

Three Essays

on Bayesian Shrinkage Methods

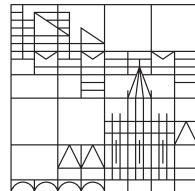
Dissertation

zur Erlangung des akademischen Grades eines Doktors der
Wirtschaftswissenschaften (Dr.rer.pol.)

vorgelegt von Frey, Christoph

an der

Universität
Konstanz



Sektion Politik - Recht - Wirtschaft

Fachbereich Wirtschaftswissenschaften

Tag der mündlichen Prüfung: 12. Januar 2017

1. Referent: Professor Dr. Winfried Pohlmeier
2. Referent: Professor Gary Koop, Ph.D.

*So eine Arbeit wird eigentlich nie fertig, man muss sie für fertig erklären,
wenn man nach Zeit und Umständen das möglichste getan hat.*

Johann Wolfgang von Goethe, Caserta, den 16. März 1787.

Danksagung

Mein Dank gilt an dieser Stelle einer Reihe von Personen, die einen großen Anteil an der Entstehung dieser Arbeit haben.

Besonders bedanke ich mich bei meinem Doktorvater, Herrn Professor Dr. Winfried Pohlmeier. Sein Vertrauen und seine Unterstützung seit meiner ersten Anstellung am Lehrstuhl als studentische Hilfskraft haben mir stets geholfen, Antworten auf meine akademischen Fragen zu finden. Ohne seine Initiative hätte ich das Projekt Promotion nicht begonnen.

Ich bedanke mich ebenfalls bei Herrn Professor Dr. Gary Koop für seine Bereitschaft, diese Arbeit zu begutachten und bei Herrn Professor Dr. Ralf Brüggemann, der als unterstützender Begleiter seit meiner ersten Statistikvorlesung immer mit Zeit und Interesse auf meine Fragen reagiert hat.

Die Arbeit am Lehrstuhl für Ökonometrie in Konstanz hat mir auch wegen meiner Kollegen viel Freude bereitet. Ich erinnere mich an eine Vielzahl von intensiven und spannenden Diskussion und möchte mich ganz besonders bei folgenden Personen bedanken: Sebastian Bayer, Roxana Halbleib, Liu Hao, Phillip Heiler, Fabian Krüger, Frieder Mokinski, Peter Schanbacher, Ruben Seiberlich, Anna Slavutskaya, Audra Virbickaité, Stefan Voigt. Ein ganz besonderer Dank geht an Jana Marečková, die als Bürokollegin und Freundin die vielen Stunden der Kleinstarbeit hautnah miterleben durfte und mir immer den nötigen Anstoß zum Weiterdenken vermittelt hat. Erwähnen möchte ich auch Verena Kretz, die als Sekretärin viele administrative Abläufe auch für mich geregelt hat und auf deren Hilfe ich mich in jedem Moment verlassen konnte.

Einen ganz wichtigen Dank richte ich auch an meine Mutter und meine Großeltern, Gertrud und Kurt Frey. Ohne Ihre Liebe und Nachsicht hätte ich diese Arbeit nicht schreiben können.

Contents

Zusammenfassung	2
Summary	6
1 Forecasting with Bayesian Vector Autoregressions estimated using Professional Forecasts	9
1.1 Introduction	10
1.2 VAR estimation using professional nowcasts	13
1.2.1 Augmenting a VAR with survey nowcasts	13
1.2.2 How to address that survey nowcasts may not be “correctly specified”?	14
1.2.3 Adding Wright’s democratic steady-state prior	16
1.3 Empirical application	17
1.3.1 Trained hyperparameters	19
1.3.2 Survey forecasts and forecast combination	20
1.4 Concluding Remarks	21
References	22
1.5 Appendix	25
1.5.1 Posterior distribution for the augmented VAR	25
1.5.2 Posterior distribution for augmented VAR with a democratic steady-state prior	26
1.5.3 Figures & Tables	28
2 Bayesian Shrinkage of Portfolio Weights	34
2.1 Introduction	35
2.2 Penalized regressions and the GMVP	37
2.2.1 The classical global minimum variance portfolio problem	37
2.2.2 Reference augmented GMVP regression	39
2.2.3 Regularization strategies for the reference augmented GMVP regression	40
2.2.4 Choosing the Prior	41
2.3 Data and horse race set-up	48

2.3.1	Data and investment set-up	48
2.3.2	Evaluation criteria	49
2.3.3	Competing frequentist strategies	50
2.4	Empirical findings	52
2.4.1	Comparison between Bayesian shrinkage methods	54
2.4.2	Robustness Checks	55
2.5	Concluding remarks	55
	References	57
2.6	Appendix	61
3	Using Analysts' Forecasts for Stock Predictions - An Entropic Tilting Approach	72
3.1	Introduction	73
3.2	Predictive powers	76
3.2.1	Analysts' forecasts	76
3.2.2	Predictive regressions	77
3.3	Methodology	78
3.3.1	Prediction model	78
3.3.2	Entropic tilting	83
3.4	Empirical Application	86
3.4.1	Data and set-up	86
3.4.2	Competing models	87
3.4.3	Evaluation criteria	88
3.4.4	Individual predictor performance	89
3.4.5	Complete model performance	90
3.5	Concluding Remarks	90
	References	92
3.6	Appendix	96
3.6.1	Estimation of the TVP-BVAR with SV using forgetting factors . . .	96
3.6.2	Figures	97
3.6.3	Tables	100
	Complete Bibliography	112
	Abgrenzung	120

List of Figures

1.1	Variance Ratio $VR(r, \rho)$	28
3.1	Spot price, 12 months forward target price and percentage of buy recommendations of the IBM stock (monthly data) between 1999 and 2015	97
3.2	Out-of-sample forecast performance results for different univariate models for the IBM stock for 2004 to 2015	98
3.3	Kernel estimates of predictive density of the IBM returns from the TVP-VAR(1) with dynamic model averaging and tilting towards the mean of monthly target price implied expected returns	99
3.4	Kernel estimates of predictive density of the IBM returns from the TVP-VAR(1) with dynamic model averaging and tilting towards the mean and variance of monthly target price implied expected returns	99

List of Tables

1.1	Data description and variable transformation	29
1.2	Prior Specifications	30
1.3	Forecasts with Different Prior Specifications: Relative Mean Squared Forecast Errors (Evaluation Sample: 1984:Q2 - 2011:Q2)	31
1.4	Forecasting with Trained Hyperparameters: Relative Mean Squared Forecast Errors (Evaluation Sample: 1990:Q4 - 2011:Q2)	32
1.5	Comparing the Forecasts with both Survey Expectations and Combined Forecasts: Relative Mean Squared Forecast Errors (Evaluation Sample: 1990:Q4 - 2011:Q2)	33
2.1	Out-of-sample results for $N = 5$ Fama/French industry portfolios with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)	61
2.2	Out-of-sample results for $N = 6$ Fama/French portfolios formed on size and book-to-market with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)	62
2.3	Out-of-sample results for $N = 25$ Fama/French portfolios formed on size and book-to-market with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)	63
2.4	Out-of-sample results for $N = 30$ Fama/French industry portfolios with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)	64
2.5	Out-of-sample results for $N = 100$ Fama/French portfolios formed on size and book-to-market with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)	65
2.6	Out-of-sample results for $N = 5$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)	66
2.7	Out-of-sample results for $N = 25$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)	67
2.8	Out-of-sample results for $N = 50$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)	68

2.9	Out-of-sample results for $N = 100$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)	69
2.10	Out-of-sample results for $N = 250$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)	70
2.11	Out-of-sample results for $N = 500$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)	71
3.1	Relative root mean squared errors between forecasted and observed spot prices for 20 Dow Jones constituents (sample: 1999 - 2015)	100
3.2	Descriptive statistics on the returns, target prices and recommendations for 20 Dow Jones constituents (sample: 1999 - 2015)	101
3.3	Out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a Bayesian VAR(1)	102
3.4	Out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility	103
3.5	Out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean of monthly target price implied expected returns	104
3.6	Out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean and variance of monthly target price implied expected returns	105
3.7	Average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a Bayesian VAR(1)	106
3.8	Average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility	107
3.9	Average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean of monthly target price implied expected returns	108
3.10	Average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean and variance of monthly target price implied expected returns	109
3.11	Out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) for various forecasting models	110
3.12	Average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) for various forecasting models	111

Zusammenfassung

... im Hinblick auf die Prognosefähigkeit, ... wird ein guter Bayesianer einen Nicht-Bayesianer schlagen, der wiederum besser sein sollte als ein schlechter Bayesianer.

(Granger, 1986, p. 16)

Die vorliegende Dissertation befasst sich mit Bayesianischen Shrinkage- und Regularisierungsmethoden, die das Ziel haben, die Auswirkungen von Schätzfehlern bei ökonomischen Prognosen und Portfolioanalysen zu minimieren. Prognosen werden im Allgemeinen auf der Grundlage verfügbarer Informationen und Annahmen (Daten, Modelle usw.) erstellt, die eine *a priori* Vermutung über den zugrundeliegenden datenerzeugenden Prozess ausdrücken. Das besondere Merkmal des Bayesianischen Ansatzes ist es, diese Vermutungen zu formalisieren und bedingte Wahrscheinlichkeitsaussagen zu den unbekannten Parametern zu machen. Konditional auf die beobachteten Daten liefert dieser Ansatz eine so genannte *posteriore* Verteilung für die unbekannten Parameter des Modells und eine so genannte *prädiktive* Verteilung für die zukünftigen Ausprägungen der Variablen von Interesse. Im Gegensatz zu Mittelwertsprognosen machen diese prädiktiven Verteilungen eine Aussage über die Unsicherheit in zukünftigen Ausprägungen. Beispielsweise kann ein Portfoliomanager damit die Unsicherheit in geschätzten Portfoliogewichten quantifizieren und Transaktionskosten minimieren, indem er sehr unsichere Positionen vermeidet. Des Weiteren könnte eine Zentralbank an den Konfidenzintervallen bei der Prognose der Inflation interessiert sein, um eine optimale Geldpolitik für zukünftiges Wirtschaftswachstum zu bestimmen.

Das Prinzip von Bayesianschen Shrinkage-Methoden besteht darin, durch die Verwendung von *a priori* und eventuell subjektiver Information Modellparameter nahe bei Null zu beschränken, um ein parameterarmes posteriores Modell zu erhalten. Modelle wie Vektorautoregressionen (VARs) haben eine große Flexibilität bei der Modellierung komplexer Beziehungen zwischen Variablen, hängen in der Regel aber von einer großen Anzahl zu schätzender Parametern ab. Ein übermäßiges Anpassen der Modellparameter auf die Daten ist daher eine unmittelbare Gefahr für die Prognosegenauigkeit. Eine künstliche

Verringerung der Anzahl der zu schätzenden Parameter durch Vorinformationen verzerrt die übrigen Parameterschätzungen, verbessert aber im Allgemeinen die Out-of-Sample Prognosefähigkeit (Banbura et al., 2010).

Alle drei Kapitel dieser Dissertation befassen sich mit der optimalen Kombination von Vorinformationen, zum Beispiel auf Grundlage externer Quellen oder theoretischer Überlegungen, mit Zeitreihenbeobachtungen der interessierenden Variablen, um diese vorherzusagen. Es ist nicht das Anliegen dieser Arbeit, konsistente und asymptotisch effiziente Modellschätzer zu finden, sondern zu erkunden, wie die erhöhte Prognosegenauigkeit der Shrinkage-Methoden zustande kommt. Während Kapitel eins und drei multivariate Vorhersagemodelle verwenden, um Zeitreihenbeobachtungen mit Umfragedaten zu kombinieren, beschäftigt sich Kapitel zwei mit Shrinkage-Methoden für Portfoliogewichte, die eine Vorvermutung über die optimale Portfoliozusammensetzung berücksichtigen. Alle drei Kapitel sind eigenständige Forschungsarbeiten, die ich während meines Promotionsstudiums an der Universität Konstanz verfasst habe. Die ersten beiden Papiere sind gemeinsam mit Koautoren entstanden, das dritte Papier habe ich in Gänze selbst erstellt. Aus Konsistenzgründen bezeichnen alle drei Papiere den bzw. die Verfasser mit „wir“.

Das erste Kapitel meiner Dissertation mit dem Titel “Forecasting with Bayesian Vector Autoregressions estimated using Professional Forecasts” ist eine gemeinsame Arbeit mit Frieder Mokinski. Wir schlagen einen Bayesianischen Shrinkage-Ansatz für Vektorautoregressionen vor, der kurzfristige Umfrageprognosen als zusätzliche Informationsquelle für die Modellparameter verwendet. Insbesondere erweitern wir den Vektor der abhängigen Variablen mit dazugehörigen Umfragebeobachtungen und behaupten, dass jede Modellvariable und ihre Umfrage in ähnlicher Weise mit den vorherigen Beobachtungen der abhängigen Variablen korreliert sind. Die Shrinkage-Idee ergibt sich sowohl aus dem Prior als auch aus den Daten: Wir schrumpfen eine Gruppe unbekannter Parameter (Regressionskoeffizienten zwischen Umfragen und vorherigen Beobachtungen der abhängigen Variablen) zu einer zweiten Gruppe unbekannter Parameter (Regressionskoeffizienten des VAR). Somit beruht das Verfahren auf dem „Erlernen“ der unbekannten Parameter der ursprünglichen Vektorautoregression aus den Umfragebeobachtungen. In einer Anwendung mit makroökonomischen Daten zeigen wir, dass ein mit Umfragedaten erweitertes VAR typischerweise kleinere mittlere quadratische Prognosefehler erzeugt als eine Reihe von Benchmarkmethoden.

Das zweite Kapitel ”Bayesian Shrinkage of Portfolio Weights“ ist gemeinsam mit meinem Doktorvater Prof. Dr. Winfried Pohlmeier entstanden. Wir schlagen hier eine alternative Strategie der Portfolioregularisierung durch eine Bayesianische Regression für das Globale Minimum Varianz Portfolio (GMVP) vor. Insbesondere repräsentieren wir die Abweichungen in den GMVP-Gewichten von einem gegebenen Referenzportfolio (z. B.

das naive $1/N$ -Portfolio) als Koeffizienten einer linearen Regression und schrumpfen sie gegen Null. Die direkte Modellierung der optimalen Portfoliogewichte durch Bayesianische Methoden vermeidet die Schätzung der Momente der Verteilung der Aktienrenditen und reduziert die Dimension des Schätzproblems deutlich. Sie ist darüber hinaus wesentlich intuitiver und ökonomisch einfacher zu interpretieren. Beispielsweise könnte der Prior auf den Portfoliogewichten dazu verwendet werden, um die Vermutung des Anlegers über das optimale Portfolio aus einer früheren Strategie wiederzugeben und Transaktionskosten zu minimieren. Oder sie können dazu dienen, Ungleichheitsbeschränkungen für die Portfoliogewichte einzuhalten, um beispielsweise keine Leerverkäufe zuzulassen. Die Verwendung des Bayesianischen Ansatzes für die Portfoliogewichte erlaubt ferner, unterschiedliche Schätzungsrisiken in unterschiedlichen Aktien durch individuelle posteriori Varianzen zu berücksichtigen. Zudem ermöglicht sie die Schätzung von Portfolios in hochdimensionalen Situationen, in denen die Anzahl der Vermögenswerte im Vergleich zur Stichprobengröße sehr groß ist. In diesem für Praktiker besonders relevanten Fall erweisen sich frequentistische Standartansätze oft als nicht durchführbar oder sie liefern unbrauchbare Ergebnisse. Wir vergleichen die vorgeschlagenen Bayesianischen Shrinkage-Strategien mit populären Ansätzen aus der Literatur und zeigen, dass unsere Schätzer zu besseren Out-of-Sample Portfolioergebnissen führen.

Schließlich verknüpfen wir im dritten Kapitel mit dem Titel „Using Analysts' Forecasts for Stock Predictions - An Entropic Tilting Approach“ die prognostizierende Verteilung für Aktienrenditen aus Bayesianischen Vektorautoregressionen mit Vorhersagen von Finanzanalysten durch eine so genannte entropische (exponentielle) Neugewichtung. Die Idee der Methode ist es, die prognostizierende Verteilung der Aktienrenditen so zu gewichten, dass eine bestimmte Momentenbedingung erfüllt wird, die hier auf der Basis durchschnittlicher Analystenprognosen erstellt wird. Wir beschränken insbesondere den Mittelwert und die Varianz der prognostizierenden Verteilung der Aktienrenditen auf den Mittelwert und die Varianz der durch den Zielpreis der Analysten implizierten erwarteten monatlichen Renditen, d.h. der einfachen Renditen zwischen dem aktuellen Aktienpreis und dem Zielpreis. Der Vorteil dieses Ansatzes ist es, dass wir modellbasierte Zeitreiheninformationen mit anderen Informationen in einer einfachen Weise unter Verwendung von analytischen Lösungen kombinieren können. Während diese Methode den Mittelwert und die Form der prognostizierenden Verteilung ändert, erzeugt sie keine besseren Parameterschätzungen für das zugrundeliegende Vorhersagemodell. Für die monatlichen Renditen der Dow-Jones-Aktien stellen wir fest, dass die Einschränkung der Varianz der Renditenverteilung besonders vorteilhaft für die Out-of-Sample Vorhersageleistung ist, da diese Varianz ein zukunftsweisendes Maß für die (Un-)Sicherheit auf dem Markt ist. Der empirische Beitrag dieses Kapitels zur Literatur ist dreifach: Erstens zeigen wir,

dass der Dissens/Konsens zwischen Finanzanalysten, denen häufig Interessenkonflikte und Absprachen vorgeworfen werden (Ramnath et al., 2008), Vorhersagekraft für die Aktienrenditen hat. Zweitens kombinieren wir das entropische Gewichten mit Prognosen eines flexiblen Bayesianischen Autoregressionssystems, das viele der wichtigen Modellaspekte bei der Renditenvorhersage berücksichtigt: zeitvariierende Parameter, stochastische Volatilität, Parameter-Shrinkage sowie eine dynamische Modellmittelung und Selektion. Drittens ist dies eine eher ungewöhnliche Studie, die, anstatt aggregierte Marktrenditen zu prognostizieren, die Vorhersagbarkeit von Renditen einzelner Aktien untersucht.

Summary

... in terms of forecasting ability, ... a good Bayesian will beat a non-Bayesian, who will do better than a bad Bayesian.
(Granger, 1986, p. 16)

This dissertation is concerned with Bayesian shrinkage methods in the context of forecasting and portfolio analysis to remedy the impact of estimation errors. Economic forecasts are generally made conditional on the available information (data, models etc.) that express an *a priori* belief about the underlying data generating process. The salient feature of the Bayesian approach is to formalize such prior beliefs and to make conditional probabilistic statements about the unknown elements. Conditional on the observed data, that is it provides a so-called *posterior* distribution for the unknown parameters of the model and a *predictive* distribution for the future outcome of the variable of interest. In contrast to single point forecasts, predictive distributions provide information about the overall uncertainty in future outcomes. For example, a portfolio manager may want to quantify the uncertainty in estimated portfolio weights in order to minimize transaction costs by avoiding positions with too much risk. Also, a central bank may be interested in confidence bands when forecasting inflation to decide the optimal monetary policy to foster economic growth in the future.

Bayesian shrinkage describes the use of informative (possibly subjective) priors that shrink parameter estimates to zero in order to obtain a parsimonious posterior model. Models that allow for a great flexibility in modeling complex dynamic relations, such as Vector Autoregressions (VARs), usually also have a great number of parameters to be estimated. Potential over-fitting the model to the data is, therefore, an immediate threat to forecast accuracy. Artificial reducing parameter uncertainty through prior information comes for the cost of a bias but generally translates into superior out-of-sample forecast performance (Banbura et al., 2010).

All three chapters of this dissertation deal with the optimal combination of prior information, coming from external sources or from economic theory, with time-series observation of the variables of interest for the sake of forecasting. It is not concerned with

finding consistent and asymptotically efficient model estimates, but explores where the gains in forecast accuracy from using shrinkage methods come from. While Chapter one and three use a multivariate setting to combine time-series observations with survey data, Chapter two deals with shrinkage priors for portfolio weights, which allow incorporating prior beliefs about the optimal asset allocations. All three chapters are standalone research papers which I have written during my Ph.D. studies at the University of Konstanz. The first two papers are written with coauthors and the last one is completely my own work. For consistency reasons, the text refers to “we” throughout the thesis.

The first chapter of my dissertation with the title “Forecasting with Bayesian Vector Autoregressions estimated using Professional Forecasts” is a joint work with Frieder Mokinski. We propose a Bayesian shrinkage approach for vector autoregressions (VAR) that uses short-term survey forecasts as an additional source of information about model parameters. In particular, we augment the vector of dependent variables by their survey nowcasts and claim that each variable modeled in the VAR and its nowcast are likely to depend on the lagged dependent variables in a similar way. Here, the idea of shrinkage comes both from the prior and from the data: We shrink one set of unknown parameters (regression of survey nowcasts on lagged dependent variables) towards a second set of unknown parameters (regression coefficients of the VAR). Thus, the method relies on ‘learning’ about the parameters of the original vector autoregression from survey nowcasts. In an application to macroeconomic data, we find that the forecasts obtained from a VAR fitted by our new shrinkage approach typically yield smaller mean squared forecast errors than the forecasts obtained from a range of benchmark methods.

The second chapter “Bayesian Shrinkage of Portfolio Weights” is coauthored with my supervisor Professor Dr. Winfried Pohlmeier. Here, we propose an alternative strategy of portfolio weight shrinkage by means of a Bayesian regression for the Global Minimum Variance Portfolio (GMVP). Specifically, we represent the weight deviations of the GMVP from a given reference portfolio (e.g. the naive $1/N$ portfolio) as coefficients of a linear regression and shrink them towards zero. Modeling the optimal portfolio weights through Bayesian priors avoids estimating the moments of the asset return distribution and substantially reduces the dimensionality of the estimation problem. It is also much more intuitive and easy to interpret in economic terms. For example, the prior can be used to reflect the investors’ views about the optimal portfolio from a previous allocation to minimize transactions costs or it can be used to incorporate inequality restrictions on the portfolio weights such as no short-selling. The use of Bayesian shrinkage priors further (i) allows accounting for different degrees of estimation risk across assets by assigning different posterior variances to each portfolio weight and (ii) it allows estimating portfolios in high dimensional settings when the number of assets is very large relative to the sample size. In

this for practitioners particularly relevant case, standard frequentist estimation approaches turn out to be infeasible or perform poorly. We compare the proposed Bayesian shrinkage strategies to popular approaches from the literature and find that the former show better out-of-sample performance based on various performance criteria.

Eventually in the third Chapter entitled “Using Analysts’ Forecasts for Stock Predictions - An Entropic Tilting Approach”, we combine predictive density forecasts for US stock returns from Bayesian vector autoregressions with financial analysts’ forecasts via *entropic (exponential) tilting*. The idea of the method is to reweight the predictive distribution to match moment conditions that are formed based on average analysts’ forecasts. In particular, we restrict the mean and variance of the predictive distribution of the asset returns to coincide with the mean and the variance of monthly target price implied expected returns, i.e. simple returns between the spot and the target price. The advantage of this approach is to combine model-based time-series information with information from other origins in a parsimonious way and by using closed-form solutions. For monthly returns of Dow Jones constituents, we find that restricting the variance of the asset returns is particularly beneficial in terms of out-of-sample performance as it provides a forward-looking measure of market (un-)certainty. The empirical contribution of this Chapter to the literature is threefold: First, we show that the (dis-)agreement among financial analysts, which have been accused to have skewed incentives due to conflicts of interest (Ramnath et al., 2008), has predictive power for asset returns. Second, we combine entropic tilting with forecasts from a large Bayesian vector autoregressive system allowing for all features recently found to be important for return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and selection. Third, we provide an uncommon study considering individual asset returns instead of aggregated market returns.

References

- BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): “Large Bayesian Vector Auto Regressions,” *Journal of Applied Econometrics*, 25, 71–92.
- GRANGER, C. W. J. (1986): “[Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts]: Comment,” *Journal of Business & Economic Statistics*, 4, 16–17.
- RAMNATH, S., S. ROCK, AND P. SHANE (2008): “The financial analyst forecasting literature: A taxonomy with suggestions for further research,” *International Journal of Forecasting*, 24, 34–75.

CHAPTER 1

Forecasting with Bayesian Vector
Autoregressions estimated using Professional
Forecasts

1.1 Introduction

Vector Autoregressions (VARs) are among the most popular tools in economic forecasting. VARs offer great flexibility in modelling the complex dynamic relations among macroeconomic variables, they are easy to estimate and can be used to generate forecasts at multiple horizons (see e.g. [Stock and Watson, 2001](#)). However, as even medium-sized VARs (10-20 variables) have several hundred parameters to estimate, potential over-fitting is an immediate threat to forecast accuracy. The literature has therefore either used VARs with only a handful of variables ([Chauvet and Potter, 2013](#); [Faust and Wright, 2013](#)), or it has resorted to Bayesian shrinkage methods ([Banbura et al., 2010](#)). Such methods include [Doan et al. \(1984\)](#)'s Minnesota prior, which assumes that each variable evolves according to a random walk, and [Wright \(2013\)](#)'s democratic steady-state prior, which uses long-run forecasts from an expert survey as prior information for the vector of unconditional means.

We build on [Wright \(2013\)](#)'s work and consider a *Bayesian shrinkage approach that additionally exploits the non-sample information in survey nowcasts*, i.e. forecasts for the current quarter or month. The idea of our approach is that the variables modeled in the VAR and their corresponding survey nowcasts are likely to depend in a similar way on the lagged dependent variables. To exploit this conjecture, we first augment the vector of dependent variables of the VAR with survey nowcasts and then express our belief of similar dependence on the lagged dependent variables through a Bayesian prior. The idea is best illustrated with a simple example: Consider a variable y_t , modeled as a univariate autoregression (AR) with a single lag, i.e. $y_t = ay_{t-1} + \varepsilon_t$, and its nowcast s_t . The augmented model is

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} y_{t-1} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix},$$

and the prior distribution favoring pairwise identical coefficients can be stated as

$$p \begin{bmatrix} a \\ (b-a) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \underline{a} \\ 0 \end{bmatrix}, \begin{bmatrix} \underline{v}_a & 0 \\ 0 & \underline{v}_\Delta \end{bmatrix} \right).$$

This prior implies $E[b] = \underline{a}$, i.e. we expect that y_t and s_t depend on y_{t-1} in the same way. Through \underline{v}_Δ we express our confidence in this conjecture. If the dependence of the survey nowcasts (s_t) on the lagged dependent variables is indeed not dissimilar from the actuals (y_t), i.e. if Δ is small, then the extra information provided through the survey nowcasts will help us pin down the parameters of the original VAR. Put differently, the shrinkage method is likely to reduce the risk of over-fitting the model to the data and we therefore expect it to provide us with more accurate forecasts.¹ Note that our

¹Taking a frequentist perspective, [Ing and Wei \(2003, Theorem 3\)](#) show that better coefficient estimates

interpretation of shrinkage differs somewhat from the above mentioned approaches: The Minnesota prior, for example, shrinks the coefficients of a vector autoregression towards a system of univariate random walks. Thus, shrinkage is directly provided through the prior. In our case, instead, shrinkage comes both from the prior and from the data: We shrink one set of unknown parameters (regression of survey nowcasts on lagged dependent variables) towards a second set of unknown parameters (regression coefficients of the VAR). Thus, the method relies on 'learning' about the parameters of the original vector autoregression from survey nowcasts.

In a forecasting application with U.S. macroeconomic and macro-financial data, we find that a ten-variable VAR estimated using our novel shrinkage approach produces forecasts that are superior to a range of benchmark methods. Specifically, we find that mean squared forecast errors (MSFEs) are typically lower with our method than with a univariate AR(1) estimated by OLS, uniformly lower than with the same VAR estimated using only the Minnesota prior, and comparable to those of survey forecasts.

The idea of similar dependence on the lagged dependent variables can be motivated in several ways: First, empirically, survey nowcasts have often been found to be very accurate predictions of the target variable (e.g. [Faust and Wright, 2013](#)). We would therefore expect that they exploit the available information in a way that resembles the true data generating process. Second, Online Appendix B.1 shows that the shrinkage target $\Delta = 0$ can alternatively be motivated from assumptions about the expectations formation process and about the time series model specification. Specifically, if (i) average expectations are formed in a fully rational manner based on an information set that includes the lagged dependent variables of the VAR, and (ii) the VAR is correctly specified, then the true value of Δ is zero. The fact that these ideal conditions are not likely to be fully satisfied in practice is one motivation to use $\Delta = 0$ as a shrinkage target instead of imposing it deterministically.

Similar approaches to incorporate information from survey nowcasts have been used in the frequentist estimation of a three-factor affine Gaussian model for U.S. Treasury yields by [Kim and Orphanides \(2012\)](#), and in the Bayesian estimation of a DSGE model by [Del Negro and Schorfheide \(2013\)](#). However, besides the different model class, a major difference is that these studies have assumed that coefficients are *exactly equal* for each pair of actuals and nowcasts. By avoiding to impose equal coefficients deterministically, our Bayesian shrinkage method reduces the risk of deteriorating forecasts by imposing restrictions that may turn out to be severely erroneous.

Recently, a number of studies have used exponential tilting ([Robertson et al., 2005](#))
(in terms of mean squared error (MSE)) asymptotically translate into superior forecasts (in terms of MSFE).

to incorporate moment restrictions - for example from survey forecasts - into predictive densities obtained from macroeconomic time series models. Exponential tilting proceeds in the following way: From the universe of densities fulfilling the moment restrictions, it chooses the one closest in terms of relative entropy to the predictive density obtained from the time series model. Using this method, Cogley et al. (2005) have considered forecasting UK inflation with moment restrictions for the mean and variance taken from fan charts of the Bank of England. Alternatively, Altavilla et al. (2014) have used survey point forecasts of short-term interest rates to adjust the forecasts of a Dynamic Nelson-Siegel model of the U.S. yield curve. Lately, Krüger et al. (2015) have employed moment restrictions which represent the mean and variance of survey nowcasts in order to modify the forecast density of a Bayesian VAR. Incorporating survey-based information through exponential tilting differs in a number of ways from our approach: First, it only exploits the survey data after model estimation. Thus, although such information is deemed informative, it is not used to learn about the data generating process but only to modify the forecast density. Second, the method makes no attempt to evaluate empirically whether the moment restrictions (obtained from survey forecasts) it imposes are likely to hold in the data. Exponential tilting therefore relies heavily on carefully selecting the 'right' moment restrictions. Our method instead lets the data decide how informative survey forecasts are about the data generating process. Eventually, exponential tilting is forecast-horizon specific, i.e. it can only be used to adjust forecasts at horizons for which moment restrictions are available. By contrast, in our method, survey nowcasts are used to shrink coefficients of a time series model that can provide forecasts at any horizon.

Baştürk et al. (2014) present another approach of incorporating survey data into a forecasting model. Specifically, they estimate a new Keynesian Phillips Curve model, using inflation expectations to facilitate estimation of the expectations mechanism. A major difference to our approach is that they effectively include survey forecasts as a regressor, whereas we model survey nowcasts as a by-product of the data generating process. Additionally, while their method is tailor-made for inflation forecasting, ours can in principle be applied to any macroeconomic variable.

The paper proceeds as follows. Chapter 1.2 introduces the methodology and the underlying econometric ideas. Chapter 3.4 presents our empirical findings and chapter 3.5 summarizes our results.

1.2 VAR estimation using professional nowcasts

1.2.1 Augmenting a VAR with survey nowcasts

Our point of departure is a standard M -variate VAR model with p lags

$$y_t = a_0 + \sum_{i=1}^p A_i y_{t-i} + \varepsilon_t, \quad (1.1)$$

where y_t is the $M \times 1$ vector of dependent variables, a_0 is an $M \times 1$ vector of intercepts, A_i is an $M \times M$ matrix of slope coefficients, and ε_t is an $M \times 1$ vector of disturbances. We augment the VAR with:

$$s_t = b_0 + \sum_{i=1}^p B_i y_{t-i} + \eta_t, \quad (1.2)$$

where s_t collects the survey nowcasts of the variables in y_t , η_t is another $M \times 1$ vector of disturbances, and $\{b_0, B_1, \dots, B_p\}$ are used in the same way as in equation (1.1). The augmented VAR reads

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{i=1}^p \begin{bmatrix} A_i \\ B_i \end{bmatrix} y_{t-i} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}. \quad (1.3)$$

Equation (1.3) states that the survey nowcasts s_t for the elements of y_t depend on the same variables $\{y_{t-1}, \dots, y_{t-p}\}$ as y_t itself, though they can have different coefficients. Estimating the augmented system (1.3) without imposing further restrictions on $\{b_0, B_1, \dots, B_p\}$, we will hardly reduce the risk of over-fitting $\{a_0, A_1, \dots, A_p\}$ to the data. By contrast, if we impose $\{b_0 = a_0, B_1 = A_1, \dots, B_p = A_p\}$, provided that the restrictions are not *too incorrect*, this may help us to pin down the parameters of the VAR. To see that, it is convenient to take a frequentist perspective for a moment. To keep things simple, we consider the same AR(1) as in the introduction and impose equal coefficients:

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a \\ a \end{bmatrix} y_{t-1} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \quad |a| < 1, \quad \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon,\eta} \\ \sigma_{\varepsilon,\eta} & \sigma_\eta^2 \end{bmatrix} \right). \quad (1.4)$$

By standard Maximum Likelihood theory, the asymptotic distribution of the parameter estimate from the augmented model is

$$\sqrt{T}(\hat{a}_{\text{aug}} - a) \xrightarrow{d} \mathcal{N} \left(0, (1 - a^2) \frac{(\sigma_\eta^2 \sigma_\varepsilon^2 - \sigma_{\varepsilon,\eta}^2)}{\sigma_\varepsilon^2 (\sigma_\eta^2 - 2\sigma_{\varepsilon,\eta} + \sigma_\varepsilon^2)} \right). \quad (1.5)$$

By contrast, the standard OLS estimation approach for the AR ($y_t = ay_{t-1} + \varepsilon_t$), which makes no use of survey nowcasts, is asymptotically distributed as

$$\sqrt{T}(\hat{a}_{std} - a) \xrightarrow{d} \mathcal{N}(0, 1 - a^2). \quad (1.6)$$

Thus, the ratio of the two asymptotic variances is

$$VR := \frac{V_a[\hat{a}_{aug}]}{V_a[\hat{a}_{std}]} = \frac{(\sigma_\eta^2 \sigma_\varepsilon^2 - \sigma_{\varepsilon,\eta}^2)}{\sigma_\varepsilon^2 (\sigma_\eta^2 - 2\sigma_{\varepsilon,\eta} + \sigma_\varepsilon^2)} = \frac{r^2(1 - \rho^2)}{r^2 - 2\rho r + 1}, \quad (1.7)$$

where $r = \sigma_\eta/\sigma_\varepsilon$ measures the imprecision of the survey nowcast as a signal about the conditional mean $E[y_t|y_{t-1}, \dots, y_{t-p}]$, and ρ is the correlation between the two disturbances ε_t and η_t . A value of VR below one means that the parameter estimate from the augmented model is asymptotically more precise than the standard OLS estimate. It is easy to show that VR can never exceed one, meaning that the estimator based on the augmented model never produces asymptotically less efficient parameter estimates. Figure 1.1 depicts VR as a function of ρ and r . It shows that gains are particularly high when r is small, i.e. if survey nowcasts tend to be relatively close to the true conditional mean, and if the correlation ρ among the two disturbances is either negative or close to one.

1.2.2 How to address that survey nowcasts may not be “correctly specified”?

In the previous section, we have derived the efficiency gain implied by the augmented model conditional on the assumption of equal coefficient matrices for the actuals y_t and survey nowcasts s_t , i.e. $b_0 = a_0, B_1 = A_1, \dots, B_p = A_p$. This is arguably a demanding assumption that is not likely to be exactly met in practice. Indeed, Online Appendix B.1 shows that sufficient conditions for it to hold are that expectations are formed in a fully rational manner based on an information set that includes the conditioning information of the correctly specified VAR.

In this section, we propose a Bayesian estimation approach, that uses equal coefficients as a shrinkage target, but does not impose them deterministically. We thus conserve some of the potential gains sketched in the previous section without running into the risk of deteriorating forecasts by imposing severely erroneous restrictions.

To express the belief that coefficients are equal, it is helpful to adjust the parametrization of the augmented VAR in equation (1.3). Specifically, we replace b_0 with $a_0 + \Delta_0$, B_1 with

$A_1 + \Delta_1, B_2$ with $A_2 + \Delta_2$, etc. such that (1.3) becomes

$$\begin{bmatrix} y_t \\ s_t \end{bmatrix} = \begin{bmatrix} a_0 \\ a_0 + \Delta_0 \end{bmatrix} + \sum_{i=1}^p \begin{bmatrix} A_i \\ A_i + \Delta_i \end{bmatrix} y_{t-i} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \quad (1.8)$$

and, for convenience, we assume $\begin{bmatrix} \varepsilon'_t & \eta'_t \end{bmatrix}' \sim \mathcal{N}(0, \Sigma)$. Using the new parameterization, we specify a multivariate normal prior distribution for $\{a_0, A_1, \dots, A_p, \Delta_0, \Delta_1, \dots, \Delta_p\}$. Given that we assume that all prior covariances are zero, it suffices to define the marginal prior distribution for each element of the aforementioned matrices and vectors. Denoting by $A_i^{k,l}$ and $\Delta_i^{k,l}$ the (k,l) -cell of A_i and Δ_i respectively, the marginal priors are

$$p(A_i^{k,l}) \sim \mathcal{N}\left(\underline{A}_i^{k,l}, \lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2\right), \quad \text{with } \underline{A}_i^{k,l} = d_k \text{ if } k = l \wedge i = 1, \text{ and } \underline{A}_i^{k,l} = 0 \text{ otherwise,} \quad (1.9)$$

$$p(\Delta_i^{k,l}) \sim \mathcal{N}\left(0, \zeta^2 (\lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2)\right), \quad (1.10)$$

$$p(a_0) \sim \mathcal{N}(0, \kappa \cdot I_M), \quad (1.11)$$

$$p(\Delta_0) \sim \mathcal{N}(0, \kappa \cdot I_M), \quad (1.12)$$

where $\kappa \rightarrow \infty$. The joint prior distribution is the product of the independent marginals. We complete the specification by assuming a diffuse prior distribution for Σ that is independent from the prior distribution of the remaining model parameters: $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$.

Next, we discuss the prior element-by-element. The prior for $\{a_0, A_1, \dots, A_p\}$ is a variant of the Minnesota prior (Doan et al., 1984) that has been used by Wright (2013). While being diffuse about the vector of intercepts a_0 , it is informative about the matrices of slope parameters $\{A_1, \dots, A_p\}$. By setting all prior means except for the first lag of the dependent variable to zero, it expresses the belief that the variables are generated from univariate AR(1) processes.² In the specification of the prior variances in equation (1.9), the hyperparameter λ governs the overall tightness of the prior for A_1, \dots, A_p : If $\lambda = 0$, the prior expresses that we are absolutely certain about the prior means. If, by contrast, $\lambda \rightarrow \infty$, the prior becomes diffuse. The factor $1/i^2$ implies that the prior gets tighter, the higher the lag we consider. It thus reflects the belief that more distant lags play a minor role. Finally, the ratio σ_k^2/σ_l^2 accommodates differences in the scale and variability of the different variables. As we do not have a good prior guess about the term, we follow common practice and proxy σ_k^2 by the residual variance of an AR(1) regression for the k -th variable.

The prior for $\{\Delta_0, \Delta_1, \dots, \Delta_p\}$ is centered at zero, reflecting that we expect the

²This contrasts with Doan et al. (1984), who have suggested a random walk prior with $d_1 = \dots = d_M = 1$. Their specification makes perfect sense when time series are modeled in levels, but it is inappropriate for the stationary variables we consider (see e.g. Banbura et al., 2010).

coefficients to be equal for the actuals y_t and their survey nowcasts s_t . By specifying the prior variances of the Δ_i 's relative to the corresponding elements of $\{a_0, A_1, \dots, A_p\}$, we obtain a parsimonious way to express our confidence in equal coefficients.³ Details about the posterior distribution are given in Appendix 1.5.1.

1.2.3 Adding Wright's democratic steady-state prior

Wright (2013) suggests using long-term survey forecasts to form a prior for the unconditional mean of the variables involved in a VAR. The underlying idea is that professional forecasters should realize shifts in time series endpoints well before they can be inferred from realizations of the process. Villani (2009) outlines the Bayesian estimation of a VAR where a prior is specified for the unconditional mean instead of the vector of intercepts as in section 1.2.2. We extend his approach to the augmented VAR. To implement a prior for the unconditional mean, we set up the following steady-state representation of the augmented VAR:

$$\begin{bmatrix} y_t - \psi \\ s_t - \psi^+ \end{bmatrix} = \sum_{i=1}^p \begin{bmatrix} A_i \\ A_i + \Delta_i \end{bmatrix} (y_{t-i} - \psi) + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \quad (1.13)$$

where $\psi = E[y_t]$ and $\psi^+ = E[s_t]$. Equation (1.13) has been obtained by subtracting

$$\underbrace{\begin{bmatrix} E[y_t] \\ E[s_t] \end{bmatrix}}_{=[\psi' \psi^{+ \prime}]'} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{i=1}^p \begin{bmatrix} A_i \\ A_i + \Delta_i \end{bmatrix} \underbrace{E[y_{t-i}]}_{=\psi}$$

from equation (1.8).

Parameterizing $\psi^+ = \psi + \Delta_\psi$, we specify a multivariate normal prior for $\{\psi, A_1, \dots, A_p, \Delta_\psi, \Delta_1, \dots, \Delta_p\}$. Denoting by $A_i^{k,l}$ and $\Delta_i^{k,l}$ the (k,l)-cell of A_i and Δ_i respectively and by

³Similar to the specification of the prior variance in the Minnesota prior, the $1/i^2$ term implies that survey respondents consider higher lags as less important. Following the advice of an anonymous referee, we have alternatively considered dropping the $1/i^2$ term, finding very similar results in terms of forecast accuracy in our empirical application.

ψ_k and Δ_ψ^k the k-th entry of ψ and Δ_ψ , we set

$$p(A_i^{k,l}) \sim \mathcal{N}(\underline{A}_i^{k,l}, \lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2) \\ \text{with } \underline{A}_i^{k,l} = d_k \text{ if } k = l \wedge i = 1, \text{ and } \underline{A}_i^{k,l} = 0 \text{ otherwise,} \quad (1.14)$$

$$p(\Delta_i^{k,l}) \sim \mathcal{N}(0, \zeta^2 (\lambda^2/i^2 \cdot \sigma_k^2/\sigma_l^2)), \quad (1.15)$$

$$p(\psi_j) \sim \mathcal{N}(\underline{\psi}_j, \lambda_0^2), \quad (1.16)$$

$$p(\Delta_\psi) \sim \mathcal{N}(0, \zeta_0^2 \cdot \lambda_0^2). \quad (1.17)$$

Once again, provided that we assume that the prior covariances are zero, the joint prior can be obtained by multiplying the marginals. With regard to the elements of A_i and Δ_i , the prior is identical to section 1.2.2, but instead of being diffuse about the vector of intercepts, it uses an informative prior for the vector of unconditional means ψ and for the difference vector Δ_ψ . Following Wright (2013), we set the elements of $\underline{\psi}_j$ to the most recent average long-term survey forecasts.⁴ The hyperparameter λ_0 governs the tightness of the prior for ψ , and thus reflects how optimistic we are about the informativeness of the long-term forecasts. Eventually, ζ_0 expresses our confidence in the equality of ψ and ψ^+ , where ψ^+ is the unconditional mean implied by the survey nowcasts. The specification is completed by assuming an independent diffuse prior for Σ , $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$. Details about the posterior distribution can be found in Appendix 1.5.2.

1.3 Empirical application

In this section, we evaluate the forecasts of a ten-variable quarterly VAR(4) that is estimated using our novel approach. As in Wright (2013), our model features eight U.S. macroeconomic variables, a short-term and a long-term yield. To produce the forecasts, we use real-time data from the *Philadelphia Federal Reserve Bank's Real-Time Data Set for Macroeconomists* and *average* survey forecasts from its quarterly *Survey of Professional Forecasters* (SPF). Table 1.1 gives details about the data and how we have processed it.

We conduct the following forecasting experiment: Each period from the second quarter of 1984 through the second quarter of 2011, we re-estimate the VAR on an expanding real-time data window, and produce point forecasts at horizons of one, four, eight and twelve quarters using the iterated approach outlined in Carriero et al. (2015). This approach implies drawing a simulated sample from the posterior predictive density and using the sample mean as a point forecast. To avoid throwing away time series information, we use

⁴For example, for the CPI inflation rate we use the forecasts with a ten year horizon collected by the Philadelphia Federal Reserve's Survey of Professional Forecasters.

an estimation window with atypical design: Whereas the time series of actuals (y_t) starts in the second quarter of 1962, the time series of survey nowcasts (s_t) only begins in the fourth quarter of 1968. An additional complication arises from the fact that our vector of survey nowcasts (s_t) comprises only six of the ten variables included in the VAR.⁵ Online Appendix B.2 modifies our approach to this setting.

In what follows, we try to discern the impact of the different sets of non-sample information by considering alternative specifications of the prior given in equations (1.14 - 1.17) of section 1.2.3. Table 1.2 shows the details: Specification M has the structure of Doan et al. (1984)'s Minnesota prior and ignores all survey information. W adds Wright (2013)'s democratic steady-state prior and thus additionally exploits the long-run survey forecasts. S extends W by using the non-sample information provided through the survey nowcasts. Finally, $S2$ sets the prior variances of the difference parameters to very low values and thus virtually imposes that the slope and unconditional mean parameters are exactly identical for the elements of y_t and of s_t .

Below, we study the forecasts for real GDP growth, GDP deflator inflation, CPI inflation, industrial production growth, the three-month Treasury bill rate and the unemployment rate. We evaluate the forecasts by their MSFE, specifying as the forecast target the value recorded in the second vintage following the quarter, to which the prediction refers. Benchmark forecasts are generated from an AR(1) model, which is estimated by OLS. The AR(1) is often found to be a tough competitor to more complex forecasting models (Chauvet and Potter, 2013; Del Negro and Schorfheide, 2013).⁶

Table 1.3 reports the results of the forecasting experiment. Its key message is that specifications S and $S2$ produce better forecasts for most variables and horizons than all the benchmarks we consider. This result highlights that it pays off in terms of forecast accuracy to exploit the additional information provided through the survey nowcasts. A few more points are notable in Table 1.3: First, in terms of its MSFE, the OLS-VAR(4) is typically inferior to the OLS-AR(1). As the AR(1) model is nested in the VAR(4), this deterioration is likely to reflect over-fitting. Second, the Minnesota prior (M) turns out to improve the VAR forecasts, yet only to a level that is comparable to that of the OLS-AR(1). Third, adding the democratic steady-state prior as in specification W increases the forecast precision (relative to M) for the long-run inflation forecasts, but turns out to make little difference for the remaining variables and horizons.⁷ Fourth, in most cases, augmenting

⁵Note that the time series of survey nowcasts for CPI inflation only starts in 1981:Q3. To obtain pre-1981:Q3 survey nowcasts of CPI inflation, we use an imputation regression based on survey nowcasts of GDP Deflator inflation. For details see Table 1.1.

⁶As an alternative, following Wright (2013), we have considered the forecasts of an AR(p) model with the lag length selected by the BIC. We found that, on average, the AR(1) was harder to beat.

⁷To understand the differences between our results and those of Wright (2013), it is important to note that we use different long-term survey forecasts. Whereas he uses data from the Blue Chip Survey

the VAR with survey nowcasts as in specification S gives superior forecasts. The strongest improvements are obtained for the two inflation series (with a relative gain above 50 percent for GDP deflator inflation on the longest horizon) and for the unemployment rate. For real GDP growth, industrial production growth and the Treasury bill yield, the improvements are less profound but still visible. Fifth, adjusting the prior to rely even more on the survey nowcasts, specification $S2$ gives an additional improvement in predictive ability. This is indirect evidence for our initial guess that survey nowcasts and actuals depend in a very similar way on the lagged dependent variables.

To test if a method improves significantly over the OLS-AR(1), we apply the test for equal finite sample predictive ability proposed by [Giacomini and White \(2006\)](#).⁸ While the test results support that the OLS-VAR(4) tends to produce inferior forecasts, the predictive ability of specifications M and W is rarely significantly different from the OLS-AR(1). By contrast, the forecasts of specifications S and $S2$ are significantly superior at all horizons for the two inflation rates and the unemployment rate. Moreover, the two specifications significantly improve over the AR(1) at longer forecast horizons for industrial production growth and the 3-month T-Bill yield.

1.3.1 Trained hyperparameters

So far, we have considered four alternative specifications of the set of prior hyperparameters $\{\lambda, \zeta, \lambda_0, \zeta_0\}$, finding that their choice strongly affects forecasting performance: On an evaluation sample spanning from the second quarter of 1984 through the second quarter of 2011, we found that stronger shrinkage, i.e. smaller parameter values, typically implied

that has collected long-term forecasts of all the ten variables twice a year since 1984, the SPF's ten-year forecasts are available for only four variables and start in 1991:Q4 earliest. It is therefore not surprising that he finds a much larger improvement in predictive ability from the democratic steady-state prior than we do.

⁸Note that the test we use differs from the one employed by [Wright \(2013\)](#). Whereas we use a test for finite sample predictive ability of alternative forecast methods ([Giacomini and White, 2006](#)), he uses a test for equal population level predictive ability that is suitable for nested forecast models ([Clark and West, 2007](#)). We prefer our test for two reasons: First, our test is far more demanding with respect to the extent of forecast improvement. The test used by Wright (2013) only requires the richer model to produce better forecasts at population level. It therefore ignores estimation uncertainty that is likely to deteriorate the forecasts of the richer model relative to the nested model in a finite sample context. As the samples encountered in real-world macroeconomic forecasting applications are not even close to the population level, we consider the finite sample context as more relevant. Second, the finite sample test allows for comparisons among different forecasting methods, i.e. combinations of a forecasting model and an estimation strategy, whereas the population level test can only distinguish models (because estimation is irrelevant at the population level).

Due to the expanding estimation window, the asymptotics presented in [Giacomini and White \(2006\)](#) are not valid in our context. In favor of using the method with expanding estimation windows anyway, [Clark and McCracken \(2015\)](#) show in a simulation study that the test has reasonable size properties.

Note also that the use of real-time data may invalidate the asymptotics of tests for equal predictive ability such as the one we use, for details see [Clark and McCracken \(2009b\)](#).

better forecasting performance. Despite the promising result, a valid criticism is the arbitrary choice of the hyperparameters values.

To address this concern, we have considered choosing the hyperparameters based on a training sample and evaluating the performance of this prior specification on a subsequent evaluation sample. Specifically, we think ourselves back to 1990:Q4. Using all data available at that time, we evaluate the pseudo out-of-sample forecasts obtained from each possible combination of the following hyperparameters values: $\lambda = \{.01, .05, .1, .15, .2\}$, $\zeta = \{.01, .1, .5, 1, 2, 10\}$, $\lambda_0 = .5$, $\zeta_0 = \{.01, .1, .5, 1, 2, 10\}$, where we need to fix λ_0 because the data on long-term survey forecasts only start in 1991:Q4.⁹ To choose a single best specification, we use a criterion that aggregates the forecast performance across several variables and horizons. In the spirit of Wright (2013), we compute for each variable-horizon combination the relative MSFE versus the *AR(1)* model, and aggregate by averaging across variables and forecast horizons (considering only the six variables and four horizons evaluated in Table 1.3). We find that the criterion prefers the following specification, which we subsequently denote by T : $\lambda^* = 0.1$, $\zeta^* = 0.01$, $\lambda_0^* = 0.5$ and $\zeta_0^* = 0.01$. This is the tightest specification available with respect to ζ and ζ_0 , the two hyperparameters that relate to the survey nowcasts, but not with respect to λ , the hyperparameter that governs the tightness of the Minnesota prior. Based on prior specification T , we start generating real-time out-of-sample forecasts with the 1990:Q4 real-time data vintage. Table 1.4 summarizes the results of the forecasting experiment: The four specifications M , W , S , and $S2$ perform similarly on this shorter evaluation sample as on the full sample considered in the previous section: The tightest variant $S2$ typically provides the best forecasts. Interestingly, the trained specification T roughly performs at eye-level with the best specification ($S2$), indicating that the real-time choice of hyperparameters works pretty well.

1.3.2 Survey forecasts and forecast combination

In this section, we compare the forecasts from our method, using prior specification T , to two additional benchmarks: The SPF survey forecasts themselves, and different linear combinations of the survey forecasts and the Bayesian VAR forecasts. Contrary to the previous evaluations, due to the limited availability of the survey data, we can only consider forecasts at horizons of one, two, three, and four quarters.

The comparison of the model-based forecasts with survey-forecasts raises some intricate timing issues: For a fair comparison, the two methods should have similar information sets available. To illustrate the difficulty, we consider the one-quarter ahead forecast for

⁹ The value of 0.5 roughly coincides with the specification that Wright (2013) infers from his training sample (0.557 in terms of our specification of the prior) using a richer data set of survey long-run forecasts.

the growth of real GDP in 1990:Q4: The latest information used by the VAR refers to 1990:Q3, whereas (i) the one-quarter ahead survey forecast produced by the quarter-mid of 1990:Q3 has only limited information about the 1990:Q3 data and (ii) the survey nowcast made in 1990:Q4 has extra-information (relative to the VAR) about the ongoing quarter, such as the industrial production growth in 1990:M10. Here, we follow Wright (2013) and use the one-quarter ahead survey forecast, thus putting the survey forecasts at a slight information disadvantage relative to the VAR.

Despite this disadvantage, Table 1.5 shows that survey forecasts are a tough competitor to our method. Considering the two inflation series, the gain from using the survey forecast is considerable with respect to GDP deflator inflation and moderate for CPI inflation. Considering the remaining four series, the table suggests that the two methods roughly perform at eye-level with a slight edge for our method. It should be kept in mind that even though our method cannot clearly beat survey forecasts, it has the advantage of providing forecasts at any horizon and any point in time.

The head-to-head race among our method and the survey forecasts suggests that we may benefit from forecast combinations. We consider three approaches with pseudo real-time updates of the forecast weights:

1. The **MSFE approach** weighs the two forecasts according to the inverse of their MSFE.
2. The **Granger and Ramanathan (1984) approach** obtains weights by regressing the realization on the two forecasts, subject to the restriction the regression coefficient sum to unity.
3. The $\frac{1}{N}$ **approach** weighs each forecast by 0.5.

The results are also found in Table 1.5: The first insight is that the different weighting approaches perform similarly, allowing no uniform ranking across the variables and horizons. Moreover, the MSFE of the combined forecast is typically marginally higher than the MSFE of the better individual forecast. This is a typical result in forecast combination experiments (e.g. Krüger, 2014) and suggests that without reliable ex-ante knowledge of the relative performance of the two forecast methods, combination is an advisable strategy.

1.4 Concluding Remarks

In this paper, we have proposed a Bayesian shrinkage method for VARs that uses both long- and short-run survey forecasts as non-sample information. Our empirical application

has shown that the method typically improves forecast accuracy relative to approaches that do not use such (non-sample) information. The shrinkage approach is easy to implement and it can be transferred to other types of time-series models, such as the non-linear class of vector STAR models (e.g. [Schleer, 2015](#)).

References

- ALTAVILLA, C., R. GIACOMINI, AND G. RAGUSA (2014): “Anchoring the Yield Curve Using Survey Expectations,” ECB Working Paper No. 1632, ECB.
- BAŞTÜRK, N., C. ÇAKMAKLI, S. P. CEYHAN, AND H. K. VAN DIJK (2014): “Posterior-Predictive Evidence on Us Inflation Using Extended New Keynesian Phillips Curve Models with Non-Filtered Data,” *Journal of Applied Econometrics*, 29, 1164–1182.
- BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): “Large Bayesian Vector Auto Regressions,” *Journal of Applied Econometrics*, 25, 71–92.
- BAUWENS, L., M. LUBRANO, AND J. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*, Advanced Texts in Econometrics, Oxford University Press: Oxford, UK.
- CARRIERO, A., T. E. CLARK, AND M. MARCELLINO (2015): “Bayesian VARs: Specification Choices and Forecast Accuracy,” *Journal of Applied Econometrics*, 30, 46–73.
- CHAUVET, M. AND S. POTTER (2013): “Forecasting Output,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier: Amsterdam, NL, vol. 2.
- CLARK, T. E. AND M. W. McCACKEN (2009b): “Tests of Equal Predictive Ability With real-time Data,” *Journal of Business & Economic Statistics*, 27, 441–454.
- (2015): “Nested forecast model comparisons: A new approach to testing equal accuracy,” *Journal of Econometrics*, 186, 160–177.
- CLARK, T. E. AND K. D. WEST (2007): “Approximately Normal Tests for Equal Predictive Accuracy in Nested Models,” *Journal of Econometrics*, 138, 291–311.
- COGLEY, T., S. MOROZOV, AND T. J. SARGENT (2005): “Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system,” *Journal of Economic Dynamics and Control*, 29, 1893–1925.
- CROUSHORE, D. (1993): “Introducing: The Survey of Professional Forecasters,” *Business Review*, 6, 3–15.
- CROUSHORE, D. AND T. STARK (2001): “A real-time data set for macroeconomists,” *Journal of Econometrics*, 105, 111–130.
- DEL NEGRO, M. AND F. SCHORFHEIDE (2013): “DSGE Model-Based Forecasting,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier: Amsterdam, NL, vol. 2.

- DOAN, T., R. LITTERMAN, AND C. SIMS (1984): “Forecasting and Conditional Projection using Realistic Prior Distributions,” *Econometric Reviews*, 3, 1–100.
- FAUST, J. AND J. H. WRIGHT (2013): “Forecasting Inflation,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier: Amsterdam, NL, vol. 2.
- GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*, Wiley: Hoboken, NJ.
- GIACOMINI, R. AND H. WHITE (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74, 1545–1578.
- GRANGER, C. W. J. AND R. RAMANATHAN (1984): “Improved Methods of Combining Forecasts,” *Journal of Forecasting*, 3, 197–204.
- ING, C.-K. AND C.-Z. WEI (2003): “On Same-Realization Prediction in an Infinite-Order Autoregressive Process,” *Journal of Multivariate Analysis*, 85, 130–155.
- KIM, D. AND A. ORPHANIDES (2012): “Term Structure Estimation with Survey Data on Interest Rate Forecasts,” *Journal of Financial & Quantitative Analysis*, 47, 241–272.
- KRÜGER, F. (2014): “Combining Density Forecasts under various Scoring Rules: An Analysis of UK Inflation,” Working Paper, Heidelberg Institute for Theoretical Studies.
- KRÜGER, F., T. E. CLARK, AND F. RAVAZZOLO (2015): “Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts,” *Journal of Business & Economic Statistics*, 1–48.
- ROBERTSON, J. C., E. W. TALLMAN, AND C. H. WHITEMAN (2005): “Forecasting Using Relative Entropy,” *Journal of Money, Credit and Banking*, 37, 383–401.
- SCHLEER, F. (2015): “Finding Starting-Values for the Estimation of Vector STAR Models,” *Econometrics*, 3, 65–90.
- STOCK, J. H. AND M. W. WATSON (2001): “Vector Autoregressions,” *The Journal of Economic Perspectives*, 15, 101–115.
- VILLANI, M. (2009): “Steady-State Priors for Vector Autoregressions,” *Journal of Applied Econometrics*, 24, 630–650.
- WRIGHT, J. H. (2013): “Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior,” *Journal of Applied Econometrics*, 28, 762–776.

1.5 Appendix

1.5.1 Posterior distribution for the augmented VAR

To derive the posterior distribution of the model of section 1.2.2, we put (1.8) in matrix notation:

$$\begin{bmatrix} Y & S \end{bmatrix} = X \begin{bmatrix} A & A + \Delta \end{bmatrix} + \begin{bmatrix} E & H \end{bmatrix}, \quad (1.18)$$

where $Y = [y_1 \dots y_T]', S = [s_1 \dots s_T]', X = [x_1 \dots x_T]',$
 $x_t = [1 \ y'_{t-1} \ \dots \ y'_{t-p}]', A' = [a_0 \ A_1 \ \dots \ A_p], \Delta' = [\Delta_0 \ \Delta_1 \ \dots \ \Delta_p],$
 $E = [\varepsilon_1 \ \dots \ \varepsilon_T]',$ and $H = [\eta_1 \ \dots \ \eta_T]'$. Vectorizing the matrix representation column-by-column, we obtain

$$\underbrace{\begin{bmatrix} \text{vec}(Y) \\ \text{vec}(S) \end{bmatrix}}_y = \underbrace{\begin{bmatrix} I_M \otimes X & 0 \\ I_M \otimes X & I_M \otimes X \end{bmatrix}}_Z \underbrace{\begin{bmatrix} \text{vec}(A) \\ \text{vec}(\Delta) \end{bmatrix}}_\beta + \underbrace{\begin{bmatrix} \text{vec}(E) \\ \text{vec}(H) \end{bmatrix}}_\epsilon, \quad (1.19)$$

where $V[\epsilon] = \Sigma \otimes I_T$. The vectorized representation has the structure of a multivariate linear seemingly unrelated regression (SUR) model, such that we can use standard results outlined e.g. in Geweke (2005, p.162ff) for its Bayesian estimation.

Given our normal-diffuse prior of the form $p(\beta) \sim \mathcal{N}(\underline{\beta}, V_\beta)$, $\Sigma \propto |\Sigma|^{-2(2M+1)/2}$, the full conditional posterior of β is

$$p(\beta | Y_T, \Sigma) \sim \mathcal{N}(\bar{\beta}, \bar{V}_\beta), \quad (1.20)$$

with $\bar{V}_\beta = (V_\beta^{-1} + Z'(\Sigma^{-1} \otimes I_T)Z)^{-1}$ and $\bar{\beta} = \bar{V}_\beta (V_\beta^{-1} \underline{\beta} + Z'(\Sigma^{-1} \otimes I_T)y)$. And the full conditional posterior of Σ is

$$p(\Sigma | Y_T, \beta) \sim \mathcal{IW}(T, U) \quad (1.21)$$

where $U = ([Y \ S] - X[A \ A + \Delta])'([Y \ S] - X[A \ A + \Delta])$, and \mathcal{IW} is the inverted Wishart distribution (see Bauwens et al., 1999, section A.2.6). Thus, we can use the Gibbs sampler to obtain draws from the posterior distribution, iterating between (1.20) and (1.21).

1.5.2 Posterior distribution for augmented VAR with a democratic steady-state prior

Below we discuss the posterior distribution of the model of section 1.2.3. First, it proves helpful to put the model in matrix notation as

$$\begin{bmatrix} Y_\psi & S_{\psi^+} \end{bmatrix} = X_\psi \begin{bmatrix} \Lambda & \Lambda + \Delta_\Lambda \end{bmatrix} + \begin{bmatrix} E & H \end{bmatrix} \quad (1.22)$$

where $Y_\psi = [y_1 - \psi \dots y_T - \psi]', S_{\psi^+} = [s_1 - \psi^+ \dots s_T - \psi^+]'$, $X_\psi = [x_{\psi 1} \dots x_{\psi T}]'$, $x_{\psi t} = [y'_{t-1} - \psi' \dots y'_{t-p} - \psi']'$, $\Lambda' = [A_1 \dots A_p]'$, $\Delta'_\Lambda = [\Delta_1 \dots \Delta_p]'$, $E = [\varepsilon_1 \dots \varepsilon_T]'$, and $H = [\eta_1 \dots \eta_T]'$. Vectorizing the matrix representation column-by-column, we obtain

$$\underbrace{\begin{bmatrix} \text{vec}(Y_\psi) \\ \text{vec}(S_{\psi^+}) \end{bmatrix}}_y = \underbrace{\begin{bmatrix} I_M \otimes X_\psi & 0 \\ I_M \otimes X_\psi & I_M \otimes X_\psi \end{bmatrix}}_Z \underbrace{\begin{bmatrix} \text{vec}(\Lambda) \\ \text{vec}(\Delta_\Lambda) \end{bmatrix}}_\beta + \underbrace{\begin{bmatrix} \text{vec}(E) \\ \text{vec}(H) \end{bmatrix}}_\epsilon, \quad (1.23)$$

where as before $V[\epsilon] = \Sigma \otimes I_T$. Noting that conditional on ψ and Δ_ψ , the model has the structure of a multivariate linear seemingly unrelated regression (SUR) model, we can use standard results for its Bayesian estimation. The full conditional posterior of β is

$$p(\beta|Y_T, \Sigma, \psi, \Delta_\psi) \sim \mathcal{N}(\bar{\beta}, \bar{V}_\beta), \quad (1.24)$$

with $\bar{V}_\beta = V_\beta^{-1} + Z'(\Sigma^{-1} \otimes I_T)Z$ and $\bar{\beta} = \bar{V}_\beta (V_\beta^{-1} \beta + Z'(\Sigma^{-1} \otimes I_T)y)$. And the full conditional posterior of Σ is

$$p(\Sigma|Y_T, \Lambda, \Delta_\Lambda, \psi, \Delta_\psi) \sim \mathcal{IW}(T, U) \quad (1.25)$$

where $U = ([Y_\psi \ S_{\psi^+}] - X_\psi [\Lambda \ \Lambda + \Delta_\Lambda])' ([Y_\psi \ S_{\psi^+}] - X_\psi [\Lambda \ \Lambda + \Delta_\Lambda])$.

To obtain the full conditional posterior of $\Psi = [\psi' \ \Delta'_\psi]'$, we rewrite the steady-state representation of the VAR (1.13) as

$$\underbrace{y_t - \sum_{i=1}^p A_i y_{t-i}}_{y_t^d} = \underbrace{\left(I_M - \sum_{i=1}^p A_i \right)}_{\Pi_1} \psi + \varepsilon_t, \quad (1.26)$$

$$\underbrace{s_t - \sum_{i=1}^p (A_i + \Delta_i) y_{t-i}}_{s_t^d} = \underbrace{\left(I_M - \sum_{i=1}^p (A_i + \Delta_i) \right)}_{\Pi_2} \psi + I_M \Delta_\psi + \eta_t, \quad (1.27)$$

and put it into matrix notation as

$$\underbrace{\begin{bmatrix} y_1^d \\ s_1^d \\ \vdots \\ y_T^d \\ s_T^d \end{bmatrix}}_{y_d} = \underbrace{\begin{bmatrix} \Pi_1 & 0 \\ \Pi_2 & I_M \\ \vdots & \\ \Pi_1 & 0 \\ \Pi_2 & I_M \end{bmatrix}}_{Z_d} \Psi + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \eta_1 \\ \vdots \\ \varepsilon_1 \\ \eta_1 \end{bmatrix}}_{\epsilon_d}, \text{ where } V[\epsilon_d] = I_T \otimes \Sigma. \quad (1.28)$$

Conditional on Λ , Δ_Λ , and Σ , equation (1.28) is a multivariate linear regression of y_d on Z_d , such that the posterior distribution is obtained from standard SUR results as

$$p(\Psi | Y_T, \Sigma, \Lambda, \Delta_\Lambda) \sim \mathcal{N}(\bar{\Psi}, \bar{V}_\Psi), \quad (1.29)$$

where $\bar{V}_\Psi = (\underline{V}_\Psi^{-1} + Z_d'(I_T \otimes \Sigma^{-1})Z_d)^{-1}$ and $\bar{\Psi} = \bar{V}_\Psi (\underline{V}_\Psi^{-1} \underline{\Psi} + Z_d'(I_T \otimes \Sigma^{-1})y_d)$.

To obtain draws from the posterior distribution, Villani (2009) suggests using a three-block Gibbs sampler that iterates between (1.24), (1.25) and (1.29).

1.5.3 Figures & Tables

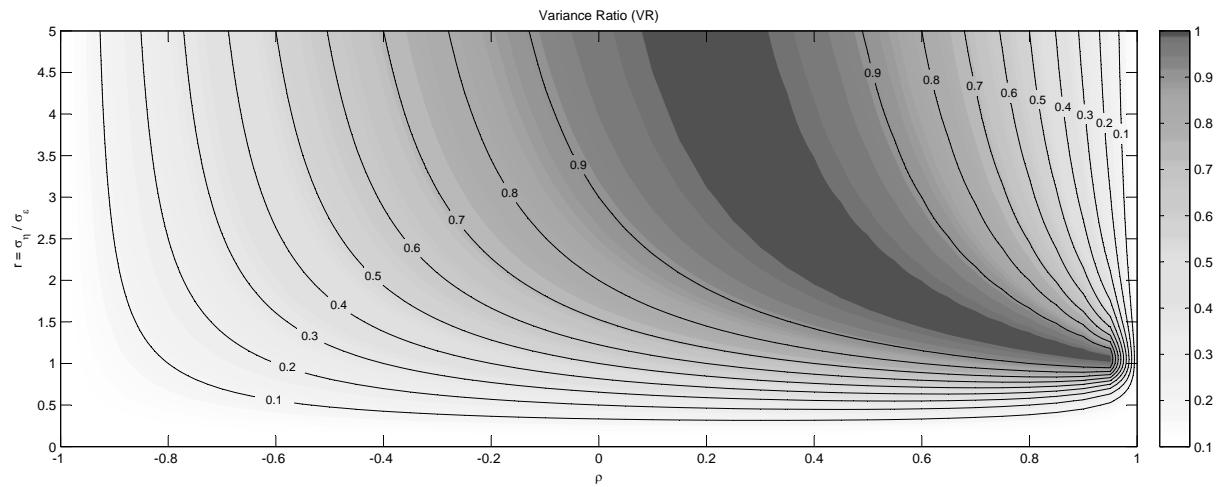


Figure 1.1: Variance Ratio $VR(r, \rho)$

Table 1.1: Data description and variable transformation

Variable (x_t)	Transformation used in VAR (y_t) ¹	Real-time data ² (earliest vintage used)	Original data frequency ³	Survey nowcasts starting in 1968:Q4 ⁴	Long-term forecasts w. 10y horizon
1 Real GDP	$100((x_t/x_{t-1})^4 - \frac{\text{yes}}{(1984:Q2)})$		Quarterly	yes	yearly fr. 1992:Q1
2 GDP deflator	$400(\ln x_t - \ln x_{t-1})^{\text{yes}}_{(1984:Q2)}$		Quarterly	yes	-
3 CPI	$100((x_t/x_{t-1})^4 - \frac{\text{yes}}{(1994:Q3)^5})$		Monthly	yes ⁶	quarterly fr. 1991:Q4
4 Industrial production	$400(\ln x_t - \ln x_{t-1})^{\text{yes}}_{(1984:Q2)}$		Monthly ⁷	yes	-
5 Nonresidential fixed investment	$400(\ln x_t - \ln x_{t-1})^{\text{yes}}_{(1984:Q2)}$		Quarterly	-	-
6 Real personal consumption expenditure	$400(\ln x_t - \ln x_{t-1})^{\text{yes}}_{(1984:Q2)}$		Quarterly	-	-
7 Housing starts	x_t	$\frac{\text{yes}}{(1984:Q2)}$	Monthly ⁶	yes	-
8 Unemployment rate	x_t	$\frac{\text{yes}}{(1984:Q2)}$	Monthly	yes	-
9 10y Treasury bond yield ⁸	x_t	-	Daily	-	yearly fr. 1992:Q1
10 3m Treasury bill yield ⁸	x_t	-	Daily	-	yearly fr. 1992:Q1

Notes:

¹ We transform real GDP and CPI to *geometric* growth rates, because in the two cases the survey forecasts refer to growth rates instead of index levels. Thus, we make sure that the variables in the VAR and the survey nowcasts use identical definitions.

² Source of the real-time data: Philadelphia Federal Reserve Bank's Real-Time Data Set for Macroeconomists (<http://www.phil.frb.org/research-and-data/real-time-center/real-time-data/>, see Croushore and Stark (2001)).

³ We obtain quarterly time series by averaging the monthly or daily observations. Note that growth rates are computed after averaging across the high-frequency observations in levels.

⁴ Source of the Survey forecasts: Philadelphia Federal Reserve Bank's quarterly Survey of Professional Forecasters (SPF, <http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/>, see Croushore (1993)).

⁵ We extract the appropriate data for earlier pseudo real-time periods from the 1994:Q3 vintage.

⁶ CPI has only been part of the survey since 1981:Q3. We impute values based on a regression of the average CPI inflation nowcast on the average GDP deflator inflation nowcast and an intercept based on post 1981:Q2 data. Alternatively, acknowledging the real-time nature of our forecasting experiment, we have considered imputing the average GDP deflator nowcast without making an adjustment to it. The empirical results show only a very slight deterioration of the forecasts, suggesting that the imputation method plays only a minor role for forecast accuracy.

⁷ Vintages are monthly, we extract the quarter-middle vintage.

⁸ Downloaded from Thomson Reuters Datastream.

Table 1.2: Prior Specifications

	$\mathbf{A}_1, \dots, \mathbf{A}_p$ $[p(A_i^{k,l}) \sim \mathcal{N}\left(\underline{A}_i^{k,l}, \frac{\lambda^2}{i^2} \frac{\sigma_k^2}{\sigma_l^2}\right)]$			$\Delta_1, \dots, \Delta_p$ $[p(\Delta_i^{k,l}) \sim \mathcal{N}\left(0, \zeta^2 \frac{\lambda^2}{i^2} \frac{\sigma_k^2}{\sigma_l^2}\right)]$		
Spec.	$\underline{A}_i^{k,l} _{i \neq 1} \vee$	$\underline{A}_1^{k,k}$	λ	$Pr(\cdot) \approx 0.95$	ζ	$Pr(\cdot) \approx 0.95$
M	0	d_k^\dagger	0.2 ^{††}	$A_1^{k,k} \in [d_k \pm 0.4]$	1000	$\Delta_1^{k,k} \in [\pm 400]$
W	"	"	"	"	"	"
S	"	"	"	"	0.100	$\Delta_1^{k,k} \in [\pm 0.04]$
$S2$	"	"	"	"	0.001	$\Delta_1^{k,k} \in [\pm 0.0004]$

	ψ $[p(\psi_j) \sim \mathcal{N}\left(\underline{\psi}_j, \lambda_0^2\right)]$			Δ_ψ $[p(\Delta_{\psi_j}) \sim \mathcal{N}\left(0, \zeta_0^2 \cdot \lambda_0^2\right)]$	
Spec.	$\underline{\psi}_j$	λ_0	$Pr(\cdot) \approx 0.95$	ζ_0	$Pr(\cdot) \approx 0.95$
M	0	10^5	$\psi_k \in [\pm 2 \times 10^5]$	1	$\Delta_{\psi_k} \in [\pm 2 \times 10^5]$
W	$l_k // 0^*$	$0.5 // 10^{5*}$	$\psi_k \in [l_k \pm 1] // [\pm 2 \times 10^5]^*$	2×10^5	"
S	"	"	"	0.2	$\Delta_{\psi_k} \in [\pm 0.2]$
$S2$	"	"	"	0.001	$\Delta_{\psi_k} \in [\pm 0.001]$

Note: The table presents the different prior specifications employed in the forecasting experiment of section 3.4. l_k is the mean long-term forecast of variable k . The * indicates that for the variables and points in time, for which no long-term forecasts l_k are available, we use the value after the // to specify the prior distribution.

[†] Following Wright (2013), we set $d_k = 0$ for each real variable (real GDP, nonresidential fixed investment, and real personal consumption expenditures), and $d_k = 0.8$ for the nominal variables.

^{††} According to Carriero et al. (2015), this choice for the tightness parameter of the Minnesota prior is common in the Bayesian VAR forecasting literature. In a forecast experiment that uses macroeconomic data, the authors find that the optimal value of λ is close to 0.2.

Table 1.3: Forecasts with Different Prior Specifications:
Relative Mean Squared Forecast Errors (Evaluation Sample: 1984:Q2 - 2011:Q2)

Horizon	MSFE		Relative MSFE to the OLS-AR1			
	OLS-AR1	OLS-VAR	<i>M</i>	<i>W</i>	<i>S</i>	<i>S2</i>
Real GDP growth						
<i>h</i> = 1	4.615	1.599**	1.188	1.158	0.910	0.887
<i>h</i> = 4	5.732	1.338**	1.190	1.148	0.932	0.882
<i>h</i> = 8	5.807	1.339**	1.101	1.082	0.942	0.922
<i>h</i> = 12	5.855	0.972	0.920	0.921	0.937	1.011
GDP deflator inflation						
<i>h</i> = 1	1.443	1.164	0.928	0.933	0.940	0.846*
<i>h</i> = 4	2.359	1.196	0.948	0.911	0.781***	0.625***
<i>h</i> = 8	4.091	1.385	0.965	0.840	0.610***	0.438***
<i>h</i> = 12	5.084	1.204	0.954	0.759	0.498***	0.332***
CPI inflation						
<i>h</i> = 1	5.021	1.095	0.919	0.919	0.835	0.772*
<i>h</i> = 4	6.218	1.388	1.059	1.055	0.818*	0.734**
<i>h</i> = 8	6.893	1.586	1.128	1.039	0.716***	0.656***
<i>h</i> = 12	7.411	1.295	1.037	0.895	0.604***	0.592***
IP growth						
<i>h</i> = 1	14.772	1.697**	1.266*	1.264*	1.153	1.159
<i>h</i> = 4	25.180	1.301**	1.163	1.149	0.997	0.933
<i>h</i> = 8	25.825	1.140	0.969	0.957	0.906*	0.929**
<i>h</i> = 12	26.250	0.968	0.899	0.886	0.941*	0.983
Three-month Treasury bill yield						
<i>h</i> = 1	0.252	2.091**	1.072	1.092	1.150	1.128
<i>h</i> = 4	2.301	1.345	0.994	1.014	0.990	0.901
<i>h</i> = 8	5.744	1.118	0.984	0.987	0.850	0.742
<i>h</i> = 12	7.942	1.071	1.022	1.004	0.756*	0.568**
Unemployment rate						
<i>h</i> = 1	0.092	0.819	0.606*	0.612*	0.598*	0.644**
<i>h</i> = 4	0.885	0.930	0.834	0.850	0.766**	0.774**
<i>h</i> = 8	2.433	1.023	0.951	0.969	0.818*	0.747***
<i>h</i> = 12	3.444	0.883	0.780***	0.794***	0.715***	0.685***

Note: The table reports results from a pseudo real-time out-of-sample forecasting experiment. Column 'MSFE' holds the mean squared forecast error of the AR(1) model at different forecast horizons and for different variables. The columns titled 'Relative MSFE' show the ratio of the MSFE of different alternative forecasting methods to the MSFE of the AR(1) model. For each method, we test whether it has lower MSFE than the AR(1) by the test proposed by [Giacomini and White \(2006\)](#). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level. The methods OLS-AR1 and OLS-VAR refer to an AR(1) and the ten-variable quarterly VAR(4) estimated using OLS, respectively. The prior specifications *M*, *W*, *S*, *S2* are outlined in table 1.2 and on page 18.

Table 1.4: Forecasting with Trained Hyperparameters:
Relative Mean Squared Forecast Errors (Evaluation Sample: 1990:Q4 - 2011:Q2)

Horizon	MSFE		Relative MSFE to the OLS-AR1				
	OLS-AR1	OLS-VAR	<i>M</i>	<i>W</i>	<i>S</i>	<i>S2</i>	<i>T</i>
Real GDP growth							
<i>h</i> = 1	5.249	1.458*	1.186	1.154	0.873	0.835	0.830
<i>h</i> = 4	5.988	1.446**	1.291**	1.243*	1.014	0.924	0.890
<i>h</i> = 8	6.187	1.331	1.161	1.140	1.024	0.946	0.930
<i>h</i> = 12	6.367	0.904	0.895	0.894	0.985	0.961	0.957
GDP deflator inflation							
<i>h</i> = 1	1.314	1.013	0.891	0.891	0.899	0.816*	0.841*
<i>h</i> = 4	2.313	1.029	0.907	0.857	0.736**	0.612***	0.690***
<i>h</i> = 8	4.088	0.968	0.836	0.673	0.524***	0.413***	0.458***
<i>h</i> = 12	5.422	0.940	0.875	0.621**	0.446***	0.325***	0.372***
CPI inflation							
<i>h</i> = 1	5.540	1.060	0.895	0.892	0.814	0.736*	0.739*
<i>h</i> = 4	6.586	1.168	1.025	1.023	0.781	0.674***	0.685***
<i>h</i> = 8	7.307	1.147	0.983	0.883	0.684**	0.612***	0.634***
<i>h</i> = 12	8.797	1.018	0.966	0.803	0.593***	0.529***	0.563***
IP growth							
<i>h</i> = 1	17.752	1.167	1.129	1.129	1.029	1.025	1.080
<i>h</i> = 4	27.972	1.243	1.186	1.166	1.022	0.937	0.932
<i>h</i> = 8	27.980	1.163	1.026	1.011	0.965	0.951*	0.942**
<i>h</i> = 12	29.185	0.911	0.866	0.853*	0.956*	0.971*	0.965*
Three-month Treasury bill yield							
<i>h</i> = 1	0.242	1.641	0.992	1.007	1.026	0.951	0.847
<i>h</i> = 4	2.465	1.103	0.851	0.875	0.855	0.743*	0.736**
<i>h</i> = 8	6.099	0.948	0.824	0.829	0.694**	0.585**	0.627***
<i>h</i> = 12	8.330	1.009	0.896	0.875	0.582***	0.434***	0.509***
Unemployment rate							
<i>h</i> = 1	0.110	0.651	0.586*	0.589*	0.580*	0.622**	0.641**
<i>h</i> = 4	1.052	0.940	0.888	0.907	0.814*	0.817**	0.821**
<i>h</i> = 8	2.821	1.123	1.065	1.085	0.909	0.824**	0.822**
<i>h</i> = 12	3.889	0.839**	0.822***	0.838**	0.769***	0.729***	0.760***

Note: The table reports results of a pseudo real-time out-of-sample forecasting experiment. Relative to table 1.3, to evaluate the different forecast methods, this experiment evaluates a smaller sample of forecasts produced from 1990:Q4 through 2011:Q2 in pseudo real-time. The reason is that the prior specification *T* uses hyperparameters that have been trained on a sample extending from 1984:Q2 through 1990:Q3. Column 'MSFE' holds the mean squared forecast error of the AR(1) model at different forecast horizons and for different variables. The columns titled 'Relative MSFE' show the ratio of the MSFE of different alternative forecasting methods to the MSFE of the AR(1) model. For each method, we test whether it has lower MSFE than the AR(1) by the test proposed by Giacomini and White (2006). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level. The methods OLS-AR1 and OLS-VAR refer to an AR(1) and the ten-variable quarterly VAR(4) estimated using OLS, respectively. The prior specifications *M*, *W*, *S*, *S2* are outlined in table 1.2 and on page 18.

Table 1.5: Comparing the Forecasts with both Survey Expectations and Combined Forecasts:

Relative Mean Squared Forecast Errors (Evaluation Sample: 1990:Q4 - 2011:Q2)

Horizon	OLS-AR1	MSFE		Relative MSFE to the OLS-AR1		
		VAR, specification T	Survey Forecasts	Combined Forecasts (VAR+Surveys)		
				Inv. MSFE	G.R.	$\frac{1}{N}$
Real GDP growth						
$h = 1$	5.249	0.830	0.831	0.785	0.810	0.783
$h = 2$	6.148	0.849	0.848	0.814	0.847	0.813
$h = 3$	6.019	0.859	0.950	0.881	0.923	0.881
$h = 4$	5.988	0.890	1.006	0.925	0.937	0.929
GDP deflator inflation						
$h = 1$	1.314	0.841*	0.592**	0.625**	0.602**	0.636**
$h = 2$	1.555	0.662***	0.533***	0.534***	0.534***	0.541***
$h = 3$	1.810	0.681***	0.525***	0.550***	0.506***	0.556***
$h = 4$	2.313	0.690***	0.495***	0.549***	0.483***	0.555***
CPI inflation						
$h = 1$	5.540	0.739*	0.695	0.690*	0.710*	0.688*
$h = 2$	7.092	0.618**	0.567**	0.580**	0.579**	0.581**
$h = 3$	6.505	0.688**	0.625*	0.641**	0.636**	0.642**
$h = 4$	6.586	0.685***	0.622**	0.638***	0.626**	0.640***
IP growth						
$h = 1$	17.752	1.080	1.161	1.079	1.073	1.081
$h = 2$	26.781	1.004	0.966	0.953	0.966	0.953
$h = 3$	27.284	0.931	1.009	0.948	0.963	0.948
$h = 4$	27.972	0.932	1.002	0.951	0.971	0.952
Three-month Treasury bill yield						
$h = 1$	0.242	0.847	1.086	0.869*	0.897	0.873*
$h = 2$	0.796	0.814*	0.949	0.824***	0.868*	0.825***
$h = 3$	1.550	0.775**	0.956	0.820***	0.841**	0.825***
$h = 4$	2.465	0.736**	0.987	0.819**	0.793*	0.829**
Unemployment rate						
$h = 1$	0.110	0.641**	0.897	0.664**	0.667*	0.702**
$h = 2$	0.336	0.703**	0.738	0.675**	0.728**	0.681**
$h = 3$	0.668	0.771**	0.768*	0.734**	0.815*	0.738**
$h = 4$	1.052	0.821**	0.824*	0.791***	0.898	0.795***

Note: The table reports results of a pseudo real-time out-of-sample forecasting experiment. Relative to table 1.3, this experiment uses a smaller sub-sample of realizations, which spans from 1990:Q4 through 2011:Q2, to evaluate the different methods. The reason is that the forecast combination methods and the trained VAR prior ('VAR, specification T ') require a training sample. Note that the VAR prior is trained only once, whereas forecast combination weights are re-estimated recursively. Column 'MSFE' holds the mean squared forecast error of the AR(1) model at different forecast horizons and for different variables. The columns title 'Relative MSFE' show the ratio of the MSFE of different alternative forecasting methods to the MSFE of the AR(1) model. For each method, we test whether it has lower MSFE than the AR(1) by the test proposed by Giacomini and White (2006). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

CHAPTER 2

Bayesian Shrinkage of Portfolio Weights

2.1 Introduction

Estimation risk belongs to the great challenges of empirical portfolio modeling. In the standard Markowitz framework optimal portfolio weights for an N -dimensional asset space are functions of the mean and the variance of the underlying asset returns. The estimation of N mean and $N(N+1)/2$ variance-covariance parameters yields extremely noisy portfolio weights estimates with large standard errors. The resulting estimated portfolios show poor out-of-sample performance (Jobson and Korkie, 1980), extreme short positions and little diversification (Michaud, 1989). Additionally, small changes in the parameter estimates can lead to drastic changes in portfolio weights (Best and Grauer, 1991).

To mitigate these shortcomings and to robustify portfolio estimates against too large estimation noise, the literature has proposed a range of alternative strategies such as (i) robust reformulations of the original optimization problem (Goldfarb and Iyengar, 2003), (ii) resampled efficient portfolios (Fletcher and Hillier, 2001), (iii) common factor portfolios (Brandt et al., 2009), (iv) norm-constrained portfolios (Jagannathan and Ma, 2003) and (v) shrinkage portfolios.¹

Shrinkage methods have received considerable attention in the recent past because they can be generalized in many dimensions and because their statistical foundations are well understood. Shrinkage methods are proposed either for the moments of the asset returns (Jorion, 1986; Ledoit and Wolf, 2003; Kourtis et al., 2012) or directly for the portfolio weights by combining the estimated optimal portfolios with a fixed reference portfolio (among others Kan and Zhou, 2007; Frahm and Memmel, 2010; Pollak, 2011; Tu and Zhou, 2011). The distinction between imposing norm-constraints on the portfolio weights and shrinkage of the variance-covariance matrix can be regarded as somewhat artificial, because both lead to an identical portfolio allocation in many cases (DeMiguel et al., 2009a; Fan et al., 2012). Examples of L_1 and L_2 norm-constraint portfolios can be found in Brodie et al. (2009), Li (2015), Goto and Xu (2015), Yen (2016).

This study proposes an alternative strategy of portfolio weight shrinkage by means of a Bayesian regression for the global minimum variance portfolio (GMVP). Specifically, we represent the weight deviations of the GMVP from a given reference portfolio as coefficients of a linear regression and shrink them towards zero by means of Bayesian regularization techniques. This allows us to estimate and shrink the portfolio weights directly and to impose restrictions on the weights by an appropriate choice of the prior distribution. As reference portfolio (the shrinkage target) we choose in the following the naive $1/N$ -portfolio because the $1/N$ -strategy has been shown to be very competitive in terms of out-of-sample portfolio performance (DeMiguel et al., 2009b). Thus, our approach incorporates the

¹Brandt (2010) provides a comprehensive review.

GMVP and the $1/N$ -portfolio as borderline cases: While a completely uninformative (flat) prior on the portfolio weights in our regression setting yields the plug-in estimator of the GMVP, a completely informative (deterministic) prior leads to the $1/N$ strategy. In the context of our shrinkage approach, the strong performance of the $1/N$ -strategy in many out-of-sample comparisons can therefore be explained by the fact that the performance gains due to the complete elimination of estimation noise outweigh the losses resulting from relying on a suboptimal portfolio strategy which ignores information on the correlation structures of the asset returns.

The choice of the optimal shrinkage parameter is a crucial problem of any shrinkage estimator. Often this is done via cross-validation or bootstrapping. Alternatively, the shrinkage intensity can be selected by an appropriate criterion linked to the objective function: DeMiguel et al. (2013) for example propose to tune the shrinkage parameter according to portfolio performance measures such as the Sharpe ratio. Despite the performance gains from an optimized shrinkage intensity, choosing a single parameter for the shrinkage combination of N assets is very restrictive. Instead, hierarchical shrinkage priors, which are Bayesian versions of the penalized least squares estimators such as the lasso or ridge, account for different degrees of estimation risk across assets by assigning different posterior variances to each portfolio weight. This translates into asset specific shrinkage intensities. In this way, these priors offer a data-driven method to learn about which assets are particularly important for the portfolio performance and which assets carry greater estimation uncertainty that should lead to stronger shrinkage towards $1/N$. Most importantly, the great advantage of these priors is their applicability in high dimensional settings when the number of assets is very large relative to the sample size. In this for practitioners particularly relevant case, standard frequentist estimation approaches turn out to be infeasible or show a comparatively poor performance.

Another positive feature of our Bayesian regression approach is the great flexibility of specifying a prior distribution directly on the portfolio weights and their deviations from the reference portfolio. This strategy is much more intuitive and easy to interpret in economic terms than imposing restrictions on the first and/or second moments of the asset returns. For example, the prior can be used to reflect the investors' views about the optimal portfolio from a previous allocation to minimize transactions costs or it can be used to incorporate inequality restrictions on the portfolio weights such as no short-selling.

Representing the portfolio choice problem in terms of an estimation problem of a linear regression model is well-known. Britten-Jones (1999) proposes a regression approach for the tangency portfolio and Kempf and Memmel (2006) as well as Fan et al. (2012) show that the plug-in estimator for the GMVP weights can also be obtained by means of a linear regression. More recently, Li (2015) provides a regression representation of the

mean-variance portfolio. Our approach differs from the regression representation mentioned above by (i) directly integrating the shrinkage target in the regression formulation, thus allowing to shrink all N portfolio asset weights simultaneously, and (ii) by using the Bayesian framework as the regularization strategy. In this paper, we focus on the regression representation of the GMVP, but extensions to other portfolio strategies are straightforward.

In the following, we investigate the impact of various penalization strategies and compare different Bayesian and popular frequentist shrinkage methods in an empirical application using US equity returns. We find that most of the portfolios from Bayesian shrinkage approaches reveal superior out-of-sample performance based on various criteria like the out-of-sample portfolio standard deviation, certainty equivalent, the Sharpe ratio or the weight turnover. Especially, hierarchical shrinkage methods such as the Bayesian lasso and Bayesian elastic net, which allow for variable shrinkage intensities among portfolio weights, show strong out-of-sample performance. The results hold for different asset classes and are stronger for high dimensional portfolios.

The paper is organized as follows. Section 2.2 introduces the reference portfolio augmented regression approach for the GMVP and discusses a variety of Bayesian regularization strategies. Section 2.3 summarizes the set-up of our empirical horse race and Section 2.4 presents the empirical findings. Section 2.5 concludes and gives an outlook on further generalizations.

2.2 Penalized regressions and the GMVP

2.2.1 The classical global minimum variance portfolio problem

Consider an investor with N risky financial assets who chooses her portfolio weights such that the risk of her portfolio in terms of its variance is minimized. Let $r_t = (r_{t,1}, \dots, r_{t,N})'$ be the $(N \times 1)$ return vector at time t with mean vector μ and variance-covariance matrix Σ . The vector of GMVP weights is defined as the solution to the minimization problem

$$\omega_g = \arg \min_{\omega, \omega' \iota_N = 1} \omega' \Sigma \omega = \frac{\Sigma^{-1} \iota_N}{\iota_N' \Sigma^{-1} \iota_N}, \quad (2.1)$$

where ω is the N -dimensional weight vector satisfying the summing-up constraint $\omega' \iota_N = 1$ and ι_N is an N -dimensional vector of ones. The minimization problem in (2.1) can be represented equivalently by minimizing the population moment of the quadratic form

$$\omega_g = \arg \min_{\omega, \omega' \iota_N = 1} E [(\omega' (r_t - \mu))^2]. \quad (2.2)$$

If r_t follows an *iid*-process, an unbiased and consistent estimate of Σ is obtained by the sample covariance matrix $\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})(r_t - \bar{r})'$ with $\bar{r} = (\frac{1}{T} \sum_{s=1}^T r_{s,1}, \dots, \frac{1}{T} \sum_{s=1}^T r_{s,N})'$. Replacing Σ in (2.1) by its sample counterpart yields the plug-in estimator for the GMVP weights:

$$\hat{\omega}_g = \frac{\hat{\Sigma}^{-1} \iota_N}{\iota_N' \hat{\Sigma}^{-1} \iota_N}. \quad (2.3)$$

Similarly, the plug-in estimator can be obtained by minimizing the sample counterpart of (2.2):

$$\hat{\omega}_g = \arg \min_{\omega, \omega' \iota_N = 1} \frac{1}{T} \sum_{t=1}^T (\omega'(r_t - \bar{r}))^2. \quad (2.4)$$

Thus, the empirical minimization problem (2.4) turns out to be a classical restricted least squares problem. [Kempf and Memmel \(2006\)](#) also show that the weights of the plug-in estimator can be obtained by least squares estimation of

$$r_{t,N} = \mu_g + \sum_{i=1}^{N-1} \omega_i (r_{t,N} - r_{t,i}) + \varepsilon_t, \quad t = 1, \dots, T. \quad (2.5)$$

The adding-up constraint is imposed in the regression equation such that the weight for the N -th asset, taken here as the reference asset on the left side of the equation, results from $\omega_N = 1 - \sum_{i=1}^{N-1} \omega_i$. In (2.5), the intercept term corresponds to the expected return of the GMVP, $\mu_g = \mu' \omega_g$. By construction, the variance of the regression equation equals the variance of the minimum variance portfolio as $\sigma_\varepsilon^2 = \sigma_g^2 = \omega_g' \Sigma \omega_g$.

Following [Brodie et al. \(2009, eq. 2\)](#), a natural way to impose norm-constraints on the portfolio weights is by defining the minimum variance problem as a penalized least squares problem:

$$\hat{\omega}_g(\lambda) = \arg \min_{\omega, \iota_N' \omega = 1} \frac{1}{T} \sum_{t=1}^T (\omega'(r_t - \bar{r}))^2 + \lambda J(\omega), \quad (2.6)$$

where $\lambda \geq 0$ is the penalty term chosen by the investor (econometrician) and $J(\omega)$ denotes a general penalty function. An obvious choice is the L_2 -norm penalty given by $J(\omega) = \|\omega\|_2^2 = \sum_{i=1}^N \omega_i^2$. It is well-known that L_2 -penalization of the least squares problem leads to the ridge estimator ([Hoerl and Kennard, 1970](#)). However, as λ goes to infinity, estimating (2.6) by the ridge estimator would shrink all portfolio weights towards zero except for the weight of the reference asset, which would converge to one due to the adding-up constraint. The resulting estimates would therefore always depend arbitrarily on the choice of the reference asset.

2.2.2 Reference augmented GMVP regression

A simple way to impose shrinkage on all N portfolio weights simultaneously is to augment the asset space by an additional reference portfolio $r_{t,0}$. Let $\tilde{r}_t = (r_{t,0}, r_{t,1}, \dots, r_{t,N})'$ be the $(N+1)$ -dimensional return vector augmented by the return of the reference portfolio $r_{t,0} = w_0' r_t$, where the N -dimensional weight vector $\omega_0 = (\omega_{1,0}, \dots, \omega_{N,0})'$ is assumed to be given. An obvious benchmark case for such a reference portfolio is the naive $1/N$ -portfolio with $w_0 = \frac{1}{N}\boldsymbol{\iota}_N$.

The investment problem can now be reformulated such that the investor chooses the weights for the N assets and the weight for the reference portfolio, i.e. the investor chooses the weights of the assets such that they optimally deviate from the reference portfolio to minimize risk. The original GMVP optimization problem from (2.1) in terms of the reference augmented asset space with $N+1$ portfolio weights is given by

$$\tilde{\omega}_g = \arg \min_{\tilde{\omega}, \iota_{N+1}' \tilde{\omega} = 1} \tilde{\omega}' \tilde{\Sigma} \tilde{\omega} = \frac{\tilde{\Sigma}^+ \iota_{N+1}}{\iota_{N+1}' \tilde{\Sigma}^+ \iota_{N+1}},$$

where $\tilde{\Sigma}^+$ denotes the Moore-Penrose pseudo-inverse of $V[\tilde{r}_t] = \tilde{\Sigma}$, which has to be used here because of the singularity of $\tilde{\Sigma}$ since $r_{t,0}$ is a linear combination of the original return vector. Similar to (2.5), the weights $\tilde{\omega}_i$ represent regression parameters of a regression of $r_{t,0}$ on the deviations $r_{t,0} - r_{t,i}$ for $i = 1, \dots, N$, such that the adding-up constraint for the reference portfolio and the original assets holds.

Obviously the risk minima for the N -dimensional asset space and the index augmented asset space coincide, i.e. $\sigma_g^2(w) = \sigma_g^2(\tilde{w})$, as by index-augmentation no additional risk reduction can be achieved. However, we will see in the following that by regressing the reference return on the vector of individual asset return deviations from $r_{t,0}$, the regression coefficients now correspond to (standardized) weight deviations of the GMVP from the reference portfolio. The new augmented regression formulation is given by

$$r_{t,0} = \tilde{\mu}_g + \sum_{i=1}^N \tilde{\omega}_i (r_{t,0} - r_{t,i}) + \tilde{\varepsilon}_t, \quad t = 1, \dots, T, \quad (2.7)$$

where $\tilde{\omega}_i$ are the weight deviations in the index regression and $\tilde{\omega}_0 = 1 - \sum_{i=1}^N \tilde{\omega}_i$. Using this adding-up constraint for the index augmented model and the original model, i.e. $\sum_{i=0}^N \tilde{\omega}_i = \sum_{i=1}^N \omega_i = 1$, the overall portfolio return from all $N+1$ assets at time t is given then

by

$$\begin{aligned}
 r_{t,p} &= \tilde{\omega}' \tilde{r}_t = \tilde{\omega}_0 \cdot r_{t,0} + \sum_{i=1}^N \tilde{\omega}_i r_{t,i} \\
 &= \left(1 - \sum_{j=1}^N \tilde{\omega}_j\right) r_{t,0} + \sum_{i=1}^N \tilde{\omega}_i r_{t,i} \\
 &= \sum_{i=1}^N \omega_{i,0} r_{t,i} - \sum_{i=1}^N \omega_{i,0} r_{t,i} \sum_{j=1}^N \tilde{\omega}_j + \sum_{i=1}^N \tilde{\omega}_i r_{t,i} \\
 &= \sum_{i=1}^N \underbrace{\left(\tilde{\omega}_i + \omega_{i,0} \left(1 - \sum_{j=1}^N \tilde{\omega}_j\right)\right)}_{=: \omega_i} r_{t,i} = \omega' r_t,
 \end{aligned}$$

where the last equality results from the fact that the same portfolio return can be obtained by directly investing in all N original assets. Therefore, the original weights of the GMVP are related to the weights of the index augmented model as follows:

$$\omega_i = \tilde{\omega}_i + \omega_{i,0} \left(1 - \sum_{j=1}^N \tilde{\omega}_j\right) = \tilde{\omega}_i + \omega_{i,0} \cdot \tilde{\omega}_0, \quad \forall i = 1, \dots, N \quad (2.8)$$

and as $\omega_i = \tilde{\omega}_i + \frac{1}{N} - \frac{1}{N} \sum_{j=1}^N \tilde{\omega}_j \forall i$ when the equally weighted portfolio is used as the reference. Once estimates for $\tilde{\omega}_i$ are available by estimating (2.7), the actual portfolio weights ω_i are easily calculated using (2.8).

2.2.3 Regularization strategies for the reference augmented GMVP regression

By applying a regularization strategy to the coefficients in equation (2.7), we now shrink all weights of the N original assets. Equation (2.8) guarantees that the adding-up constraint for the original weights is satisfied without requiring to select a specific asset as the reference asset as assumed in (2.5). For the index augmented representation in (2.7), an L_2 -penalization is also feasible. In this case, the singularity of $\tilde{\Sigma}$ is irrelevant. For the limiting case of penalization ($\lambda \rightarrow \infty$) the estimates for $\tilde{\omega}_i$ vanish and the index augmented regression model produces the reference portfolio. For the case of no penalization ($\lambda = 0$) the OLS estimator yields the plug-in GMVP estimates. Therefore, the L_2 -penalized index augmented GMVP nests the conventional GMVP and the reference portfolio as special cases.

We propose to estimate the portfolio weights in (2.7) within a Bayesian regression

framework for several reasons: First, there exists a close relationship between frequentist shrinkage approaches and Bayesian estimation methods. In particular, specific prior choices yield posterior estimates from a Bayesian regression that coincide with frequentist solutions and can therefore be interpreted as the Bayesian analogues to frequentist shrinkage approaches (see e.g. [Fahrmeir et al. \(2010\)](#) for a unifying perspective). Second, the use of a prior distribution for the portfolio weights as a shrinkage device provides great flexibility. While shrinkage can be obtained through different hyperparameters, the impact of their choice can be diminished either by means of an empirical Bayes approach or by means of hierarchical modeling, avoiding an arbitrary choice of the penalization parameter by using the data to determine the optimal shrinkage intensity. The choice of the prior distribution also allows to incorporate a range of additional constraints on the investment strategy, for example no short-selling. While certain priors may not yield analytical posterior results, numerical results can be easily obtained by Markov Chain Monte Carlo (MCMC) methods. Obtaining a whole posterior distribution instead of having only an optimal point estimate for the portfolio weights has several advantages. For example, it can be used for hypothesis testing about the optimal investment strategy or to quantify estimation uncertainty.

Note that contrary to other shrinkage methods in the literature (e.g. [Brodie et al., 2009](#)) our approach is defined in terms of shrinking the deviations from the reference portfolio towards zero. Therefore, it is reasonable to assume zero means for the priors of the regression coefficients $\tilde{\omega}$ in the augmented model (2.7). Then, only the variance of the prior determines the magnitude of deviations from the reference portfolio. By choosing the mean of the prior such that the reference portfolio serves as the benchmark in our Bayesian framework also has an intuitive theoretic explanation: In order to decide in favor of a deviation from the reference portfolio, the investor must obtain empirical evidence of its inferiority. This evidence results from the data in form of the likelihood and leads to a revision of the benchmark choice reflected in the posterior estimates.

2.2.4 Choosing the Prior

We now describe different prior choices and their implications for the optimal portfolio weights. We will consider priors with unbounded support leaving the domain of the posterior estimates of the portfolio weights unrestricted and priors with bounded support such that the domain of weight estimates is restricted.

Note that the prior is specified for the weight deviation from the reference portfolio. There are no domain and summing-up restrictions on the weight deviations imposed by the prior, because equation (2.8) ensures proper portfolio weights ex-ante. Therefore, we can use a multivariate normal setting for the augmented regression model and avoid the use of special prior distributions such as the Dirichlet distribution.

i. Ridge type priors

A standard choice for the prior in linear regression models is the combination of a normal distribution for the regression parameters and an inverse-gamma distribution for the error term variance $\tilde{\sigma}^2$, because it can produce uninformative as well as very subjective priors depending on the prior hyperparameter choices (e.g. Bauwens and Korobilis, 2013), i.e.

$$\tilde{\omega} | \tilde{\sigma}^2 \sim \mathcal{N}(0, \tilde{\sigma}^2 \underline{V}) \quad \text{and} \quad \tilde{\sigma}^2 \sim \mathcal{IG}(\underline{\nu}, \underline{s}), \quad (2.9)$$

where the zero mean of the normal prior above results from centering the weight deviations around zero such that the weights are centered around the $1/N$ weights (the reference portfolio). Under normality of the likelihood, the posterior mean is of the form

$$\bar{\omega} = (\underline{V}^{-1} + X'X)^{-1} X'y,$$

where y is the vector of the index returns and X the matrix of explanatory variables consisting of the return differences from (2.7) centered around their mean. Here, \underline{V} determines the extent of shrinkage towards the naive portfolio. A common choice is $\underline{V} = \tau I_N$, yielding the ridge regression prior. This only requires to specify one hyperparameter τ instead of $N(N+1)/2$ prior variance parameters leading to a posterior mean for $\tilde{\omega}$ of the form equivalent to the classical ridge regression estimator:

$$\bar{\omega} = \left(X'X + \frac{1}{\tau} I_N \right)^{-1} X'y. \quad (2.10)$$

We use four different specifications for τ in the ridge regression in the empirical application:

1. *Naive $1/N$ portfolio*: We use the naive $1/N$ portfolio as the benchmark strategy. It is obtained with a deterministic prior using $\underline{V} = \tau I_N$, for $\tau \rightarrow 0$.
2. *GMVP*: The classical plug-in estimator for the GMVP serves as a second benchmark. It is obtained by using a non-informative prior with $\underline{V} = \tau I_N$, for $\tau \rightarrow \infty$.
3. *Lindley-Smith ridge prior*: A data-driven choice of the shrinkage parameter proposed by Lindley and Smith (1972) takes the form

$$\tau_{LS} = \left(\frac{(T - N) \cdot (N + 2)}{(T + 2)} \cdot \frac{\hat{\sigma}^2}{\hat{\omega}' \hat{\omega}} \right)^{-1},$$

with $\hat{\sigma}^2 = (y - X\hat{\omega})'(y - X\hat{\omega})/T$ and $\hat{\omega}$ is the OLS estimate of (2.7). For $N < T$, this shrinkage parameter corrects for non-orthogonality in the data and avoids an arbitrary choice of τ by allowing the data to estimate it. For $N \rightarrow T$ from below, the

shrinkage increases towards the naive $1/N$ portfolio. This prior is only applicable for $N < T$.

4. *Dimensionality-dependent shrinkage prior:* A simple way to account for the necessity to reduce shrinkage with increasing sample size and to enhance shrinkage with a larger portfolio dimension that avoids additional estimation noise by a data driven choice of the shrinkage parameter is given by

$$\tau_{\frac{T}{N}} = \frac{T}{N}. \quad (2.11)$$

When plugging (2.11) into equation (2.10), the larger N is relative to T , the shrinkage increases towards the naive $1/N$ portfolio. This prior is parameter free and also applicable when the number of observations is smaller than the number of assets.

ii. Empirical Bayes priors

While the ridge regression prior implies no correlation between the individual weights, which might be specifically unrealistic in the context of portfolio optimization, the empirical Bayes or g -prior replaces the identity matrix with the information matrix of the data in equation (2.10), i.e. $\underline{V} = \tau(X'X)^{-1}$. Although this prior contradicts Bayes' theorem as it depends on the data, this property might be beneficial for the investor who wants to use market information to decide on the deviation from the reference portfolio. The posterior mean reflects some form of James-Stein-type of shrinkage of the plug-in estimator of the GMVP weights:

$$\bar{\omega} = \frac{\tau}{\tau + 1} \hat{\omega}. \quad (2.12)$$

Again, $\tau \rightarrow \infty$ yields the empirical GMVP estimates and $\tau \rightarrow 0$ implies $\tilde{\omega} = 0$, producing the naive $1/N$ -portfolio. In our empirical application we use two different empirical Bayes specifications for τ :

5. *Judge-Brock empirical Bayes prior:* A convenient prior for empirical Bayes estimators of linear models is proposed in [Judge and Bock \(1978\)](#). This prior is easy to implement and leads in our application to a stronger penalization compared to the Lindley-Smith prior:

$$\tau_{JB} = \frac{(y - X\hat{\omega})'(y - X\hat{\omega})}{T}, \quad \xi^2 = \frac{\hat{\omega}'\hat{\omega}}{\text{tr}((X'X)^{-1})} - \hat{\sigma}^2,$$

with $\hat{\sigma}^2 = (y - X\hat{\omega})'(y - X\hat{\omega})/T$ and $\hat{\omega}$ is the OLS estimate of (2.7).

6. *Dimensionality-dependent empirical Bayes shrinkage prior:* This shrinkage strategy is similar to the dimensionality-dependent shrinkage strategy described for ridging.

We simply apply the shrinkage parameter $\tau_{\frac{T}{N}} = \frac{T}{N}$ from (2.11) to (2.12). Since $\lim_{T \rightarrow \infty} \tau_{\frac{T}{N}} / (\tau_{\frac{T}{N}} + 1) = 1$ and $\lim_{N \rightarrow \infty} \tau_{\frac{T}{N}} / (\tau_{\frac{T}{N}} + 1) = 0$, the shrinkage strategy operates in the same way as in the ridging case.

iii. Flexible shrinkage priors

The ridge prior and the empirical Bayes prior both depend on a single shrinkage intensity τ for all $\tilde{\omega}$'s. While this choice is arbitrary, assigning individual shrinkage intensities τ_i 's for each weight (asset) can be achieved by drawing the τ_i 's from a common hyperprior distribution within a Gibbs sampler. As this adds a second layer to the model, the following approaches can be labeled as *hierarchical shrinkage priors*. Here, the investor avoids a subjective parameter choice and uses instead the data through the hyperprior distribution to determine asset specific shrinkage intensities. In this sense, it can also be seen as a Bayesian alternative to the multivariate or adaptive shrinkage idea of [Golosnoy and Okhrin \(2007\)](#).

7. *Hierarchical ridge*: The prior for each coefficient is given by $\tilde{\omega}_i \sim \mathcal{N}(0, \tau_i)$ for $i = 1, \dots, N$. Using $\tau_i \sim \mathcal{IG}(q_1, q_2) \forall i$ with $q_1 = q_2 = 0.0001$ establishes a flat prior for each τ_i . By weight specific prior variances, hierarchical ridging allows for weight specific penalization and corresponds to the generalized ridge estimator. The Gibbs sampler for this approach is taken from [Bauwens and Korobilis \(2013\)](#):

- (i) Draw $\tau_i | \tilde{\omega}_i$ from $\mathcal{IG}(q_1 + 1, q_2 + \tilde{\omega}_i^2)$.
- (ii) Draw $\tilde{\sigma}^2 | \tilde{\omega}, X$ from $\mathcal{IG}(T, (y - X\tilde{\omega})'(y - X\tilde{\omega}))$.
- (iii) Draw $\tilde{\omega} | \tau_1, \dots, \tau_N, \tilde{\sigma}^2, X$ from $\mathcal{N}(\bar{\tilde{\omega}}, \bar{V})$, where $\bar{V} = (\underline{V}^{-1} + \tilde{\sigma}^{-2} X'X)^{-1}$ with $\underline{V} = \text{diag}(\tau_1, \dots, \tau_N)$ and $\bar{\tilde{\omega}} = \bar{V}X'y$.

8. *Bayesian lasso*: The L_2 -penalization of the ridge regression stabilizes the portfolio weights but leaves the portfolio dimension unchanged. [Brodie et al. \(2009\)](#) show that an active subset of assets from a given portfolio universe can be selected by imposing a L_1 -norm penalization (the "lasso") on the portfolio weights. The lasso idea goes back to [Tibshirani \(1996\)](#) who proposes an L_1 -norm penalization for the coefficients in a linear regression model. Applied to our regression model (2.7) the objective function for an L_1 -norm penalization takes the form

$$\arg \min_{\tilde{\omega}} \sum_{t=1}^T (y_t - X'_t \tilde{\omega})^2 + \lambda \sum_{i=1}^N |\tilde{\omega}_i|, \quad (2.13)$$

such that the lasso automatically restricts a subset of the portfolio weights to $1/N$. The Bayesian analog to the frequentist lasso was proposed by [Park and Casella](#)

(2008), who note that the lasso estimator is equivalent to the posterior mode of the Bayes estimate for a Laplace prior given by

$$p(\tilde{\omega}|\tilde{\sigma}^2) \sim \prod_{i=1}^N \frac{\lambda}{2\sqrt{\tilde{\sigma}^2}} \exp\left(-\frac{\lambda}{2\sqrt{\tilde{\sigma}^2}}|\tilde{\omega}_i|\right).$$

The Laplace density can be written as a mixture of normal distributions which allows for a simple Gibbs sampler representation for the model.² The complete model is given by

$$\begin{aligned} y_t|X_t, \tilde{\omega}, \tilde{\sigma}^2 &\sim \mathcal{N}(X'_t \tilde{\omega}, \tilde{\sigma}^2 I_T), \\ \tilde{\omega}|\tilde{\sigma}^2, \tau_1^2, \dots, \tau_N^2 &\sim \mathcal{N}(0, \tilde{\sigma}^2 D_\tau), \quad D_\tau = \text{diag}(\tau_1^2, \dots, \tau_N^2), \\ \tau_1^2, \dots, \tau_N^2 &\sim \prod_{i=1}^N \frac{\lambda^2}{2} \exp\left(-\frac{\lambda\tau_i^2}{2}\right) d\tau_i^2, \end{aligned}$$

where $\tilde{\sigma}^2 > 0$ and $\tau_i^2 > 0 \forall i$ and $\tilde{\sigma}^2 \propto 1/\tilde{\sigma}^2$. Note that the Bayesian lasso allows for weight specific shrinkage through the prior parameters τ_i^2 , while overall shrinkage is determined by the hyperparameter λ . These weight specific priors offer a data-driven method to learn about the relevance of specific assets for the portfolio performance and do not solely rely on a single value penalty parameter for all N assets. In contrast to this, the hyperparameter λ is in charge of the overall extent of shrinkage, such that the lasso shrinks the weight deviations to zero to obtain $1/N$ -weights in the limiting case.

9. *Bayesian elastic net*: In the context of portfolio optimization the elastic net regularization (Zou and Hastie, 2005) is an attractive generalization of the conventional lasso as it combines ridging and lassoing. It can therefore account for both, high collinearity of the asset returns as well as for sparse portfolio selection. For the frequentist case, the lasso objective function (2.13) is augmented by an additional L_2 -norm penalty term:

$$\arg \min_{\tilde{\omega}} \sum_{t=1}^T (y_t - X'_t \tilde{\omega})^2 + \lambda_1 \sum_{i=1}^N |\tilde{\omega}_i| + \lambda_2 \sum_{i=1}^N \tilde{\omega}_i^2. \quad (2.14)$$

²Note that the conditioning on $\tilde{\sigma}^2$ ensures an unimodal posterior distribution.

The Bayesian set-up for the elastic net is given by

$$y_t | X_t, \tilde{\omega}, \tilde{\sigma}^2 \sim \mathcal{N}(X'_t \tilde{\omega}, \tilde{\sigma}^2 \mathbf{I}_T), \\ \tilde{\omega} | \tilde{\sigma}^2, \tau_1^2, \dots, \tau_N^2 \sim \mathcal{N}(0, \tilde{\sigma}^2 \mathbf{D}_\tau), \quad \mathbf{D}_\tau = \text{diag}((\tau_1^{-2} + \lambda_2)^{-1}, \dots, (\tau_N^{-2} + \lambda_2)^{-1}) \\ \tau_1^2, \dots, \tau_N^2 \sim \prod_{i=1}^N \frac{\lambda_1^2}{2} \exp\left(-\frac{\lambda_1 \tau_i^2}{2}\right) d\tau_i^2,$$

where $\tilde{\sigma}^2 > 0$, $\tau_i^2 > 0 \forall i$ and $\tilde{\sigma}^2 \propto 1/\tilde{\sigma}^2$. Using the frequentist approach, [Yen \(2016\)](#) shows that a weighted L_1 - and L_2 -norm penalization of the GMVP weights significantly improves the out-of-sample portfolio performance.

10. *Bayesian lasso of turnover:* The L_1 -penalization for deviations from zero in the lasso can easily be generalized to a penalization for deviations from any real value δ_i :

$$\arg \min_{\tilde{\omega}} \sum_{t=1}^T (y_t - X'_t \tilde{\omega})^2 + \lambda \sum_{i=1}^N |\tilde{\omega}_i - \delta_i|, \quad (2.15)$$

where we set δ_i to the estimated portfolio weight deviations from the reference portfolio in the previous period between asset i and $1/N$. The idea here is to minimize transactions for a long-term investor and to reduce the periodical weight turnover.

11. *Bayesian elastic net of turnover:* The Bayesian elastic net of turnover penalizes the deviation from the previous allocation (or from any given real value $\delta_i \neq 0$) even stronger through an additional L_2 penalty:

$$\arg \min_{\tilde{\omega}} \sum_{t=1}^T (y_t - X'_t \tilde{\omega})^2 + \lambda_1 \sum_{i=1}^N |\tilde{\omega}_i - \delta_i| + \lambda_2 \sum_{i=1}^N (\tilde{\omega}_i - \delta_i)^2, \quad (2.16)$$

Again, we set the δ_i 's to the weight deviations of the previous period, which leads to a greater penalization when the deviations from the reference portfolio changes a lot from one period to the next. This reduces long-term transaction costs for the investor by avoiding too frequent periodical weight turnover resulting from instable estimates due to the randomness of the estimated portfolio weights. The regularization of the estimated portfolio weights is threefold. The lasso part stabilizes the weights across time, while ridging alleviates the impact of strong collinearities between returns on the portfolio stability. By choosing $\delta_i = 0$ for the estimates of the first period, the estimated weights are also anchored to the weights of the reference portfolio.

For the models introduced in this section analytical expressions for the posterior are

infeasible. But the posteriors can easily be estimated using a Gibbs sampler. Details about the algorithms can be found e.g. in [Kyung et al. \(2010\)](#) and [Korobilis \(2013\)](#).

The posterior estimates from the Bayesian regularization approaches never set portfolio weights exactly to zero similar to the ridge, but they yield posterior estimates very close to zero in many cases. Sparse portfolios can be obtained by using credibility intervals around the posterior mean or mode estimates to decide about deleting assets from the optimal portfolio.

iv. Priors with bounded support

Within our Bayesian framework, priors with bounded support allow us to restrict the range of the portfolio weight posterior estimates ex-ante. An obvious choice of priors with bounded support is the Beta-distribution or the truncated Normal distribution. For example, by choosing appropriate hyperparameters for the prior, short-selling constraints can be easily implemented.

In addition, a prior with bounded support allows us to impose inequality constraints such as $a < H\tilde{\omega} < b$ for a given non-singular matrix H . The interval $[a, b]$ determines the degree of variability in the portfolio composition a priori. The tighter the interval is around zero, the more we shrink the weights towards the reference. Here, we apply a zero-centered truncated Normal distribution for each weight deviation. [Geweke \(1996\)](#) proposes an algorithm to sample each $\gamma_i = \{H\tilde{\omega}\}_{(i)}$ conditional on all other $\{\gamma \setminus \gamma_i\}$ individually.³ As we can transform the weight deviations from the reference portfolio $\tilde{\omega}$ in a linear fashion to the actual portfolio weights, this approach allows for a general portfolio optimization under any linear restrictions.

The crucial choice here is how to set the bounds of the interval, because they determine the extent of shrinkage and hence possible performance improvements. For our horse race comparison below we use a no short-sale variant:

12. *Independent Truncated Normal prior on $[-1/N, 1]$:* This truncation allows no short selling when the naive $1/N$ portfolio is used as the reference portfolio.

Other criteria than the no short-selling constraint can be used to restrict the range of the posterior estimates. For example, the bounds can be linked to the objective of the investor such as the out-of-sample certainty equivalent (e.g. [Okhrin and Schmid, 2007](#), eq. 9). The choice of the bounds can also be motivated by minimizing any given information criterion such as the Bayesian information criteria (BIC).

³The density function of the truncated normal distribution is given by $f(\tilde{\omega}_i | a, b, \mu, \tilde{\sigma}) = \frac{\frac{1}{\sigma} \phi\left(\frac{\tilde{\omega}_i - \mu}{\tilde{\sigma}}\right)}{\Phi\left(\frac{b - \mu}{\tilde{\sigma}}\right) - \Phi\left(\frac{a - \mu}{\tilde{\sigma}}\right)}$ for all $\tilde{\omega}_i \in [a, b]$ and $f(\tilde{\omega}_i | a, b, \mu, \tilde{\sigma}) = 0$ otherwise.

2.3 Data and horse race set-up

In what follows, we try to discern the impact of the different shrinkage techniques presented in Section 2.2 on the optimal portfolio composition and performance. We will compare them to several portfolio shrinkage techniques from the literature and evaluate their performance in an out-of-sample horse race for various empirical data sets and for different portfolio dimensions. For each set-up, we chose the equally weighted $1/N$ portfolio as the reference portfolio in the augmented regression in (2.7).

2.3.1 Data and investment set-up

Our first empirical application is based on the return data provided on the website of Kenneth R. French.⁴ We consider five data sets with different numbers of assets (N): 5 and 30 industry portfolios; 6, 25 and 100 portfolios formed on size and book-to-market. We investigate monthly average value-weighted returns. The time range of our selected data sets is from January 1953 to December 2015. Our analysis is based on excess returns which are obtained by subtracting the corresponding one-month T-Bill rate from the asset returns, which is also available on the website.

As a robustness check, we investigate the performance of our novel Bayesian regularization strategies also using monthly US equity data from Thomson Reuters Datastream. For this, at each point in time from 01/2001 to 12/2015 we select 500 random stocks from the US market constituents list⁵ that have a complete return history over the last 60 months (5 years) and a return future over the next five years.⁶ From these 500 assets, we consider portfolio sizes of 5, 25, 50, 100, 250 and 500. The different equity data sets differ only in their portfolio dimension and are more homogeneous in terms of their construction compared to the different Fama-French portfolios. Therefore, they are more appropriate for making any inference on the effect of portfolio dimension on the portfolio performance. Moreover, the Datastream equity data allows us to perform the analysis for high-dimensional settings in which the portfolio dimension strongly exceeds the sample size.

We apply a rolling window approach with $h = 60$ months (5 years) of data for the estimation in each step, i.e. at time t , we use the last 60 data points from $t - 59$ until t to obtain the estimates and the corresponding portfolio weights. Parallel to re-estimation within the rolling window we rebalanced all portfolios every month. In general, the

⁴See <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html> for further details about the construction of the data sets.

⁵The Datastream code (mnemonic) is for the constituents list is LTOTMKUS.

⁶We adopt this 'forward-looking' restriction from [Ledoit and Wolf \(2014\)](#). Although it is not a requirement for a real life investment experiment, it is standard practices in the related literature for out-of-sample portfolio evaluation.

empirical exercise should not be viewed as an investment experiment with a rolling window optimization set-up for one single investor with no memory, but rather as a strategy for a group of investors with the same strategy who enter the market at different points in time. It is therefore a mean comparison of investment strategies for all investors.

To construct the out-of-sample portfolio returns, we use posterior means of the weight deviations for models 1.-6. and 12. For the Bayesian lasso variants we use the posterior mode results. We then obtain a series of $(T - h)$ out-of-sample portfolio returns.⁷ For the Datastream equity data, we use gross out-of-sample returns for the evaluation. This is so because the portfolio constituents change at each point of the estimation. For the Fama/French portfolio data, we calculate net returns after transaction costs. These are obtained from the wealth process for strategy s given by

$$W_{t+1}(s) = W_t(s) (1 + r_{t,p}(s)) \left(1 - c \cdot \sum_{j=1}^N |\hat{\omega}_{t+1,j}(s) - \hat{\omega}_{t+,j}(s)|\right), \quad (2.17)$$

where $\hat{\omega}_{t,j}(s)$ is the optimal weight at time t for asset j given strategy s and $\hat{\omega}_{t+,j}(s)$ is the actual portfolio weight before rebalancing in $t + 1$. We start with an initial capital of $W_1(s) = 1$ and set the proportional transactions cost c equal to 50 basis points per transaction as assumed e.g. in DeMiguel et al. (2009b). We then calculate simple returns from $W_{t+1}(s)$ and use these to calculate the out-of-sample mean and variance for the portfolio evaluation.

2.3.2 Evaluation criteria

For each strategy, $s = 1, \dots, 17$, we report various performance criteria using the out-of-sample portfolio returns of each strategy:

- a. *Mean out-of-sample portfolio return* ($\hat{\mu}$): $\hat{\mu}(s) = \frac{1}{T-h} \sum_{t=h+1}^T r_{t,p}(s)$.
- b. *Standard deviation* ($\hat{\sigma}$): $\hat{\sigma}(s) = \sqrt{\frac{1}{T-h-1} \sum_{t=h+1}^T (r_{t,p}(s) - \hat{\mu}(s))^2}$. The out-of-sample standard deviation is the objective function of a GMVP investor.
- c. *Sharpe ratio* (SR): The Sharpe ratio is given by $\widehat{\text{SR}}(s) = \hat{\mu}(s)/\hat{\sigma}(s)$.
- d. *Certainty equivalent* (CE): The CE is defined as $\widehat{\text{CE}}(s) = \hat{\mu}(s) - \frac{\gamma}{2} \hat{\sigma}^2(s)$, where γ is the risk aversion of the investor. Following DeMiguel et al. (2009b) we set $\gamma = 1$. While the GMVP is theoretically only optimal for a completely risk averse investor, evaluating the performance of different strategies under a loss function for an investor

⁷All calculations were performed with MATLAB 2015a. The program code is available upon request from the author.

who is willing to take some risks provides valuable information on the robustness of each strategy.

- e. *Return loss (RL)*: We define the return loss as in DeMiguel et al. (2009b), who calculate the lost portfolio return relative to the naive $1/N$ portfolio due to transaction costs. The return loss is defined as the additional return needed for strategy s to perform as well as the $1/N$ strategy in terms of the Sharpe ratio, i.e. return loss(s) = $\frac{\hat{\mu}(\frac{1}{N})}{\hat{\sigma}(\frac{1}{N})} \cdot \hat{\sigma}(s) - \hat{\mu}(s)$.

We also analyze the portfolio weights and provide some summary statistics. We compute the following criteria at each point in time and then report mean values for each strategy over time:

- f. *Turnover (TO)*: The turnover is a measure to investigate the required trading of each strategy. This is important to the investor, because every trade imposes some trading costs. It is $\widehat{\text{TO}}(s) = \frac{1}{T-h} \sum_{t=h+1}^T \sum_{j=1}^N (|\hat{\omega}_{t+1,j}(s) - \hat{\omega}_{t,j}(s)|)$, where $\hat{\omega}_{t,j}(s)$ is the j -th asset's optimal weight at time t given strategy s and $\hat{\omega}_{t,j}(s)$ is again the actual portfolio weight before rebalancing in $t+1$. For example, assume there are two stocks only, both are worth 100. The investor diversifies equally and puts 50% of her wealth in both stocks. If in the next period asset one increases by 10% and asset two decreases by 10%, the value of the portfolio will remain 100, but the actual weights of the portfolio are 0.55 for asset one and 0.45 for asset two. Hence, to obtain an equally weighted portfolio again, trading is required.
- g. *Min*: minimum portfolio weight
- h. *Max*: maximum portfolio weight
- i. *Mean absolute deviation from $1/N$ (MAD)*: $\text{MAD}_t(s) = \sum_{j=1}^N |\hat{\omega}_{t,j}(s) - \frac{1}{N}|$.

2.3.3 Competing frequentist strategies

Finally, we compare the Bayesian shrinkage strategies introduced in Section 2.2 with five popular shrinkage portfolio strategies from the literature.

i. Variance-covariance shrinkage strategies

- 13. *Ledoit and Wolf (2003)* propose to shrink the variance-covariance matrix of the asset returns towards a single-index covariance matrix. This yields

$$\hat{\Sigma}_{LW} = \frac{\kappa}{T} \hat{F} + \left(1 - \frac{\kappa}{T}\right) \hat{\Sigma},$$

where \hat{F} is an estimator for the covariance matrix from a single market index model. The shrinkage intensity is of the form $\kappa = \frac{p-r}{c}$, where p measures the error on the sample covariance matrix and c accounts for the misspecification of the single-index model. r measures the covariance between the estimation errors of \hat{F} and $\hat{\Sigma}$.

14. *Kourtis et al. (2012)* propose a direct shrinkage method for the inverse of the variance-covariance matrix of the asset returns, because this is the actual ingredient in the optimal GMVP weights in equation (2.3). It is given by

$$\hat{\Sigma}_K^{-1} = c_1 \hat{\Sigma}^{-1} + c_2 \hat{F}^{-1},$$

where \hat{F} is again a single market-index covariance matrix and c_1 and c_2 are chosen to minimize the out-of-sample variance via cross-validation.

ii. Weight shrinkage strategies

Rather similar in spirit to our approach are the next two weight shrinkage strategies which combine the naive portfolio with the weights of the GMVP.

15. *Frahm and Memmel (2010)* propose a portfolio strategy that minimizes the out-of-sample portfolio variance. It is given by

$$\hat{\omega}_{FM} = \kappa \omega_{eq} + (1 - \kappa) \hat{\omega},$$

where $\kappa = \min(\kappa_s, 1)$ and $\kappa_s = \frac{N-3}{T-N+2} \cdot \frac{1}{\hat{\tau}}$ and $\hat{\tau} = \frac{\omega'_{eq} \hat{\Sigma} \omega_{eq} - \hat{\omega}' \hat{\Sigma} \hat{\omega}}{\omega'_{eq} \hat{\Sigma} \omega_{eq}}$ is the estimated relative loss of the equally weighted portfolio. Since this portfolio is a convex linear combination of the $1/N$ portfolio and the GMVP based on the sample variance-covariance matrix $\hat{\Sigma}$, it is not feasible for high dimensional portfolios when $N > T$.

16. *Pollak (2011)* investigates the statistical difference between the naive portfolio and the GMVP and proposes

$$\hat{\omega}_P = \omega_{eq} g(D(\hat{\omega}, \omega_{eq})) + \hat{\omega}(1 - g(D(\hat{\omega}, \omega_{eq}))),$$

where $g(x) = 1/(1 + b \cdot x)$ and $D(\hat{\omega}, \omega_{eq}) = |\hat{\omega} - \omega_{eq}|/s$ and s is chosen to be the largest sample standard deviation of the weights obtained by a bootstrap procedure. While *Pollak (2011)* chooses $b = 0.5$, we choose b as the value between zero and ten that minimizes the out-of-sample portfolio variance given in (e.g. *Okhrin and Schmid, 2007*, eq. 9): $V[\hat{\omega}'_P r_{t+1}] = \text{tr} [\hat{\Sigma} \cdot V[\omega_P]] + \hat{\mu}' V[\omega_P] \hat{\mu} + E[\omega_P]' \hat{\Sigma} E[\omega_P]$, which incorporates the uncertainty in $\hat{\omega}_P$. Here, the function $g(\cdot)$ measures the 'distance' between the portfolios, it goes to one if they are statistically non-distinguishable,

i.e. when the norm between the two approaches zero, and it approaches zero if the opposite is true.

iii. Factor model strategy

Factor models are commonly used in the finance literature. Here, we use the 3-factor model by Fama and French (1993) to estimate the variance-covariance matrix.

17. The *Fama-French 3-factor model* is defined by

$$r_{t,i} = \alpha_1(r_{t,m} - r_{t,f}) + \alpha_2 \text{SMB}_t + \alpha_3 \text{HML}_t + \varepsilon_t, \quad (2.18)$$

where $(r_{t,m} - r_{t,f})$ describes the excess return of the market over the risk free rate, SMB_t is composed as the average returns on three small portfolios minus the average returns on three big portfolios. In particular, it defines a zero-cost portfolio that is long in stocks with a small market capitalization and short in stocks with a large market capitalization. HML_t comprises a zero-cost portfolio that is long in stocks with a high book-to-market value and short in low book-to market stocks. The Fama-French factors are constructed using the 6 value-weighted portfolios formed on size and book-to-market.⁸ Writing equation (2.18) in matrix form yields $R = F \cdot A + \varepsilon$ with $F = ((r_m - r_f), \text{SMB}, \text{HML})'$ and $A = (\alpha_1, \alpha_2, \alpha_3)'$. The estimator for the variance-covariance matrix for the 3-factor model is defined by

$$\hat{\Sigma}_{3FF} = \hat{A} \hat{\Sigma}_F \hat{A}' + \hat{D},$$

where $\hat{\Sigma}_F$ denotes the sample covariance matrix of the three factors and \hat{D} represents a diagonal matrix that contains the estimated variances of the variances of OLS residuals. The portfolio weights for the 3-factor model are obtained by using $\hat{\Sigma}_{3FF}$ as a plug-in estimator for the GMVP. Note that among the 17 models, the three-factor model is the only one which incorporates additional information on covariates/factors to determine the empirical portfolio weights.

2.4 Empirical findings

The results for our out-of-sample forecasting horse race based on the corresponding entire available sample size are given in Tables 2.1 - 2.11 of the Appendix. In order to check the robustness of our findings, we performed in addition the same analysis for two subsamples.

⁸A more detailed definition of the factors can be found http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

These additional results for the sub-period 1/2001 - 1/2006, representing market conditions with comparatively high average return and low volatility and the sub-period 1/2006 - 1/2012, capturing more turbulent times with low return rates and high volatilities are relegated to a separated web-appendix.

In general, the performance is better for the French/Fama (FF) data sets than for the stock portfolios, which can be explained by the random choice of the included assets. However, we do not see any qualitative differences in our findings depending on the nature of the data (FF-portfolios sorted on size and book-to-market, FF-industry portfolios, equity portfolios). For the sake of brevity, we refrain from discussing the findings w.r.t. the nature of the data explicitly.

Generally, we can state that no portfolio rule outperforms any other strategy significantly on all types of data and for all portfolio dimensions and time horizons considered. But we find clear performance variations across different portfolio dimensions. Relative to our estimation window of $h = 60$ we distinguish between small portfolios, $N = 5$ and $N = 6$, large portfolios with $25 \leq N \leq 50 < h = 60$, where the degrees of freedom are small but estimation by conventional methods is still feasible, and high-dimensional portfolios with $h = 60 < N$.

Concentrating on the three major performance measures (SR, CE and RL, given in columns 3 to 5 in the tables) we find that the weight shrinkage method proposed by [Pollak \(2011\)](#) outperforms all other competitors in 8 out of 9 cases for the small portfolios. Here, many Bayesian shrinkage rules can also significantly outperform the $1/N$ portfolio in terms of out-of-sample SR and out-of-sample CE.

For larger portfolio dimensions ($N = 25, 30, 50, 100, 250, 500$) our Bayesian regression shrinkage strategies dominate all alternative strategies substantially. Among the conventional frequentist strategies (13) - (17) none is particularly salient. For the large portfolios ($N = 25, 30, 50$) in 10 out of 12 comparisons the best portfolio is a Bayesian shrinkage strategy, namely the hierarchical ridge, strategy (7), and the Lasso approaches accounting for turnover, i.e. strategies (10) and (11), are the best performing variants. The direct shrinkage method of the inverse of the covariance method by [Kourtis et al. \(2012\)](#) performs best for the FF 30 industry portfolios in terms of SR and RL.

The findings for the high-dimensional portfolios are very similar to the ones for the large portfolios. One of the Bayesian strategies performs best in 10 out of 12 comparisons. There seems to be no systematic reasoning for the two winning alternative strategies, as we find the highest SR for the Ledoit-Wolf covariance shrinkage method for equity portfolios with $N = 250$ and the best CE for the naive $1/N$ strategy in the case of the equity portfolios with $N = 100$.

In many cases, but not all, the naive $1/N$ portfolio can be outperformed by several

shrinkage strategies. It is not surprising that the probability of significantly outperforming the $1/N$ strategy decreases with the portfolio dimension as the naive strategy performance is relatively better compared to the other strategies with increasing portfolio dimension.

The weight characteristics reveal that all Bayesian and frequentist shrinkage methods produce portfolio weights that often vary more than $1/N$ but always less than GMVP. This is also reflected by the turnover values for the Fama/French portfolios. Only the no-short sale portfolio is able to generate slightly lower turnover than the naive $1/N$ portfolio, but shows otherwise a very similar performance to it. Eventually, the three factor strategy, although having an information advantage, does not outperform the other strategies for any portfolio dimension.

2.4.1 Comparison between Bayesian shrinkage methods

The performance of the fixed ridge shrinkage strategies depends heavily on the performance of the GMVP, i.e. they also perform poorly when the GMVP estimates are poor and the penalty is too weak so that their shrinkage towards the $1/N$ -strategy is too small. However, the simple ridge estimator with dimensionality-dependent shrinkage parameter performs comparatively well for high-dimensional portfolios. In this case, the penalty is strong, so that the ridge estimates come close to the naive weights but still account partly for the correlation structure of the returns. In almost all cases, the empirical Bayes prior outperforms the ridge strategy, yielding evidence for the information gain from using the data to build the prior. In general, the strategies using flexible shrinkage intensities (determined by the data) show superior performance, especially for larger portfolio dimensions. For the French/Fama data, the Bayesian lasso of turnover and the Bayesian elastic net of turnover outperform the other models.⁹ While these strategies reduce the amount of portfolio rebalancing, we have to note that these two models have an information advantage to the other strategies by using the portfolio weight of the last period. For the stock portfolios, the Bayesian elastic net seems to be the strongest portfolio rule.

In general, shrinkage towards the weights of the $1/N$ strategy leads to a reduction of the variation in the estimated portfolio weights compared to noisy unpenalized plug-in GMVP estimates. Therefore, it is not surprising that the performance improvement in terms of SR and CE is due to the reduction of the out-of-sample standard deviation of the portfolio returns. The shrinkage strategies profit from the reduction of estimation noise more than they lose by moving towards the naive, theoretically sub-optimal strategy.

⁹Note that we omitted these two models for the equity data sets due to the random choice of the assets in every period. The weight of the previous period, most likely in another asset, has therefore no relevant information for the current allocation.

2.4.2 Robustness Checks

Our findings remain robust when several dimensions of our set-up are changed.¹⁰ In our sub-sample analysis we also investigate the portfolio performances in calm market times (2002 to 2006) and in crisis times (2006 to 2012). The results reveal that the rankings for the calm period and the more volatile period are basically the same. However, because the number of out-of-sample predictions are now considerably smaller compared to the predictions for the entire sample, the portfolio performance of all strategies (including $1/N$) deteriorates.

We further estimated the portfolios with a rebalancing frequency of six months, thus keeping the estimated portfolios fixed for six periods and only updating the portfolio every half year. Again, the results are very similar to the results with monthly rebalancing. While the portfolio performance measures differ substantially for some data sets, the ordering of the models remains stable.

Finally, we estimated the portfolios using $h = 120$ past observations. The results are very similar to the results presented here for $h = 60$.

2.5 Concluding remarks

Stabilizing portfolio weights is a challenging task of empirical portfolio choice. Numerous strategies have been proposed to overcome the problem of too high estimation noise. In this paper, we propose a new class of models based on flexible Bayesian regularization strategies for the portfolio weights. Our approach allows for a range of stabilization strategies including the imposition of short sale constraints, the reduction of turnover cost or imposition of a bounded support for the portfolio weights. Further, our approach is applicable for high dimensional settings when the number of assets exceeds the number of observations ($N > T$). Applying Bayesian lasso variants in the context of portfolio optimization ultimately offers a novel field of applications for these models.

In a horse race based on portfolios with different number of assets ranging from 5 to 500, our class of models is shown to be highly competitive and beats a wide range of models proposed in the literature independent of the performance criterion chosen. We see that particularly hierarchical shrinkage methods for the portfolio weights show strong out-of-sample performance.

Future work needs to examine our approach in the light of alternative strategies to choose the penalty parameters, which could either be based on the out-of-sample performance of the estimators measured for example by the certainty equivalent. Moreover, our model class can be further extended to alternative penalization and other portfolio

¹⁰In the interest of parsimony, these results are delegated to a separated web-appendix.

strategies, e.g. the tangency portfolio and the mean-variance portfolio. Finally, with the Bayesian approach presented it will also be interesting not only to look at the point estimates of the performance measures, but also on their distributions across time.

References

- BAUWENS, L. AND D. KOROBILIS (2013): “Bayesian methods,” in *Handbook Of Research Methods And Applications In Empirical Macroeconomics*, ed. by N. Hashimzade and M. A. Thornton, Edward Elger Publishing, Handbooks of Research Methods and Applications series, chap. 16, 363–380.
- BEST, M. J. AND R. R. GRAUER (1991): “On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results,” *Review of Financial Studies*, 4, 315–342.
- BRANDT, M. (2010): “Portfolio Choice Problems,” in *Handbook of Financial Econometrics, Volume 1: Tools and Techniques*, ed. by Y. Ait-Sahalia and L. Hansen, Amsterdam: North Holland, 269–336.
- BRANDT, M. W., P. SANTA-CLARA, AND R. VALKANOV (2009): “Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns,” *Review of Financial Studies*, 22, 3411–3447.
- BRITTEN-JONES, M. (1999): “The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights,” *The Journal of Finance*, 54, 655–671.
- BRODIE, J., I. DAUBECHIES, C. D. MOL, D. GIANNONE, AND I. LORIS (2009): “Sparse and stable Markowitz portfolios,” *Proceedings of the National Academy of Sciences*, 106, 12267–12272.
- DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009a): “A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms,” *Management Science*, 55, 798–812.
- DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009b): “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?” *Review of Financial Studies*, 22, 1915–1953.
- DEMIGUEL, V., A. MARTIN-UTRERA, AND F. J. NOGALES (2013): “Size matters: Optimal calibration of shrinkage estimators for portfolio selection,” *Journal of Banking & Finance*, 37, 3018–3034.
- FAHRMEIR, L., T. KNEIB, AND S. KONRATH (2010): “Bayesian regularisation in structured additive regression: A unifying perspective on shrinkage, smoothing and predictor selection,” *Statistics and Computing*, 20, 203–219.

- FAMA, E. F. AND K. R. FRENCH (1993): “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.
- FAN, J., J. ZHANG, AND K. YU (2012): “Vast Portfolio Selection With Gross-Exposure Constraints,” *Journal of the American Statistical Association*, 107, 592–606.
- FLETCHER, J. AND J. HILLIER (2001): “An Examination of Resampled Portfolio Efficiency,” *Financial Analysts Journal*, 57, 66–74.
- FRAHM, G. AND C. MEMMEL (2010): “Dominating estimators for minimum-variance portfolios,” *Journal of Econometrics*, 159, 289–302.
- GEWEKE, J. F. (1996): “Bayesian Inference for Linear Models Subject to Linear Inequality Constraints,” in *Modelling and Prediction Honoring Seymour Geisser*, ed. by J. C. Lee, W. O. Johnson, and A. Zellner, Springer New York, 248–263.
- GOLDFARB, D. AND G. IYENGAR (2003): “Robust Portfolio Selection Problems,” *Mathematics of Operations Research*, 28, 1–38.
- GOLOSNOY, V. AND Y. OKHRIN (2007): “Multivariate Shrinkage for Optimal Portfolio Weights,” *The European Journal of Finance*, 13, 441–458.
- GOTO, S. AND Y. XU (2015): “Improving Mean Variance Optimization through Sparse Hedging Restrictions,” *Journal of Financial and Quantitative Analysis*, 50, 1415–1441.
- HOERL, A. E. AND R. W. KENNARD (1970): “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, 12, 55–67.
- JAGANNATHAN, R. AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *The Journal of Finance*, 58, 1651–1684.
- JOBSON, J. D. AND B. KORKIE (1980): “Estimation for Markowitz Efficient Portfolios,” *Journal of the American Statistical Association*, 75, 544–554.
- JORION, P. (1986): “Bayes-Stein Estimation for Portfolio Analysis,” *Journal of Financial and Quantitative Analysis*, 21, 279–292.
- JUDGE, G. AND M. BOCK (1978): *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Studies in Mathematical and Managerial Economics, North-Holland Publications.
- KAN, R. AND G. ZHOU (2007): “Optimal Portfolio Choice with Parameter Uncertainty,” *Journal of Financial and Quantitative Analysis*, 42, 621–656.

- KEMPF, A. AND C. MEMMEL (2006): “Estimating the Global Minimum Variance Portfolio,” *Schmalenbach Business Review (SBR)*, 58, 332–348.
- KOROBILIS, D. (2013): “Hierarchical shrinkage priors for dynamic regressions with many predictors,” *International Journal of Forecasting*, 29, 43–59.
- KOURTIS, A., G. DOTSIS, AND R. N. MARKELLOS (2012): “Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix,” *Journal of Banking & Finance*, 36, 2522–2531.
- KYUNG, M., J. GILL, M. GHOSH, AND G. CASELLA (2010): “Penalized regression, standard errors, and Bayesian lassos,” *Bayesian Analysis*, 5, 369–411.
- LEDOIT, O. AND M. WOLF (2003): “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” *Journal of Empirical Finance*, 10, 603–621.
- (2008): “Robust performance hypothesis testing with the Sharpe ratio,” *Journal of Empirical Finance*, 15, 850–859.
- (2011): “Robust Performances Hypothesis Testing With the Variance,” *Wilmott*, 2011, 86–89.
- (2014): “Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks,” Working paper No. 137, Department of Economics, University of Zurich.
- LI, J. (2015): “Sparse and Stable Portfolio Selection with Parameter Uncertainty,” *Journal of Business & Economic Statistics*, 33, 381–392.
- LINDLEY, D. V. AND A. F. M. SMITH (1972): “Bayes Estimates for the Linear Model,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 1–41.
- MICHAUD, R. O. (1989): “The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal?” *Financial Analysts Journal*, 45, 31–42.
- OKHRIN, Y. AND W. SCHMID (2007): “Comparison of different estimation techniques for portfolio selection,” *AStA Advances in Statistical Analysis*, 91, 109–127.
- PARK, T. AND G. CASELLA (2008): “The Bayesian Lasso,” *Journal of the American Statistical Association*, 103, 681–686.

- POLLAK, I. (2011): “Weight shrinkage for portfolio optimization,” in *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 37–40.
- TIBSHIRANI, R. (1996): “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.
- TU, J. AND G. ZHOU (2011): “Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies,” *Journal of Financial Economics*, 99, 204–215.
- YEN, Y. (2016): “Sparse Weighted-Norm Minimum Variance Portfolios,” *Review of Finance*, 20, 1259–1287.
- ZOU, H. AND T. HASTIE (2005): “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320.

2.6 Appendix

Table 2.1: Out-of-sample results for $N = 5$ Fama/French industry portfolios with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)

Method	Returns					Weights			
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	TO	min	max	MAD
(1) $1/N$	6.69	14.83	0.45	5.59	0.00	0.19	0.20	0.20	0.00
(2) GMVP	7.01	13.20***	0.53	6.14	-12.71	0.83	-1.23	1.59	0.53
(3) Ridge (τ_{LS})	7.01	13.20***	0.53	6.14	-12.71	0.83	-1.23	1.59	0.53
(4) Ridge ($\tau = h/N$)	7.01	13.19***	0.53	6.14	-12.79	0.82	-1.21	1.58	0.53
(5) Emp. Bayes (τ_{JB})	6.97	14.35***	0.49***	5.94*	-5.97	0.19	-0.09	0.49	0.07
(6) Emp. Bayes ($\tau = h/N$)	7.04	13.12***	0.54**	6.18*	-13.46	0.72	-1.12	1.48	0.49
(7) Hierarchical ridge	6.54	13.28***	0.49*	5.66	-6.69	0.57	-1.02	1.52	0.42
(8) Bay. lasso	6.57	13.37***	0.49*	5.68	-6.54	0.49	-0.88	1.38	0.34
(9) Bay. elastic net	6.71	14.77	0.45*	5.62	-0.62	0.20	-0.21	0.68	0.02
(10) Bay. lasso of TO	7.23	13.62***	0.53**	6.31*	-13.12	0.75	-0.94	1.34	0.51
(11) Bay. elastic net of TO	7.30	13.66***	0.53**	6.36*	-13.66	0.70	-0.87	1.27	0.49
(12) No short sale	6.97	14.25***	0.49**	5.95	-6.54	0.18	0.05	0.33	0.09
(13) Ledoit/Wolf (2003)	7.01	13.16***	0.53	6.14*	-12.90	0.81	-1.19	1.54	0.52
(14) Kourtis et al. (2012)	6.61	13.12**	0.50	5.75	-8.39	0.65	-1.06	1.47	0.43
(15) Frahm/Memmel (2010)	6.93	13.14***	0.53	6.07	-12.07	0.71	-1.12	1.49	0.48
(16) Pollak (2011)	7.10	13.21***	0.54	6.23*	-13.81	0.51	-1.03	1.30	0.37
(17) GMVP (3FF Factors)	6.86	14.32***	0.48**	5.83	-4.85	0.18	0.09	0.34	0.05

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio returns net of transaction costs (50 basis points per trade) for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent, RL stands for return loss in terms of Sharpe ratio relative to the $1/N$ portfolio and TO stands for turnover. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the $1/N$ benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the mean absolute distance relative to $1/N$ (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.2: Out-of-sample results for $N = 6$ Fama/French portfolios formed on size and book-to-market with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)

Method	Returns					Weights			
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	TO	min	max	MAD
(1) $1/N$	7.42	16.41	0.45	6.07	0.00	0.18	0.17	0.17	0.00
(2) GMVP	8.70	13.00***	0.67**	7.86	-33.85	1.63	-1.47	2.03	0.67
(3) Ridge (τ_{LS})	8.70	13.00***	0.67**	7.85	-33.85	1.63	-1.46	2.03	0.67
(4) Ridge ($\tau = h/N$)	8.66	12.97***	0.67**	7.82	-33.55	1.57	-1.41	1.93	0.65
(5) Emp. Bayes (τ_{JB})	7.91	14.91***	0.53***	6.80*	-13.99	0.22	-0.30	0.70	0.15
(6) Emp. Bayes ($\tau = h/N$)	8.69	12.92***	0.67**	7.86	-34.22	1.36	-1.32	1.86	0.61
(7) Hierarchical ridge	8.69	12.98***	0.67**	7.85	-33.90	1.01	-1.18	1.56	0.46
(8) Bay. lasso	8.23	12.93***	0.64**	7.39	-28.56	0.96	-0.97	1.39	0.42
(9) Bay. elastic net	7.42	16.24***	0.46	6.10	-0.89	0.36	-0.44	0.79	0.07
(10) Bay. lasso of TO	8.74	13.24***	0.66**	7.87	-33.09	1.36	-0.92	1.22	0.57
(11) Bay. elastic net of TO	8.67	13.23***	0.66**	7.80	-32.29	1.26	-0.83	1.17	0.55
(12) No short sale	7.02	15.26***	0.46	5.86	-1.47	0.17	0.07	0.27	0.08
(13) Ledoit/Wolf (2003)	8.45	12.95***	0.65*	7.61	-31.14	1.46	-1.30	1.95	0.60
(14) Kourtis e al. (2012)	7.97	12.94***	0.62*	7.14	-25.50	1.20	-1.11	1.60	0.49
(15) Frahm/Memmel (2010)	8.68	12.93***	0.67**	7.85	-34.00	1.35	-1.30	1.85	0.60
(16) Pollak (2011)	8.78	12.94***	0.68**	7.94	-35.12	1.13	-1.29	1.83	0.54
(17) GMVP (3FF Factors)	7.43	15.38***	0.48**	6.25	-5.69	0.17	0.07	0.32	0.06

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio returns net of transaction costs (50 basis points per trade) for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent, RL stands for return loss in terms of Sharpe ratio relative to the $1/N$ portfolio and TO stands for turnover. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the $1/N$ benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the mean absolute distance relative to $1/N$ (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.3: Out-of-sample results for $N = 25$ Fama/French portfolios formed on size and book-to-market with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)

Method	Returns					Weights			
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	TO	min	max	MAD
(1) $1/N$	7.64	17.10	0.45	6.18	0.00	0.23	0.04	0.04	0.00
(2) GMVP	7.39	14.52***	0.51	6.33	-10.78	4.68	-1.26	1.30	0.37
(3) Ridge (τ_{LS})	7.39	14.51***	0.51	6.34	-10.83	4.67	-1.26	1.30	0.37
(4) Ridge ($\tau = h/N$)	7.61	13.94***	0.55	6.64	-16.55	4.19	-1.09	1.14	0.34
(5) Emp. Bayes (τ_{JB})	8.20	14.22***	0.58**	7.19	-22.09	1.03	-0.55	0.60	0.15
(6) Emp. Bayes ($\tau = h/N$)	8.07	13.59***	0.59	7.15	-23.96	2.45	-0.88	0.93	0.26
(7) Hierarchical ridge	8.20	12.76***	0.64*	7.38	-29.91	1.64	-0.81	1.06	0.16
(8) Bay. lasso	7.96	12.43***	0.64*	7.19	-28.85	1.74	-0.52	0.67	0.16
(9) Bay. elastic net	7.67	16.93***	0.45**	6.24	-1.26	0.22	0.01	0.07	0.00
(10) Bay. lasso of TO	8.82	12.91***	0.68*	7.99*	-36.56	2.76	-0.46	0.54	0.24
(11) Bay. elastic net of TO	8.76	12.78***	0.69**	7.94	-36.54	2.29	-0.37	0.42	0.21
(12) No short sale	7.73	16.02***	0.48**	6.44	-6.79	0.21	0.02	0.07	0.02
(13) Ledoit/Wolf (2003)	8.14	12.79***	0.64	7.32	-29.04	2.63	-0.66	0.79	0.23
(14) Kourtis et al. (2012)	7.68	12.92***	0.59	6.84	-22.86	2.12	-0.65	0.74	0.18
(15) Frahm/Memmel (2010)	7.89	15.14***	0.52**	6.75	-13.48	0.61	-0.32	0.40	0.10
(16) Pollak (2011)	8.11	14.03***	0.58	7.12	-22.06	1.85	-0.87	0.91	0.21
(17) GMVP (3FF Factors)	7.77	15.69***	0.50**	6.54	-9.05	0.21	-0.01	0.11	0.02

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio returns net of transaction costs (50 basis points per trade) for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent, RL stands for return loss in terms of Sharpe ratio relative to the $1/N$ portfolio and TO stands for turnover. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the $1/N$ benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the mean absolute distance relative to $1/N$ (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.4: Out-of-sample results for $N = 30$ Fama/French industry portfolios with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)

Method	Returns					Weights			
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	TO	min	max	MAD
(1) $1/N$	6.64	15.91	0.42	5.37	0.00	0.33	0.03	0.03	0.00
(2) GMVP	3.99	15.03	0.27	2.87	27.32	2.70	-0.60	0.72	0.18
(3) Ridge (τ_{LS})	4.00	15.02	0.27	2.87	27.28	2.70	-0.60	0.72	0.18
(4) Ridge ($\tau = h/N$)	4.16	14.70	0.28	3.08	23.73	2.60	-0.56	0.68	0.18
(5) Emp. Bayes (τ_{JB})	6.73	13.74***	0.49	5.78	-11.93	0.73	-0.25	0.36	0.07
(6) Emp. Bayes ($\tau = h/N$)	5.38	13.27***	0.41	4.50	1.88	1.43	-0.39	0.49	0.12
(7) Hierarchical ridge	6.09	12.11***	0.50*	5.35	-12.40	0.98	-0.32	0.44	0.08
(8) Bay. lasso	6.23	11.84***	0.53**	5.52	-15.42	1.10	-0.22	0.32	0.08
(9) Bay. elastic net	6.68	15.69***	0.43*	5.45	-1.64	0.33	0.02	0.05	0.00
(10) Bay. lasso of TO	5.11	12.83***	0.40	4.29	2.90	1.82	-0.24	0.33	0.13
(11) Bay. elastic net of TO	5.60	12.47***	0.45*	4.83	-4.80	1.57	-0.19	0.28	0.12
(12) No short sale	7.11	14.64***	0.49**	6.04*	-11.96	0.32	0.01	0.06	0.02
(13) Ledoit/Wolf (2003)	5.88	11.86***	0.50*	5.18	-11.18	1.43	-0.24	0.34	0.11
(14) Kourtis et al. (2012)	6.53	12.18***	0.54*	5.79	-17.32	1.25	-0.27	0.37	0.09
(15) Frahm/Memmel (2010)	6.61	15.64**	0.42	5.39	-1.02	0.33	-0.01	0.08	0.01
(16) Pollak (2011)	5.86	13.63***	0.43	4.93	-2.05	1.25	-0.39	0.49	0.10
(17) GMVP (3FF Factors)	6.84	13.93***	0.49**	5.87	-12.32	0.31	-0.01	0.10	0.02

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio returns net of transaction costs (50 basis points per trade) for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent, RL stands for return loss in terms of Sharpe ratio relative to the $1/N$ portfolio and TO stands for turnover. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the $1/N$ benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the mean absolute distance relative to $1/N$ (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.5: Out-of-sample results for $N = 100$ Fama/French portfolios formed on size and book-to-market with $h = 60$ months estimation window size (sample: 1/1953 - 12/2015)

Method	Returns					Weights			
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	TO	min	max	MAD
(1) $1/N$	7.66	17.26	0.44	6.17	0.00	0.33	0.01	0.01	0.00
(2) GMVP	NA	NA	NA	NA	NA	NA	NA	NA	NA
(3) Ridge (τ_{LS})	NA	NA	NA	NA	NA	NA	NA	NA	NA
(4) Ridge ($\tau = h/N$)	7.70	14.77***	0.52**	6.61	-13.74	0.53	-0.02	0.04	0.01
(5) Emp. Bayes (τ_{JB})	NA	NA	NA	NA	NA	NA	NA	NA	NA
(6) Emp. Bayes ($\tau = h/N$)	NA	NA	NA	NA	NA	NA	NA	NA	NA
(7) Hierarchical ridge	6.56	13.86***	0.47	5.60	-4.88	2.66	-0.50	0.58	0.06
(8) Bay. lasso	6.58	12.61***	0.52**	5.78	-11.75	1.95	-0.18	0.23	0.04
(9) Bay. elastic net	7.70	17.09***	0.45**	6.24	-1.32	0.32	0.00	0.02	0.00
(10) Bay. lasso of TO	6.94	14.40***	0.48**	5.90	-6.52	3.94	-0.17	0.20	0.08
(11) Bay. elastic net of TO	7.58	13.43***	0.56	6.67	-19.37	3.00	-0.12	0.13	0.06
(12) No short sale	7.74	16.21***	0.48**	6.42	-6.45	0.31	0.00	0.02	0.00
(13) Ledoit/Wolf (2003)	5.63	12.64***	0.45	4.83	-0.22	2.62	-0.18	0.24	0.06
(14) Kurtis et al. (2012)	6.98	13.22***	0.53	6.10	-13.28	2.28	-0.20	0.22	0.04
(15) Frahm/Memmel (2010)	NA	NA	NA	NA	NA	NA	NA	NA	NA
(16) Pollak (2011)	NA	NA	NA	NA	NA	NA	NA	NA	NA
(17) GMVP (3FF Factors)	7.49	15.44***	0.49	6.30	-7.67	0.30	-0.02	0.06	0.01

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio returns net of transaction costs (50 basis points per trade) for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent, RL stands for return loss in terms of Sharpe ratio relative to the $1/N$ portfolio and TO stands for turnover. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the $1/N$ benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the mean absolute distance relative to $1/N$ (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.6: Out-of-sample results for $N = 5$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)

Method	Returns					Weights		
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	min	max	MAD
(1) 1/ N	9.60	26.50	0.36	6.09	0.00	0.20	0.20	0.00
(2) GMVP	6.94	18.86***	0.37	5.16	-1.23	-0.32	0.97	0.18
(3) Ridge (τ_{LS})	6.94	18.87**	0.37	5.16	-1.26	-0.32	0.97	0.18
(4) Ridge ($\tau = h/N$)	6.94	18.86**	0.37	5.16	-1.24	-0.32	0.97	0.18
(5) Emp. Bayes (τ_{JB})	8.56	25.23***	0.34	5.38	6.95	0.07	0.38	0.03
(6) Emp. Bayes ($\tau = h/N$)	7.14	19.05**	0.37	5.33	-2.88	-0.28	0.91	0.17
(7) Hierarchical ridge	8.19	19.27**	0.43*	6.34	-14.56	-0.29	0.99	0.16
(8) Bay. lasso	7.87	19.26**	0.41*	6.02	-10.70	-0.27	0.92	0.15
(9) Bay. elastic net	8.49	20.52**	0.41*	6.38	-12.60	-0.17	0.82	0.08
(10) No short sale	7.73	22.33***	0.35	5.24	4.30	0.04	0.34	0.09
(13) Ledoit/Wolf (2003)	6.66	18.73**	0.36	4.90	1.56	-0.32	0.97	0.19
(14) Kourtis et al. (2012)	7.28	19.32**	0.38*	5.41	-3.29	-0.34	0.92	0.16
(15) Frahm/Memmel (2010)	7.33	19.11**	0.38	5.50	-4.85	-0.29	0.93	0.17
(16) Pollak (2011)	9.25	21.31***	0.43**	6.98	-18.35	-0.11	0.71	0.09
(17) GMVP (3FF Factors)	7.30	18.55**	0.39*	5.58	-6.90	-0.12	0.87	0.16

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio gross returns for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent and RL stands for return loss in terms of Sharpe ratio relative to the 1/ N portfolio. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the 1/ N benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the the mean absolute distance relative to 1/ N (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.7: Out-of-sample results for $N = 25$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)

Method	Returns					Weights		
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	min	max	MAD
(1) 1/ N	6.41	19.74	0.32	4.46	0.00	0.04	0.04	0.00
(2) GMVP	5.86	14.27**	0.41*	4.84	-14.78	-0.29	0.52	0.09
(3) Ridge (τ_{LS})	5.86	14.26**	0.41*	4.84	-14.80	-0.29	0.52	0.09
(4) Ridge ($\tau = h/N$)	5.87	14.24**	0.41*	4.86	-14.98	-0.29	0.52	0.09
(5) Emp. Bayes (τ_{JB})	6.51	16.41***	0.40*	5.16	-14.22	-0.07	0.17	0.02
(6) Emp. Bayes ($\tau = h/N$)	6.02	13.73***	0.44*	5.08	-18.79	-0.20	0.38	0.06
(7) Hierarchical ridge	6.96	12.65***	0.55**	6.16	-34.32	-0.22	0.52	0.06
(8) Bay. lasso	5.79	12.86***	0.45*	4.96	-19.37	-0.18	0.39	0.05
(9) Bay. elastic net	5.72	18.08***	0.32	4.09	1.75	-0.03	0.09	0.01
(10) No short sale	5.69	17.02***	0.33	4.24	-1.99	0.01	0.06	0.02
(13) Ledoit/Wolf (2003)	6.07	12.19***	0.50*	5.33	-25.38	-0.14	0.40	0.06
(14) Kourtis et al (2012)	6.02	13.45***	0.45*	5.11	-19.85	-0.18	0.34	0.05
(15) Frahm/Memmel (2010)	5.78	16.38***	0.35	4.44	-5.54	-0.06	0.19	0.02
(16) Pollak (2011)	6.70	16.02*	0.42*	5.41	-17.98	-0.12	0.25	0.03
(17) GMVP (3FF Factors)	5.92	12.71***	0.47*	5.11	-21.49	-0.09	0.37	0.05

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio gross returns for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent and RL stands for return loss in terms of Sharpe ratio relative to the 1/ N portfolio. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the 1/ N benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the the mean absolute distance relative to 1/ N (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.8: Out-of-sample results for $N = 50$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)

Method	Returns					Weights		
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	min	max	MAD
(1) 1/ N	8.14	20.34	0.40	6.07	0.00	0.02	0.02	0.00
(2) GMVP	3.40	21.72	0.16	1.05	63.40	-0.51	0.61	0.11
(3) Ridge (τ_{LS})	3.27	21.70	0.15	0.91	64.94	-0.51	0.61	0.11
(4) Ridge ($\tau = h/N$)	3.81	20.17	0.19	1.77	51.15	-0.41	0.51	0.10
(5) Emp. Bayes (τ_{JB})	7.47	18.72***	0.40	5.72	0.21	-0.10	0.16	0.02
(6) Emp. Bayes ($\tau = h/N$)	5.52	17.85*	0.31	3.92	19.52	-0.27	0.34	0.06
(7) Hierarchical ridge	7.46	12.08***	0.62**	6.73	-31.48	-0.16	0.40	0.04
(8) Bay. lasso	5.13	12.29***	0.42*	4.38	-2.57	-0.11	0.24	0.03
(9) Bay. elastic net	7.18	19.08***	0.38	5.36	5.45	-0.01	0.05	0.00
(10) No short sale	7.18	17.79***	0.40	5.60	-0.79	0.01	0.03	0.01
(13) Ledoit/Wolf (2003)	4.85	11.52***	0.42*	4.18	-2.83	-0.08	0.24	0.03
(14) Kourtis et al (2012)	5.34	14.06***	0.38	4.36	3.37	-0.12	0.18	0.03
(15) Frahm/Memmel (2010)	8.14	20.34	0.40	6.07	0.00	0.02	0.02	0.00
(16) Pollak (2011)	6.62	17.85***	0.37	5.03	6.27	-0.23	0.30	0.04
(17) GMVP (3FF Factors)	3.69	11.66***	0.32	3.01	11.69	-0.05	0.22	0.03

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio gross returns for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent and RL stands for return loss in terms of Sharpe ratio relative to the 1/ N portfolio. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the 1/ N benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the the mean absolute distance relative to 1/ N (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.9: Out-of-sample results for $N = 100$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)

Method	Returns					Weights		
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	min	max	MAD
(1) 1/ N	6.41	19.37	0.33	4.54	0.00	0.01	0.01	0.00
(2) GMVP	NA	NA	NA	NA	NA	NA	NA	NA
(3) Ridge (τ_{LS})	NA	NA	NA	NA	NA	NA	NA	NA
(4) Ridge ($\tau = h/N$)	3.90	12.28***	0.32	3.14	2.06	-0.03	0.04	0.01
(5) Emp. Bayes (τ_{JB})	NA	NA	NA	NA	NA	NA	NA	NA
(6) Emp. Bayes ($\tau = h/N$)	NA	NA	NA	NA	NA	NA	NA	NA
(7) Hierarchical ridge	2.47	12.42***	0.20	1.70	19.70	-0.16	0.31	0.03
(8) Bay. lasso	3.62	11.27***	0.32	2.99	1.32	-0.08	0.13	0.02
(9) Bay. elastic net	5.79	18.06***	0.32	4.16	2.27	-0.00	0.02	0.00
(10) No short sale	5.66	16.62***	0.34*	4.27	-1.83	0.00	0.02	0.00
(13) Ledoit/Wolf (2003)	2.17	9.93***	0.22	1.68	13.41	-0.05	0.13	0.02
(14) Kourtis et al. (2012)	3.07	11.56***	0.27	2.41	9.03	-0.06	0.07	0.01
(15) Frahm/Memmel (2010)	NA	NA	NA	NA	NA	NA	NA	NA
(16) Pollak (2011)	NA	NA	NA	NA	NA	NA	NA	NA
(17) GMVP (3FF Factors)	0.94	10.77***	0.09	0.36	31.46	-0.03	0.12	0.02

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio gross returns for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent and RL stands for return loss in terms of Sharpe ratio relative to the 1/ N portfolio. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the 1/ N benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the the mean absolute distance relative to 1/ N (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.10: Out-of-sample results for $N = 250$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)

Method	Returns					Weights		
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	min	max	MAD
(1) 1/ N	6.48	19.88	0.33	4.50	0.00	0.00	0.00	0.00
(2) GMVP	NA	NA	NA	NA	NA	NA	NA	NA
(3) Ridge (τ_{LS})	NA	NA	NA	NA	NA	NA	NA	NA
(4) Ridge ($\tau = h/N$)	5.11	12.47***	0.41	4.34	-12.63	-0.02	0.02	0.00
(5) Emp. Bayes (τ_{JB})	NA	NA	NA	NA	NA	NA	NA	NA
(6) Emp. Bayes ($\tau = h/N$)	NA	NA	NA	NA	NA	NA	NA	NA
(7) Hierarchical ridge	4.95	9.50***	0.52**	4.50	-22.25	-0.04	0.08	0.01
(8) Bay. lasso	5.24	10.38***	0.51*	4.70	-22.35	-0.04	0.04	0.01
(9) Bay. elastic net	6.20	18.68***	0.33	4.45	-1.34	0.00	0.01	0.00
(10) No short sale	6.57	17.13***	0.38**	5.11	-11.93	0.00	0.01	0.00
(13) Ledoit/Wolf (2003)	4.87	9.33***	0.52**	4.44	-22.00	-0.02	0.06	0.01
(14) Kourtis et al. (2012)	5.44	11.13***	0.49	4.82	-21.82	-0.03	0.02	0.01
(15) Frahm/Memmel (2010)	NA	NA	NA	NA	NA	NA	NA	NA
(16) Pollak (2011)	NA	NA	NA	NA	NA	NA	NA	NA
(17) GMVP (3FF Factors)	4.18	9.80***	0.43	3.70	-11.87	-0.01	0.05	0.01

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio gross returns for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent and RL stands for return loss in terms of Sharpe ratio relative to the 1/ N portfolio. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the 1/ N benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the the mean absolute distance relative to 1/ N (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

Table 2.11: Out-of-sample results for $N = 500$ equity assets portfolio with $h = 60$ months estimation window size (sample: 1/2001 - 12/2015)

Method	Returns					Weights		
	$\hat{\mu}$	$\hat{\sigma}$	SR	CE	RL	min	max	MAD
(1) 1/ N	7.05	19.69	0.36	5.11	0.00	0.00	0.00	0.00
(2) GMVP	NA	NA	NA	NA	NA	NA	NA	NA
(3) Ridge (τ_{LS})	NA	NA	NA	NA	NA	NA	NA	NA
(4) Ridge ($\tau = h/N$)	4.86	12.54***	0.39	4.07	-4.47	-0.01	0.01	0.00
(5) Emp. Bayes (τ_{JB})	NA	NA	NA	NA	NA	NA	NA	NA
(6) Emp. Bayes ($\tau = h/N$)	NA	NA	NA	NA	NA	NA	NA	NA
(7) Hierarchical ridge	5.42	9.62***	0.56*	4.95	-23.69	-0.02	0.03	0.00
(8) Bay. lasso	5.46	10.43***	0.52*	4.91	-20.67	-0.02	0.02	0.00
(9) Bay. elastic net	6.61	18.55***	0.36	4.89	0.33	0.00	0.00	0.00
(10) No short sale	6.67	17.08***	0.39	5.21	-6.63	0.00	0.00	0.00
(13) Ledoit/Wolf (2003)	5.15	9.24***	0.56*	4.72	-22.06	-0.01	0.03	0.00
(14) Kourtis et al. (2012)	5.82	11.11***	0.52*	5.20	-22.12	-0.01	0.01	0.00
(15) Frahm/Memmel (2010)	NA	NA	NA	NA	NA	NA	NA	NA
(16) Pollak (2011)	NA	NA	NA	NA	NA	NA	NA	NA
(17) GMVP (3FF Factors)	3.76	9.88***	0.38	3.27	-2.64	-0.01	0.02	0.00

Note: The table reports annualized mean results (in percent) for out-of-sample portfolio gross returns for the rolling window one-step ahead portfolio optimization exercise. $\hat{\mu}$ stands for the average returns, $\hat{\sigma}$ stands for the standard deviation of the returns, SR stands for Sharpe ratio, CE stands for certainty equivalent and RL stands for return loss in terms of Sharpe ratio relative to the 1/ N portfolio. For the CE, we assume a risk aversion of $\gamma = 1$. For each method, we test the difference in variances (standard deviations) and Sharpe ratios from the 1/ N benchmark by the bootstrap methodology proposed in [Ledoit and Wolf \(2011\)](#) and [Ledoit and Wolf \(2008\)](#) respectively. To test the difference in certainty equivalents, we use the methodology described by [DeMiguel et al. \(2009b\)](#) on page 1929. One/two/three asterisks denote rejection of the null hypothesis of equal Sharpe ratios/variances/certainty equivalents at the ten/five/one percent test level. We also report mean out-of-sample values for the minimum and maximum weights as well as the the mean absolute distance relative to 1/ N (MAD). GMVP (3FF Factors) refers to the GMVP estimated using the 3 Fama/French factors.

CHAPTER 3

Using Analysts' Forecasts for Stock
Predictions - An Entropic Tilting Approach

3.1 Introduction

Predicting stock returns is a popular exercise for professionals and academics in finance. It is a challenging task because of estimation and model uncertainty, because of a substantial unpredictable component (shocks) in future stock returns and because successful forecasting models may be exploited by other market participants causing trading behavior that destroys the prediction power of the model. Recent empirical findings suggest that the equity premium is predictable when accounting for model and estimation uncertainty through time-varying parameters, stochastic volatility and by using Bayesian predictive regressions (Zellner and Chetty, 1965; Barberis, 2000; Pettenuzzo and Ravazzolo, 2016). While the amount of predictability from variables such as the dividend-earnings ratio is limited in terms of out-of-sample R², it can translate into substantial portfolio gains for an expected utility investor (Johannes et al., 2014).¹

Most prediction models relate the future stock returns to past observations of asset specific variables such as the dividend yield or to general macroeconomic indicators such as inflation. However, the market stock price is forward-looking as it reflects the expectations of market participants about the future cash flows of the company. It is natural to expect that using forward-looking information should also be beneficial for stock return predictions. A recent example is Metaxoglou et al. (2016), who improve equity premium forecasts by using option prices. In this paper, we use professional analyst's forecasts to improve asset return predictions. We do so by reweighting the predictive return distributions by a method called *entropic (exponential) tilting* to incorporate the information contained in analysts' forecasts, such as target prices. The idea of this method is to modify the predictive density of the asset returns to match certain moment conditions that are formed based on average analysts' forecasts. The advantage of this approach is that we can combine model-based time-series information with external, forward-looking information in a parsimonious way using closed-form solutions.

Professional security analysts provide market analyses, make earnings forecasts and give investment advise by providing twelve months forward target prices and stock recommendations. Their reports and opinions can have major short- and long-run impact on stock prices (e.g. Irvine, 2003) and provide a powerful way to disseminate financial information to market participants.² And a vast literature exists about the shortcomings of analyst forecasts revealing skewed incentives (conflicts of interest), herding and biases to please clients that may create market inefficiencies.³

¹See Rapach and Zhou (2013) for a recent overview on return predictability.

²The US Bureaus of Labor statistics report over 275,000 thousand financial analysts jobs in 2014 (see <http://www.bls.gov/ooh/business-and-financial/financial-analysts.htm>, checked on 19.08.2015).

³See Ramnath et al. (2008) for an in-depth review of the analyst forecast literature.

In this study, we do not evaluate the accuracy of the analysts' forecasts, but we use the (dis-)agreement in the analysts' forecast to regularize the predictive return distribution. In particular, we restrict the mean and variance of the predictive distribution to coincide with the mean and the variance of monthly target price implied expected returns, i.e. simple returns between the current spot and the target price. While we find that restricting the variance of the asset returns is beneficial in terms of out-of-sample performance, as it provides a forward looking measure for (un-)certainty in the market, only restricting the mean has no particular forecasting power. Target prices are usually higher than current spot prices and so there is an upward bias in the target price implied expected returns which is not beneficial for the forecast performance.

Of course we could also include the analyst information simply as another predictor variable in the predictive regressions, but this would add further parameters and more estimation noise to the prediction problem. Using the analyst information instead in a tilting framework, we only change the shape of the predictive distribution by reweighting it and hence do not require the data to formalize the relationship between asset returns and target prices.

Another econometric contribution of this paper is the use of a large Bayesian vector autoregressive system to formalize the predictive relationship between asset returns and a great number of predictor variables. To overcome the computational burden that arises in a recursive forecasting exercise, we adopt the so-called *forgetting factors* approach of Koop and Korobilis (2013) which allows for all the features recently found to be important to find significant return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and variable selection. The idea of forgetting factors is to approximate the conditional error term variance in a Kalman filter type estimation of the time-varying model, reducing the computational burden significantly. While Dangl and Halling (2012) used a similar forgetting factor approach for return prediction, we will combine this method with a Minnesota-type prior that restricts the parameter matrix to deal with the case of many predictors.

It is also worth mentioning that market excess returns, such as the S&P500, might not be the ideal candidate to find predictability, because of their aggregation over many sectors and industries. In this paper instead, we will investigate the predictability of individual assets by looking at a cross-section of Dow Jones index constituents. We will investigate if these returns are predictable using company-specific characteristics like the book-to-market ratio or by using market and economic indicators such as inflation. Out-of-sample studies for cross-sectional returns are limited and a few exceptions among others are Avramov (2002) and Rapach et al. (2015). While these studies use factor and industry portfolio returns, we fill the gap and go down to the level of individual equity asset predictability.

Figure 3.1 serves as an illustrative example. It shows the IBM spot price, the mean twelve months forward target price and the percentage of buy recommendations from all recommendations (buy, sell, hold) of the IBM stock. While the correlation between the spot and the target price is almost perfect (0.9721), spot price and buy recommendations are negatively correlated (-0.6470). In the plot the target price is almost always higher than the current spot price, indicating an upward bias in these forecasts. The only times when the two plots coincide is after price drops when the spot price starts to increase again, i.e. in late 2002, 2007 and 2014. That is, financial analyst might not be able to predict trend changes (regime switches) but they are able to forecast the price direction. This is in line with the more volatile buy recommendations, which only on average indicate the stock price direction.

While this pattern is similar also for the other 19 stocks considered here, comparing the target price with the observed spot price twelve months ahead gives a mixed picture. Table 3.1 provides the root mean squared forecast errors (RMSFE) between the two for the 20 stocks and compares it to the two year historical mean. Only for 11 out of the 20 stocks the target prices were better forecasts than the historical mean, e.g. for the IBM stock.

Figure 3.2 gives an idea about the predictive power of individual predictors against a benchmark intercept only model by plotting three out-of-sample performance measures for the IBM stock. Considering the first panel showing the cumulative sum of squared forecast errors differentials, we see that mostly all predictors fail to outperform the intercept only model, especially in the financial crisis. The only exception is the log return of the mean analyst twelve months forward target prices. Note that values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. The target price itself only has some predictive power between 2008 and 2010 (square markers). This might suggest that not the level of the predicted target price matters, but that the timing of the shifts, represented by the target price implied returns, have predictive power. Also, we see that the average analyst recommendations have no particular predictive power for the IBM stock.

The paper is now organized as follows. Section 3.2 gives an overview about the findings from other authors trying to exploit analysts' forecasts and reviews the state of the art on return predictability. Section 3.3 describes the applied Bayesian vector autoregression model, which relies on the forgetting factor approach by Koop and Korobilis (2013) for large systems. We then introduce the concept of entropic tilting and explain how we translate the analyst forecasts into moment conditions for the predictive return distribution. Section 3.4 summarizes the set-up of the empirical study and presents the results. Section 3.5 concludes and gives an outlook on further generalizations.

3.2 Predictive powers

3.2.1 Analysts' forecasts

A great number of studies document the investment value of sell-side analyst research: Barber et al. (2001); Green (2006) report portfolio benefits and abnormal returns from simple trading strategies that go long in stocks with favorable analyst recommendations and short in stocks with unfavorable ratings. Jegadeesh et al. (2004) instead finds that the predictive power of the level of consensus analyst recommendations varies substantially across assets and proposes to use the changes or revisions in consensus analyst recommendations instead. Jegadeesh and Kim (2006) applies this to build portfolio strategies that are long (short) in stocks with positive (negative) recommendation revisions yielding significant positive abnormal risk-adjusted returns against the market. Cvitanić et al. (2006) further shows that such simple long-short portfolios can be outperformed by more complicated multi-period expected utility maximizing strategies that are estimated using consensus analysts recommendations. This is not the case for He et al. (2013), who find no significant portfolio returns from a Black and Litterman (1992) strategy incorporating analyst recommendations for the Australian stock market.

The evidence for return predictability from target prices is even more mixed: While Brav and Lehavy (2003) show that target prices are generally informative about future stock prices and that there are substantial short-term market reactions in the stock price to target changes, Bonini et al. (2010) find little evidence for target price accuracy using different metrics measuring the prediction error between the target price and the current, twelve months forward and in-between spot prices. This is in line with Bradshaw et al. (2012) who find an average target price premium over the spot price of 15 percent and who report that only two thirds of the target prices are met by the stock price at some time during the forecast horizon in their sample. Lin et al. (2016) instead consider changes in target prices, but also find no evidence that institutional trading activity following the direction of the target price revisions yields higher risk-adjusted out-of-sample returns. Only Da and Schaumburg (2011) find that aggregating stocks across sectors according to their twelve month forward target price implied expected return, i.e. simple return between the current and the target price, yields significant risk-adjusted abnormal returns for different long-short portfolio.

However, despite this discouraging evidence, none of the studies above operate in a Bayesian framework using predictive distributions and none consider a tilting framework in which the analyst's forecasts serve as a shrinkage target instead of just being another predictor variable.

3.2.2 Predictive regressions

Kandel and Stambaugh (1987) propose a vector autoregression formulation to jointly model the dynamics of asset returns and its predictor:

$$r_{t+1} = a_r + b_r x_t + \varepsilon_{r,t+1}, \quad (3.1)$$

$$x_{t+1} = a_x + b_x x_t + \varepsilon_{x,t+1}, \quad (3.2)$$

where x_t is the explanatory variable (for example the dividend yield) and r_t is the asset return and $\varepsilon_t = (\varepsilon_{r,t}, \varepsilon_{x,t})'$ $\stackrel{iid}{\sim} \mathcal{N}(0, \Sigma)$. Here, equation (3.2) is needed to model the times-series dynamics of the predictor variable through an autoregressive process. This model has been used by various authors to investigate the predictability of the equity premium and to build portfolios using the predictive return distribution.

Kandel and Stambaugh (1996) estimate the model using Bayesian priors that center the mean of the slope coefficient in (3.1) to zero, implying no predictability and a weak form of market efficiency. Using a one-period model with power utility, they find that not accounting for the variability in the regressor can decrease the annual certainty equivalent up to 3%.

Pástor (2000) extends the model in (3.1) - (3.2) to a multivariate regression for N risky assets and apply a Normal inverse Wishart prior for the model mean and variance of the asset returns. They interpret the prior as a form to describe the disbelief of the investor in the given prediction/asset-pricing model. Importantly, they find that the optimal portfolio weights tend to be less extreme when using the joint predictive distribution of the asset returns by integrating out the estimated model parameters.

While these studies are concerned with investments for one period, Barberis (2000) investigates the impact of changes in the prediction variables to the asset allocation in the long-term. As the variance of the cumulative log returns grows less than linear in the investment horizon, stocks appear less risky and so an investor should allocate more funds to stocks when the investment horizon increases. However such an over-allocating to stocks disappears when accounting for estimation uncertainty and possible structural breaks in the parameters of the prediction model (Pettenuzzo and Timmermann, 2011).

Stock returns may be predicted by company specific characteristics such as the dividend-price ratio or they may be macroeconomic indicators. Given the great number of potential predictor variables, shrinkage and model averaging methods are natural candidates to reduce model risk in predictive regressions. For example, Avramov (2002) considers 14 predictor variables and extend the predictive system in (3.1) - (3.2). The author averages over all $2^{14} = 16384$ distinct model to predict the asset returns by weighting each model by its posterior weight. Similar examples for Bayesian model averaging

(BMA) in the literature include Cremers (2002) and Wachter and Warusawitharana (2015). Furthermore, Pettenuzzo and Ravazzolo (2016) propose a Bayesian density combination approach over 15 predictor variables that, instead of minimizing a statistical loss function, maximizes the certainty equivalent of a power utility investor and report robust out-of-sample predictability across various performance measures.

Studies using time-varying coefficients in prediction models have also become numerous in recent years: While Welch and Goyal (2008) find no evidence for significant in-sample and out-of-sample predictability of 14 variables on the equity premium using a rolling window estimation, Dangl and Halling (2012) find predictability in one-step-ahead forecasts of monthly S&P 500 returns by using Bayesian predictive regressions with time-varying parameters and model averaging over a range of predictor variables. Using time-varying coefficients as well as stochastic volatility but only two possible predictor variables, Johannes et al. (2014) document statistically and economically significant portfolio benefits for a dynamic rebalancing power utility investor using Bayesian predictive regressions. Very recently, Feng and Polson (2016) propose to perform a *predictive* cross-validation approach to select the right amount of shrinkage that provides optimal posterior point estimates instead of marginalizing out the model parameters to obtain a full predictive return distribution.

Very related to this paper is Metaxoglou et al. (2016), who are the first to apply entropic tilting to predict the US equity premium. For this, the authors rely on option data to form corresponding moment conditions. In particular, the authors use the second moment of the risk-neutral density implied by option prices on the S&P500 to tilt predictive distributions obtained from OLS model combinations with individual predictors, where the combination weights are chosen such that the model minimizes a discounted mean squared forecast error (Rapach and Zhou, 2013). For monthly returns and one-step ahead forecasts, they find that the tilted forecasts outperform an historical average based on various performance measures and for various sub-samples.

3.3 Methodology

3.3.1 Prediction model

In this study we follow the literature and use vector autoregressions (VARs) to model the relationship between asset returns and predictor variables. Extending the VAR given in

equation (3.1) - (3.2) to a system with K predictors reads as follows:

$$\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a + \sum_{i=1}^p A_i \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t, \quad t = 1, \dots, T, \quad (3.3)$$

where r_t is the excess return of a particular stock, $x_t = [x_{1,t}, \dots, x_{K,t}]'$ is a $K \times 1$ vector of predictor variables and $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma)$.

The number of parameters in the system in (3.3) grows quickly with the number of included predictor variables. We therefore focus on two restrictions to reduce estimation noise. Since every VAR(p) system can be written in VAR(1) companion form, we restrict $p = 1$. Second, in (3.1) - (3.2) both, the excess return and the predictor variable, only depend on their own lag, but not on the lag of the other variable. We follow this and restrict the system such that r_t depends on the entire x_{t-1} vector but $x_{k,t}$, $1 \leq k \leq K$, only depends on its own lag $x_{k,t-1}$. Compactly, the resulting model is of the form

$$y_t = (r_t, x_t)' = a + A_1 y_{t-1} + \varepsilon_t, \quad (3.4)$$

where $a = (a_r, a_{x_1}, \dots, a_{x_K})'$ and $A_1 = \begin{pmatrix} 0 & A_1^{1,2} & A_1^{1,3} & \cdots & A_1^{1,K+1} \\ 0 & A_1^{2,2} & 0 & \cdots & 0 \\ \vdots & \ddots & A_1^{3,3} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & A_1^{K+1,K+1} \end{pmatrix}$. The zero restriction follows from (3.1) - (3.2) using multiple predictor variables. Usually, the correlation between the return and its first lag is very low, supporting the restriction $A_1^{1,1} = 0$. All other variables are supposed to follow an autoregressive process of order 1. To implement these restrictions *softly* on the slope coefficient matrix A_1 we use a variant of the Minnesota prior (Doan et al., 1984). Further specifying independent marginal normal priors for each parameter yields the joint prior distribution through multiplication of the independent

marginals. That is

$$p(a) \sim \mathcal{N}(0, \zeta \times I_{(K+1 \times K+1)}) , \quad (3.5)$$

$$p(A_1^{1,1}) \sim \mathcal{N}(0, \varrho \times 1) , \quad (3.6)$$

$$p(A_1^{1,k}) \sim \mathcal{N}\left(0, \zeta \times \frac{\sigma_r^2}{\sigma_{x_k}^2}\right) , \quad k = 1, \dots, K \quad (3.7)$$

$$p(A_1^{k,1}) \sim \mathcal{N}\left(0, \varrho \times \frac{\sigma_{x_k}^2}{\sigma_r^2}\right) , \quad k = 2, \dots, K , \quad (3.8)$$

$$p(A_1^{k,l}) \sim \mathcal{N}\left(\underline{A}_1^{k,l}, \varrho \times \frac{\sigma_{x_k}^2}{\sigma_{x_l}^2}\right) , \quad k = 2, \dots, K, \quad l = 2, \dots, K \quad (3.9)$$

with $\underline{A}_1^{k,l} = d_k$ if $k = l$, and $\underline{A}_1^{k,l} = 0$ otherwise.

Following Frey and Mokinski (2016), we set $d_k = 0$ for each real variable, and $d_k = 0.8$ for the nominal variables. Further we fix $\varrho = 10^{-4}$ and $\zeta = 0.2$, a common choice for the tightness parameter of the Minnesota prior in the Bayesian VAR forecasting literature according to Carriero et al. (2015). Note that the prior in (3.7) is centered around zero implying no predictability. Finally, the ratios $\sigma_{x_k}^2 / \sigma_r^2 \forall k$ and $\sigma_{x_k}^2 / \sigma_{x_l}^2 \forall k, l$ account for differences in the scale and variability of the different predictor variables. $\sigma_{x_k}^2 \forall k$ and also σ_r^2 are approximated by the residual variances of an AR(1) regression for k-th variable and the asset return. The specification is completed by assuming an independent diffuse prior for Σ , $p(\Sigma) \propto |\Sigma|^{-2(2M+1)/2}$.⁴

Time-varying Bayesian VAR and stochastic volatility (TVP-BVAR with SV)

The literature provides various examples favoring equity prediction models with time-varying parameters (TVP) (Dangl and Halling, 2012), stochastic volatility (SV) (Johannes et al., 2014) and Bayesian model averaging techniques (Pettenuzzo and Ravazzolo, 2016). To evaluate the predictive performance, for example marginal likelihoods for individual models have to be easily available without great computational costs at each point of the forecasting period. While this may be so for simple constant parameter models through the use of conjugate priors, they are almost infeasible to obtain for large VAR models such as given in (3.3) with many parameters. The latter require informative priors to reduce estimation noise which rely on Markov Chain Monte Carlo (MCMC) methods for estimation at each point in time with typical tens of thousands of simulation draws to ensure convergence.

The same is true for time-varying parameter models with stochastic volatility that not

⁴Posterior results for the full model are obtained in a standard fashion and are omitted here for parsimony. The interested reader is referred to Koop and Korobilis (2010).

only require Kalman filtering for the regression coefficients but also computational costly sampling methods for the error term variances. To overcome the computational burden that arises in a recursive forecasting exercise, we adopt the so-called *forgetting factors* approach of [Koop and Korobilis \(2013\)](#) which also allows for all the features to model return predictability: Time-varying parameters, stochastic volatility, parameter shrinkage as well as dynamic model averaging and variable selection. Forgetting factors are used in state space models to allow for a moderate variation of the predictive variance over time. Let us consider a time-varying VAR version of (3.4) with stochastic volatility which can be expressed as follows:

$$y_t = a_t + A_{1,t} y_{t-1} + \varepsilon_t, \quad (3.10)$$

$$A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t, \quad (3.11)$$

where $A_t = [a_t \ A_{1,t}]$ is time-indexed for every parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of each other for all t and s . Here, ϕ is an unknown parameter governing the mean of A_t . While $\phi = 1$ implies a random walk behavior, $\phi = 0$ implies a random behavior of each A_t around \underline{A}_0 . Here, we will use the means of the Minnesota prior described in the previous section to specify \underline{A}_0 . Since ϕ adds another layer to the prediction model, the restrictions imposed on the coefficient matrix are relaxed compared to the constant coefficient model.

Typically, the estimation of the system (3.10) - (3.11) relies on MCMC techniques. Given the initial conditions A_0 , Σ_0 and Ω_0 , it involves drawing A_t conditional on Σ_t and Ω_t (e.g. through a Kalman filter), then drawing Σ_t conditional on A_t and Ω_t , the sampling Ω_t given A_t and Σ_t and eventually drawing further parameters given conditional on A_t , Σ_t , and Ω_t for all t . This is computationally demanding as it involves simulating Σ_t , and Ω_t for every $t = 1, \dots, T$. The idea of the forgetting factors here is to avoid simulating Ω_t recursively for each t . Instead, we avoid using Ω_t in the Kalman filter by approximating the one-step ahead predictor variance of $A_t|y^{t-1} \sim \mathcal{N}(A_{t|t-1}, P_{t|t-1})$, i.e. $P_{t|t-1}$, by the variance of the filtered estimator $A_{t-1}|y^{t-1} \sim \mathcal{N}(A_{t-1|t-1}, P_{t-1|t-1})$, i.e. $P_{t-1|t-1}$, divided by a *forgetting factor* $\lambda \in [0, 1]$. That is $P_{t|t-1} = P_{t-1|t-1}/\lambda$.⁵ Then, Ω_t is approximated by $(\lambda^{-1} - 1)P_{t-1|t-1}$. From this we can see that $\lambda = 1$ implies a constant coefficient model. Eventually, Σ_t is estimated recursively through an exponential weighted moving average using a decay factor κ between $\hat{\Sigma}_{t-1}$ and the variance-covariance matrix of filtered Kalman residuals, i.e. $\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1 - \kappa) \hat{\varepsilon}_t \hat{\varepsilon}'_t$, where $\hat{\varepsilon}_t = y_t - A_{t|t}[1 \ y_{t-1}]$ is obtained in the Kalman filter.⁶

⁵For textbook explanations of the Kalman filtering technique the reader is referred to for example [Durbin and Koopman \(2012\)](#).

⁶The details of the estimation of the model can be found in the Appendix 3.6.1.

The specification of the model involves a set of parameters, namely λ , κ and ϕ , that have to be defined by the prior, either through an hierarchical hyperprior, an empirical Bayes estimator or a search over a grid of possible values. Here, we estimate the model for every parameter combination over a grid and then choose the model with the highest predictive density over the recent past. We also consider an average over all models with different hyperparameter values.

Similar to [Koop and Korobilis \(2013\)](#), the *dynamic model selection and averaging* technique is performed over different priors and not different sets of predictor variables. The idea follows [Raftery et al. \(2010\)](#). In particular, the weights for model j , which comes from the j -th combination of λ , κ and ϕ , at time t using all the information up to $t - 1$ are given by

$$\omega_{t|t-1,j} = \omega_{t-1|t-1,j}^\alpha / \sum_{j=1}^J \omega_{t-1|t-1,j}^\alpha, \text{ and} \quad (3.12)$$

$$\omega_{t|t,j} = \omega_{t|t-1,j} p_j(y_t | y^{t-1}) / \sum_{j=1}^J \omega_{t|t-1,j} p_j(y_t | y^{t-1}), \quad (3.13)$$

where $p_j(y_t | y^{t-1})$ is the predictive likelihood of model j evaluated at y_t and $\alpha = 0.99$ is a decay factor governing the weighting of past observations. For monthly data, this value implies that the observations from about two years ago only receive approximately 80 percent of the weight of the observation in $t - 1$. We note that dynamic model weights imply a different treatment of every model in each period leading to different averaging results and also may lead to a different forecasting model selection in each period. Following [Koop and Korobilis \(2013\)](#), we perform model averaging across different prior parameter values. That is, $\lambda \in \{0.97, 0.98, 0.99, 1\}$, $\kappa \in \{0.94, 0.96, 0.98\}$ and $\phi \in \{0, 0.5, 0.75, 1\}$. This results in 48 models based on different model parameters from which we either select the best performing one or average across all of them.⁷

Eventually, we are interested in the marginal predictive distribution of the asset return r_t . This is a main advantage of the Bayesian approach ([Klein and Bawa, 1976](#); [Barberis, 2000](#)). The predictive distribution is obtained from the joint predictive density function of r_{t+1} and $\Theta_t = [A_t, \Sigma_t, \Omega_t]$ by integrating over all values of Θ_t . This is

$$f(r_{t+1} | y^t) = \int f(r_{t+1}, \Theta_t | y^t) d\Theta_t = \int f(r_{t+1} | y^t, \Theta_t) p(\Theta_t | y^t) d\Theta_t, \quad (3.14)$$

where $y^t = \{y_1, \dots, y_t\}$ is the collection of all past observations used for estimation. This function is independent of the unknown parameters and is in fact something like the

⁷The reader is referred to [Koop and Korobilis \(2013\)](#) for more details about the forecasting set-up and model selection.

average over all possible values for Θ_t . Numerically, it is obtained by simulating I draws from the posterior distribution and making a prediction \hat{r}_{t+1} for every posterior draw.

3.3.2 Entropic tilting

In addition to traditional point forecasts, the recent literature has considered probabilistic (or “density”) forecasts of macroeconomic and financial variables. In contrast to point forecasts, the latter provide information on various possible scenarios and thus quantify the uncertainty surrounding the future. In the Bayesian methodology, predictive distributions for the variables of interest are easily obtained by integrating out the parameter uncertainty from the likelihood function (evaluated at a future (predicted) realization) times the posterior distribution.

Similar to point forecasts, density forecasts can also be combined to form a merged model that upholds the strengths of each of its components. This can be achieved for example by mixing two densities or by reweighting a forecasted density according to another model; ensuring that the new mixture model is well defined. Entropic tilting is a non-parametric method to combine time-series model forecasts with information from other origins. We now explain the method in more detail.

Suppose at time t we want to make a forecast h periods ahead for a $N \times 1$ vector of interest r_{t+h} , in our case a vector of out-of-sample excess stock returns. Denote by $f_{t,h} := \{r_{t+h,i}\}_{i=1}^I$, where $r_{t+h} \in \mathbb{R}^N$ and $N \geq 1$, a baseline sample from the predictive return distribution $p(r_{t+h}|r^t)$, i.e. a discrete sample of I (MCMC) draws of the h -step ahead forecasts. These draws can either come from a closed-form analytical expression of the predictive density $f_{t,h}$ or might be simulated. It also may depend on estimated parameters.

We now want to incorporate additional information about the return r_{t+h} , which was not used to generate the base sample, in the form of M moment conditions on the function $g(r_{t+h}) : \mathbb{R}^N \rightarrow \mathbb{R}^M$ in the following sense:

$$\mathbb{E}[g(r_{t+h})] = \bar{g}_t, \quad (3.15)$$

where $\bar{g}_t \in \mathbb{R}^M$ and $M, N \geq 1$. For example $g(r_{t+h}) = r_{t+h}$ imposes that the mean of r_{t+h} is equal to \bar{g}_t and $g(r_{t+h}) = (r_{t+h} - \mathbb{E}(r_{t+h}))^2$ sets the variance equal to it. \bar{g}_t can be formed from various origins: [Giacomini and Ragusa \(2014\)](#) use an Euler equation to specify \bar{g}_t , [Altavilla et al. \(2014\)](#); [Krüger et al. \(2015\)](#) use survey forecasts and [Metaxoglou et al. \(2016\)](#) adopt option-implied information for \bar{g}_t .

In general under the base density $f_{t,h}$, the moments of $g(r_{t+h})$ are not equal to \bar{g}_t :

$$\mathbb{E}_{f_{t,h}} [g(r_{t+h})] = \int g(r_{t+h}) f_{t,h}(r_{t+h}) dr_{t+h} \neq \bar{g}_t. \quad (3.16)$$

Instead, entropic tilting describes finding the density $\tilde{f}_{t,h}$ out of the set of densities that fulfill the moment condition in (3.15) that is closest to the base density in terms of the Kullback-Leibler divergence measure. This is formalized in the following proposition.

Proposition 3.3.1

If a solution $\tilde{f}_{t,h}(r)$ to the constrained minimization

$$\min_{\tilde{f}_{t,h} \in \mathcal{F}} \mathbb{E}_{\tilde{f}_{t,h}} \left[\log \frac{\tilde{f}_{t,h}(r)}{f_{t,h}(r)} \right] = \int \log \frac{\tilde{f}_{t,h}(r)}{f_{t,h}(r)} \tilde{f}_{t,h}(r) dr, \quad (3.17)$$

$$\text{s.t. } \mathbb{E}_{\tilde{f}_{t,h}} [g(r)] = \int g(r) \tilde{f}_{t,h}(r) dr = \bar{g}_t, \quad (3.18)$$

exists, then it is unique and it is given by

$$\tilde{f}_{t,h}^*(r) = f_{t,h}(r) \exp \left(\gamma_{t,h}^{*''} g(r) \right) / \int \exp \left(\gamma_{t,h}^{*''} g(r) \right) f_{t,h}(r) dr, \quad (3.19)$$

$$\gamma_{t,h}^* = \arg \min_{\gamma_{t,h}} \int f_{t,h}(r) \exp \left(\gamma_{t,h}' (g(r) - \bar{g}_t) \right) dr. \quad (3.20)$$

Proof. The proof is given in [Giacomini and Ragusa \(2014\)](#) Proposition 1 on page 147. \square

While $\tilde{f}_{t,h}^*(r)$ is generally not of a known form, the entropic tilting problem can also be interpreted as finding a new sets of weights $\pi_{t,h}^*$ in t for the base h -step ahead density $f_{t,h}(r)$ that satisfy the moment condition. For a sample of I draws from the base predictive density, the expectation in (3.17) is

$$\mathbb{E}_{\tilde{f}_{t,h}} \left[\log \frac{\tilde{f}_{t,h}(r)}{f_{t,h}(r)} \right] = \sum_{i=1}^I \tilde{\pi}_i \log \left(\frac{\tilde{\pi}_i}{\pi_i} \right) \stackrel{\pi_i=1/I}{=} \log I + \sum_{i=1}^I \tilde{\pi}_i \log (\tilde{\pi}_i), \quad (3.21)$$

where π_i , $i = 1, \dots, I$, are the original weights for the base density usually equal to $1/I$. Following [Robertson et al. \(2005\)](#), imposing the condition (3.18) via $\mathbb{E}_{\tilde{f}_{t,h}} [g(r)] =$

$\sum_{i=1}^I \tilde{\pi}_i g(r_{t,i})$ yields the tilting solution from (3.19) and (3.20) as

$$\pi_i^* = \frac{\exp(\gamma_{t,h}' g(r_{t+h,i}))}{\sum_{i=1}^I \exp(\gamma_{t,h}' g(r_{t+h,i}))}, \quad (3.22)$$

$$\gamma_{t,h}^* = \arg \min_{\gamma_{t,h}} \sum_{i=1}^I \exp(\gamma_{t,h}'(g(r_{t+h,i}) - \bar{g}_t)). \quad (3.23)$$

Equation (3.22) ensures that all elements of the new weight vector $\pi_{t,h}^*$ are positive and sum up to one. $\gamma_{t,h}^*$ in (3.23) has dimension M (the number of moment conditions) and can easily be found by a Lagrangian optimization.

The moment condition in (3.15) restricts the set of possible candidate densities. Hence, the usefulness or uncertainty about the additional information is not measured. Moreover, the more moment conditions exists, the smaller is the set of candidate distributions.

Entropic tilting also has a shrinkage interpretation (Robertson et al., 2005, p. 394): Given a certain mean condition on the target random variable, imposing higher moment conditions that *shrink* the variance of the target variable to zero, sets the mean automatically to the imposed target mean. In other words, \bar{g}_t can be interpreted as a shrinkage target for the entire predictive return distribution, achieved through re-weighting every single draw of it and that changes its moments. This can also be seen from considering the following example: Let y follow a bivariate normal distribution with $f(y) = N(\theta, \Sigma)$ and impose the restriction that the mean of the second variable y_2 is μ_2 and its variance is Ω_{22} . Then it follows that the tilted distribution is also normal $\tilde{f}^*(y) = N(\mu, \Omega)$ and the mean of y_1 is given by

$$\mu_1 = \theta_1 + \Sigma_{22}^{-1} \Sigma_{12} (\mu_2 - \theta_2) = \lambda \theta_1 + (1 - \lambda) \underbrace{\left(\theta_1 + \frac{\Sigma_{22}^{-1} \Sigma_{12} (\mu_2 - \theta_2)}{1 - \lambda} \right)}_{=: \tilde{\theta}_1}. \quad (3.24)$$

Here, $\tilde{\theta}_1$ is the shrinkage target that depends implicitly on the distance between the mean condition for second variable μ_2 and the true mean θ_2 .

Entropic tilting only changes the location and shape of predictive return distribution, but it does not foster better parameter estimates for the underlying prediction model. It is therefore not equivalent to putting an informative prior on the mean and variance of the prediction model centered at the analysts' forecasts. Also, it does not imply a structural relationship between the asset returns and the analysts' forecasts as the approach of Frey and Mokinski (2016), who augment a Bayesian VAR system by survey nowcasts and impose parameter restrictions between the original variables in the VAR and the added equations for the nowcasts through the prior.

3.4 Empirical Application

3.4.1 Data and set-up

All data used in the empirical exercise was obtained from Thomson Reuters Datastream. We will investigate the forecast performance of various prediction models for 20 Dow Jones constituents for which the Institutional Brokers Estimate System (I/B/E/S) database provide target prices and analyst recommendations. We use monthly observations because the I/B/E/S summary data is aggregated on this frequency. Historical target price data from I/B/E/S is available from April 1999 to October 2014. The initial estimation window has size $h = 60$ and hence pseudo out-of-sample evaluation period starts in April 2004. We only consider one-step ahead predictions in this study. For each stock, we compute logarithmic returns (including dividends) and subtract the 3-month T-bill rate to obtain excess returns. For the choice of the predictor variables, we partly follow [Welch and Goyal \(2008\)](#); [Pettenuzzo and Ravazzolo \(2016\)](#) and (i) consider firm specific *fundamentals* such as the log dividend yield, the log earnings price ratio, the log dividend-payout ratio, and the book-to-market ratio, (ii) market and economic measures such as the 3-month T-bill rate, the yield on long-term government bonds, the market excess return and CPI inflation.

Thomson Reuters I/B/E/S database provides detailed and consensus estimates featuring up to 26 forecast measures for more than 70,000 companies in more than 90 countries worldwide. While, the summary files contain a monthly snapshot of each company followed by sell-side analysts whose brokerage firm provides data to I/B/E/S, the detail files offer forecasts from individual analysts. We are mainly interested in two variables:

1. **Price targets:** The mean and standard deviation of the projected price level forecasted by professional analysts with a 12-month time horizons. At each point in time, we also consider the entire vector of target prices from individual analysts.
2. **Recommendations summary:** The mean and standard deviation of analysts' recommendation based on a five point standardized scale (strong buy = 1, buy, hold, sell, strong sell = 5) as well as the total number of recommendations, the number of up- and downgrade revisions and the percentage of buy, hold and sell recommendations.

We will use the target price returns and variance as well as the consensus analyst recommendations and their revisions as predictor variables. In particular, target prices will be used to calculate (i) monthly forward target price implied expected return, i.e. simple returns between the spot and the twelve months forward target price at each point t divided by twelve, and (ii) monthly target price implied expected return variances. For this, we used the detail history I/B/E/S files that contain target prices of individual

analysts. We first calculate monthly forward target price implied expected returns for every individual analyst and then use the mean and variance of these returns as first and second moment restrictions for the entropic tilting exercise.

Table 3.2 reports the descriptive statistics on the returns, expected target returns and recommendations for the 20 Dow Jones constituents used in this study. While we note that the mean and standard deviations of the logarithmic returns differ substantially across assets, the target price and recommendation characteristics are very similar: All mean forward target price implied expected returns are positive, indicating an upward bias in the target prices compared to the spot prices. The standard deviation of the expected target price returns is generally slightly lower in scale than the standard deviation of the logarithmic stock returns. Moreover, the upward bias can also be seen in the mean recommendations, which is always smaller than three, given the five point scale. A lower score indicates more buy recommendations. The company with the least number target prices of 11 and number of recommendations with 16 is DuPont (DD) and the company with the greatest number of target prices, 28, and recommendations, 39, is Intel (INTC).

3.4.2 Competing models

We consider a number of different models to distinguish the effects, considering different sets of predictor variables, models with constant and time-varying coefficients as well as mean and variance tilted models. These are

1. [AR1] Autoregressive model of order one for the return process of each asset.
2. [VAR-Full] Bayesian vector autoregressive model of order one with an uninformative prior for all parameters (full specification).
3. [VAR-Minnesota] Bayesian vector autoregressive model of order one with the Minnesota prior given in (3.5) - (3.9) to impose the full model for the return and an autoregressive model for all other variables.
4. [TVPVAR-SV-DMA] Time-varying parameter model with stochastic volatility and using forgetting factors and dynamic model averaging over different prior parameters as described in Section 3.3.1.
5. [TVPVAR-SV-DMS] Time-varying parameter model with stochastic volatility and using forgetting factors and dynamic model selection over different prior parameters as described in Section 3.3.1.
6. [TVPVAR-SV-DMAm] TVP-BVAR with SV using dynamic model averaging with mean tilting using the monthly target price implied expected returns.

7. [TVPVAR-SV-DMAm/v] TVP-BVAR with SV using dynamic model averaging with mean and variance tilting using the monthly target price implied expected returns.
8. [TVPVAR-SV-DMSm] TVP-BVAR with SV using dynamic model selection with mean tilting using the monthly target price implied expected returns.
9. [TVPVAR-SV-DMSm/v] TVP-BVAR with SV using dynamic model selection with mean and variance tilting using the monthly target price implied expected returns.
10. [Bayesian lasso] The Bayesian lasso ([Park and Casella, 2008](#)) is a shrinkage for univariate regressions based on the Laplace prior that can be used to impose an L_1 -norm penalization on the regression coefficients to shrink them to zero. It has been shown to be a strong forecasting device ([Korobilis, 2013](#)) and is used here as a benchmark model.

3.4.3 Evaluation criteria

We consider two evaluation criteria in this study. To evaluate the entire predictive accuracy of point forecasts we use the out-of-sample R^2 . For model j , it is given by

$$R_{OoS,j}^2 = 1 - \frac{\sum_{i=h+1}^T e_{j,i}^2}{\sum_{i=h+1}^T e_{0,i}^2}, \quad (3.25)$$

where $e_{0,i} = y_i^1 - \hat{y}_{0,i}^1 = r_i - \hat{r}_{0,i}$ denotes the forecast error of a simple mean or intercept only model ($r_i = a + \varepsilon_i$) and $e_{j,i}$ the forecast error in the returns of model j at time i and h denotes the end-point of the initial estimation period. Note that we only evaluate the forecast error with respect to the asset return and not with regard to all predictor variables. All errors are obtained by averaging over the (marginal) predictive density function of the asset returns. Values above zero indicate that model j produces lower forecasts error than the intercept only model. Second, to evaluate the predictive distribution, we consider the average log score differential (LSD) given by

$$LSD_{j,t} = \frac{\sum_{i=h+1}^t (LS_{j,i} - LS_{0,i})}{\sum_{i=h+1}^t LS_{0,i}}, \quad (3.26)$$

where where $LS_{j,i}$ is log predictive score of model j at time i . Again, values above zero indicate that a given model j shows better forecast performance than the benchmark model, while negative values suggest the opposite.

3.4.4 Individual predictor performance

Before heading to the main analysis using all predictors, we start by looking at models with only a single predictor in order to investigate which ones have particular forecasting power. For this, we apply a constant parameter model and a time-varying model with stochastic volatility to estimate the system given in (3.1) - (3.2).⁸ Tables 3.3 to 3.6 provide the results for the out-of-sample R² and Tables 3.7 to 3.10 provide the results in terms of the log predictive scores. Bold numbers show positive performance measures, indicating that the one predictor model outperforms the simple intercept only model. We test statistical significance using the Diebold and Mariano (1995) t-tests with a null hypothesis of equal average forecasting ability.

The results can be summarized as follows: (i) No predictor variable shows significant forecast performance using a constant parameter model (Tables 3.3 and 3.7). (ii) Using the time-varying parameter model with stochastic volatility increases forecast performance for all predictors, most profoundly for the general economic and market indicators (Tables 3.4 and 3.8). This may be due to the model flexibility which can reflect the nature of asset returns (e.g. heteroskedasticity) much better (see also Johannes et al., 2014). (iii) Tilting the predictive distribution from the TVP-BVAR towards the target price implied expected returns does not increase forecast performance (Tables 3.5 and 3.9) significantly. (iv) However, tilting the predictive distribution from the TVP-BVAR to posses the mean and variance of the target price implied expected returns does indeed increase forecast performance, but only significantly for a couple of assets and predictors, especially for the models using the target price return and variance as the predictor variables (Tables 3.6 and 3.10).

To illustrate the effect of the tilting, we plot the baseline and the tilted predictive densities for the IBM stock returns at different times. Figure 3.3 compares the baselines density against the mean tilted density for the TVPVAR(1) model. It is obvious that the baseline and tilted densities are very similar to each other, only in 2010 and 2012 the mode of tilted distribution seems to be closer to the actual outcome. The similarity in the two distributions thus explains the lack of forecast improvements from mean tilting. The picture changes when looking at the tilted density of the mean and variance tilted models. In Figure 3.4, we see that the tilted distribution is much more concentrated around the actual outcome in calm market times (2006, 2012) and more flat and has fatter tails than the baseline density in crises times between 2008 and 2010. In the entire sample, the tilted densities are more often stronger concentrated around the true observation than the baseline. This concentration in the density comes from the agreement of the analyst about the future target price and does not imply that the analysts make unbiased forecasts.

⁸For the latter, the underlying model parameters are set to $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$.

3.4.5 Complete model performance

None of the individual predictor models outperformed the intercept only model consistently across all assets. Instead, the individual predictor exercise revealed that especially the time-varying parameters and the stochastic volatility in connecting with the mean and variance tilting can produce significantly better forecasts. Therefore, we now put all predictors in the model and perform dynamic model averaging and selection using the full VAR system for the 13 variables (12 predictors plus the asset return). This leads to the main results of this paper which can be found in Tables 3.11 and 3.12. While the AR(1) model cannot outperform the mean model, the full BVAR system with uninformative priors clearly gives worst forecasts. This deterioration is likely to reflect over-fitting from the great number of estimable parameters. The Minnesota prior (shrinking many parameters to zero here) instead improves forecast performance, although not significantly. Its performance is similar to the Bayesian lasso. The time-varying models with dynamic model averaging and selection clearly improve the forecast performance further. However, the difference between model averaging and selection is only marginal. Again, tilting the mean of the predictive distribution towards the target price implied return does not improve the forecast further. This is only achieved when also tilting the variance towards the target price implied return variance, giving significant better forecasts for a wide range of assets.

3.5 Concluding Remarks

In this paper we demonstrate that financial analyst forecasts can have predictive power for equity assets of the Dow Jones Industrial index using a novel entropic tilting approach to combine time-series forecasts with analysts' information. We find that the extent of predictability varies across assets and that using Bayesian vector autoregressions with time-varying coefficients, stochastic volatility and model averaging and selection among priors improves return predictions across all assets. While tilting the mean of the predictive distribution towards the target price implied expected returns did not improve forecast performance significantly, tilting the variance of the predictive return distribution towards the implied expected target return variance produced forecasts outperforming a simple intercept only model. This may be explained by the fact that the tilted densities are more often stronger concentrated around the true outcome than the baseline density. In other words, the agreement among analysts reduces the predictive variance of the asset returns. Contrary, the disagreement among the analysts may be an indicator for future market uncertainties.

Using entropic tilting has several advantages: It can incorporate any kind of (forward-

looking) information into a predictive regression framework in a parsimonious way without increasing estimation noise. Hence, it is a regularization approach that is suitable in high-dimensional settings, even portfolio problems. Notably, by using the analyst information in a tilting framework we only change the simulated draws of the predictive distribution and hence we do not require the data to formalize and estimate the true relationship between asset returns and analyst target prices. However, while in this way the dimensionality of the forecasting model, in this case of a Bayesian vector autoregression, is unchanged, we do not account for the stochastic nature of the tilting information.

This opens the door for possible extensions: While one could develop a tilting framework that not only considers the set of predictive distributions that strictly fulfill the moment conditions, one could try to search among all possible predictive distributions and then minimize the distance of the target moments given a statistic or economic criteria. Second, one could also apply the tilting framework to predictive portfolio weight regressions. For example, the framework of (Frey and Pohlmeier, 2016) may be used to incorporate external knowledge directly about the portfolio weights, i.e. a certain target allocation, instead of tilting the underlying return process. Finally, the tilting approach is also applicable for panel VAR systems that model various assets simultaneously. Tilting the joint predictive return distributions may then be used in a portfolio allocation problem of an expected utility maximizing investor.

References

- ALTAVILLA, C., R. GIACOMINI, AND G. RAGUSA (2014): “Anchoring the Yield Curve Using Survey Expectations,” ECB Working Paper No. 1632, ECB.
- AVRAMOV, D. (2002): “Stock return predictability and model uncertainty,” *Journal of Financial Economics*, 64, 423–458.
- BARBER, B., R. LEHAVY, M. McNICHOLS, AND B. TRUEMAN (2001): “Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns,” *The Journal of Finance*, 56, 531–563.
- BARBERIS, N. (2000): “Investing for the Long Run when Returns are Predictable,” *The Journal of Finance*, 55, 225–264.
- BLACK, F. AND R. LITTERMAN (1992): “Global Portfolio Optimization,” *Financial Analysts Journal*, 48, 28–43.
- BONINI, S., L. ZANETTI, R. BIANCHINI, AND A. SALVI (2010): “Target Price Accuracy in Equity Research,” *Journal of Business Finance & Accounting*, 37, 1177–1217.
- BRADSHAW, M. T., L. D. BROWN, AND K. HUANG (2012): “Do Sell-Side Analysts Exhibit Differential Target Price Forecasting Ability?” *Review of Accounting Studies*, 18, 930–955.
- BRAV, A. AND R. LEHAVY (2003): “An Empirical Analysis of Analysts’ Target Prices: Short-Term Informativeness and Long-Term Dynamics,” *The Journal of Finance*, 58, 1933–1967.
- CARRIERO, A., T. E. CLARK, AND M. MARCELLINO (2015): “Bayesian VARs: Specification Choices and Forecast Accuracy,” *Journal of Applied Econometrics*, 30, 46–73.
- CREMERS, K. (2002): “Stock return predictability: a Bayesian model selection perspective,” *The Review of Financial Studies*, 15, 1223–1249.
- CVITANIĆ, J., A. LAZRAK, L. MARTELLINI, AND F. ZAPATERO (2006): “Dynamic Portfolio Choice with Parameter Uncertainty and the Economic Value of Analysts’ Recommendations,” *Review of Financial Studies*, 19, 1113–1156.
- DA, Z. AND E. SCHAUUMBURG (2011): “Relative Valuation and Analyst Target Price Forecasts,” *Journal of Financial Markets*, 14, 161–192.
- DANGL, T. AND M. HALLING (2012): “Predictive regressions with time-varying coefficients,” *Journal of Financial Economics*, 106, 157–181.

- DIEBOLD, F. X. AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 20, 134–144.
- DOAN, T., R. LITTERMAN, AND C. SIMS (1984): “Forecasting and Conditional Projection using Realistic Prior Distributions,” *Econometric Reviews*, 3, 1–100.
- DURBIN, J. AND S. KOOPMAN (2012): *Time Series Analysis by State Space Methods*, Oxford Statistical Science, Oxford University Press, 2 ed.
- FENG, G. AND N. G. POLSON (2016): “Regularizing Bayesian Predictive Regressions,” *arXiv:1606.01701 [stat]*.
- FREY, C. AND F. MOKINSKI (2016): “Forecasting with Bayesian Vector Autoregressions Estimated Using Professional Forecasts,” *Journal of Applied Econometrics*, 31, 1083–1099.
- FREY, C. AND W. POHLMEIER (2016): “Bayesian Shrinkage of Portfolio Weights,” Available at SSRN 2730475.
- GIACOMINI, R. AND G. RAGUSA (2014): “Theory-Coherent Forecasting,” *Journal of Econometrics*, 182, 145–155.
- GIACOMINI, R. AND H. WHITE (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74, 1545–1578.
- GREEN, T. C. (2006): “The Value of Client Access to Analyst Recommendations,” *Journal of Financial and Quantitative Analysis*, 41, 1–24.
- HE, P. W., A. GRANT, AND J. FABRE (2013): “Economic value of analyst recommendations in Australia: an application of the Black–Litterman asset allocation model,” *Accounting & Finance*, 53, 441–470.
- IRVINE, P. J. (2003): “The Incremental Impact of Analyst Initiation of Coverage,” *Journal of Corporate Finance*, 9, 431–451.
- JEGADEESH, N., J. KIM, S. D. KRISCHE, AND C. M. C. LEE (2004): “Analyzing the Analysts: When Do Recommendations Add Value?” *The Journal of Finance*, 59, 1083–1124.
- JEGADEESH, N. AND W. KIM (2006): “Value of Analyst Recommendations: International Evidence,” *Journal of Financial Markets*, 9, 274–309.
- JOHANNES, M., A. KORTEWEG, AND N. POLSON (2014): “Sequential Learning, Predictability, and Optimal Portfolio Returns,” *The Journal of Finance*, 69, 611–644.

- KANDEL, S. AND R. STAMBAUGH (1996): “On the Predictability of Stock Returns: An Asset-Allocation Perspective,” *Journal of Finance*, 51, 385–424.
- KANDEL, S. AND R. F. STAMBAUGH (1987): “Long-horizon returns and short-horizon models,” CRSP Working paper 222, University of Chicago.
- KLEIN, R. AND V. BAWA (1976): “The Effect of Estimation Risk on Optimal Portfolio Choice,” *Journal of Financial Economics*, 3, 215–231.
- KOOP, G. AND D. KOROBILIS (2010): “Bayesian Multivariate Time Series Methods for Empirical Macroeconomics,” *Foundations and Trends® in Econometrics*, 3, 267–358.
- (2013): “Large time-varying parameter VARs,” *Journal of Econometrics*, 177, 185–198.
- KOROBILIS, D. (2013): “Hierarchical shrinkage priors for dynamic regressions with many predictors,” *International Journal of Forecasting*, 29, 43–59.
- KRÜGER, F., T. E. CLARK, AND F. RAVAZZOLO (2015): “Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts,” *Journal of Business & Economic Statistics*, 1–48.
- LIN, S., H. TAN, AND Z. ZHANG (2016): “When Analysts Talk, Do Institutional Investors Listen? Evidence from Target Price Changes,” *Financial Review*, 51, 191–223.
- METAXOGLOU, K., D. PETTENUZZO, AND A. SMITH (2016): “Option-Implied Equity Premium Predictions via Entropic Tilting,” Available at SSRN 2754246.
- PARK, T. AND G. CASELLA (2008): “The Bayesian Lasso,” *Journal of the American Statistical Association*, 103, 681–686.
- PÁSTOR, Ľ. (2000): “Portfolio Selection and Asset Pricing Models,” *The Journal of Finance*, 55, 179–223.
- PETTENUZZO, D. AND F. RAVAZZOLO (2016): “Optimal Portfolio Choice Under Decision-Based Model Combinations,” *Journal of Applied Econometrics*, forthcoming.
- PETTENUZZO, D. AND A. TIMMERMANN (2011): “Predictability of stock returns and asset allocation under structural breaks,” *Journal of Econometrics*, 164, 60–78.
- RAFTERY, A. E., M. KÁRNÝ, AND P. ETTLER (2010): “Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill,” *Technometrics*, 52, 52–66.

- RAMNATH, S., S. ROCK, AND P. SHANE (2008): “The financial analyst forecasting literature: A taxonomy with suggestions for further research,” *International Journal of Forecasting*, 24, 34–75.
- RAPACH, D., J. STRAUSS, J. TU, AND G. ZHOU (2015): “Industry interdependencies and cross-industry return predictability,” *Available at SSRN 2566541*.
- RAPACH, D. AND G. ZHOU (2013): “Chapter 6 - Forecasting Stock Returns,” Elsevier, vol. 2, Part A of *Handbook of Economic Forecasting*, 328–383.
- ROBERTSON, J. C., E. W. TALLMAN, AND C. H. WHITEMAN (2005): “Forecasting Using Relative Entropy,” *Journal of Money, Credit and Banking*, 37, 383–401.
- WACHTER, J. A. AND M. WARUSAWITHARANA (2015): “What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio,” *Journal of Econometrics*, 186, 74–93.
- WELCH, I. AND A. GOYAL (2008): “A Comprehensive Look at The Empirical Performance of Equity Premium Prediction,” *Review of Financial Studies*, 21, 1455–1508.
- ZELLNER, A. AND V. CHETTY (1965): “Prediction and Decision Problems in Regression Models from the Bayesian Point of View,” *Journal of the American Statistical Association*, 60, 608–616.

3.6 Appendix

3.6.1 Estimation of the TVP-BVAR with SV using forgetting factors

We consider the model

$$y_t = a_t + A_{1,t} y_{t-1} + \varepsilon_t, \quad (3.27)$$

$$A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t, \quad (3.28)$$

where $A_t = [a_t \ A_{1,t}]$ is an unknown state vector, \underline{A}_0 is some initial condition for each t and also $A_0 = \underline{A}_0$, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$ with initial condition Σ_0 , $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ with initial condition Ω_0 and ε_t and u_s are independent of each other for all t and s . To estimate the mode, we run a Kalman filter for $t = 1, \dots, T$ as follows:⁹

I. Prediction step:

1. Set $A_{t|t-1} = \phi A_{t-1|t-1} + (1 - \phi) \underline{A}_0$.
2. Set $P_{t|t-1} = \frac{1}{\lambda} P_{t-1|t-1}$
where for $t = 1$ we set $A_{0|0} = \underline{A}_0$ and $P_{0|0} = \underline{P}_0$.

II. Update step:

1. Calculate $\tilde{\varepsilon}_t = y_t - a_{t|t-1} + A_{t|t-1} y_{t-1}$.
2. Calculate $\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1 - \kappa) \tilde{\varepsilon}_t' \tilde{\varepsilon}_t$ with $\hat{\Sigma}_1 = \kappa \Sigma_0$.
3. Estimate $A_{t|t} = A_{t|t-1} + P_{t|t-1} [1 y_{t-1}]' \left(\hat{\Sigma}_t + [1 y_{t-1}] P_{t|t-1} [1 y_{t-1}]' \right)^{-1} \tilde{\varepsilon}_t$.
4. Calculate $P_{t|t} = P_{t|t-1} + P_{t|t-1} [1 y_{t-1}]' \left(\hat{\Sigma}_t + [1 y_{t-1}] P_{t|t-1} [1 y_{t-1}]' \right)^{-1} P_{t|t-1}$.

The one-step ahead predictive density of the VAR model is then analytically available from the Kalman filter as

$$p(y_t | y^t) \sim \mathcal{N}\left([1 \ y_{t+1}] A_{t+1|t}, \hat{\Sigma}_{t+1} + [1 \ y_{t+1}] A_{t+1|t} [1 \ y_{t+1}]'\right). \quad (3.29)$$

⁹The algorithm is taken and amended from the technical appendix of the working paper version of Koop and Korobilis (2013).

3.6.2 Figures

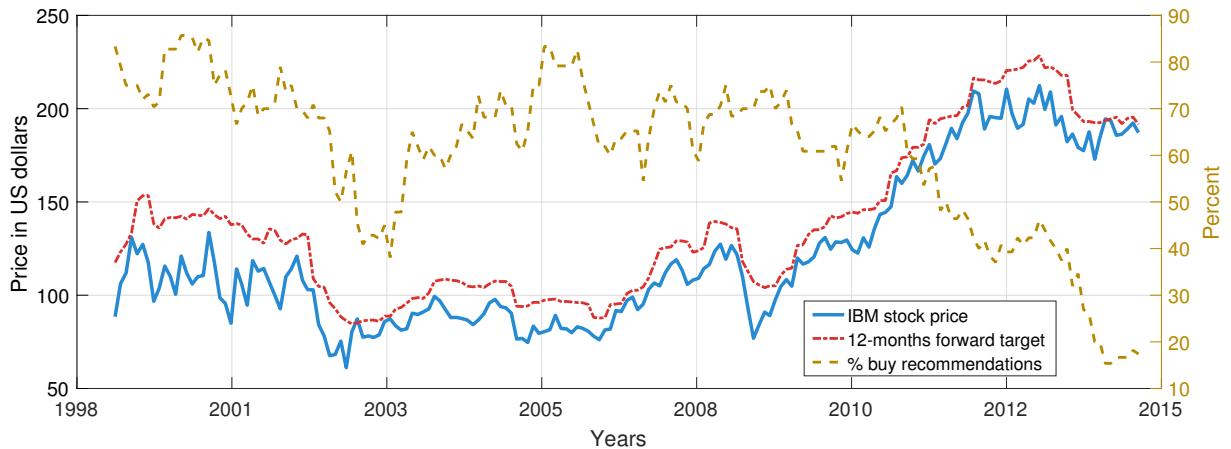


Figure 3.1: The figure shows IBM spot price, the mean 12 months forward target price and the percentage of buy recommendations from all recommendations (buy, sell, hold) of the IBM stock between 1999 and 2015.

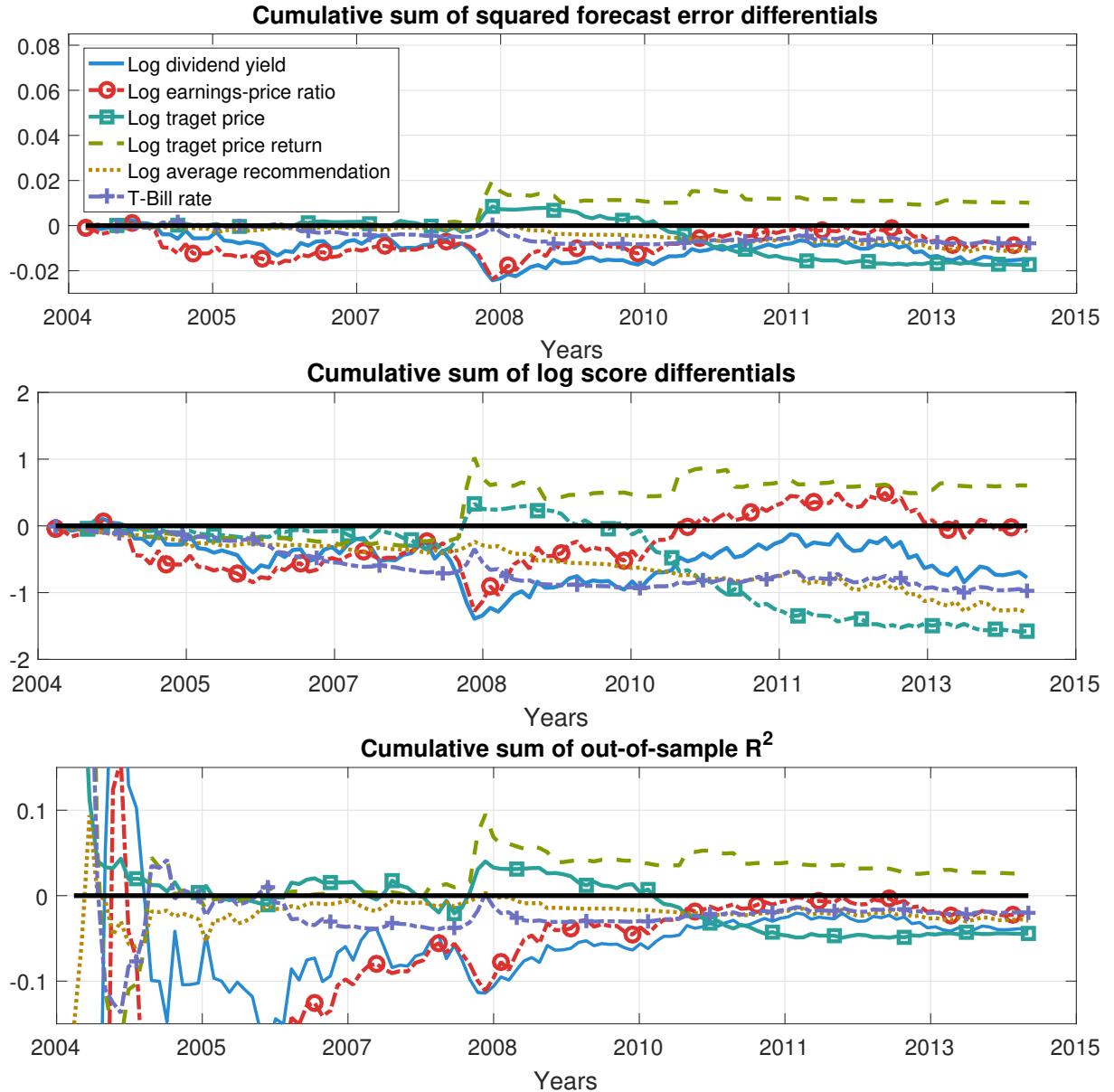


Figure 3.2: The figure provides out-of-sample forecast performance results for different univariate models for the IBM stock for 2004 to 2015. The top panel shows the cumulative sum of squared forecast errors of the benchmark mean model minus the sum of squared forecast errors for six univariate models with different regressors, i.e. for model m this is $\text{CSSED}_{m,t} = \sum_{i=S+1}^t (e_{0,i}^2 - e_{m,i}^2)$. Each model is estimated from a linear regression of monthly excess returns on an intercept and a lagged predictor variable, i.e. $r_t = \alpha + \beta x_{t-1} + \varepsilon_t$. The middle panel shows the cumulative sum of log predictive scores of the six models minus the sum of log predictive scores of the benchmark mean model. The bottom panel shows the cumulative sum of out-of-sample R^2 values of each of the six univariate models. For all three panels it holds that values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite.

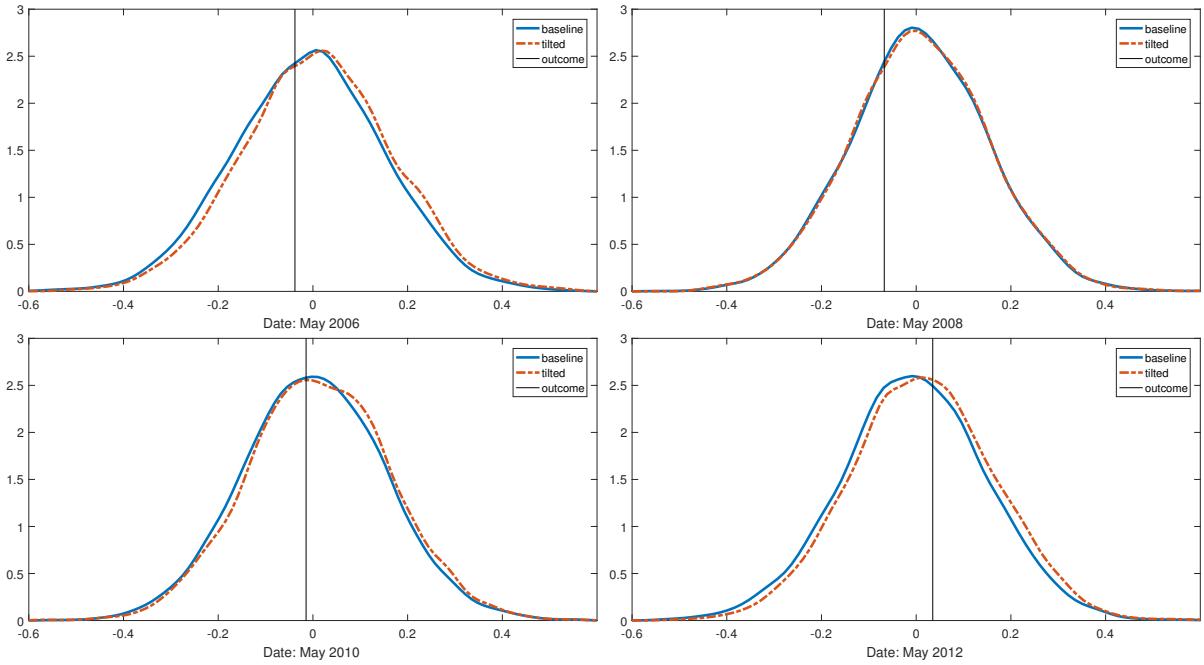


Figure 3.3: The figure shows the kernel of the predictive density of the IBM returns from the TVPVAR(1) model with dynamic model averaging and mean tilting towards the target price implied expected return at different times. The black horizontal line indicates the actual outcome return)

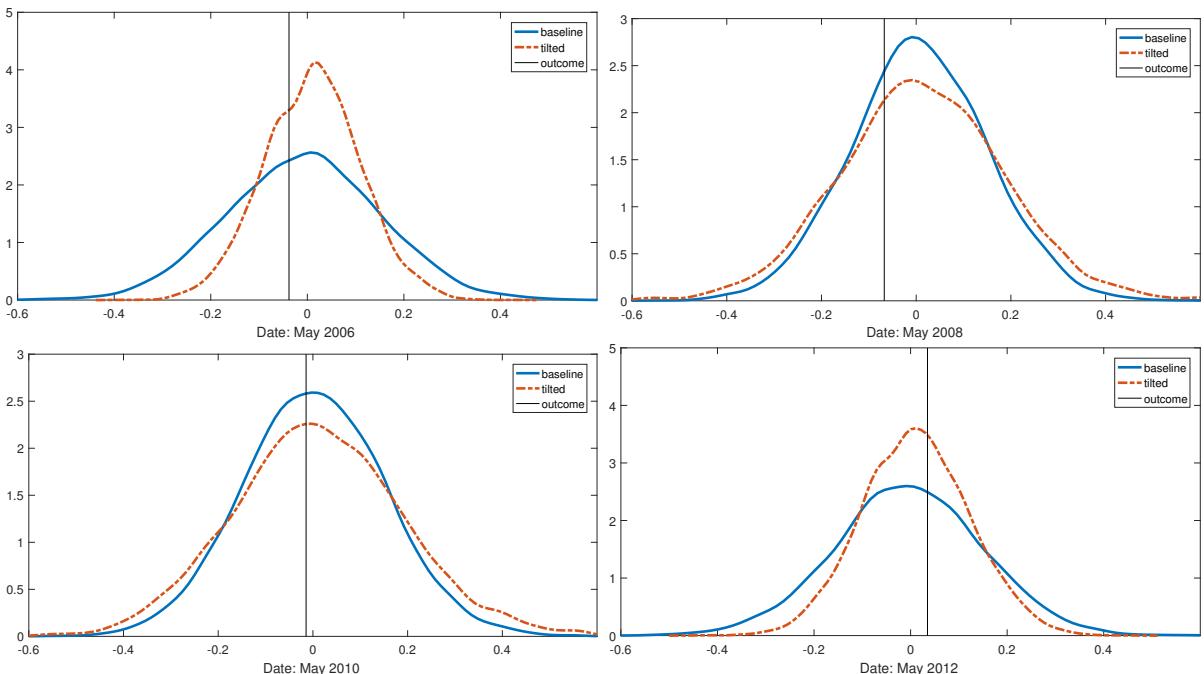


Figure 3.4: The figure shows the kernel of the predictive density of the IBM returns from the TVPVAR(1) model with dynamic model averaging with tilting towards the mean and variance of the target price implied expected returns at different times. The black horizontal line indicates the actual outcome return)

3.6.3 Tables

Table 3.1: Relative root mean squared errors between forecasted and observed spot prices for 20 Dow Jones constituents (sample: 1999 - 2015)

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
rRMSFE	1.33**	0.74	1.45***	0.93	0.87	0.96	1.09	1.24**	1.23	0.83
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
rRMSFE	1.36	0.94*	0.85	0.69***	1.25*	1.65***	1.01	0.82	1.24	0.84

Note: The table displays relative root mean squared errors between observed spot price twelve months ahead and the mean 12-month forward target price as well as the two year historical average for 20 Dow Jones constituents between 1999 and 2015. Values lower than one indicate that the target price generates superior forecast performance. For each stock, we test whether the target price forecast has lower MSFE than the average price forecast by the test proposed by Giacomini and White (2006). One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.2: Descriptive statistics on the returns, target prices and recommendations for 20 Dow Jones constituents (sample: 1999 - 2015)

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Mean log ret	-0.52	2.01	-1.79	0.26	0.41	0.43	-0.05	-0.18	-0.35	0.27
Std log return	12.00	14.13	21.28	9.52	8.95	10.13	6.03	8.27	8.74	8.22
# price targets	14.38	25.19	13.17	16.61	16.18	13.99	13.11	11.81	13.33	18.39
Mean exp ret	1.44	1.44	2.52	1.12	1.03	1.08	0.95	1.30	1.31	1.10
Std exp ret	10.72	7.49	25.95	6.08	7.04	7.20	4.57	5.79	6.07	6.37
# RECs	18.40	32.33	20.19	21.36	23.18	20.09	17.73	16.48	17.07	25.76
Mean RECs	2.35	2.13	2.19	2.36	2.26	2.29	2.11	2.40	2.02	2.11
Std RECs	0.38	0.38	0.58	0.35	0.38	0.25	0.25	0.28	0.34	0.25
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Mean log ret	-0.13	0.14	0.24	0.27	-0.24	-0.09	0.14	0.39	0.10	0.40
Std log return	11.57	7.76	5.26	6.54	7.88	8.90	5.78	7.25	5.76	7.91
# price targets	28.66	17.20	14.66	14.92	15.53	23.65	13.34	14.89	17.98	20.19
Mean exp ret	1.34	0.95	0.71	1.10	0.97	1.53	0.86	0.95	1.06	1.26
Std exp ret	7.91	5.29	3.68	5.40	5.91	6.91	3.80	4.52	4.16	6.62
# RECs	39.53	23.10	23.78	20.60	24.28	32.73	18.77	19.98	26.02	27.22
Mean RECs	2.18	2.18	2.10	2.16	2.42	1.91	2.09	1.95	2.05	2.29
Std RECs	0.29	0.26	0.25	0.26	0.36	0.26	0.21	0.24	0.25	0.25

Note: The table reports descriptive statistics on the returns, expected target returns and recommendations for 20 Dow Jones constituents. It reports the mean and standard deviation of the logarithmic monthly returns, the mean number of available target prices, the mean and variance of the monthly forward target price implied expected return, i.e. simple returns between the spot and the twelve month forward target price at each point t divided by 12, constructed from individual analyst data, the mean number of recommendations as well as the mean and standard deviation of the recommendations based on the 1 (strong buy) to 5 (strong sell) scale. Mean returns and standard deviations are multiplied by 100. Target prices and recommendations are obtained from I/B/E/S Datastream.

Table 3.3: Forecast performance in terms of out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a Bayesian VAR(1)

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-6.90	0.20	-3.46	-0.68	-1.24	-0.51	-0.04	-0.20	-0.44	-0.13
Log EPR	-1.97	0.13	-0.82	-0.30	-5.98	-0.06	-0.22	-0.16	-0.32	-0.62
Log DPR	-0.13	-0.04	-0.05	-0.09	-1.09	0.13	-0.10	0.25	-0.13	-0.18
BMR	-0.15	0.21	-1.18	-0.14	-0.13	-0.13	-0.73	-0.24	0.00	-0.06
3M Tbill rate	-0.10	0.09	-0.07	0.33	0.06	-0.18	-0.36	-0.28	-0.11	-0.17
Market return	-0.09	0.22	-0.02	-0.06	-0.05	-0.12	-0.09	-0.21	-0.11	-0.57
LT yield	-0.22	0.03	0.24	-0.15	-0.35	-0.27	-0.42	-0.28	-0.09	-0.29
CPI inflation	-0.15	0.18	-0.10	-0.14	-0.13	-0.19	-0.30	-0.19	-0.20	0.14
Log TPR	-0.05	-0.16	-0.58	-0.45	-0.10	-0.13	-0.30	-0.17	-0.50	-0.20
Log TPV	-0.24	0.15	-0.12	-0.17	-0.12	-0.53	-0.16	-0.19	-0.28	0.28
Log REC	-0.54	0.01	-0.08	-0.48	-1.10	-0.31	-0.52	-0.72	-0.22	-0.08
Log REC return	8.06***	-0.36	-0.09	0.09	-0.32	-0.01	8.06***	8.06***	-0.11	-0.02
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	-0.35	-0.20	-0.46	-0.39	-0.31	-0.22	-0.43	-0.06	-0.23	-0.01
Log EPR	-0.23	-0.19	-0.05	-0.51	-0.51	-0.11	0.02	-0.03	-0.44	-0.08
Log DPR	-0.10	-0.45	-0.45	-0.50	-0.12	0.09	-1.70	0.35	-0.26	-0.03
BMR	-0.12	-0.35	-0.15	-0.48	-0.12	-0.28	0.03	-0.06	-0.20	-0.06
3M Tbill rate	-0.36	-0.41	-0.64	-0.62	-0.32	-0.32	-0.13	-0.09	-0.17	0.27
Market return	-0.39	-0.23	-0.07	-0.37	-0.16	-0.18	-0.52	-0.19	-0.23	-0.20
LT yield	-0.39	-0.38	-0.25	-0.46	-0.40	-0.39	-0.73	-0.19	-0.32	-0.28
CPI inflation	-0.14	-0.16	-0.06	-0.33	-0.22	-0.15	-0.40	-0.10	-0.14	-0.02
Log TPR	-0.02	-0.24	-0.03	0.03	-0.21	-0.00	-0.16	0.05	-0.26	0.01
Log TPV	-0.18	-0.15	-0.07	-0.52	-0.18	-0.12	-0.48	-0.24	-0.24	0.11
Log REC	-0.88	-0.38	-0.25	-1.07	-0.38	-0.38	-1.83	-1.16	-0.23	-0.34
Log REC return	-0.59	-0.20	-0.05	-0.06	-0.03	-0.19	-0.14	-0.60	-0.15	-0.12

Note: The table provides forecast performance results in terms of mean out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. The benchmark model is a simple mean model. For each asset, we estimate a Bayesian VAR system with constant coefficients using the Minnesota prior outlined in section 3 for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a + A_1 \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.4: Forecast performance in terms of out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-6.49	0.68	-3.12	-0.33	-0.89	-0.14	0.36	0.26	-0.15	0.21
Log EPR	-1.52	0.37	-0.44	-0.14	-5.60	0.06	-0.10	-0.01	-0.08	-0.25
Log DPR	-0.07	0.36	0.32	0.39	-0.95	0.39	0.36	0.63	-0.12	0.04
BMR	0.30	0.28	-0.99	-0.12	0.21	0.22	-0.56	0.14	0.17	-0.02
3M Tbill rate	0.21	0.30	0.26	0.55	0.39	0.27	-0.26	-0.09	-0.03	-0.06
Market return	-0.04	0.67	0.06	0.13	0.03	0.36	0.03	0.07	0.28	-0.11
LT yield	-0.08	0.42	0.59	0.23	-0.29	0.00	-0.11	-0.24	0.06	-0.22
CPI inflation	0.13	0.66	-0.09	0.25	0.12	-0.12	-0.07	-0.16	0.07	0.55
Log TPR	0.47	0.35	-0.41	-0.26	0.38	0.24	0.31	0.36	0.18	0.01
Log TPV	0.30	0.90	0.44	-0.08	0.24	-0.00	0.12	0.15	0.39	0.77
Log REC	-0.46	0.43	-0.03	-0.25	-0.81	0.11	-0.23	-0.25	-0.09	-0.05
Log REC return	8.55***	0.10	0.32	0.41	-0.21	0.11	8.34***	8.13***	0.22	0.20
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	-0.30	-0.11	-0.26	-0.22	-0.03	0.06	-0.18	0.11	-0.11	0.12
Log EPR	0.26	-0.06	-0.02	-0.06	-0.48	0.04	0.24	0.44	-0.35	0.12
Log DPR	-0.10	-0.38	-0.33	-0.31	-0.00	0.46	-1.48	0.79	-0.14	0.27
BMR	0.27	-0.28	-0.08	-0.42	0.06	-0.19	0.19	0.21	0.01	0.07
3M Tbill rate	0.05	0.02	-0.55	-0.23	0.09	0.02	0.12	0.22	-0.02	0.58
Market return	0.04	0.06	0.05	-0.18	-0.15	-0.09	-0.27	0.10	0.23	0.16
LT yield	-0.35	-0.11	-0.04	-0.34	-0.38	-0.20	-0.32	-0.09	-0.11	-0.17
CPI inflation	0.06	-0.09	-0.03	-0.13	-0.14	0.16	-0.01	0.05	-0.04	0.04
Log TPR	0.44	0.46	0.23	0.78	0.17	0.43	0.28	0.47	-0.09	0.10
Log TPV	0.26	0.06	0.02	-0.01	0.33	0.39	0.10	0.56	-0.16	0.16
Log REC	-0.66	-0.20	-0.01	-0.60	-0.05	0.09	-1.42	-0.74	-0.01	-0.13
Log REC return	-0.13	0.06	0.19	0.42	0.20	0.20	0.13	-0.51	-0.10	0.13

Note: The table provides forecast performance results in terms of mean out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. The benchmark model is a simple mean model. For each asset, we estimate a Bayesian VAR system with time-varying coefficients and stochastic volatility for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a_t + A_{1,t} \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$, $A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t$, where $A_t = [a_t \ A_{1,t}]$ is

time-index for every single parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of one each other for all t and s . We estimate the model using forgetting factors with the following parameter values: $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.5: Forecast performance in terms of out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean of monthly target price implied expected returns

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-6.45	0.38	-3.35	-0.57	-1.24	-0.11	0.47	0.00	-0.05	0.10
Log EPR	-1.74	0.50	-0.60	-0.03	-5.50	0.26	0.10	-0.13	-0.04	-0.51
Log DPR	0.36	0.20	0.37	-0.01	-0.75	0.27	-0.09	0.52	0.17	0.34
BMR	0.06	0.67	-0.73	-0.11	0.41	0.24	-0.67	-0.13	0.25	-0.01
3M Tbill rate	0.32	0.51	0.37	0.80	0.35	-0.13	0.12	-0.21	-0.04	-0.12
Market return	0.13	0.31	0.15	0.25	0.21	0.23	0.18	-0.10	0.16	-0.49
LT yield	0.23	0.50	0.53	0.36	0.09	0.09	0.05	-0.20	0.38	-0.20
CPI inflation	0.27	0.72	-0.05	0.24	-0.01	0.21	-0.19	-0.08	0.28	0.48
Log TPR	0.20	0.25	0.15	-0.42	-0.03	0.76	0.54	0.45	-0.02	0.61
Log TPV	-0.02	0.35	0.76	0.31	0.60	-0.47	-0.14	0.52	0.52	0.84
Log REC	-0.10	0.33	0.29	0.01	-0.79	0.11	-0.51	-0.56	0.09	0.43
Log REC return	8.59***	-0.28	0.18	0.63	0.15	0.30	8.40***	8.36***	0.24	0.38
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	0.05	0.32	-0.43	-0.14	0.06	0.28	0.12	0.43	0.17	0.05
Log EPR	-0.19	-0.03	0.04	-0.09	-0.25	-0.07	0.20	0.26	-0.05	-0.01
Log DPR	0.38	-0.29	-0.08	-0.05	0.38	0.22	-1.54	0.41	0.17	0.27
BMR	0.40	-0.16	0.04	-0.42	-0.06	-0.25	0.07	0.39	-0.04	0.21
3M Tbill rate	0.18	-0.16	-0.14	-0.52	0.09	-0.08	0.03	0.10	0.21	0.76
Market return	0.08	0.12	-0.00	-0.17	0.24	-0.17	-0.50	-0.03	0.08	0.24
LT yield	0.04	-0.37	0.30	-0.43	-0.09	0.11	-0.45	0.22	-0.10	0.13
CPI inflation	0.14	0.30	0.24	-0.04	-0.12	-0.04	0.01	-0.10	-0.10	0.01
Log TPR	0.10	-0.07	0.07	0.88	0.16	0.85	0.73	0.35	-0.04	0.62
Log TPV	0.02	-0.11	0.48	-0.20	0.71	0.48	0.21	0.36	0.03	0.59
Log REC	-0.81	-0.19	-0.10	-0.96	-0.30	-0.13	-1.79	-0.83	0.10	0.10
Log REC return	-0.57	0.05	0.18	0.44	0.09	-0.14	0.29	-0.32	0.26	0.40

Note: The table provides forecast performance results in terms of mean out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. The benchmark model is a simple mean model. For each asset, we estimate a Bayesian VAR system with time-varying coefficients and stochastic volatility for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a_t + A_{1,t} \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$, $A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t$, where $A_t = [a_t \ A_{1,t}]$ is time-index for every single parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of one each other for all t and s . We estimate the model using forgetting factors with the following parameter values: $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$. The mean of the predictive distribution is tilted towards the mean of the monthly forward target price implied expected returns. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.6: Forecast performance in terms of out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean and variance of monthly target price implied expected returns

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-6.77	0.60	-2.88	0.10	-0.33	-0.17	0.30	0.34	-0.34	0.93
Log EPR	-1.76	0.38	0.25	0.70	-5.48	0.21	0.66	1.00	0.85	-0.51
Log DPR	-0.01	0.57	0.12	0.39	-0.50	0.99	0.06	0.30	0.66	0.93
BMR	0.43	1.30**	-0.72	0.76	0.70	0.62	0.27	0.93	0.28	0.42
3M Tbill rate	0.13	0.84	1.05*	1.33**	1.23**	0.53	-0.19	-0.05	0.38	-0.12
Market return	0.99	0.34	1.08*	0.33	0.34	0.68	0.61	0.59	0.03	-0.16
LT yield	-0.10	0.49	1.09*	0.51	0.66	-0.21	0.02	0.42	0.23	0.59
CPI inflation	-0.09	0.24	0.64	1.03*	0.75	0.23	0.67	0.62	0.11	1.09*
Log TPR	0.96	0.88	-0.56	-0.03	1.26**	-0.09	0.95	-0.08	0.38	0.88
Log TPV	0.40	0.53	0.67	1.24**	0.46	0.47	1.30**	0.49	0.75	1.25**
Log REC	-0.16	1.20**	0.07	0.27	-0.67	0.54	0.53	0.25	0.20	0.99
Log REC return	8.28***	0.61	0.79	0.52	0.48	1.01*	8.49***	8.09***	0.04	0.04
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	0.01	0.98	0.11	0.57	0.13	0.27	-0.13	0.99	0.37	0.52
Log EPR	-0.17	0.67	0.38	0.57	0.61	0.69	0.96	0.31	0.39	0.44
Log DPR	0.14	0.56	0.50	0.22	0.87	1.21**	-0.64	1.16*	0.75	0.11
BMR	0.75	0.17	0.79	0.58	0.90	0.69	1.13*	0.74	0.53	0.92
3M Tbill rate	0.51	0.15	0.16	0.51	0.13	0.26	0.54	0.06	0.52	0.66
Market return	0.66	0.44	0.09	0.29	0.55	0.73	0.20	0.30	0.16	0.10
LT yield	0.31	-0.06	-0.22	0.42	0.65	0.12	-0.55	0.14	0.23	0.13
CPI inflation	-0.05	0.74	0.61	0.36	0.90	1.01*	0.68	0.76	0.72	0.43
Log TPR	1.07*	0.64	0.06	0.46	1.23**	0.29	1.24**	0.46	0.33	0.20
Log TPV	0.38	0.03	1.40**	0.37	0.10	0.39	0.11	-0.01	0.32	0.76
Log REC	-0.54	-0.01	0.92	-0.30	0.41	0.09	-0.75	-0.17	-0.21	0.13
Log REC return	0.06	-0.03	0.29	0.56	0.06	0.35	0.78	-0.14	0.66	0.36

Note: The table provides forecast performance results in terms of mean out-of-sample R^2 for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. The benchmark model is a simple mean model. For each asset, we estimate a Bayesian VAR system with time-varying coefficients and stochastic volatility for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a_t + A_{1,t} \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$, $A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t$, where $A_t = [a_t \ A_{1,t}]$ is time-index for every single parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of one each other for all t and s . We estimate the model using forgetting factors with the following parameter values: $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$. The mean and variance of the predictive distribution are tilted towards the mean and variance of the monthly forward target price implied expected returns. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.7: Forecast performance in terms of average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a Bayesian VAR(1)

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-2.02	0.04	7.14***	-0.44	-0.23	-0.39	-0.00	-0.08	-0.25	-0.05
Log EPR	-0.10	-0.01	0.55*	-0.16	-0.26	-0.06	-0.07	-0.05	-0.13	-0.15
Log DPR	-0.09	-0.03	-0.28	-0.05	-0.33	0.05	0.02	0.16	-0.03	-0.06
BMR	-0.03	0.03	0.91*	-0.07	-0.04	-0.06	-0.11	-0.07	0.06	-0.03
3M Tbill rate	0.01	-0.01	-0.46	0.33	0.12	-0.09	-0.06	-0.01	0.05	-0.05
Market return	-0.06	0.04	-0.20	-0.03	-0.03	-0.09	0.01	-0.11	-0.03	-0.08
LT yield	-0.09	0.00	-1.00	-0.03	-0.10	-0.18	-0.09	-0.08	0.01	-0.08
CPI inflation	-0.12	0.04	-0.20	-0.10	-0.07	-0.17	-0.06	-0.09	-0.09	0.02
Log TPR	-0.02	-0.05	0.67*	-0.26	-0.05	-0.10	-0.08	-0.06	-0.26	-0.07
Log TPV	-0.19	0.01	-0.20	-0.09	-0.06	-0.48	-0.04	-0.09	-0.14	0.07
Log REC	-0.44	-0.04	-0.23	-0.27	-0.28	-0.26	-0.11	-0.29	-0.09	-0.03
Log REC return	0.00	-0.16	-0.28	0.08	-0.12	-0.02	0.00	0.00	-0.03	0.01
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	-0.08	-0.00	-0.08	-0.07	-0.11	-0.06	-0.05	-0.03	-0.05	-0.00
Log EPR	-0.08	-0.04	-0.03	-0.01	-0.17	-0.03	-0.01	-0.03	-0.08	-0.05
Log DPR	-0.01	-0.08	-0.06	-0.07	-0.03	0.02	-0.19	0.13	-0.04	0.01
BMR	-0.06	-0.05	-0.01	-0.07	-0.04	-0.08	0.01	-0.03	-0.04	-0.02
3M Tbill rate	-0.09	-0.07	-0.08	-0.05	-0.09	-0.08	-0.04	-0.01	-0.01	0.18
Market return	-0.08	-0.04	0.03	-0.05	-0.06	-0.06	0.05	-0.04	-0.04	-0.04
LT yield	-0.09	-0.08	-0.06	-0.08	-0.11	-0.11	-0.11	-0.05	-0.07	-0.08
CPI inflation	-0.06	-0.05	-0.03	-0.05	-0.09	-0.05	-0.06	-0.05	-0.04	-0.01
Log TPR	-0.03	-0.07	-0.02	-0.01	-0.08	-0.02	-0.04	-0.00	-0.06	0.02
Log TPV	-0.07	-0.02	-0.02	-0.07	-0.06	-0.03	-0.07	-0.07	-0.05	0.04
Log REC	-0.18	-0.07	-0.06	-0.14	-0.11	-0.10	-0.21	-0.23	-0.06	-0.10
Log REC return	-0.02	-0.06	-0.03	-0.02	0.00	-0.07	-0.03	-0.14	-0.04	-0.04

Note: The table provides forecast performance results in terms of average log predictive score differentials between the benchmark mean model and a single regressor model for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. For each asset, we estimate a Bayesian VAR system with constant coefficients using the Minnesota prior outlined in section 3 for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a + A_1 \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.8: Forecast performance in terms of average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-1.98	0.31	7.19***	0.01	0.18	-0.36	0.48	-0.05	-0.06	0.16
Log EPR	0.03	0.11	0.68	0.00	0.03	0.28	0.26	0.15	-0.03	0.35
Log DPR	0.31	0.21	-0.12	0.30	-0.23	0.07	0.42	0.42	0.22	0.09
BMR	-0.02	0.34	1.25**	0.03	0.08	-0.02	0.12	0.14	0.23	0.32
3M Tbill rate	0.48	0.33	-0.39	0.35	0.57	0.17	0.15	0.31	0.52	0.28
Market return	0.31	0.23	0.16	0.34	-0.01	-0.04	0.42	0.20	0.43	0.18
LT yield	0.15	0.19	-0.95	0.22	0.14	0.23	-0.05	0.07	0.04	0.27
CPI inflation	0.17	0.53	0.13	0.14	0.02	0.24	0.00	0.12	0.28	0.35
Log TPR	0.30	0.24	1.17*	0.48	0.10	0.46	0.34	0.63	0.05	0.52
Log TPV	0.17	0.62	0.42	0.69	0.05	-0.40	0.38	0.30	0.40	0.49
Log REC	0.04	0.42	0.13	0.04	-0.03	0.07	0.31	-0.21	0.18	0.47
Log REC return	0.27	0.24	0.17	0.51	0.12	0.24	0.40	0.05	0.45	0.10
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	-0.06	0.09	0.34	0.14	-0.07	0.33	0.27	0.14	0.11	0.12
Log EPR	0.21	0.17	0.46	0.14	-0.01	0.18	0.47	0.36	0.31	0.41
Log DPR	0.43	0.17	0.31	0.01	0.23	0.06	-0.07	0.46	0.19	0.14
BMR	0.27	0.01	0.16	0.02	0.29	0.05	0.35	-0.03	-0.02	0.36
3M Tbill rate	0.01	0.23	0.21	0.16	0.11	0.00	0.11	0.29	0.08	0.28
Market return	0.11	0.07	0.09	-0.00	0.35	0.08	0.38	0.15	0.32	0.11
LT yield	0.14	0.11	0.40	0.22	0.25	0.11	0.24	0.40	0.17	-0.03
CPI inflation	0.43	0.24	0.41	0.18	0.40	0.21	-0.02	-0.05	0.04	0.28
Log TPR	0.25	0.40	0.02	0.19	0.47	0.57	0.51	0.35	0.20	0.24
Log TPV	0.05	0.19	0.58	0.28	0.23	0.28	0.49	-0.05	0.29	0.19
Log REC	0.15	0.24	0.23	0.18	-0.06	0.15	0.12	0.00	0.04	0.11
Log REC return	0.17	0.07	-0.01	0.00	0.31	0.40	0.39	0.25	0.33	0.28

Note: The table provides forecast performance results in terms of average log predictive score differentials between the benchmark mean model and a single regressor model for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. For each asset, we estimate a Bayesian VAR system with time-varying coefficients and stochastic volatility for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a_t + A_{1,t} \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$, $A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t$,

where $A_t = [a_t \ A_{1,t}]$ is time-index for every single parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of one each other for all t and s . We estimate the model using forgetting factors with the following parameter values: $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.9: Forecast performance in terms of average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean of monthly target price implied expected returns

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-1.65	0.45	7.57***	-0.15	0.01	-0.02	0.23	0.09	0.06	0.17
Log EPR	-0.03	0.52	0.66	-0.11	0.10	0.18	0.30	0.12	0.04	0.20
Log DPR	0.30	0.44	-0.07	0.17	-0.26	0.30	0.42	0.28	0.06	0.02
BMR	0.03	0.08	1.22**	-0.01	0.47	0.28	0.18	0.07	0.40	0.18
3M Tbill rate	0.08	0.19	-0.33	0.39	0.23	-0.05	-0.00	0.48	0.59	0.04
Market return	0.29	0.24	0.16	0.40	0.12	0.08	0.36	0.27	0.07	0.33
LT yield	0.09	0.38	-0.73	0.13	0.34	0.24	-0.02	0.23	0.15	0.40
CPI inflation	0.24	0.37	-0.12	0.23	0.20	0.22	0.01	0.01	0.13	0.21
Log TPR	0.35	0.62	1.17*	0.20	0.59	0.22	0.23	0.35	-0.07	0.21
Log TPV	0.36	0.53	0.32	-0.01	0.84	0.40	0.76	0.29	-0.02	0.72
Log REC	-0.03	0.08	-0.07	0.11	-0.13	-0.21	-0.02	0.21	0.13	0.26
Log REC return	0.13	-0.11	-0.15	0.50	-0.10	-0.02	0.11	0.39	0.51	0.47
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	0.25	0.16	0.30	0.45	-0.01	0.36	0.39	0.33	0.05	0.36
Log EPR	0.11	0.13	0.23	0.28	-0.03	0.10	0.46	0.31	0.45	0.16
Log DPR	0.15	0.28	0.06	0.31	-0.02	0.05	0.09	0.34	0.10	0.35
BMR	0.19	0.47	0.04	-0.05	0.47	0.34	0.36	0.05	0.47	-0.01
3M Tbill rate	0.14	0.45	0.37	0.40	0.27	0.29	0.48	0.00	0.11	0.68
Market return	0.12	0.21	0.13	0.36	0.45	0.33	0.29	0.19	0.17	0.40
LT yield	0.21	0.05	0.03	-0.01	-0.02	0.24	-0.07	0.05	-0.02	0.33
CPI inflation	0.35	0.37	0.34	0.24	0.42	0.18	0.42	0.35	0.31	0.43
Log TPR	0.68	-0.06	0.81	0.03	0.56	0.18	0.57	0.56	0.10	0.71
Log TPV	0.55	0.74	0.67	0.27	0.60	0.21	0.36	0.14	0.69	0.88
Log REC	-0.11	0.34	0.32	0.08	0.13	0.07	0.34	0.18	0.34	0.21
Log REC return	-0.01	-0.00	0.06	0.21	0.14	0.38	0.09	0.18	0.15	0.25

Note: The table provides forecast performance results in terms of average log predictive score differentials between the benchmark mean model and a single regressor model for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. For each asset, we estimate a Bayesian VAR system with time-varying coefficients and stochastic volatility for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a_t + A_{1,t} \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$, $A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t$, where $A_t = [a_t \ A_{1,t}]$ is time-index for every single parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of one each other for all t and s . We estimate the model using forgetting factors with the following parameter values: $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$. The mean of the predictive distribution is tilted towards the mean of the monthly forward target price implied expected returns. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.10: Forecast performance in terms of average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) using a TVP-BVAR(1) with stochastic volatility and entropic tilting towards the mean and variance of monthly target price implied expected returns

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
Log DY	-1.41	0.06	7.66***	-0.21	0.52	0.04	0.70	0.06	-0.13	0.28
Log EPR	0.69	0.64	1.14*	0.15	0.58	0.49	0.49	0.75	0.96	-0.02
Log DPR	1.05*	0.08	-0.20	1.03*	0.54	0.51	0.08	0.88	0.10	0.55
BMR	0.83	0.20	1.98***	0.65	0.37	0.87	0.16	0.00	0.68	0.68
3M Tbill rate	0.49	0.75	-0.38	0.93	0.74	0.79	0.94	0.05	0.22	0.86
Market return	0.94	1.07*	0.33	0.70	0.64	0.42	0.03	0.07	0.64	0.01
LT yield	0.07	1.17*	-0.01	0.96	0.09	0.65	0.95	-0.05	0.02	0.71
CPI inflation	-0.05	0.72	0.27	0.54	0.61	0.97	0.03	0.43	0.83	0.64
Log TPR	0.12	-0.04	1.86***	0.41	0.04	0.86	1.18*	1.21**	0.16	0.28
Log TPV	0.74	0.87	0.15	0.77	0.69	-0.14	1.42**	0.67	0.98	1.51**
Log REC	-0.05	0.58	0.84	0.24	0.73	-0.12	0.15	0.33	1.09*	0.67
Log REC return	0.36	0.24	0.84	1.24**	0.76	0.45	0.69	1.04*	0.58	0.54
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
Log DY	1.05*	0.13	0.23	-0.07	0.52	0.79	1.05*	0.19	0.93	0.60
Log EPR	0.71	0.49	0.33	0.21	0.15	0.74	1.08*	0.36	0.80	0.87
Log DPR	0.53	0.28	0.45	0.10	0.05	0.68	0.52	0.61	0.17	0.92
BMR	0.95	0.43	0.13	0.25	0.48	0.18	0.41	0.63	0.39	0.67
3M Tbill rate	0.55	0.93	0.52	0.16	0.12	0.85	0.99	0.05	0.22	1.08*
Market return	0.59	0.44	0.88	0.12	-0.03	0.21	0.58	0.62	-0.04	0.74
LT yield	0.72	0.39	0.24	0.64	1.04*	0.34	0.98	0.28	0.31	0.07
CPI inflation	0.38	0.38	0.92	1.03*	0.43	1.02*	-0.02	0.24	0.80	0.59
Log TPR	0.90	0.19	0.37	1.36**	1.01*	0.84	0.57	0.13	0.70	1.11*
Log TPV	0.83	0.11	1.26**	0.98	0.29	1.18*	1.41**	0.41	0.04	0.87
Log REC	0.86	0.03	-0.06	0.44	-0.10	0.28	0.00	0.92	0.47	0.07
Log REC return	0.47	0.46	0.24	0.43	0.82	0.66	0.37	0.99	0.30	0.19

Note: The table provides forecast performance results in terms of average log predictive score differentials between the benchmark mean model and a single regressor model for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. For each asset, we estimate a Bayesian VAR system with time-varying coefficients and stochastic volatility for the monthly excess returns on an intercept and a lagged predictor variable, i.e. $\begin{bmatrix} r_t \\ x_t \end{bmatrix} = a_t + A_{1,t} \begin{bmatrix} r_{t-i} \\ x_{t-i} \end{bmatrix} + \varepsilon_t$, $t = 1, \dots, T$, $A_t = \phi A_{t-1} + (1 - \phi) \underline{A}_0 + u_t$, where $A_t = [a_t \ A_{1,t}]$ is time-index for every single parameter, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_t)$, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Omega_t)$ and ε_t and u_s are independent of one each other for all t and s . We estimate the model using forgetting factors with the following parameter values: $\lambda = 0.99$, $\kappa = 0.96$ and $\phi = 0.5$. The mean and variance of the predictive distribution are tilted towards the mean and variance of the monthly forward target price implied expected returns. Further, DY is the dividend yield, PR is the earnings-price ratio, DPR is the dividend-price-ratio, BMR is the book-to-market ratio, LT is longterm yield, TPR is the target price return, TPV the target price variance and REC stands for recommendations. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.11: Forecast performance in terms of out-of-sample R² for 20 Dow Jones constituents (sample: 2004 - 2015) for various forecasting models

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
AR1	-0.10	0.09	-0.07	0.33	0.06	-0.18	-0.36	-0.28	-0.11	-0.17
VAR-Full	-1.16	-1.57	-1.78	-1.25	-0.57	-1.08	-1.86	-0.50	-0.33	-0.71
VAR-Minnesota	0.74	0.67	0.88	0.39	0.65	0.11	0.47	-0.09	0.33	0.22
TVPVAR-DMA	0.85	0.67	-0.03	0.39	0.91	0.30	0.04	0.55	0.27	0.39
TVPVAR-DMS	0.62	0.37	-0.09	0.15	0.85	0.00	-0.20	0.43	0.06	0.24
TVPVAR-DMAm	0.89	0.70	-0.02	0.41	0.94	0.35	0.09	0.60	0.31	0.41
TVPVAR-DMAm/v	1.91***	0.80	0.02	0.72	1.55**	0.88	1.04*	1.15*	0.75	0.98
TVPVAR-DMSm	0.63	0.38	-0.07	0.20	0.87	0.02	-0.19	0.47	0.08	0.28
TVPVAR-DMSm/v	1.66**	1.30**	0.31	1.05*	1.77**	0.79	0.53	0.83	0.34	0.55
Bayesian lasso	0.85	0.74	0.00	0.46	0.95	0.34	0.05	0.56	0.28	0.44
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
AR1	-0.36	-0.41	-0.64	-0.62	-0.32	-0.32	-0.13	-0.09	-0.17	0.27
VAR-Full	-1.41	-2.36	-2.06	-1.24	-0.90	-2.02	-1.96	-1.37	-0.68	0.10
VAR-Minnesota	0.47	0.26	-0.43	-0.30	-0.19	0.35	0.44	0.08	-0.02	0.75
TVPVAR-DMA	0.39	0.50	-0.29	0.06	0.70	-0.24	0.48	0.34	0.15	0.55
TVPVAR-DMS	0.15	0.39	-0.31	-0.11	0.43	-0.30	0.35	0.12	0.14	0.27
TVPVAR-DMAm	0.44	0.53	-0.28	0.08	0.74	-0.21	0.52	0.39	0.20	0.58
TVPVAR-DMAm/v	1.17*	0.52	0.59	0.22	1.23**	0.04	0.89	1.07*	0.34	0.86
TVPVAR-DMSm	0.18	0.41	-0.27	-0.08	0.45	-0.26	0.40	0.14	0.16	0.30
TVPVAR-DMSm/v	0.47	0.92	-0.28	0.89	1.28**	-0.24	0.51	0.53	0.56	0.67
Bayesian lasso	0.41	0.59	-0.21	0.16	0.75	-0.22	0.50	0.40	0.23	0.59

Note: The table provides forecast performance results in terms of mean out-of-sample R² for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. The benchmark model is a simple mean model. For each asset, we estimate various Bayesian VAR systems described sections 3.3 and 3.4. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Table 3.12: Forecast performance in terms of average log predictive score differentials for 20 Dow Jones constituents (sample: 2004 - 2015) for various forecasting models

Stock	AA	AAPL	AIG	AXP	BA	CAT	KO	DD	GE	HD
AR1	0.01	-0.01	-0.46	0.33	0.12	-0.09	-0.06	-0.01	0.05	-0.05
VAR-Full	-0.91	-1.29	-2.29	0.01	-1.31	-1.24	-0.93	-1.78	-0.74	-0.41
VAR-Minnesota	0.10	0.32	0.25	0.77	0.42	0.08	0.29	0.75	0.15	0.32
TVPVAR-DMA	0.21	0.74	-0.11	0.75	0.28	0.73	0.56	0.72	0.85	0.01
TVPVAR-DMS	-0.06	0.57	-0.28	0.49	0.27	0.47	0.44	0.71	0.63	-0.03
TVPVAR-DMAm	0.21	0.76	-0.07	0.76	0.32	0.77	0.60	0.77	0.87	0.03
TVPVAR-DMAm/v	0.80	1.83***	0.72	1.83***	0.54	1.31**	0.62	1.56**	1.52**	0.96
TVPVAR-DMSm	-0.02	0.60	-0.27	0.53	0.30	0.51	0.48	0.73	0.66	-0.01
TVPVAR-DMSm/v	0.88	1.55**	0.11	0.87	0.48	1.05*	0.82	1.56**	0.90	0.55
Bayesian lasso	0.23	0.84	-0.04	0.85	0.32	0.83	0.56	0.79	0.93	0.04
Stock	INTC	IBM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
AR1	-0.09	-0.07	-0.08	-0.05	-0.09	-0.08	-0.04	-0.01	-0.01	0.18
VAR-Full	-1.35	-1.32	-0.73	-1.66	-2.09	-2.04	-0.29	-0.48	-0.06	-1.03
VAR-Minnesota	0.60	-0.03	0.48	0.73	0.13	0.03	0.51	0.72	0.48	0.90
TVPVAR-DMA	0.86	0.43	0.68	0.69	0.74	0.08	0.42	0.61	0.92	1.02*
TVPVAR-DMS	0.82	0.25	0.60	0.60	0.62	-0.04	0.30	0.42	0.87	0.96
TVPVAR-DMAm	0.89	0.45	0.72	0.73	0.77	0.09	0.45	0.62	0.95	1.04*
TVPVAR-DMAm/v	1.95***	1.45**	1.13*	0.69	1.33**	0.31	0.66	0.97	1.03*	1.84***
TVPVAR-DMSm	0.85	0.29	0.64	0.60	0.66	-0.04	0.32	0.46	0.91	0.98
TVPVAR-DMSm/v	1.01*	1.18*	0.83	1.14*	1.64**	0.40	0.46	0.80	1.28**	1.32**
Bayesian lasso	0.96	0.51	0.76	0.75	0.82	0.11	0.44	0.64	0.98	1.10*

Note: The table provides forecast performance results in terms of average log predictive score differentials between the benchmark mean model and a single regressor model for 20 Dow Jones constituents (sample: 2004 - 2015) with a one month forecast horizon. For each asset, we estimate various Bayesian VAR systems described sections 3.3 and 3.4. Values above zero indicate that a given predictor has better forecast performance than the benchmark model, while negative values suggest the opposite. All values are multiplied by 100. We test statistical significance in the average loss between the each model and a simple mean model using the Diebold and Mariano (1995) test. One/two/three asterisks denote rejection of the null hypothesis of equal predictive ability at the ten/five/one percent test level.

Complete Bibliography

- ALTAVILLA, C., R. GIACOMINI, AND G. RAGUSA (2014): “Anchoring the Yield Curve Using Survey Expectations,” ECB Working Paper No. 1632, ECB.
- AVRAMOV, D. (2002): “Stock return predictability and model uncertainty,” *Journal of Financial Economics*, 64, 423–458.
- BAŞTÜRK, N., C. ÇAKMAKLI, S. P. CEYHAN, AND H. K. VAN DIJK (2014): “Posterior-Predictive Evidence on Us Inflation Using Extended New Keynesian Phillips Curve Models with Non-Filtered Data,” *Journal of Applied Econometrics*, 29, 1164–1182.
- BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): “Large Bayesian Vector Auto Regressions,” *Journal of Applied Econometrics*, 25, 71–92.
- BARBER, B., R. LEHAVY, M. McNICHOLS, AND B. TRUEMAN (2001): “Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns,” *The Journal of Finance*, 56, 531–563.
- BARBERIS, N. (2000): “Investing for the Long Run when Returns are Predictable,” *The Journal of Finance*, 55, 225–264.
- BAUWENS, L. AND D. KOROBILIS (2013): “Bayesian methods,” in *Handbook Of Research Methods And Applications In Empirical Macroeconomics*, ed. by N. Hashimzade and M. A. Thornton, Edward Elger Publishing, Handbooks of Research Methods and Applications series, chap. 16, 363–380.
- BAUWENS, L., M. LUBRANO, AND J. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*, Advanced Texts in Econometrics, Oxford University Press: Oxford, UK.
- BEST, M. J. AND R. R. GRAUER (1991): “On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results,” *Review of Financial Studies*, 4, 315–342.
- BLACK, F. AND R. LITTERMAN (1992): “Global Portfolio Optimization,” *Financial Analysts Journal*, 48, 28–43.
- BONINI, S., L. ZANETTI, R. BIANCHINI, AND A. SALVI (2010): “Target Price Accuracy in Equity Research,” *Journal of Business Finance & Accounting*, 37, 1177–1217.
- BRADSHAW, M. T., L. D. BROWN, AND K. HUANG (2012): “Do Sell-Side Analysts Exhibit Differential Target Price Forecasting Ability?” *Review of Accounting Studies*, 18, 930–955.

- BRANDT, M. (2010): “Portfolio Choice Problems,” in *Handbook of Financial Econometrics, Volume 1: Tools and Techniques*, ed. by Y. Ait-Sahalia and L. Hansen, Amsterdam: North Holland, 269–336.
- BRANDT, M. W., P. SANTA-CLARA, AND R. VALKANOV (2009): “Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns,” *Review of Financial Studies*, 22, 3411–3447.
- BRAV, A. AND R. LEHAVY (2003): “An Empirical Analysis of Analysts’ Target Prices: Short-Term Informativeness and Long-Term Dynamics,” *The Journal of Finance*, 58, 1933–1967.
- BRITTON-JONES, M. (1999): “The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights,” *The Journal of Finance*, 54, 655–671.
- BRODIE, J., I. DAUBECHIES, C. D. MOL, D. GIANNONE, AND I. LORIS (2009): “Sparse and stable Markowitz portfolios,” *Proceedings of the National Academy of Sciences*, 106, 12267–12272.
- CARRIERO, A., T. E. CLARK, AND M. MARCELLINO (2015): “Bayesian VARs: Specification Choices and Forecast Accuracy,” *Journal of Applied Econometrics*, 30, 46–73.
- CHAUVET, M. AND S. POTTER (2013): “Forecasting Output,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier: Amsterdam, NL, vol. 2.
- CLARK, T. E. AND M. W. McCACKEN (2009b): “Tests of Equal Predictive Ability With real-time Data,” *Journal of Business & Economic Statistics*, 27, 441–454.
- (2015): “Nested forecast model comparisons: A new approach to testing equal accuracy,” *Journal of Econometrics*, 186, 160–177.
- CLARK, T. E. AND K. D. WEST (2007): “Approximately Normal Tests for Equal Predictive Accuracy in Nested Models,” *Journal of Econometrics*, 138, 291–311.
- COGLEY, T., S. MOROZOV, AND T. J. SARGENT (2005): “Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system,” *Journal of Economic Dynamics and Control*, 29, 1893–1925.
- CREMERS, K. (2002): “Stock return predictability: a Bayesian model selection perspective,” *The Review of Financial Studies*, 15, 1223–1249.
- CROUSHORE, D. (1993): “Introducing: The Survey of Professional Forecasters,” *Business Review*, 6, 3–15.

- CROUSHORE, D. AND T. STARK (2001): “A real-time data set for macroeconomists,” *Journal of Econometrics*, 105, 111–130.
- CVITANIĆ, J., A. LAZRAK, L. MARTELLINI, AND F. ZAPATERO (2006): “Dynamic Portfolio Choice with Parameter Uncertainty and the Economic Value of Analysts’ Recommendations,” *Review of Financial Studies*, 19, 1113–1156.
- DA, Z. AND E. SCHAUMBURG (2011): “Relative Valuation and Analyst Target Price Forecasts,” *Journal of Financial Markets*, 14, 161–192.
- DANGL, T. AND M. HALLING (2012): “Predictive regressions with time-varying coefficients,” *Journal of Financial Economics*, 106, 157–181.
- DEL NEGRO, M. AND F. SCHORFHEIDE (2013): “DSGE Model-Based Forecasting,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier: Amsterdam, NL, vol. 2.
- DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009a): “A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms,” *Management Science*, 55, 798–812.
- DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009b): “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?” *Review of Financial Studies*, 22, 1915–1953.
- DEMIGUEL, V., A. MARTIN-UTRERA, AND F. J. NOGALES (2013): “Size matters: Optimal calibration of shrinkage estimators for portfolio selection,” *Journal of Banking & Finance*, 37, 3018–3034.
- DIEBOLD, F. X. AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 20, 134–144.
- DOAN, T., R. LITTERMAN, AND C. SIMS (1984): “Forecasting and Conditional Projection using Realistic Prior Distributions,” *Econometric Reviews*, 3, 1–100.
- DURBIN, J. AND S. KOOPMAN (2012): *Time Series Analysis by State Space Methods*, Oxford Statistical Science, Oxford University Press, 2 ed.
- FAHRMEIR, L., T. KNEIB, AND S. KONRATH (2010): “Bayesian regularisation in structured additive regression: A unifying perspective on shrinkage, smoothing and predictor selection,” *Statistics and Computing*, 20, 203–219.

- FAMA, E. F. AND K. R. FRENCH (1993): “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.
- FAN, J., J. ZHANG, AND K. YU (2012): “Vast Portfolio Selection With Gross-Exposure Constraints,” *Journal of the American Statistical Association*, 107, 592–606.
- FAUST, J. AND J. H. WRIGHT (2013): “Forecasting Inflation,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier: Amsterdam, NL, vol. 2.
- FENG, G. AND N. G. POLSON (2016): “Regularizing Bayesian Predictive Regressions,” *arXiv:1606.01701 [stat]*.
- FLETCHER, J. AND J. HILLIER (2001): “An Examination of Resampled Portfolio Efficiency,” *Financial Analysts Journal*, 57, 66–74.
- FRAHM, G. AND C. MEMMEL (2010): “Dominating estimators for minimum-variance portfolios,” *Journal of Econometrics*, 159, 289–302.
- FREY, C. AND F. MOKINSKI (2016): “Forecasting with Bayesian Vector Autoregressions Estimated Using Professional Forecasts,” *Journal of Applied Econometrics*, 31, 1083–1099.
- FREY, C. AND W. POHLMEIER (2016): “Bayesian Shrinkage of Portfolio Weights,” Available at SSRN 2730475.
- GEWEKE, J. F. (1996): “Bayesian Inference for Linear Models Subject to Linear Inequality Constraints,” in *Modelling and Prediction Honoring Seymour Geisser*, ed. by J. C. Lee, W. O. Johnson, and A. Zellner, Springer New York, 248–263.
- GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*, Wiley: Hoboken, NJ.
- GIACOMINI, R. AND G. RAGUSA (2014): “Theory-Coherent Forecasting,” *Journal of Econometrics*, 182, 145–155.
- GIACOMINI, R. AND H. WHITE (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74, 1545–1578.
- GOLDFARB, D. AND G. IYENGAR (2003): “Robust Portfolio Selection Problems,” *Mathematics of Operations Research*, 28, 1–38.
- GOLOSNOY, V. AND Y. OKHRIN (2007): “Multivariate Shrinkage for Optimal Portfolio Weights,” *The European Journal of Finance*, 13, 441–458.

- GOTO, S. AND Y. XU (2015): “Improving Mean Variance Optimization through Sparse Hedging Restrictions,” *Journal of Financial and Quantitative Analysis*, 50, 1415–1441.
- GRANGER, C. W. J. AND R. RAMANATHAN (1984): “Improved Methods of Combining Forecasts,” *Journal of Forecasting*, 3, 197–204.
- GRANGER, C. W. J. (1986): “[Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts]: Comment,” *Journal of Business & Economic Statistics*, 4, 16–17.
- GREEN, T. C. (2006): “The Value of Client Access to Analyst Recommendations,” *Journal of Financial and Quantitative Analysis*, 41, 1–24.
- HE, P. W., A. GRANT, AND J. FABRE (2013): “Economic value of analyst recommendations in Australia: an application of the Black–Litterman asset allocation model,” *Accounting & Finance*, 53, 441–470.
- HOERL, A. E. AND R. W. KENNARD (1970): “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, 12, 55–67.
- ING, C.-K. AND C.-Z. WEI (2003): “On Same-Realization Prediction in an Infinite-Order Autoregressive Process,” *Journal of Multivariate Analysis*, 85, 130–155.
- IRVINE, P. J. (2003): “The Incremental Impact of Analyst Initiation of Coverage,” *Journal of Corporate Finance*, 9, 431–451.
- JAGANNATHAN, R. AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *The Journal of Finance*, 58, 1651–1684.
- JEGADEESH, N., J. KIM, S. D. KRISCHE, AND C. M. C. LEE (2004): “Analyzing the Analysts: When Do Recommendations Add Value?” *The Journal of Finance*, 59, 1083–1124.
- JEGADEESH, N. AND W. KIM (2006): “Value of Analyst Recommendations: International Evidence,” *Journal of Financial Markets*, 9, 274–309.
- JOBSON, J. D. AND B. KORKIE (1980): “Estimation for Markowitz Efficient Portfolios,” *Journal of the American Statistical Association*, 75, 544–554.
- JOHANNES, M., A. KORTEWEG, AND N. POLSON (2014): “Sequential Learning, Predictability, and Optimal Portfolio Returns,” *The Journal of Finance*, 69, 611–644.
- JORION, P. (1986): “Bayes-Stein Estimation for Portfolio Analysis,” *Journal of Financial and Quantitative Analysis*, 21, 279–292.

- JUDGE, G. AND M. BOCK (1978): *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Studies in Mathematical and Managerial Economics, North-Holland Publications.
- KAN, R. AND G. ZHOU (2007): “Optimal Portfolio Choice with Parameter Uncertainty,” *Journal of Financial and Quantitative Analysis*, 42, 621–656.
- KANDEL, S. AND R. STAMBAUGH (1996): “On the Predictability of Stock Returns: An Asset-Allocation Perspective,” *Journal of Finance*, 51, 385–424.
- KANDEL, S. AND R. F. STAMBAUGH (1987): “Long-horizon returns and short-horizon models,” CRSP Working paper 222, University of Chicago.
- KEMPF, A. AND C. MEMMEL (2006): “Estimating the Global Minimum Variance Portfolio,” *Schmalenbach Business Review (SBR)*, 58, 332–348.
- KIM, D. AND A. ORPHANIDES (2012): “Term Structure Estimation with Survey Data on Interest Rate Forecasts,” *Journal of Financial & Quantitative Analysis*, 47, 241–272.
- KLEIN, R. AND V. BAWA (1976): “The Effect of Estimation Risk on Optimal Portfolio Choice,” *Journal of Financial Economics*, 3, 215–231.
- KOOP, G. AND D. KOROBILIS (2010): “Bayesian Multivariate Time Series Methods for Empirical Macroeconomics,” *Foundations and Trends® in Econometrics*, 3, 267–358.
- (2013): “Large time-varying parameter VARs,” *Journal of Econometrics*, 177, 185–198.
- KOROBILIS, D. (2013): “Hierarchical shrinkage priors for dynamic regressions with many predictors,” *International Journal of Forecasting*, 29, 43–59.
- KOURTIS, A., G. DOTSIS, AND R. N. MARKELLOS (2012): “Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix,” *Journal of Banking & Finance*, 36, 2522–2531.
- KRÜGER, F. (2014): “Combining Density Forecasts under various Scoring Rules: An Analysis of UK Inflation,” Working Paper, Heidelberg Institute for Theoretical Studies.
- KRÜGER, F., T. E. CLARK, AND F. RAVAZZOLO (2015): “Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts,” *Journal of Business & Economic Statistics*, 1–48.
- KYUNG, M., J. GILL, M. GHOSH, AND G. CASELLA (2010): “Penalized regression, standard errors, and Bayesian lassos,” *Bayesian Analysis*, 5, 369–411.

- LEDOIT, O. AND M. WOLF (2003): “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” *Journal of Empirical Finance*, 10, 603–621.
- (2008): “Robust performance hypothesis testing with the Sharpe ratio,” *Journal of Empirical Finance*, 15, 850–859.
- (2011): “Robust Performances Hypothesis Testing With the Variance,” *Wilmott*, 2011, 86–89.
- (2014): “Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks,” Working paper No. 137, Department of Economics, University of Zurich.
- LI, J. (2015): “Sparse and Stable Portfolio Selection with Parameter Uncertainty,” *Journal of Business & Economic Statistics*, 33, 381–392.
- LIN, S., H. TAN, AND Z. ZHANG (2016): “When Analysts Talk, Do Institutional Investors Listen? Evidence from Target Price Changes,” *Financial Review*, 51, 191–223.
- LINDLEY, D. V. AND A. F. M. SMITH (1972): “Bayes Estimates for the Linear Model,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 1–41.
- METAXOGLOU, K., D. PETTENUZZO, AND A. SMITH (2016): “Option-Implied Equity Premium Predictions via Entropic Tilting,” Available at SSRN 2754246.
- MICHAUD, R. O. (1989): “The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal?” *Financial Analysts Journal*, 45, 31–42.
- OKHRIN, Y. AND W. SCHMID (2007): “Comparison of different estimation techniques for portfolio selection,” *AStA Advances in Statistical Analysis*, 91, 109–127.
- PARK, T. AND G. CASELLA (2008): “The Bayesian Lasso,” *Journal of the American Statistical Association*, 103, 681–686.
- PÁSTOR, Ľ. (2000): “Portfolio Selection and Asset Pricing Models,” *The Journal of Finance*, 55, 179–223.
- PETTENUZZO, D. AND F. RAVAZZOLO (2016): “Optimal Portfolio Choice Under Decision-Based Model Combinations,” *Journal of Applied Econometrics*, forthcoming.
- PETTENUZZO, D. AND A. TIMMERMANN (2011): “Predictability of stock returns and asset allocation under structural breaks,” *Journal of Econometrics*, 164, 60–78.

- POLLAK, I. (2011): “Weight shrinkage for portfolio optimization,” in *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 37–40.
- RAFTERY, A. E., M. KÁRNÝ, AND P. ETTLER (2010): “Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill,” *Technometrics*, 52, 52–66.
- RAMNATH, S., S. ROCK, AND P. SHANE (2008): “The financial analyst forecasting literature: A taxonomy with suggestions for further research,” *International Journal of Forecasting*, 24, 34–75.
- RAPACH, D., J. STRAUSS, J. TU, AND G. ZHOU (2015): “Industry interdependencies and cross-industry return predictability,” *Available at SSRN 2566541*.
- RAPACH, D. AND G. ZHOU (2013): “Chapter 6 - Forecasting Stock Returns,” Elsevier, vol. 2, Part A of *Handbook of Economic Forecasting*, 328–383.
- ROBERTSON, J. C., E. W. TALLMAN, AND C. H. WHITEMAN (2005): “Forecasting Using Relative Entropy,” *Journal of Money, Credit and Banking*, 37, 383–401.
- SCHLEER, F. (2015): “Finding Starting-Values for the Estimation of Vector STAR Models,” *Econometrics*, 3, 65–90.
- STOCK, J. H. AND M. W. WATSON (2001): “Vector Autoregressions,” *The Journal of Economic Perspectives*, 15, 101–115.
- TIBSHIRANI, R. (1996): “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.
- TU, J. AND G. ZHOU (2011): “Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies,” *Journal of Financial Economics*, 99, 204–215.
- VILLANI, M. (2009): “Steady-State Priors for Vector Autoregressions,” *Journal of Applied Econometrics*, 24, 630–650.
- WACHTER, J. A. AND M. WARUSAWITHARANA (2015): “What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio,” *Journal of Econometrics*, 186, 74–93.
- WELCH, I. AND A. GOYAL (2008): “A Comprehensive Look at The Empirical Performance of Equity Premium Prediction,” *Review of Financial Studies*, 21, 1455–1508.

- WRIGHT, J. H. (2013): “Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior,” *Journal of Applied Econometrics*, 28, 762–776.
- YEN, Y. (2016): “Sparse Weighted-Norm Minimum Variance Portfolios,” *Review of Finance*, 20, 1259–1287.
- ZELLNER, A. AND V. CHETTY (1965): “Prediction and Decision Problems in Regression Models from the Bayesian Point of View,” *Journal of the American Statistical Association*, 60, 608–616.
- ZOU, H. AND T. HASTIE (2005): “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320.

Abgrenzung

Das erste Kapitel “Forecasting with Bayesian Vector Autoregressions estimated using Professional Forecasts” ist in Zusammenarbeit mit Frieder Mokinski entstanden. Meine individuelle Leistung bei der Erstellung dieser Arbeit beträgt 40 Prozent.

Das zweite Kapitel “Bayesian Shrinkage of Portfolio Weights” ist in Zusammenarbeit mit Professor Dr. Winfried Pohlmeier entstanden. Der Anteil meiner eigenen Leistung bei der Erstellung dieser Arbeit beträgt 75 Prozent.

Ich versichere hiermit, dass ich das dritte Kapitel “Using Analysts’ Forecasts for Stock Predictions - An Entropic Tilting Approach” selbstständig und ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfemittel erstellt habe.