2006), 'The conditional capm does not explain asset-pricing anomalies', *Journal of Financial Economics* **82**(2), 289–314.

Ley, E. & Steel, M. F. (2011), *Mixtures of g-priors for Bayesian Model Averaging with economic application*, The World Bank.

Li, J. & Tsiakas, I. (2017), 'Equity premium prediction: The role of economic and statistical constraints', *Journal of Financial Markets* **36**, 56–75.

Lin, Q. (2018), 'Technical analysis and stock return predictability: An aligned approach', *Journal of Financial Markets* **38**, 103–123.

Liu, Y. & Xie, T. (2019), 'Machine learning versus econometrics: prediction of box office', *Applied Economics Letters* **26**(2), 124–130.

Loucks, D., Van Beek, E., Stedinger, J., Dijkman, J. & Villars, M. (2005), 'Model sensitivity and uncertainty analysis', *Water Resources Systems Planning and Management* **12**(1), 255–290.

Ludvigson, S. C. & Ng, S. (2007), 'The empirical risk–return relation: A factor analysis approach', *Journal of Financial Economics* **83**(1), 171–222.

Maillard, S., Roncalli, T. & Teïletche, J. (2010), 'The properties of equally weighted risk contribution portfolios', *The Journal of Portfolio Management* **36**(4), 60–70.

Makridakis, S., Hyndman, R. J. & Petropoulos, F. (2020), 'Forecasting in social settings: The state of the art', *International Journal of Forecasting* **36**(1), 15–28.

Mallows, C. (1973), 'Some comments on cp', *Technometrics* **15**(1), 661–675.

Mausser, H. & Romanko, O. (2014), 'Computing equal risk contribution portfolios', *Journal of Research and Development* **58**(4), 5: 1–5: 12.

McCracken, M. W. & Ng, S. (2016), 'Fred-md: A monthly database for macroeconomic research', *Journal of Business and Economic Statistics* **34**(4), 574–589.

Meligkotsidou, L., Panopoulou, E., Vrontos, I. D. & Vrontos, S. D. (2021), 'Out-of-sample equity premium prediction: A complete subset quantile regression approach', *The European Journal of Finance* **27**(1-2), 110–135.

Moral-Benito, E. (2015), 'Model averaging in economics: An overview', *Journal of Economic Surveys* **29**(1), 46–75.

Morgan, J. (1996), Reuters (1996) riskmetrics-technical document, Report.

Moritz, B. & Zimmermann, T. (2016), 'Tree-based conditional portfolio sorts: The relation between past and future stock returns', *Available at SSRN 2740751* .

Narayan, P. K., Narayan, S., Phan, D. H. B., Thuraisamy, K. S. & Tran, V. T. (2017), 'Credit quality implied momentum profits for islamic stocks', *Pacific-Basin Finance Journal* **42**, 11–23.

Neely, C. J., Rapach, D. E., Tu, J. & Zhou, G. (2014), 'Forecasting the equity risk premium: the role of technical indicators', *Management Science* **60**(7), 1772–1791.

Neffelli, M. (2018), 'Target matrix estimators in risk-based portfolios', *Risks* **6**(4), 125.

Ni, Z.-X., Wang, D.-Z. & Xue, W.-J. (2015), 'Investor sentiment and its nonlinear effect on stock returns—new evidence from the chinese stock market based on panel quantile regression model', *Economic Modelling* **50**, 266–274.

Pantaleo, E., Tumminello, M., Lillo, F. & Mantegna, R. N. (2011), 'When do improved covariance matrix estimators enhance portfolio optimization? an empirical comparative study of nine estimators', *Quantitative Finance* **11**(7), 1067–1080.

Pesaran, M. H. & Pick, A. (2011), 'Forecast combination across estimation windows', *Journal of Business and Economic Statistics* **29**(2), 307–318.

Pesaran, M. H., Pick, A. & Pranovich, M. (2013), 'Optimal forecasts in the presence of structural breaks', *Journal of Econometrics* **177**(2), 134–152.

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Petropoulos, F., Hyndman, R. J. & Bergmeir, C. (2018), 'Exploring the sources of uncertainty: Why does bagging for time series forecasting work?', *European Journal of Operational Research* **268**(2), 545–554.

Rapach, D. E., Ringgenberg, M. C. & Zhou, G. (2016), 'Short interest and aggregate stock returns', *Journal of Financial Economics* **121**(1), 46–65.

Rapach, D. E., Strauss, J. K. & 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.

Rapach, D. E. & Zhou, G. (2020), 'Time-series and cross-sectional stock return forecasting: New machine learning methods', *Machine Learning for Asset Management: New Developments and Financial Applications* **1**(1), 1–33.

Rossi, B. & Inoue, A. (2012), 'Out-of-sample forecast tests robust to the choice of window size', *Journal of Business and Economic Statistics* **30**(3), 432–453.

Salisu, A. A., Isah, K. & Akanni, L. O. (2019), 'Improving the predictability of stock returns with bitcoin prices', *The North American Journal of Economics and Finance* **48**, 857–867.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

Smith, G. (2018), 'Step away from stepwise', *Journal of Big Data* **5**(1), 1–12.

Tian, J. & Anderson, H. M. (2014), 'Forecast combinations under structural break uncertainty', *International Journal of Forecasting* **30**(1), 161–175.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58**(1), 267–288.

Timmermann, A. (2006), 'Forecast combinations', *Handbook of Economic Forecasting* **1**, 135–196.

Vendrame, V., Guermat, C. & Tucker, J. (2018), 'A conditional regime switching capm', *International Review of Financial Analysis* **56**, 1–11.

Wang, K. Q. (2002), 'Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio', *Journal of Empirical Finance* **9**(2), 133–169.

Wang, K. Q. (2003), 'Asset pricing with conditioning information: A new test', *The Journal of Finance* **58**(1), 161–196.

Wang, Y., Hao, X. & Wu, C. (2020), 'Forecasting stock returns: A time-dependent weighted least squares approach', *Journal of Financial Markets* **53**, 100568.

Young, C. (2018), 'Model uncertainty and the crisis in science', *Socius: Sociological Research for a Dynamic World* **4**, 1–7.

Zhang, H., He, Q., Jacobsen, B. & Jiang, F. (2020), 'Forecasting stock returns with model uncertainty and parameter instability', *Journal of Applied Econometrics* **35**(5), 629–644.

Zhang, Y., Chu, G. & Shen, D. (2021), 'The role of investor attention in predicting stock prices: The long short-term memory networks perspective', *Finance Research Letters* **38**, 101484.

Ziyadi, M. & Al-Qadi, I. L. (2019), 'Model uncertainty analysis using data analytics for life-cycle assessment (lca) applications', *The International Journal of Life Cycle Assessment* **24**(5), 945–959.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society* **67**(2), 301–320.