

ESSAYS IN EMPIRICAL ASSET PRICING

Irina Pimenova

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Francis J. DiTraglia, Assistant Professor of Economics

Graduate Group Chairperson

Jesús Fernández-Villaverde, Professor of Economics

Dissertation Committee

Francis J. DiTraglia, Assistant Professor of Economics

Frank Schorfheide, Professor of Economics

Xu Cheng, Associate Professor of Economics

ESSAYS IN EMPIRICAL ASSET PRICING

© COPYRIGHT

2018

Irina Pimenova

This work is licensed under the
Creative Commons Attribution
NonCommercial-ShareAlike 4.0
International License (CC BY-NC-SA 4.0)

To view a copy of this license, visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

ACKNOWLEDGEMENT

The five years of graduate school have been a period of intense learning for me, not only in the scientific arena, but also on a personal level. I would not have finished the graduate school without the amazing people I have been surrounded by.

First and foremost I would like to thank my advisor, Professor Francis DiTraglia, for his fundamental role in my doctoral work. I greatly appreciate his patient guidance, encouragement, and advice he has provided throughout my time as his student to make my Ph.D. experience productive and stimulating.

I am also grateful to my other dissertation committee members, Professors Xu Cheng and Frank Schorfheide, for their guidance and insightful comments. Without their help, this dissertation could not be completed.

I am greatly indebted to Professor Jakub Jurek, who introduced me to the exciting world of empirical finance and taught me how good applied research is done. I have learned a lot from him, and the summers that I have spent working with him have been the highlights of my graduate school experience.

I would like to thank my friends in the economics department for all the great times that we have shared. I am particularly grateful to Hanna Wang for her infinite patience and support.

Lastly, I would like to thank my family. I am grateful to my parents for all their love and encouragement and for being there for me since day one. I thank my sister for being my role model despite being younger than me. Without them, this thesis would never have been written.

ABSTRACT

ESSAYS IN EMPIRICAL ASSET PRICING

Irina Pimenova

Francis J. DiTraglia

In this dissertation, I revisit two problems in empirical asset pricing. In Chapter 1, I propose a methodology to evaluate the validity of linear asset pricing factor models under short sale restrictions using a regression-based test. The test is based on the revised null hypothesis that intercepts obtained from regressing excess returns of test assets on factor returns, usually referred to as alphas, are non-positive. I show that under short sale restrictions a much larger set of models is supported by the data than without restrictions. In particular, the Fama-French five-factor model augmented with the momentum factor is rejected less often than other models. In Chapter 2, I investigate patterns of equity premium predictability in international capital markets and explore the robustness of common predictive variables. In particular, I focus on predictive regressions with multiple predictors: dividend-price ratio, four interest rate variables, and inflation. To obtain precise estimates, two estimation methods are employed. First, I consider all capital markets jointly as a system of regressions. Second, I take into account uncertainty about which potential predictors forecast excess returns by employing spike-and-slab prior. My results suggest evidence in favor of predictability is weak both in- and out-of-sample and limited to a few countries. The strong predictability observed on the U.S. market is rather exceptional. In addition, my analysis shows that considering model uncertainty is essential as it leads to a statistically significant increase of investors' welfare both in- and out-of-sample. On the other hand, the welfare increase associated with considering capital markets jointly is relatively modest. However, it leads to reconsider the relative importance of predictive variables because the variables that are statistically significant predictors in the country-specific regressions are insignificant when the capital markets are studied jointly. In particular, my results suggest that the

in-sample evidence in favor of the interest rate variables, that are believed to be among the most robust predictors by the literature, is spurious and is mostly driven by ignoring the cross-country information. Conversely, the dividend-price ratio emerges as the only robust predictor of future stock returns.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	iii
ABSTRACT	iv
LIST OF TABLES	ix
LIST OF ILLUSTRATIONS	x
CHAPTER 1 : Short Alpha: Linear Factor Model Diagnostics Under Short Sale	
Constraints	1
1.1 Introduction	1
1.2 Econometric Methods	4
1.3 Simulation: Comparing Finite-Sample Performance	14
1.4 Application:	
Can Short Sale Restrictions Help to Explain Returns?	18
1.5 Robustness Check: Transaction Costs	25
1.6 Conclusion	27
1.7 Appendix	29
CHAPTER 2 : International Return Predictability	
2.1 Introduction	31
2.2 Methodological Framework	37
2.3 Data Overview	43
2.4 In-Sample Analysis	48
2.5 Out-of-sample Analysis	61
2.6 Extensions	68
2.7 Conclusion	83

2.8 Appendix	87
BIBLIOGRAPHY	96

LIST OF TABLES

TABLE 1 :	Simulation Results for Mean Test	17
TABLE 2 :	Results for Sorts Based on Factors	21
TABLE 3 :	Results for Sorts Based on Anomalies	23
TABLE 4 :	Results for Industry Portfolios	25
TABLE 5 :	Required Transaction Costs	26
TABLE 6 :	Simulation for Mean Test with GARCH Disturbances	29
TABLE 7 :	Summary Statistics of International Returns	46
TABLE 8 :	Summary Statistics of Predictive Variables	47
TABLE 9 :	Models with Highest Posterior Odds	50
TABLE 10 :	Bayesian Estimates of Slope Parameters (Basic Specification).	53
TABLE 11 :	Bayesian Estimates of Inclusion Probabilities (Basic Specification)	54
TABLE 12 :	In-Sample Asset Allocation	59
TABLE 13 :	In-Sample Certainty Equivalent Returns	60
TABLE 14 :	R^2_{OOS} (Basic Specification)	63
TABLE 15 :	Out-of-Sample Asset Allocation (Basic Specification)	67
TABLE 16 :	Out-of-Sample Certainty Equivalent Returns (Basic Specification) .	67
TABLE 17 :	Bayesian Estimates of Inclusion Probabilities (Borrowing from Cross- Section)	73
TABLE 18 :	R^2_{OOS} (Borrowing From Cross-Section)	76
TABLE 19 :	Out-of-Sample Asset Allocation (Borrowing From Cross-Section) .	77
TABLE 20 :	Out-of-Sample Certainty Equivalent Returns (Borrowing From Cross- Section)	78
TABLE 21 :	Bayesian Estimates of Inclusion Probabilities (Stochastic Volatility)	83
TABLE 22 :	R^2_{OOS} (Stochastic Volatility)	84
TABLE 23 :	Out-of-Sample Asset Allocation (Stochastic Volatility)	85

LIST OF ILLUSTRATIONS

FIGURE 1 :	Illustration of Moment Selection	7
FIGURE 2 :	Posterior Distribution of Slope Coefficients (Basic Specification) .	49
FIGURE 3 :	Net-SSE	66
FIGURE 4 :	Posterior Distribution of Slope Coefficients (Borrowing from Cross- Section)	72
FIGURE 5 :	Posterior Distribution of Slope Coefficients (Stochastic Volatility)	81

CHAPTER 1 : Short Alpha: Linear Factor Model Diagnostics Under Short Sale Constraints

1.1. Introduction

Standard econometric tests for evaluating the validity of a linear asset pricing factor model with traded factors focus on intercepts, usually referred to as alphas, obtained from regressing excess returns of test assets on factor returns.¹² The alphas are expected to be jointly zero if factors are mean-variance efficient, and the model is rejected if alphas are not zero. However, these tests ignore a crucial assumption: mean-variance efficiency only implies zero alphas in the absence of market frictions such as short sale restriction, taxes, transaction costs, liquidity constraints, etc. The factor model may thus be rejected not because the inspected factors are not mean-variance efficient, but rather due to the market frictions. Very few papers re-consider the implications of mean-variance efficiency for factor model testing in the presence of market frictions.

In this chapter I explore one of the most prominent market frictions: the short sale restriction. A short sale is the sale of an asset that is not owned by the seller. This transaction is risky as the potential loss is unlimited, expensive due to the lending fees, and in some cases outright infeasible because of scarce lending markets. Consequently many investors either can not or do not want to sell short and are thus short sale constrained. Short-sale constraints are an important factor in determining asset prices as shown, for example, in a survey by Rubinstein (2004).³ If short sales are constrained, negative alphas cannot be arbitrated away. Hence, negative alphas are consistent with mean-variance efficiency under a short sale restriction as shown by DeRoos and Nijman (2001) and AitSahlia et al. (2017).

¹A recent survey of the econometrics underlying mean-variance efficiency tests can be found in Sentana (2009).

²“Factor model” here is defined as a set of traded factors on the right hand side in the regression. The model is considered to be valid if the factors are mean-variance efficient.

³There is large literature on the impact of short sale constraints on asset prices. The theoretical work includes Miller (1977), Jarrow (1980), Diamond and Verrecchia (1987), Allen et al. (1993), Duffie et al. (2002), Hong and Stein (2003), etc. The empirical work includes Chen et al. (2002), Ofek and Richardson (2003), Nagel (2005), Saffi and Sigurdsson (2010), Beber and Pagano (2013) and others.

This leads me to propose a model validity test under the short sale constraint based on the null hypothesis that alphas are non-positive.

The null hypothesis is an inequality restriction and is hence more difficult to test than an equality-based null, so I employ the moment inequality literature, a strand of literature in econometrics focusing on testing procedures exploring nulls that contain inequalities. In particular, I apply two testing procedures: one suggested by Andrews and Soares (2010) (AS) and one by Romano et al. (2014) (RSW). The difference between the two procedures is explained later in the text. In addition to applying the methods, I compare their finite sample properties in simulation, as this has not been done before. My findings suggest that neither test exhibits size distortions in small samples and that both tests are robust to conditional heteroscedasticity. I also find that the RSW is slightly less powerful than the AS test.

Using these tests I evaluate several factor models (defined here as a combination of factors). Among others, I examine the two classic models: the CAPM with a single market factor proposed by Lintner (1965) and Sharpe (1964) and the Fama-French three factor model of Fama and French (1993). I also include some more recent models such as the q -factor model of Hou et al. (2014a) and the Fama-French five-factor model of Fama and French (2015b). As the test assets I use sets of testing portfolios labeled as “pricing anomalies”.

I show that under short sale restrictions a much larger set of models is supported by the data than without restrictions. In particular, the five factor model augmented with the momentum factor is not rejected for multiple pricing anomalies when the short sale restrictions are taken into account. I also find some support for the four-factor model that include the Fama-French three factors and the momentum factor. These findings suggest that momentum factor is not redundant. In contrast, I strongly reject the CAPM and the original Fama-French three-factor model and find limited support for the q -factor model for almost all sets of test assets. These models are rejected even under the short sale restriction because they produce large, positive alphas for portfolios that include small companies.

One may hypothesize that the rejection is due to the transaction costs associated with small portfolios. However, I show that the decision about which models are rejected or not are robust after taking into account transaction costs. In particular, the CAPM and the original Fama-French three-factor model require assuming very large transaction costs to justify the mis-pricing.

This chapter contributes to the literature on mean-variance efficiency under short sale restriction. The literature on short sale restrictions in the context of factor models is limited to the recent paper by Dugalic and Patino (2017), who build a CAPM-like structural model with short sale constraints and obtain an equation that links asset-specific excess returns with the market equity premium. However, the derived alphas are only valid if there is a single risk factor, market factor, and extending this setup for other models is challenging because it requires building a new structural model for each combination of factors. In contrast, my approach can be easily applied to any set of traded factors.

This chapter is the first to consider mean-variance efficiency testing under short sale restrictions in the context of factor models, however, some papers studied it in different contexts. In particular, DeRoos and Nijman (2001) and Li et al. (2003) test whether U.S. investors can extend their efficient set by investing in emerging markets when accounting for short sale restrictions. Ait-Sahalia et al. (2017), Tang et al. (2010) and Elton et al. (2006) explore the efficiency of portfolios offered in 401(k) plans under short sale restrictions. These papers usually focus on one or two test assets, while testing the validity of factor models requires using from ten to thirty test assets. Consequently, factor models testing requires tests that have high power even if the number of test assets is large. For this reason, the testing procedures I use differ from the ones used in earlier papers. I explain later in the chapter how the two tests I consider maintain high power.

More generally, this chapter also contributes to the literature that compares asset pricing factor models by testing them against pricing anomalies. The most recent study by Hou et al. (2017) uses hundreds of significant anomalies as testing portfolios and finds the q -

factor model of and the Fama-French five-factor model to be the best performing models. As mentioned previously, this strand of literature does not take into account short sale restrictions.

This chapter is organized as follows. In the second section, I present the econometric framework and discuss the two testing procedures. Next I compare the finite sample performance of the two types of tests in a simulation experiment. In the fourth section I apply the tests to several linear factor models and multiple sets of test assets. Some concluding remarks are offered in the final section.

1.2. Econometric Methods

This section presents the econometric framework, discusses the intuition behind inequality testing and introduces two alternative procedures that can be used to test the null hypothesis based on the papers by Andrews and Soares (2010) (AS) and Romano et al. (2014) (RSW).

1.2.1. Econometric Framework and Notation

Consider a linear asset pricing factor model:

$$R_{i,t} = \alpha_i + \sum_{k=1}^K \beta_{i,k} f_{k,t} + \varepsilon_{i,t} \quad \forall i = 1, \dots, N, \quad \forall t = 1, \dots, T. \quad (1.1)$$

The excess return over a risk free rate on an asset i , $R_{i,t}$ (also called risk premium), is linearly related to K traded risk factors, $f_{k,t}$, with the sensitivity to a factor k given by $\beta_{i,k}$ (beta). The alpha α_i represents how much extra return an asset i is expected to deliver in addition to the reward for risk represented by factors. The return in each period is affected by an idiosyncratic shock $\varepsilon_{i,t}$. These shocks are jointly distributed with zero mean and a nonsingular covariance matrix Σ_ε .

For convenience, the model in (1.1) is re-written in matrix form. Stacking the returns for each factor k at period t in a $K \times 1$ vector f_t and factor loadings for for each asset i into

$K \times 1$ vector β_i yields:

$$R_{i,t} = \alpha_i + f_t^\top \beta_i + \varepsilon_{i,t}$$

Stacking returns on all assets in period t into a $N \times 1$ vector R_t , all idiosyncratic shocks at period t into vector ε_t , and all factor loadings β_i into a $N \times K$ matrix B yields:

$$R_t = \alpha + Bf_t + \varepsilon_t, \tag{1.2}$$

The model in (1.2) can be estimated by equation-by-equation OLS. OLS estimation in this case is equivalent to estimating a system of unrelated equations because the right-hand-side variables are the same for all test assets.

It is helpful to introduce some notation. The $N \times 1$ estimated vector of alphas is denoted $\hat{\alpha}$, and the $N \times N$ matrix $\hat{\Sigma}_\alpha$ is a consistent estimator of the covariance matrix of the scaled vector of alphas, $\sqrt{T}\hat{\alpha}$. The i th diagonal element of the covariance matrix is denoted by $\hat{\sigma}_{\alpha,i}^2$. I use bootstrap in the next section to construct the distribution of the test statistic under the null. Let R denote the number of bootstrap samples generated for N assets observed over T periods. For each bootstrap sample r , estimates of alphas $\hat{\alpha}_r$ and the covariance matrix $\hat{\Sigma}_{\alpha,r}$ are obtained. The significance level of the test is denoted by a .

1.2.2. Null Hypothesis and Test Statistic

The null hypothesis of factor mean-variance efficiency under short sale restriction is that all alphas are non-positive:

$$\begin{aligned} \mathcal{H}_0 : \alpha_i \leq 0 \quad \forall i = 1, \dots, N \\ \mathcal{H}_1 : \alpha_i > 0 \text{ for some } i \end{aligned} \tag{1.3}$$

The form of the null hypothesis in (1.3) was introduced by DeRoos and Nijman (2001), who approached the problem from the perspective of a stochastic discount factor, and later

verified by AitSahlia et al. (2017), who used direct portfolio optimization.

The intuition behind the null hypothesis in (1.3) is based on the idea of arbitrage. Assume that the alpha α_i is negative, so it delivers lower expected returns than other assets with a similar risk profile. An arbitrage strategy to exploit this mis-pricing involves selling the asset i short and buying a portfolio of factors. Consequently, if investors face short sale restrictions for some assets and can not short sell the asset i , they cannot exploit the arbitrage strategy, so the mis-pricing may persist with alpha remaining negative. Note also that a combination of factors can be mean-variance efficient if the short sale restriction is in place and not mean-variance efficient otherwise.

The test statistic for the null in (1.3) is motivated by the likelihood ratio test under the assumption of normality but can also be applied if the assumption of normality does not hold as in case of returns. This quasi-likelihood ratio (QLR), or Wald test statistic, first introduced by Kudo (1963), takes the following form:

$$W = T \inf_{\alpha_0 \leq 0} (\hat{\alpha} - \alpha_0)^\top \hat{\Sigma}_\alpha^{-1} (\hat{\alpha} - \alpha_0) = T (\hat{\alpha} - \hat{\alpha}(\mathbb{R}_N^-))^\top \hat{\Sigma}_\alpha^{-1} (\hat{\alpha} - \hat{\alpha}(\mathbb{R}_N^-)) \quad (1.4)$$

where $\hat{\alpha}(\mathbb{R}_N^-)$ is a restricted estimate of alpha under the non-positivity constraint. The restriction on vector $\alpha_0 \leq 0$ is to be understood component-wise. The Wald test statistic in (1.4) can be seen as a weighted distance between the alphas, estimated with and without the restriction of all elements being non-positive. Alphas that are originally “large” and positive have more weight on the value of the test statistic because they are more likely to be “far” from the corresponding non-positive estimator.

To develop the intuition behind the Wald statistic in (1.4), consider a simplified example with two test assets. The covariance matrix of alphas is known and equal to an identity matrix $\Sigma_\alpha = I$. The identity covariance matrix yields two simplifications: 1. the Wald test statistic is equal to a sum of squared deviations from zero and 2. under the diagonal covariance matrix the restricted estimate of alphas under non-positivity constraint is equal

to the unrestricted estimate with positive elements set to zero (see Figure 1). In contrast, under a non-diagonal covariance matrix, not only the positive elements are set to zero but also the negative elements may take a new value.

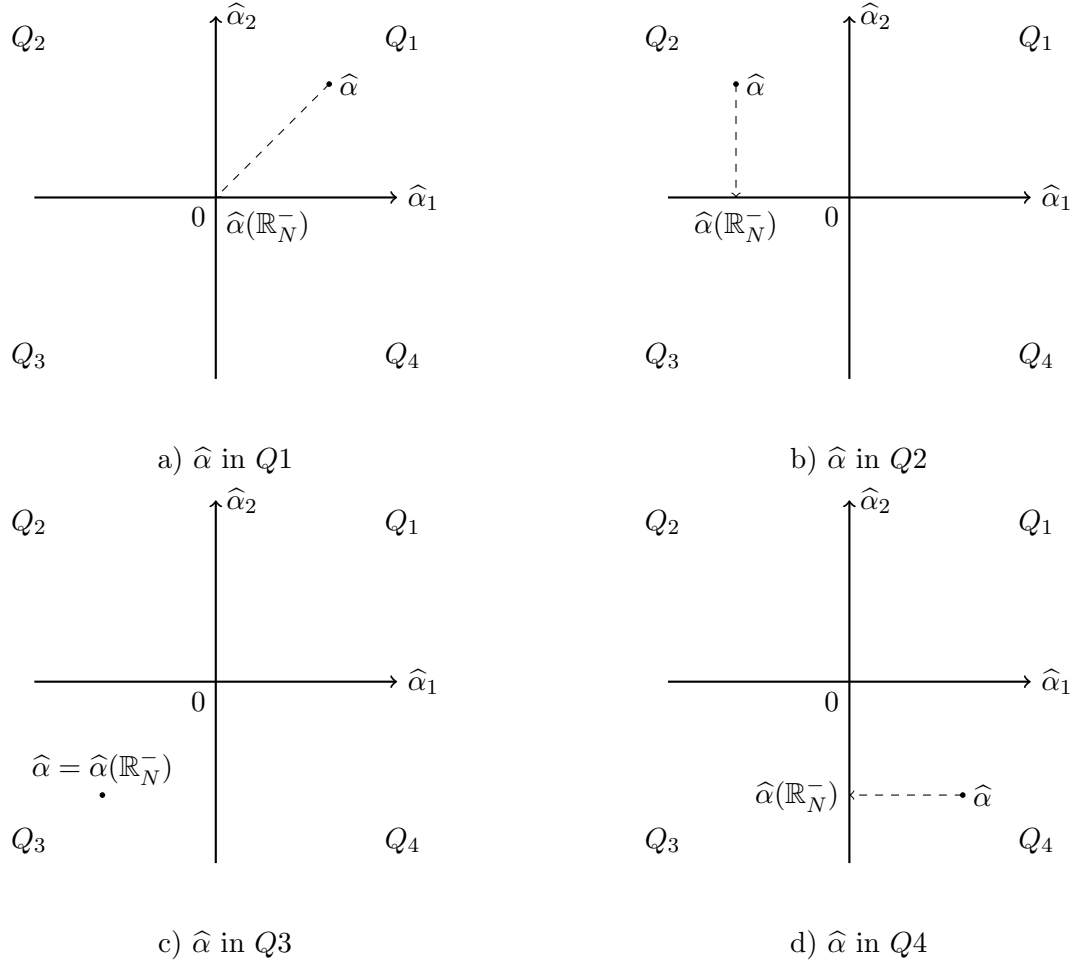


Figure 1: Illustration of Moment Selection

Four panels illustrate unrestricted estimates $\hat{\alpha}$ and restricted estimates under the non-positivity constraints $\hat{\alpha}(\mathbb{R}_N^-)$ under the identity covariance matrix.

Let Q_1, Q_2, Q_3 and Q_4 denote four quadrants in the two-dimensional space as in the Figure

1. Note that the Wald test statistic in (1.4) and its asymptotic distribution is given by:

$$W = \begin{cases} 0 & \text{if } \hat{\alpha} \in Q_1 \\ \hat{\alpha}_1^2 \sim \chi_1^2, & \text{if } \hat{\alpha} \in Q_2 \\ \hat{\alpha}_1^2 + \hat{\alpha}_2^2 \sim \chi_2^2, & \text{if } \hat{\alpha} \in Q_3 \\ \hat{\alpha}_2^2 \sim \chi_1^2, & \text{if } \hat{\alpha} \in Q_4 \end{cases} \quad (1.5)$$

The distributional result in (1.5), which was obtained under identity covariance matrix, can be generalized for any covariance matrix: given the number of non-negative elements p , the distribution of the test statistic is χ^2 with p degrees of freedom. However, the number of non-negative elements p is unknown. The unconditional distribution of the test statistic in (1.4) under the null is then a mixture of χ^2 distributions with weights given by the probability that the estimated vector is in a given quadrant (conditional on the covariance matrix). Kodde and Palm (1986) introduced a testing procedure based on this intuitively appealing idea and showed that the test is asymptotically valid.

The Kodde and Palm (1986) test works great if the number of dimensions is small but shows poor power properties if the dimensionality is large (which is the case for factor models). As the number of non-negative elements p is actually unknown, the test is based on the assumption that all inequalities can be binding. Suppose that the estimated alphas are in the Q_2 as in panel b in Figure 1 above with $\hat{\alpha}_1$ "far" from zero. The inequality $\hat{\alpha}_1 \leq 0$ is not binding and does not influence the value of the test statistic. If the distribution is constructed under the assumption that this inequality is potentially binding, it would produce a high critical value.

The problem of low power is aggravated as dimensionality increases. With higher N this test considers more and more areas which are not relevant for the case at hand, but do influence the critical value. Suppose $N - 1$ alphas are negative and "far" from the zero boundary, and only one alpha is positive, so that the $N - 1$ inequalities do not contribute to

the value of the test statistic. If the critical value equally depends on negative and positive alphas, it would be larger with $N - 1$ negative alphas than it would be if the negative alphas were absent. For big N , the critical value would be especially large, leading to low power.

The solution to the low power problem in large N setting is to somehow isolate the inequalities that are “far” from binding, and treat them differently than potentially binding inequalities - an approach called moment-selection.

1.2.3. Two Tests: AS and RSW

I consider two testing methodologies based on the idea of moment selection proposed by Andrews and Soares (2010) and Romano et al. (2014). The two methods differ in how they identify non-binding inequalities and use this information. Andrews and Soares (2010) completely eliminate inequalities that are “far” from binding based on the t statistic when constructing the distribution under the null. Romano et al. (2014) build a confidence interval with a “small” significance level to identify inequalities that are “far” from binding, and shift the null hypothesis, so that these constraints have a limited contribution to the critical value. Other tests that implicitly or explicitly use the information were suggested by Romano and Shaikh (2008), Andrews and Guggenberger (2009), Canay (2010) and Andrews and Barwick (2012). The overview of the inequality-constrained inference can be found in Silvapulle and Sen (2011).

AS Test

Andrews and Soares (2010) and Andrews and Barwick (2012) suggest to eliminate inequalities that are “far” from binding when constructing the distribution of the test statistic in (1.4) under the null. The suggested procedure consists of two steps: 1. determine inequalities that are “far” from binding and 2. find the distribution of the test statistic using only inequalities that are potentially binding.

The authors use a t -statistic to identify inequalities that are “far” from binding. In the

context of this chapter the inequality is “far” from binding if the estimated alpha i is negative and large relative to some cut-off parameter:

$$t_i = -\frac{\hat{\alpha}_i}{\hat{\sigma}_{\alpha_i}/\sqrt{T}} > \kappa. \quad (1.6)$$

A suitable choice of κ is based on Bayesian Information Criterion (BIC) $\kappa_{BIC} = [\ln(T)]^{1/2}$.⁴ The authors show in the simulation that this choice of the cut-off parameter results in a test with good finite-sample properties.

The distribution is constructed via bootstrap after eliminating all alphas that satisfy the inequality. After generating a bootstrap sample, all assets that satisfy the restriction are eliminated from further consideration. The critical value is obtained as a $1 - a$ quantile of the bootstrapped distribution.

Algorithm for the AS testing procedure:

0. Estimate the linear factor model in (1.1) by OLS to obtain $\hat{\alpha}$ and $\hat{\Sigma}$. Calculate the Wald test statistic in (1.4).
1. Find the “far” from binding inequalities.
 - (a) Calculate t -statistic for each alpha as in (1.6)
 - (b) Compare the value of the t -statistic with some cut-off value κ (e.g. κ_{BIC}) and find the inequalities that are “far” from binding.
2. Estimate the distribution of the test statistic under the null and the corresponding critical value at the significance level a .

⁴The performance of the procedure in finite samples is determined by the choice of the cut-off parameter κ , so the actual size may be different from the nominal significance level. Andrews and Barwick (2012) introduced a refined procedure with size-correction and a data-dependent tuning parameter κ , which results in a more powerful test. The determination of the size-correction parameter is computationally very complex, so the authors have to limit their attention to the case where $N \leq 10$, which is too restrictive for the factor model application where the number of test assets is usually much larger. For this reason, this test is not considered in this chapter.

- (a) Simulate R bootstrap samples each of size T .
- (b) For each bootstrap sample r , estimate the linear factor model in (1.1) by OLS to obtain $\hat{\alpha}_r$ and $\hat{\Sigma}_r$.
- (c) Eliminate the elements in $\hat{\alpha}_r$ and $\hat{\Sigma}_r$ that correspond to alphas that satisfy the condition 1.6.
- (d) Calculate the bootstrapped value of the Wald statistic as in (1.4), re-centered at the estimated value of alphas in the original sample:

$$W_r = T \inf_{\alpha_0 \leq 0} (\hat{\alpha}_r - \hat{\alpha} - \alpha_0)^\top \hat{\Sigma}_\alpha^{-1} (\hat{\alpha}_r - \hat{\alpha} - \alpha_0)$$

- (e) Find the critical value as $1 - a$ quantile.

RSW Test

Romano et al. (2014) suggested a two-step test. In the first step, a confidence region is constructed at some “small” significance level b to determine which inequalities are “far” from binding. This information is then used at the second step to estimate the distribution of the test statistic. The procedure is designed to take into account that the actual value of alpha may not be inside the confidence region, which is achieved by a Bonferonni-type correction.

At the first step, an upper confidence rectangle for the vector of alphas is constructed at the nominal level $1 - b$ in order to determine which inequalities hold. The confidence region is based on an inverted max t -statistic:

$$t^{max} = \max_{1 \leq i \leq N} t_i = \max_{1 \leq i \leq N} \frac{\alpha_i - \hat{\alpha}_i}{\hat{\sigma}_{\alpha_i}^2 / \sqrt{T}} \quad (1.7)$$

The goal is to find a cut-off value d , such that for the true mean α , the probability to

observe at least this value is $Pr(t^{max} \leq d) = 1 - b$. The distribution of the max t -statistic is constructed via bootstrap, and the multiplier d is obtained as a $1 - b$ quantile of the distribution. The confidence region based on this bootstrap limits the possible values of the true α . With probability at least $1 - b$ the biggest possible value of the true α_i under the null hypothesis is:

$$\alpha_i^* = \min \left\{ \hat{\alpha}_i + d \frac{\hat{\sigma}_{\alpha_i}}{\sqrt{T}}, 0 \right\} \quad (1.8)$$

Stacked together, they form a vector α^* .

In the second step, the confidence region obtained in step one is used to construct the distribution of the Wald test statistic. The higher power is achieved by tightening the null hypothesis. With probability at least $1 - b$, the value of the true population alphas under the null is bounded above by α^* . Instead of computing the critical value for the least favorable scenario under the null $\alpha = 0$, we can compute the critical value using the largest possible value under the null $\alpha = \alpha^* \leq 0$. This is equivalent to shifting the null hypothesis from $\alpha \leq 0$ to $\alpha \leq \alpha^*$. The Wald test statistic in (1.4) can be modified as follows:

$$\widetilde{W} = T \inf_{\alpha_0 \leq \alpha^*} (\hat{\alpha} - \alpha_0)^\top \widehat{\Sigma}_\alpha^{-1} (\hat{\alpha} - \alpha_0) = T \inf_{\alpha_0^* \leq 0} (\hat{\alpha} + \alpha^* - \alpha_0^*)^\top \widehat{\Sigma}_\alpha^{-1} (\hat{\alpha} + \alpha^* - \alpha_0^*), \quad (1.9)$$

where $\alpha_0^* = \alpha_0 - \alpha^*$. We can now bootstrap the distribution of the Wald statistic.

This test is conservative by construction. It fails to reject the null in two cases: when either test statistic is less than or equal to the critical value, or when all values in the confidence region are negative. In order to account for the fact that alphas may actually be outside of the confidence region with probability at most b , the null hypothesis should be rejected if the value of the test statistic exceeds $1 - a + b$ quantile of the bootstrapped distribution rather than $1 - a$ quantile.

Algorithm for the RSW testing procedure:

0. Estimate the linear factor model in (1.1) by OLS to obtain $\hat{\alpha}$ and $\hat{\Sigma}$. Calculate the Wald test statistic in (1.4).

1. Construct the $1 - b$ confidence region.

(a) Simulate R bootstrap samples each of size T .

(b) For each bootstrap sample r , estimate the linear factor model in (1.1) by OLS to obtain $\hat{\alpha}_r$ and $\hat{\Sigma}_r$.

(c) For each bootstrap sample r , calculate the bootstrapped value of the max t -statistic as in (1.4) re-centered at the estimated value of alphas in the original sample:

$$t_r^{max} = \max_{1 \leq i \leq N} \frac{\hat{\alpha}_i - \hat{\alpha}_{i,r}}{\hat{\sigma}_{\alpha_i,r}^2 / \sqrt{T}}.$$

(d) Find the $1 - b$ quantile of the max t -statistic distribution d .

(e) Find the upper bound of the confidence region on the true alphas under the null α^* as in (1.8).

2. Estimate the distribution of the test statistic under the null and the corresponding critical value at the significance level a .

(a) Bootstrap the data creating R bootstrap samples.

(b) For each bootstrap sample r , calculate the bootstrapped value of the modified Wald statistic under the tighter null hypothesis as in (1.9). The bootstrapped statistic is re-centered at the estimated value of alphas in the original sample:

$$\widetilde{W}_r = T \inf_{\alpha_0^* \leq 0} (\hat{\alpha}_r - \hat{\alpha} + \alpha^* - \alpha_0^*)^\top \widehat{\Sigma}_\alpha^{-1} (\hat{\alpha} - \hat{\alpha} + \alpha^* - \alpha_0^*).$$

(c) Find the critical value as $1 - a + b$ quantile.

The properties of the testing procedure depend on the choice of the significance level b . We can find an “optimal” value of b that maximizes the weighted average power. The drawback of this approach is that finding the maximum is difficult and this can only be done in simulation. The simulation results would only be valid for specific parametric assumptions and under a specific vector of alternatives, for which the power is computed. The authors found that a reasonable value of b is $a/10$. Larger values of b result in lower power, while lower values of b do not provide a noticeable increase in power but require a larger number of bootstrap samples to accurately estimate the quantile.

1.3. Simulation: Comparing Finite-Sample Performance

This section compares finite-sample properties of the AS and the RSW tests. The simulation focuses on a testing inequality restriction imposed on means, with the null being that all the elements of the mean vector are non-positive. The tests are compared in terms of maximum null rejection probability (MNRP) and average power based on simulation. Empirical MNRPs are computed as the maximum rejection probability over all vectors of means that contain only zero and $-\infty$ entries, which makes $2^N - 1$ vectors for each dimensionality N . Average power is computed as the average rejection probability of a pre-defined set of mean vectors.

I show that the AS test demonstrates both higher empirical MNRPs and average power consistent with its asymptotic properties. The AS test exhibits a higher power even after size is matched to that of the RSW. I also find that both tests tend to reject more often if the distribution is fat-tailed but the increase of the rejection probability is relatively modest.

The simulation is based on the assumption that the returns are i.i.d., which is not the case as returns are known to exhibit volatility clustering. I repeat the simulation with conditionally heteroscedastic errors and find that the both tests are robust to conditional heteroscedasticity and demonstrate the same size and power properties. Interested readers can find more details in the appendix.

1.3.1. Simulation Setup

In terms of the considered DGPs, mean vectors and covariance matrices the design of the simulation study is similar to that of Romano et al. (2014), who, in turn, replicated the design by Andrews and Barwick (2012). Despite multiple papers running similar simulation design, the direct comparison between the AS and the RSW has not been done yet and this chapter fills the gap. I also add a simulation with conditionally heteroscedastic errors, which was not done before. In addition, I use a different sample size T to fit my empirical application. It was already shown by Andrews and Soares (2010) that the AS test based on the BIC cutoff value has a more accurate MNRP, so I use this version of the test only.

The simulation covers two DGPs: normal and Student- t with three degrees of freedom that are normalized to have unit variance. The monthly returns are known to have fat tails but are not as fat-tailed as Student- t with three degrees of freedom. It is reasonable to expect that the real distribution is somewhere between normal and Student- t with three degrees of freedom. The original studies also consider χ^2 , which is omitted from the current analysis because it is of little relevance for financial returns.

The dimensionality of the mean vector is $N = 2, 4$ and 10 . For each value of N , three covariance matrices are considered: Σ_{Zero} , Σ_{Neg} and Σ_{Pos} . Σ_{Zero} is an identity matrix. Σ_{Neg} and Σ_{Pos} are Toeplitz matrices based on the following correlation vectors: for $N = 2$, $\rho = -0.9$ for Σ_{Neg} and $\rho = 0.5$ for Σ_{Pos} ; for $N = 4$, $\rho = (-0.9, 0.7, -0.5)$ for Σ_{Neg} and $\rho = (0.9, 0.7, 0.5)$ for Σ_{Pos} ; for $N = 10$, $\rho = (-0.9, 0.8, -0.7, 0.6, -0.5, 0.4, -0.3, 0.2, -0.1)$ for Σ_{Neg} and $\rho = (0.9, 0.8, 0.7, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5)$ for Σ_{Pos} . As returns can be both negatively and positively correlated Σ_{Neg} seems to be the most relevant case.

The set of mean vectors μ , for which average power is computed, is designed for each N and covariance matrix to achieve a theoretical power envelope of 75%, 80% and 85% for $N = \{2, 4, 10\}$ respectively. For $N = 2$, the set includes 7 elements; for $N = 4$, the set includes 24 vectors; and for $N = 10$, the set includes 40 vectors. The values of the mean

vector for each combination of N and covariance matrix type can be found in section 7.1 of Andrews and Barwick (2012). The sign of the mean vector is flipped because the null is that all means are non-positive rather than non-negative as in the original paper.

Sample size is fixed to $T = 500$, which corresponds to approximately 20 years of monthly data, so that it is comparable to the sample used for the empirical application.

The results are based on 10,000 repetitions, except for $N = 10$ where 5,000 repetitions are used for the MNRP calculations. The bootstrapped values are based on 3,000 samples for the AS and both steps of the RSW. The significance level is 5%, and the significance level for the first step of the RSW procedure is 0.5% as recommended by the authors.

1.3.2. Maximum Null Rejection Probabilities

The results are can be found in the upper half of Table 1.

All procedures achieve satisfactory performance with MNRP close to the nominal size of 5%. On average, the empirical MNRP for the AS test is higher than that of the RWS test. I find it also to be true for each particular mean vector. Both tests tend to reject more often if the covariance matrix is non-diagonal: the empirical MNRP is the lowest for Ω_{zero} (except for the combination $N = 2$ and t_3 distribution). The empirical MNRPs for a fat-tailed t distribution are slightly higher than for a normal distribution.

Note that the maximum rejection probability over the explored set of mean vectors may not be equal to the maximum over all mean vectors satisfying the null because the null includes vectors that contain entries other than zero and $-\infty$.

1.3.3. Average Power

This comparison slightly favors the AS test as it has higher empirical MNRP as stated in the previous section. In addition to “raw” empirical power, the table also reports size-adjusted average power for the AS test denoted as AS(size-adj). For a given combination of dimensionality N , type of the covariance matrix Ω and distribution, the nominal significance

Test	Distr	H_0/H_1	$N = 2$			$N = 4$			$N = 10$		
			Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AS	Normal	H_0	5.4	4.8	5.2	5.7	5.2	5.5	6.6	5.6	6.2
RSW	Normal	H_0	4.8	4.3	4.8	5.0	4.7	5.0	5.6	4.9	5.7
AS	t_3	H_0	5.7	6.1	5.9	5.9	5.8	6.4	6.5	6.0	6.8
RSW	t_3	H_0	5.2	5.4	5.4	5.5	5.3	6.0	6.0	5.3	6.3
AS	Normal	H_1	66.0	68.2	70.0	63.4	67.4	75.7	60.9	64.7	79.5
AS(size-adj)	Normal	H_1	63.7	66.5	68.6	60.4	65.7	73.8	57.5	62.7	78.0
RSW	Normal	H_1	62.3	64.8	67.1	58.1	62.3	73.1	53.0	56.6	76.2
AS	t_3	H_1	68.2	70.1	71.9	65.9	69.7	77.1	64.0	68.0	80.5
AS(size-adj)	t_3	H_1	67.0	68.8	69.9	64.7	68.1	75.7	62.5	66.2	79.2
RSW	t_3	H_1	64.8	66.9	69.2	61.0	64.9	74.8	56.3	60.1	77.6

Table 1: Simulation Results for Mean Test

Sample size is $T = 500$, dimensionality is $N = \{2, 4, 10\}$. Based on 10,000 Monte-Carlo repetitions (5,000 for $N = 10$ under H_0). Critical values are computed using 3,000 bootstrap samples. The nominal significance level is 5%. AS(size-adj) denotes size-adjusted version of the AS test under the adjusted nominal significance level, so that the empirical MNRP matches that of the RSW test.

level is adjusted, so that the resulting MNRP of the AS test is exactly the same as the MNRP of the RSW test. The new significance level is then used when computing the average power of the AS test for the given combination of the simulation parameters. The values of the adjusted significance level were in the interval $[0.040, 0.045]$. When comparing the RSW test with the AS test with the tuning constant Romano et al. (2014) reported size-adjusted power of the RSW test but their results did not include $N = 10$.

Similar to empirical MNRPs, the “raw” average power achieved by the AS test is higher than that of the RWS. I find that this holds uniformly across all mean vectors. The average power of the size-adjusted AS test is also higher than that of the RSW test; however, the difference between the two is smaller. Similar to the MNRPs, the average power observed for the t distribution is higher than for normal.

1.4. Application:

Can Short Sale Restrictions Help to Explain Returns?

This section applies the AS test and the RSW test to multiple sets of test assets for a variety of factor models.

I find that the Fama-French five-factor model augmented with the momentum factor can explain more sets of test assets than more parsimonious models, which are rejected for almost all test assets. The Fama-French five factor model without momentum is rejected more often suggesting that momentum is not redundant even after controlling for profitability and investment. I also find that most models struggle explaining high returns of small stocks, which is in line with the research by Fama and French (2008).

1.4.1. Data

I use 40 years of monthly data ranging from January 1975 to December 2014 giving 480 observations.

Test Assets

As LHS variables, I consider multiple sets of test assets obtained from the Kenneth French data library. Each test asset is a portfolio obtained by combining US stock data based on some sorting algorithm. Only value-weighted portfolios are explored because an average investor is likely to invest in proportion to the market capitalization as argued by Harvey and Liu (2015). Three types of test assets are used: (1) portfolios that are based on finer sorts on the same (or closely related) variables that were used to construct the factors, (2) portfolios that are based on anomaly variables unrelated to factors, and (3) industry based allocations.

The first type of test asset includes bivariate and three-way sorts based on the same variables that were used to construct the factors. Bivariate sorts are based on size, book-to-market (B/M), operating profitability (OP) and the rate of growth of total assets (Inv) with each sort producing a set of 25 assets. Three-way sorts are obtained with the first sort based on size and the second sort on a pair of B/M, OP and Inv with each sort consisting of 32 assets.

Next, I consider the portfolios sorted on the basis of asset-pricing anomalies unrelated to factors. The anomalies include: (1) lower average returns of stocks that demonstrated good performance in the previous periods (short and long term reversals) (Carhart (1997)), (2) low average returns of stocks with large accounting accruals (Sloan (1996)), (3) the flat relation between univariate market beta and average returns (Frazzini and Pedersen (2014) and Fama and MacBeth (1973)), and (4) low average returns of stocks with high volatility (Ang et al. (2006)). All anomaly-based asset sets are produced based on two-way sorts and contain 25 portfolios each.

Finally, I apply the tests to ten, twelve and thirty portfolios based on industry.

Factors

I consider six classical factor models proposed in the literature: the capital asset pricing model (CAPM), the Fama-French three and five factor models (FF3 and FF5), the Fama-

French three factor model plus the momentum factor of Carhart (1997) (FF4), the Fama-French five factor model with momentum (FF6), and the Hou, Xue and Zhang five factor model (HXZ).⁵ The CAPM suggested by Lintner (1965) and Sharpe (1991) includes only one factor, which is the excess return on the market portfolio. The Fama-French three factor model proposed by Fama and French (1992) includes three factors: market factor, size and value. Fama and French augmented the basic three-factor with two more factors in a recent paper (Fama and French (2015b)): profitability and investment. Another factor that is often added to the FF3 and the FF5 is momentum, as in Carhart (1997), which is a powerful regressor and is mostly independent of other factors as noted by Fama and French (2016). Hou et al. (2014b)(HXZ) introduced a q -factor model, which adds alternative size, alternative profitability and alternative investment factors to the market factor. The q -factor model does not include a value factor, which the authors found to be insignificant after controlling for profitability and investment factors.

1.4.2. Findings

I apply the AS test and the RSW test to multiple sets of test assets for a variety of factor models. For comparison I add the GRS test by Gibbons et al. (1989) based on the null hypothesis that all alphas are jointly zero.

Consistent with the previous sections, p -values of the RSW are somewhat lower than those of the AS; however, the decision is usually the same. Given a more restricted null of the GRS test, it is also not surprising that its p -value is usually lower than that of inequality tests.

Sorts Based on Size, B/M, OP and Inv

Two-way Sorts

I start with two-way sorts based on size (the results can be found in Panel 1 of Table 2).

⁵I thank Lu Zhang for providing us the factors constructed in Hou et al. (2014b). Other factors are obtained from authors' webpages.

Model	AS(BIC)	RSW	GRS	AS(BIC)	RSW	GRS	AS(BIC)	RSW	GRS
Two-Way Sorts Based on Size, B/M, OP and Inv									
Panel 1	Size-B/M			Size-OP			Size-Inv		
CAPM	0.0	0.5	0	0.0	0.5	0.0	0	0.5	0
FF3	0.0	0.5	0	0.0	0.6	0.0	0	0.5	0
FF4	0.2	1.1	0	2.7	3.8	0.0	0	0.5	0
FF5	0.0	0.5	0	37.1	38.8	2.4	0	0.5	0
FF6	3.9	4.7	0	32.5	32.3	0.0	0	0.6	0
HXZ	0.0	0.5	0	23.8	24.7	6.5	0	0.5	0
Panel 2	B/M-OP			B/M-Inv			OP-Inv		
CAPM	0.0	0.6	0.1	0.0	0.5	1.1	0.0	0.5	0.0
FF3	1.7	4.1	0.6	1.9	3.0	4.8	0.0	0.5	0.1
FF4	4.5	6.3	0.0	25.2	25.7	0.0	4.2	10.0	0.0
FF5	65.5	70.4	11.7	64.3	66.9	28.3	57.9	59.2	41.8
FF6	12.1	14.0	0.0	65.9	69.5	0.0	78.9	79.6	0.6
HXZ	39.5	40.2	1.4	33.8	32.5	11.0	40.5	41.0	21.0
Three-Way Sorts Based on Size, B/M, OP and Inv									
Panel 3	Size-B/M-OP			Size-B/M-Inv			Size-Inv-OP		
CAPM	0.0	0.5	0	0.0	0.5	0.0	0.0	0.5	0.0
FF3	0.0	0.5	0	0.0	0.5	0.0	0.0	0.5	0.0
FF4	0.9	1.9	0	1.8	2.5	0.0	0.2	0.7	0.0
FF5	3.5	5.0	0	1.7	2.8	0.5	0.2	0.9	0.0
FF6	15.8	18.0	0	30.6	35.2	0.0	3.9	4.8	0.0
HXZ	2.3	2.8	0	2.7	3.0	0.7	0.9	1.5	0.1

Table 2: Results for Sorts Based on Factors

The table displays results for factor-based sorts as test assets. The tests are based on monthly data from January 1975 to December 2016. For each set of regressions, the table reports p -value for the AS test, the RSW test and the GRS test. All results are based on 10,000 bootstrap samples.

One would expect that the null should not be rejected for the FF5 and the FF6 models that use exactly the same sort to construct factors, or for the HXZ model that relies on similar indicators. For the portfolios based on the Size-OP the null hypothesis of mean-variance efficiency indeed cannot be rejected for any of the three models when short sale constraints are taken into account; however, if the short sale constraints are ignored, these models are rejected as shown by the GRS test.

The portfolios with the high and positive alphas that lead to rejection are usually "microcaps": portfolios based on stocks in the lowest 20% quantile of size. This result is to

some extent surprising given that small stocks are usually more short-sale constrained than large stocks (see, e.g., Saffi and Sigurdsson (2010)), so one would expect that imposing the short sale restriction should improve the performance of the asset pricing models. The high alphas of microcaps are in line with the findings of Fama and French (2015a) that most factor models struggle to explain the average returns of very small stocks.

For sorts that do not include size, the FF5, the FF6 and the HXZ models are not rejected by the data (Panel 2). This finding holds independent of whether or not the short sale constraints are taken into account for the FF5 and the HXZ but not for the FF6. In particular, the six-factor model predicts negative alphas for stocks with high book-to-market value but low operating profitability or low investment activity. The fact that the HXZ q -factor model, which does not include a value factor, performs on par with the five factor and the six factor models is in line with the findings of Hou et al. (2014a) and Fama and French (2015b), who observe that this factor is redundant when controlling for investment and profitability.

The more parsimonious models (FF3, FF4 and CAPM) are strongly rejected, suggesting that they do not provide a good description of expected returns of these test assets. The only exception is the original Fama-French three-factor model augmented with the momentum factor is not rejected for the B/M-Inv sort under the short sale constraints, which may suggest that adding momentum may implicitly capture some of the effects of the investment factor. The rejection is not surprising given that these models do not include factors based on profitability and investment.

Three-Way Sorts Based on Size, B/M, OP and Inv

The results for three-way sorts can be found in Panel 3 of Table 2. All models except for the FF6 are rejected, or borderline rejected, at 5% level for all three way sorts (Panel 3). Similar to the returns based on two-way sorts, this finding is driven by the microcaps. The FF6 model can partially explain the relatively high returns of the small stocks for sorts

based on size, B/M and OP or size, B/M and Inv. The model struggles explaining high returns of small stocks with low profitability and low investment activity.

Sorts Based on Anomalies

Model	AS(BIC)	RSW	GRS	AS(BIC)	RSW	GRS	AS(BIC)	RSW	GRS
Panel 1	Sorts Based on Prior Returns								
	Size-Momentum			Size-SR Reversals			Size-LR Reversals		
CAPM	0.0	0.5	0	0.0	0.5	0	0.0	0.5	0.1
FF3	0.0	0.5	0	1.8	3.0	0	1.9	2.5	2.8
FF4	0.3	0.7	0	25.6	25.2	0	4.0	5.0	0.0
FF5	0.0	0.5	0	12.9	14.3	0	42.1	41.8	10.7
FF6	2.5	3.0	0	44.8	44.4	0	50.4	52.1	0.7
HXZ	0.1	0.6	0	2.1	2.8	0	8.1	7.9	3.1
Panel 2	Sorts Based on Accruals, Market Beta and Net Share Issues								
	Size-Accruals			Size-Beta			Size-Net Share Issues		
CAPM	0.1	0.6	0.0	0.0	0.5	0.1	0.0	0.5	0
FF3	0.0	0.5	0.0	5.2	5.9	4.1	0.0	0.5	0
FF4	8.9	9.2	0.1	3.8	4.5	0.0	0.1	0.8	0
FF5	0.0	0.5	0.0	49.4	51.0	4.5	0.0	0.5	0
FF6	4.2	4.8	0.0	17.2	17.2	0.0	2.5	3.0	0
HXZ	0.0	0.5	0.0	10.5	9.9	7.8	0.0	0.5	0
Panel 3	Sorts Based on Variance and Residual Variance								
	Size-Variance			Size-Residual Variance					
CAPM	0.0	0.5	0	0.0	0.5	0			
FF3	0.0	0.5	0	0.0	0.5	0			
FF4	0.0	0.5	0	0.3	0.7	0			
FF5	1.1	1.7	0	1.0	1.5	0			
FF6	0.5	1.0	0	2.2	2.7	0			
HXZ	0.2	0.9	0	0.1	0.7	0			

Table 3: Results for Sorts Based on Anomalies

The table displays p -values for sorts based on asset-pricing anomalies. The tests are based on monthly data from January 1975 to December 2016. For each set of regressions, the table reports p -value for the AS test, the RSW test and the GRS test. All results are based on 10,000 bootstrap samples.

The results for sorts based on anomalies can be found in Table 2.

We start by considering portfolios based on prior returns. The models that include momentum factor (the FF4 and FF6) are expected to perform well for these sorts as the factor is

explicitly based on prior returns. However, the results are contradicting the expectations: all models are rejected for the portfolios explicitly based on momentum itself. Again, the biggest challenges are high expected returns of microcaps. The FF4, the FF5 and the FF6 are not rejected for portfolios based on short-term reversals if the short sale constraints are taken into account. The FF5 and the FF6 are not rejected for portfolios based on long term reversals. The fact that the FF5 that does not include momentum is not rejected seems to support the view that the momentum is redundant.

All models struggle to explain low average returns of stocks with large accounting accruals, with the possible exception of the FF4 with p -values slightly below 10% (Panel 2). Again, the problem lies with size: the microcap portfolio in each accruals quintile has a very high alpha.

The FF5 and the FF6 are not rejected for the portfolios based on beta. This finding suggest that an additional "low-beta" factor is redundant after profitability and investment activity are taken into account.

All models are rejected for portfolios based on either variance or residual variance. Again, the problem lays with microcaps: the portfolios based on small stocks with low variance have a higher return than predicted by the models.

Industry Portfolios

All sets of portfolios based on industry strongly support the FF4 and the FF6 models that are the only models including the momentum factor as can be seen from figure 4. Both the FF5 and the HXZ can not explain high returns of the sector denoted as "HiTec" that includes business equipment (computers, software, and electronic equipment). This suggests that momentum is important when explaining the returns of technology stocks.

Model	AS(BIC)	RSW	GRS	AS(BIC)	RSW	GRS	AS(BIC)	RSW	GRS
Panel 1	Industry Portfolios								
	10 Portfolios			12 Portfolios			30 Portfolios		
CAPM	0.3	0.8	16.5	0.4	0.8	12.7	0.5	0.9	40.3
FF3	0.0	0.5	1.0	0.0	0.5	0.7	0.1	0.6	5.8
FF4	14.8	15.7	1.7	20.5	20.7	0.2	69.9	70.3	0.1
FF5	0.3	1.7	4.5	1.4	2.4	1.8	2.9	4.7	9.2
FF6	22.9	23.0	1.8	31.2	31.4	0.0	84.8	88.8	0.1
HXZ	0.1	0.6	0.8	0.1	0.8	0.7	0.7	1.4	3.4

Table 4: Results for Industry Portfolios

The table displays p -values industry portfolios. The tests are based on monthly data from January 1975 to December 2016. For each set of regressions, the table reports p -value for the AS test, the RSW test and the GRS test. All results are based on 10,000 bootstrap samples.

1.5. Robustness Check: Transaction Costs

In this section, I consider the effect of transaction costs on the mean-variance efficiency. While I incorporated short sale constraints, there are, of course, other market frictions that may influence the results. In particular, some of the models above could have been rejected due to the transaction costs.

Fortunately, the transaction costs can easily be incorporated into the null hypothesis as follows:

$$\begin{aligned}
 \mathcal{H}_0 &: \alpha_i \leq \text{transactionCosts}_i \quad \forall i = 1, \dots, N \\
 \mathcal{H}_1 &: \alpha_i > \text{transactionCosts}_i \quad \text{for some } i
 \end{aligned}
 \tag{1.10}$$

The testing procedures discussed in chapter 3 can be applied to test the null hypothesis in (1.10). This null is, again, constructed based on the assumption that the transaction costs of factors are zero, which, of course, is not the case.

Estimating transaction costs is a non-trivial task as discussed by, for example, Novy-Marx and Velikov (2015). It is thus difficult to determine the appropriate value of the transaction costs to be used in the null hypothesis (1.10). Instead, I suggest to focus on determining

the transaction costs needed to reject the model and try to determine whether they are reasonable. For simplicity I set the transaction costs the same for each portfolio. Given that the AS and the RSW tests produce similar results, I present the results only for the AS test here.

	CAPM	FF3	FF4	FF5	HXZ	FF6
Panel 1	Two-Way Sorts Based on Size, B/M, OP and Inv					
Size-B/M	17	5	7	7	9	2
Size-OP	7	6	2	NA	NA	NA
Size-Inv	21	7	15	5	7	6
B/M-OP	14	3	3	NA	NA	NA
B/M-Inv	13	2	NA	NA	NA	NA
OP-Inv	20	11	2	NA	NA	NA
Panel 2	Three-Way Sorts Based on Size, B/M, OP and Inv					
Size-B/M-OP	41	10	5	5	5	NA
Size-B/M-Inv	21	8	3	2	3	NA
Size-Inv-OP	23	12	7	4	5	2
Panel 3	Sorts Based on Prior Returns					
Size-Momentum	24	31	12	37	15	4
Size-SR Reversals	15	3	NA	NA	5	NA
Size-LR Reversals	18	3	2	NA	1	NA
Panel 4	Sorts Based on Accruals, Market Beta and Net Share Issue					
Size-Accruals	7	7	0	11	14	2
Size-Beta	18	1	2	NA	1	NA
Size-Net Share Issues	20	9	10	7	9	4
Panel 5	Sorts Based on Variance and Residual Variance					
Size-Variance	49	24	28	7	10	12
Size-Residual Variance	50	19	13	5	9	5
Panel 6	Industry Portfolios					
10 Industry Portfolios	6	10	NA	8	14	NA
12 Industry Portfolios	5	10	NA	8	14	NA
30 Industry Portfolios	5	8	NA	4	8	NA

Table 5: Required Transaction Costs

The table displays levels of transaction costs in basis points, so that the model cannot be rejected by the AS test at the 10 percent level, respectively. The tests are based on monthly data from January 1975 to December 2016. All results are based on 10,000 bootstrap samples.

Table 5 presents levels of transaction costs in basis points, so that the AS test does not reject the model at 10% level. For instance, in case of the Size-B/M sort, a transaction

costs below 17bps are needed to reject the CAPM and only below 5bps to reject the FF3. If the model can not be rejected at the 10 percent level, no matter how low the transaction costs are, I put “NA”. “NA”s, of course, correspond to the p-Values higher than 10% in chapter 3.

In order to judge whether the required transaction costs are reasonable, I’ll use the estimates by Novy-Marx and Velikov (2015) as a benchmark. In particular, they estimate that the transaction costs for strategies with low turnover and annual re-balancing (such as strategies based on size, value, profitability, investment or accruals) are usually around 4-10bps. The transaction costs for mid-turnover strategies with monthly re-balancing such as momentum, long term reversals, net share issuance or volatility usually average between 20-50 bps. Finally, for high turnover strategies such as short term reversals the costs are often higher than 150bps. For example, the CAPM requires the transaction costs of 20bps to explain the alphas on OP-Inv strategies, while the real transaction costs are likely to be below 10bps.

Based on these estimates, the CAPM and the FF3 would probably be rejected even after transaction costs are taken into account because these models require relatively high costs for presumably cheap strategies (such as two- or three-way sorts). These two models also need high transaction costs in order to be not rejected relative to other models. On the other hand, the FF6 always requires relatively low transaction costs. The highest value is 12bps for size-variance sort, which is a high turnover strategy and thus probably has indeed very high transaction costs. This suggests that the FF6 may sometimes be rejected solely due to the presence of transaction costs.

1.6. Conclusion

I examine the implications of mean-variance efficiency in linear factor models under consideration of short sale restrictions and explore two testing procedures to test the validity of these models when such restrictions exist. I employ two moment inequality tests, and design a simulation study to evaluate their performance. The results show that the AS

tests have slightly better power when compared to the RSW test, but their performance is very similar. Empirically, these two types of tests produce the same qualitative results. In the empirical study, I apply the two types of tests to explore the validity of multiple linear asset pricing models. The results indicate that some model specifications, such as the Fama-French five factor model augmented with the momentum factor, are able to explain several of the widely known asset pricing anomalies if we allow for possible mis-pricing due to the short sale restrictions.

1.7. Appendix

1.7.1. Conditionally Heteroscedastic Errors

I re-run the simulation for the dimensionality $N = 10$ using conditionally heteroscedastic errors. The simulated processes follow a constant conditional correlation GARCH model, which consists of 10 univariate GARCH(1, 1) processes related to one another with a constant conditional correlation matrix. The coefficients of univariate GARCH processes are set equal to the estimates obtained from the returns of ten industry portfolios. The GARCH processes are standartized, so that the unconditional variance of each process is equal to one. The correlation matrices are the same as before.

Test	Distr	H_0/H_1	$N = 10$		
			Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AS	Normal	H_0	6.5	5.6	6.1
RSW	Normal	H_0	5.6	5.2	5.5
AS	t_3	H_0	6.6	6.1	6.6
RSW	t_3	H_0	6.0	5.6	6.1
AS	Normal	H_1	58.1	64.8	79.3
AS(size-adj)	Normal	H_1	54.9	64.1	77.9
RSW	Normal	H_1	50.2	56.6	76.0
AS	t_3	H_1	63.8	67.9	80.5
AS(size-adj)	t_3	H_1	62.3	66.8	78.9
RSW	t_3	H_1	56.1	60.1	77.7

Table 6: Simulation for Mean Test with GARCH Disturbances

Sample size is $T = 500$, dimensionality is $N = 10$. Based on 5,000 Monte-Carlo repetitions under H_0 and 10,000 under H_1 . Critical values are computed using 3,000 bootstrap samples. The nominal significance level is 5%. AS(size-adj) denotes size-adjusted version of the AS test under the adjusted nominal significance level, so that the empirical MNRP matches that of the RSW test.

The results reported in Table 6 show that both tests are robust to conditional heteroscedasticity. The resulting average power and the empirical MNRP are essentially the same as for the simulation without conditional heteroscedasticity. The only exception is the combination of normal distribution and negative correlation matrix: in this case all tests demonstrate

slightly lower power under the heteroscedastic errors. This finding implies that both tests are robust under conditional heteroscedasticity although they don't explicitly account for it.

CHAPTER 2 : International Return Predictability

2.1. Introduction

Assessing the predictability of the equity premium is one of the fundamental problems in the financial literature.¹ However, despite years of research, there is no consensus about whether the equity premium is predictable, how much it varies over time and which variables can be used to predict it.²

International markets provide rich data that can be used to deepen our understanding of return predictability. However, most studies of international return predictability examine each capital market separately. Most come to the conclusion summarized by Schrimpf (2010) that “return predictability is neither a uniform, nor a universal feature across international markets”.³ The inconclusive evidence of international predictability may be driven in part by the noise in the data. Due to arbitrage considerations, if predictability exists, then it is likely to be weak.⁴ Thus, an efficient estimation method is required to assess the predictability of international equity premium. Although it would be more efficient to consider international capital markets jointly, rather than examining each one in isolation, the literature doing so is lacking.

In this chapter, I investigate patterns of equity premium predictability in international capital markets and explore the robustness of common predictive variables. In particular, I focus on predictive regressions with multiple predictors. To obtain precise estimates, two estimation methods are employed. First, I consider all capital markets jointly as a system of regressions. Second, I take into account uncertainty about which potential predictors forecast excess returns.

¹Henceforth, I use the words “equity premium” and “return” interchangeably as standard in the literature.

²See a survey by Cochrane (2011).

³Multiple papers explore each capital market individually such as Ang and Bekaert (2002), Rapach et al. (2005), Schrimpf (2010), Neely and Weller (2000), Paye and Timmermann (2006), Giot and Petitjean (2006) among others.

⁴Wachter and Warusawitharana (2009) and Wachter and Warusawitharana (2015) explore how the prior information about the weakness of the predictability can be used to improve inference.

I treat the data as seemingly unrelated regressions (SUR), first introduced by Zellner (1962). Country-specific regressions widely used in earlier literature ignore the fact that excess returns are contemporaneously correlated across countries and treat the covariance matrix of residuals as diagonal. SUR exploits this correlation to increase the estimation precision.

I assume that an investor is uncertain about which potential predictors to include in the predictive regression, which contradicts the typical assumption that the investor chooses a set of predictors a priori. The assumption of model uncertainty is justified given the large number of potential predictors in the literature and the limited consensus about which of them forecast returns. Existing pricing theories offer little guidance about which predictors should be included in the regression, so the decision regarding the predictive power of potential regressors is often based on empirical studies, making the predictability evidence subject to data-snooping concerns.⁵ The international investor is likely to face an even higher degree of model uncertainty because most empirical studies focus only on the U.S. market. Therefore, prevailing views are likely to be biased toward regressors deemed significant for the U.S.⁶ It is, therefore, necessary to take model uncertainty into account when evaluating the statistical evidence in favor of return predictability.

I use the Bayesian spike-and-slab approach to incorporate model uncertainty.⁷ The expression “spike-and-slab” refers to the prior imposed on the coefficients. The prior represents a mixture of two normal distributions: the spike, which is centered very tightly around zero, and the slab, which represents a very flat distribution. The frequency, with which a particular predictive variable is estimated to be a draw from the slab distribution, can be used to estimate the inclusion probability. A fundamental property of this approach is selective shrinkage. The posterior means of coefficients that are found to be non-informative of future returns is strongly shrunk toward zero, thereby reducing estimation error. On the other

⁵This concern was raised by Welch and Goyal (2007), Ferson et al. (2003) and Bossaerts and Hillion (1999) among others.

⁶Dimson et al. (2008) argues that the strong performance of the U.S. market is unique; thus, focusing solely on the U.S. may have consequences similar to selection bias.

⁷See Mitchell and Beauchamp (1988), George and McCulloch (1993), George and McCulloch (1997), Ishwaran et al. (2005).

hand, the posterior means of the coefficients that predict equity premium is only slightly shrunk to zero if at all.

The spike-and-slab approach has not been previously used in the context of return predictability. The classical approach to model uncertainty is Bayesian model averaging as in Avramov (2002), Cremers (2002), and Schrimpf (2010). The Bayesian model averaging approach assigns posterior probabilities to all candidate model specifications and uses these probabilities as weights on the individual models to obtain a composite model used for forecasting. The weakness of this approach is that it requires considering all possible models, which is computationally infeasible in the context of international return predictability. In this chapter, I consider ten countries with six potential regressors, which yields $2^6 = 64$ models. In contrast, the spike-and-slab method uses Gibbs sampling to identify models with a high probability of occurrence.

I re-examine both in- and out-of-sample evidence of predictability for ten international capital markets. In particular, I explore the importance of dividend-price ratio, four interest rate variables and inflation. As a result, several conclusions emerge about international stock return predictability.

First, I find that variables that are statistically significant predictors in the country-specific regressions are insignificant when the capital markets are studied jointly. In particular, the consensus in the literature is that interest rates are among the most robust predictors for international equity premium, while the valuation ratios and, in particular, the dividend-price ratio do not predict returns.⁸ However, my results suggest that the in-sample evidence in favor of the interest rate variables is spurious and is mostly driven by ignoring the cross-country information. Conversely, the dividend-price ratio emerges as the only robust predictor of future stock returns. I, thus, show that ignoring the cross-correlation of errors could lead to erroneous inferences about the relevance of predictive variables.

⁸See, for example, Schrimpf (2010), Rapach et al. (2005), Ang and Bekaert (2002), and Hjalmarsson (2010).

Second, the evidence in favor of predictability is weak both in- and out-of-sample and limited to a few countries. In-sample, most predictors are considered to be worthless with low inclusion probabilities and low coefficient values. Out-of-sample, the posterior mean usually under-performs relative to the forecast based on the past mean. Consequently, an international investor would typically not benefit from market timing. The exceptions are investors from Sweden, the U.K., and the U.S who can increase their risk-free certainty equivalent returns (CERs) by around 0.5-1% a year relative to the naive investor. Even in these countries, however, the utility gains are mostly driven by a few short episodes of predictability.

Third, my analysis shows that considering model uncertainty is essential. An investor who is forced to ignore it and, therefore, maintains a suboptimal portfolio, perceives a substantial loss in risk-free CERs both in- and out-of-sample. In addition, incorporating model uncertainty also results in economically significant improvements of model fit out-of-sample. However, the evidence in favor of predictability weakens if model uncertainty is accounted for, consistent with earlier studies such as Avramov (2002). In particular, investors have fewer incentives to time the market when model uncertainty is incorporated.

Fourth, I find moderate support in favor of SUR. In-sample, incorporating correlation of errors leads to economically small losses in risk-free CERs relative to model uncertainty. Out-of-sample, SUR is typically beneficial but may lead to worse fit if the sample size is small. The worse fit is driven by the estimation error due to the higher number of parameters.

Finally, I consider two extensions to the basic framework: borrowing strength from cross-section and stochastic volatility. Both of these methods can yield more precise estimates and thus improve inference. However, I find that incorporating either of them does not change the key results discussed above.

The literature on international return predictability that explores capital markets jointly is

currently limited to recent paper of Penasse (2016), who focuses on the predictive power of the dividend-price ratio. In addition to SUR, he considers the constraint that equity premium is positive and borrows strength from the cross-section by pooling cross-county coefficients toward the common mean. My approach differs from that of Penasse (2016) because I explore multiple predictive variables and incorporate model uncertainty. I also consider the possibility of borrowing strength from the cross-section by pooling either the coefficients or the inclusion probabilities toward the common mean. However, my results show that the utility costs associated with ignoring this mechanism are low. This suggests that the superior out-of-sample performance found by Penasse (2016) is likely to be the result of the non-negativity constraints on the expected equity premium.

From a methodological perspective, this chapter is closely related to the literature on Bayesian model averaging. In particular, Schrimpf (2010) applies model averaging to international return data from five major capital markets, finding that the interest rate variables and the output gap are among the most prominent predictive variables. This chapter deviates from the work of Schrimpf (2010) in several ways. First, I use the spike-and-slab approach to model uncertainty rather than Bayesian model averaging. Second, I consider all countries jointly by exploiting SUR and show that the predictive power of interest rates is spurious. Moreover, I show that ignoring the cross-correlation of errors leads to welfare losses both in- and out-of-sample. Third, I increase the set of considered countries from five to ten. Finally, the analysis of Schrimpf (2010) is limited to the statistical properties of the predictive regressions and, in particular, the predicted mean return. However, the investors are likely to be more interested in economic criteria such as utility gains from better asset allocation. Moreover, poor performance of the forecast in the statistical sense need not imply poor performance in the economic sense and vice versa. This discrepancy partially arises because even statistically insignificant improvement of the equity premium forecast can result in statistically significant changes in asset allocation and, consequently, increased welfare. Another reason is that the statistical measures focus on forecasting the expected premium itself and ignoring other moments of predictive distribution and, in particular,

predicted variance. In contrast, the asset allocation decision depends on both the mean and the variance. An investor can thus benefit from the predictive regression even if it fails some statistical tests.

This chapter is organized as follows. In the second section, I introduce the theoretical framework and discuss the estimation techniques. I introduce the data in the third section and then proceed to discuss the in-sample and out-of-sample results in Sections four and five. Finally, I discuss the extensions in the sixth section. Some concluding remarks are offered in the last section.

2.2. Methodological Framework

In this section, I introduce the methodology used in the subsequent empirical analysis and explain how it relates to the economic questions raised in this chapter. I start by laying out the econometric framework of predictive regressions describing international returns, and then discuss prior beliefs that address model uncertainty. Finally, I provide an overview of the estimation procedure.

2.2.1. Econometric Framework: Predictive Regressions

In this chapter, the question of predictability is studied in the conventional framework of linear predictive regressions. Consider the following predictive regression for excess log returns:

$$r_{n,t+1} = \alpha_n + x'_{n,t}\beta_n + u_{n,t+1}, \quad n = 1, \dots, N; t = 1, \dots, T, \quad (2.1)$$

where $r_{n,t+1}$ denotes the continuously compounded excess return in country n at period $t + 1$, calculated as the difference between the continuously compounded (log) return on the market portfolio and a continuously compounded risk-free rate. Henceforth, the word “return” always refers to the continuously compounded excess return. The $K \times 1$ vector $x_{n,t}$ contains predictive variables and β_n is a $K \times 1$ vector of country-specific coefficients on the predictive variables. Note that some coefficients $\beta_{n,k}$ may be either zero or close to zero. The process in (2.1) characterizes excess return dynamics for $n = 1, \dots, N$ countries over $t = 1, \dots, T$ periods.

I treat the N regressions in (2.1) as a seemingly unrelated regressions (SUR). In particular, the innovations $u_{n,t+1}$ are jointly normal with zero mean and an $N \times N$ contemporaneous covariance matrix Σ . The innovations are i.i.d. across time t . Treating the predictive regressions as SUR leads to more efficient estimates than those obtained from single-country regressions. The coefficients obtained from single-country regressions are based on the implicit assumption that the contemporaneous covariance matrix Σ is diagonal. Given that

returns across countries are correlated, some cross-sectional information is lost when ignoring the correlation, thus making the estimates inefficient. On the contrary, SUR accounts for the typically observed correlation between returns.

The predictive regressions in (2.1) are typically complemented by a process describing the dynamics of the predictive variables. I do not consider the dynamics of the predictive matrix because it requires estimating a $(K + 1)N \times (K + 1)N$ covariance matrix of shocks to K regressors and returns in N countries. Precise estimation of such a huge matrix is infeasible, so I leave it to future research. An alternative estimation strategy would be to impose some additional constraints on the structure of the covariance matrix as was done, for example, by Penasse (2016).

2.2.2. Prior Beliefs: Accounting for Model Uncertainty with Spike-and-Slab Priors

In this subsection, I introduce the spike-and-slab approach to model uncertainty.

I consider an international investor who is uncertain about which variables should be included in (2.1). On one hand, the investor is confronted with voluminous literature on the return predictability that examines a variety of potential predictors. On the other hand, the signal-to-noise ratio in the return predicting regressions is very low, so that many predictors may have been deemed statistically significant or insignificant purely by chance. Thus, the investor is uncertain about which variables (if any) are actually helpful in predicting returns. Moreover, the set of important variables may differ across countries.

The framework adopted in this chapter is different from the classical framework of model selection. Model selection is based on the assumption that there is a single “true” model, that should be used to forecast returns. However, this means that information contained in other possible models is neglected, which makes model selection not very appealing if there is no strong evidence in favor of a single model. In the model uncertainty framework, multiple candidate models are considered potentially possible and the composite model is obtained as a weighted average of the candidate models.

The conventional approach to model uncertainty adopted by Schrimpf (2010), Avramov (2002), and Cremers (2002) is the Bayesian model averaging, which requires considering all possible candidate models and deriving a posterior probability for each of them. For example, Schrimpf (2010) considers country-specific regressions and explores $K = 9$, which implies $2^9 = 512$ different models for each country. In contrast to Schrimpf (2010), I consider multiple predictive regressions jointly. I explore $K = 6$ regressors for $N = 10$ countries, which leads to the $2^{60} = 1.5 \times 10^{18}$ candidate models. It is, thus, computationally infeasible to derive a posterior probability for each of them.

Instead, I exploit the Bayesian spike-and-slab approach following the works of Ishwaran et al. (2005), George and McCulloch (1993), and George and McCulloch (1997). The spike-and-slab approach involves using Gibbs sampler to identify a set of models with a high posterior probability of occurring. Thus, instead of exploring all candidate models, I focus on a subset of models most supported by the data. The posterior probabilities are estimated based on the frequency with which each model occurs in the Gibbs sampler.

I consider a continuous bimodal spike-and-slab proposed by Ishwaran et al. (2005). The expression “spike-and-slab” refers to the family of priors where the coefficients follow a mixture distribution consisting of two parts: a very flat distribution (the slab) and a distribution centered close to zero (the spike). In particular, each coefficient $\beta_{n,k}$ with some probability can, a priori, either be distributed so closely around zero that it is essentially zero (the spike) or its distribution can be sufficiently far from zero (the slab).

The spike-and-slab prior is mathematically formulated as follows. Each coefficient $\beta_{n,k}$ follows a normal mixture model driven by a latent binary variable $\gamma_{n,k}$.

$$\begin{aligned}
 \beta_{n,k} | \gamma_{n,k}, v_0, \tau_k^2 &\stackrel{ind}{\sim} (1 - \gamma_{n,k}) \mathcal{N}(0, v_0^2 \tau_k^2) + \gamma_{n,k} \mathcal{N}(0, \tau_k^2) \\
 \gamma_{n,k} | \pi_{n,k} &\stackrel{ind}{\sim} \text{Bernoulli}(\pi_{n,k}) \\
 \tau_k^{-2} | a_{0,k}, b_{0,k} &\stackrel{ind}{\sim} \Gamma(a_{0,k}, b_{0,k})
 \end{aligned} \tag{2.2}$$

When $\gamma_{n,k} = 0$, the prior distribution is $\beta_{n,k} | \gamma_{n,k} = 0 \sim \mathcal{N}(0, v_0^2 \tau_k^2)$. In this case, the coefficient is considered to be essentially zero and closeness to zero is achieved by setting the scaling variable v_0 to be “small”. If $\gamma_{n,k} = 0$, then the posterior value is thus strongly shrunk toward zero. In contrast, the variance τ_k^2 is “large”. Therefore, when $\gamma_{n,k} = 1$ and the prior distribution is $\beta_{n,k} | \gamma_{n,k} = 1 \sim \mathcal{N}(0, \tau_k^2)$, the prior is sufficiently diffuse. The posterior in this case is mostly driven by the value observed in the data and only slightly shrunk toward zero. This part of the prior is used to identify non-zero coefficients. Note that the variance τ_k^2 depends on the predictive variable but not on the country. Given that the coefficients in front of a certain predictive variable reflect the same relationship, it is reasonable to assume that a “big” value of a certain coefficient for one country should also be considered as “big” for another country.

The values of the “large” variance τ_k^2 as well as the probability that $\gamma_{n,k} = 0$, are inferred from the data. The prior distribution for τ_k^2 is the inverse gamma. The parameters v_0 (some small near-zero value) and the shape and scale parameters of the gamma distribution are chosen so that the unconditional variance of each coefficient has a continuous bimodal distribution with a spike at v_0 and a long continuous right tail. The binary latent variable $\gamma_{n,k}$ follows a Bernoulli distribution centered around the prior inclusion probability $\pi_{n,k}$. The probability $\pi_{n,k}$ reflects the prior probability that the corresponding coefficient $\beta_{n,k}$ is different from zero. I set the prior inclusion probability $\pi_{n,k} = 0.5$. Finally, I impose a non-informative prior on the variance Σ and the intercepts α .

2.2.3. Estimation Procedure

This section lays out the estimation procedure. Additional details can be found in Appendix 2.8.1. The model described in the previous section can be thought of as a two-stage hierarchy. The first stage is the distribution of the country-specific parameters, β , conditional on the parameters: variances τ_k^2 and prior probabilities of inclusion $\gamma_{n,k}$. The second stage is the distribution of the hyper-parameters: variances τ_n^2 and prior probabilities of inclusion $\gamma_{n,k}$.

Posterior distributions are not analytically tractable in this framework. However, posterior distributions conditional on the remaining parameters do have a closed form and, thus, can be easily sampled. I find the posterior distribution using the Gibbs sampler similar to the one employed by Ishwaran et al. (2005) and Chib and Greenberg (1994). In particular, I use a five-block Gibbs sampler. Starting from the initial parameter values based on the OLS estimates, I sample the parameters successively from the posterior distribution conditional on values of other parameters in the previous iteration. The first three blocks of the sampler correspond to the individual slope parameters, $\beta_{n,k}$, the intercepts α_n and the precision matrix Ω . The following two blocks correspond to the second stage of the hierarchy: latent inclusion parameter $\gamma_{n,k}$ and the variance hyper-parameter τ_k^2 . I generate 150,000 draws with the initial 50,000 draws discarded for burn-in.

2.2.4. Portfolio Choice and Asset Allocation

In this subsection, I describe the optimization problem of a buy-and-hold investor. This setup is later used to analyze the implications of model uncertainty and the correlation of errors for asset allocation decisions. Asset allocation decisions are made separately for each country because the returns are priced in local currency. For simplicity, the country-specific subscript n is suppressed in what follows.

The investment universe of an investor in each country consists of two assets: a market index and a risk-free asset. At the beginning of each period, the investor allocates a fraction of their funds $0 \leq w \leq 1$ in the market index. Short-selling and buying on margin are not allowed in order to avoid the expected utility being $-\infty$ (see, for example, Barberis (2000)).

The investor has a power utility function with the coefficient of relative risk aversion $A = 5$, similar to earlier papers on Bayesian asset allocation such as Shanken and Tamayo (2012). The expected utility depends on the final wealth at the next period W_{T+1} .

$$\mathbb{E}[U_{t+1}|D] = \mathbb{E} \left[\frac{W_{T+1}^{1-A}}{1-A} \middle| D \right]. \quad (2.3)$$

Without loss of generality I, set the starting wealth $W_T = 1$, so that the final wealth at the of period $T + 1$ is

$$\begin{aligned} W_{T+1} &= w \exp\{r_{T+1} + r_{n,T+1}^f\} + (1 - w) \exp\{r_{T+1}^f\} \\ &= \exp\{r_{T+1}^f\} (w \exp\{r_{T+1}\} + (1 - w)), \end{aligned} \quad (2.4)$$

where r_{T+1} is, as before, the excess log return and r_{T+1}^f is the log risk-free rate. Note that as can be seen from (2.4), the optimal allocation does not depend on the risk-free rate although the optimal level of utility level is affected by it. I set a constant risk-free rate $r_{T+1}^f = 40$ bps per month following Shanken and Tamayo (2012) and Penasse (2016).

The expectation in (2.3) is taken with respect to the predictive distribution of the returns produced by the estimation algorithm. The predictive distribution can then be used to estimate expected wealth for all potential allocations. The maximization problem of the investor can be numerically approximated as

$$w_{T+1}^* \approx \arg \text{Max} \frac{1}{L} \left[\frac{\left(w \exp\{r_{T+1}^{(l)} + r_{T+1}^f\} + (1 - w) \exp\{r_{T+1}^f\} \right)^{1-A}}{1 - A} \right],$$

where $l = 1, \dots, L$ denotes draws from the MCMC algorithm.

Kandel and Stambaugh (1996) suggests the following approximation to compute optimal weights:

$$w_{T+1}^* \approx \frac{\mathbb{E}[r_{T+1}|D]}{A \text{Var}(r_{T+1}|D)} + \frac{1}{2A}, \quad (2.5)$$

where $\mathbb{E}[r_{n,T+1}|D]$ is the posterior mean and $\text{Var}(r_{n,T+1})$ is the posterior variance. The approximation highlights that the asset allocation decision of the investor is mostly driven by the first two moments of the distribution only. I find the approximation to be quite accurate.

The certainty equivalent return over the risk-free rate is the excess return that if known, would provide the same utility as the optimal risky portfolio. CERs thus provide a convenient monetary metric to evaluate utility gain from holding an optimal portfolio. CER is a solution to the following equation

$$\frac{\exp\{CER + r_{T+1}^f\}^{1-A}}{1-A} = \mathbb{E} \left[\frac{\left(w \exp\{r_{T+1} + r_{T+1}^f\} + (1-w) \exp\{r_{T+1}^f\} \right)^{1-A}}{1-A} \middle| D \right]$$

2.3. Data Overview

In this section, I discuss the data used to estimate the predictive regressions in (2.1).

The dataset consists of monthly data for ten developed countries: Australia, Canada, France, Germany, Italy, Japan, Sweden, Switzerland, the United Kingdom, and the United States. The sample for returns covers the period February 1975 through December 2016.

The returns are obtained from Morgan Stanley Capital International's (MSCI) local currency country index. I compute excess log returns by subtracting the continuously compounded 3-month short-term interest rate from the total equity log return.

I explore five popular predictors of stock returns: dividend-price ratio (*dp*), four interest rate variables and inflation (*infl*). The interest rate variables include the Treasury bill rate (*tb*), yield spread (*ys*), relative 3-month Treasury bill rate (*rtb*), and relative 3-month Treasury long-term bond interest rate (*rbr*).

The selection of predictors is based on the previous literature and guided by economic theory. In particular, the dividend-price ratio is directly linked to future stock returns via the present value identity by Campbell and Shiller (1988)(see also Rozeff (1984) and Fama and French (1988)) and has been extensively studied in the literature. Similarly, nominal interest rates are related to the stock returns via the present value relationship as shown by Camp-

bell (1987), Hodrick (1992), and Ang and Bekaert (2006). There are multiple alternative explanations about why inflation may be useful in predicting returns. In the asset-pricing model by Brandt and Wang (2003), inflation news are linked to varying risk aversion, which in turn, is directly related to returns. Conversely, Campbell and Vuolteenaho (2004) argues that the relationship between inflation and return has a behavioral rather than a rational explanation and is due to investor irrationality. The arguments in favor of the yield spread lack theoretical motivation and are mostly driven by empirical evidence. It is one of the predictors of real economic activity (see, e.g., Estrella and Hardouvelis (1991) and Stock and Watson (2003)) and, thus, can be considered to be a measure of expected business conditions as in Fama and French (1989). In particular, Fama and French (1989), Campbell and Yogo (2006), and Torous et al. (2004) find strong evidence of predictability of future returns by the yield spread.

I proceed to describe the construction of each variable and the data sources.

Dividend-Price Ratio

The log dividend-price ratio (dp) is calculated as follows:

$$dp_t = \log \left(\frac{\sum_{j=0}^{11} D_{t-j}}{P_t} \right),$$

where D_t denotes the dividends at period t and P_t denotes the price. The dividends are obtained from comparing monthly arithmetic returns on “gross” and “price” indices provided by the MSCI:

$$D_t = P_{t-1} \times (R^{gross} - R^{price}).$$

The dividend price ratio is constructed using the dividends over the past year. Dividends at higher frequencies are impossible to use due to the strong seasonal component.⁹

⁹This definition implies that dividends either are not being re-invested or are re-invested at a zero rate. Alternative strategies would be to re-invest the dividends either at the T-bill rate or at the market rate. This issue is explored by Kojen et al. (2011), Binsbergen et al. (2010) and Chen (2009) who find that re-investment strategy may have significant implications for the return predictability regressions. I choose the no re-investment strategy for an easier comparison as other papers pursue this approach.

Short-Term Interest Rate

The short-term interest rates (tb) are 3 month Treasury bill rates retrieved from the Global Financial Database (GFD). I use the continuously compounded rate. The U.S. T-bill data are obtained from the FRED database.

Yield Spread

Yield spread (ys) is calculated as the difference between log long-term rate and log short-term rate. The long-term rate is the rate for a 10-year bond retrieved from the GFD.

Relative Rates

Similar to Schrimpf (2010) and Rapach et al. (2005) I include two relative rates: Treasury bill interest rate relative to its 12-month backward-looking moving average (rtb) and long-term bond interest rate relative to its 12-month backward-looking moving average (rbr). These rates are calculated as the rate minus its moving average over the previous twelve months.

Inflation

The annual log inflation rate is calculated based on the consumer price index obtained from the OECD database:

$$Infl_t = \log\{CPI_t\} - \log\{CPI_{t-12}\}.$$

It should be noted that the OECD reports the average monthly value for each month as opposed to the end of the month value, which should be used for the growth rate calculation. However, the obtained error is small. The CPI for Australia is available only on quarterly frequency, so I use the data from the previous quarter.

Summary statistics for log excess returns are reported in Table 7. The first three columns report annualized mean, annualized standard deviation, and annualized Sharpe ratio. The subsequent columns contain information on monthly returns. The lowest returns average

	Annual			Monthly					
	Mean	SD	Sharpe	Mean	SD	Min	Max	Skewness	Kurtosis
Australia	4.05	17.92	0.23	0.34	5.17	-54.16	18.08	-2.51	24.01
Canada	3.96	16.14	0.25	0.33	4.66	-25.05	14.01	-0.97	3.90
France	5.09	19.67	0.26	0.42	5.68	-25.26	19.44	-0.47	1.23
Germany	4.90	20.03	0.24	0.41	5.78	-28.91	18.85	-0.87	3.08
Italy	1.20	23.20	0.05	0.10	6.70	-21.26	23.12	0.12	0.73
Japan	3.24	18.40	0.18	0.27	5.31	-23.69	17.82	-0.40	1.61
Sweden	8.28	22.26	0.37	0.69	6.43	-25.26	28.98	-0.28	1.98
Switzerland	6.15	15.52	0.40	0.51	4.48	-26.69	12.48	-0.94	3.71
UK	5.72	16.85	0.34	0.48	4.86	-30.80	21.01	-0.64	4.18
US	5.71	14.95	0.38	0.48	4.31	-24.45	11.98	-0.81	2.88

Table 7: Summary Statistics of International Returns

The table displays summary statistics of log excess returns. The first three columns report annualized statistics: mean, standard deviation, and Sharpe ratio (mean divided by standard deviation). The subsequent columns report summary statistics of monthly returns. The returns are observed over the period February 1975 through December 2016.

were observed for Italy (around 1%) while the highest returns were obtained for Sweden (more than 8%). The annualized standard deviations range from 15% in the United States to more than 23% in Italy. The highest monthly Sharpe ratios were delivered by Switzerland (0.40) and the U.S. (0.40), while Italy produced the lowest Sharpe ratio of 0.05. All log returns are leptokurtic with the excess kurtosis ranging from 0.7 for Italy to 24 for Austria. Australia features an exceptionally high excess kurtosis because it was hit hard by the market crash in October 1987, resulting in the loss of 54%. Most log returns are slightly skewed to the left.

The summary statistics of regressors can be found in Table 8. The dividend-price ratio usually varies from -3 to -4 with the exception of Japan with an average of -4.4. Japan and Switzerland feature the lowest average T-bill rates of less than 2.5%, whereas most countries demonstrate an average T-bill rate between 4 and 8%. The average yield spread is usually around 0.8-1.5% with the exception of Australia with spread of only 0.29%. The T-bill and 10-year bond rates tended to decline in all countries as shown by the negative average relative rates. Inflation ranges from less than 2% in Japan and Switzerland to almost 6% in Italy. Note that all regressors are very persistent. The persistence in the case of the dividend-price ratio and the inflation rate occurs by construction as they incorporate the

regressors	dp	tb	ys	rtb	rbr	infl
Australia						
Mean	-3.20	7.77	0.21	-0.00	-0.00	4.98
SD	0.20	4.06	1.45	0.01	0.01	3.71
Persistence	0.96	0.99	0.94	0.92	0.91	0.98
Canada						
Mean	-3.61	5.81	1.00	-0.00	-0.00	3.79
SD	0.35	4.07	1.37	0.01	0.01	3.07
Persistence	0.99	1.00	0.95	0.92	0.88	0.99
France						
Mean	-3.56	5.65	1.18	-0.00	-0.00	3.86
SD	0.52	4.16	1.18	0.01	0.01	3.77
Persistence	0.99	1.00	0.95	0.93	0.92	0.99
Germany						
Mean	-3.69	3.85	1.50	-0.00	-0.00	2.28
SD	0.35	2.55	0.98	0.01	0.01	1.64
Persistence	0.97	1.00	0.95	0.93	0.90	0.97
Italy						
Mean	-3.56	7.67	1.08	-0.00	-0.00	5.94
SD	0.42	5.70	1.69	0.01	0.01	5.44
Persistence	0.98	1.00	0.93	0.88	0.94	0.99
Japan						
Mean	-4.41	2.39	1.39	-0.00	-0.00	1.69
SD	0.50	2.48	0.86	0.01	0.00	2.90
Persistence	0.99	1.00	0.96	0.94	0.85	0.96
Sweden						
Mean	-3.59	6.10	0.99	-0.00	-0.00	4.16
SD	0.46	4.31	1.37	0.01	0.01	3.95
Persistence	0.98	0.99	0.93	0.90	0.92	0.99
Switzerland						
Mean	-3.81	2.54	0.95	-0.00	-0.00	1.84
SD	0.34	2.39	1.26	0.01	0.00	2.02
Persistence	0.98	0.98	0.94	0.92	0.92	0.98
United Kingdom						
Mean	-3.28	6.51	0.86	-0.00	-0.00	5.38
SD	0.26	4.08	1.73	0.01	0.01	4.79
Persistence	0.97	0.99	0.96	0.91	0.83	0.99
United States						
Mean	-3.64	5.11	1.18	-0.00	-0.00	3.77
SD	0.45	3.67	1.42	0.01	0.01	2.78
Persistence	0.99	0.99	0.93	0.89	0.89	0.98

Table 8: Summary Statistics of Predictive Variables

The table displays summary statistics of predictive variables: mean, standard deviation, and persistence. The set of predictors comprises the short-term interest rate tb , the yield spread (ys), the short-term interest rate relative to its 12-month moving average (rtb), the long-term government bond yield relative to its 12-month moving average (rbr), annual inflation rate ($infl$), and log dividend-price ratio dp . All predictors are in logs. All variables except for the log dividend-price ratio are reported in percentages. Interest rates are annualized. The regressors are observed over the period January 1975 through November 2016.

data of the previous year.

2.4. In-Sample Analysis

In this section, I discuss both the statistical and the economic importance of accounting for model uncertainty and the correlation of errors. I start by discussing the posterior distributions of coefficients and inclusion probabilities, and then proceed to examining the economic implications that the model uncertainty and SUR have on asset allocation decisions.

2.4.1. Posterior Distribution: Do Returns Appear to be Predictable?

In this subsection, I discuss the posterior distribution of coefficient values and inclusion probabilities.

I consider several statistics to investigate the robustness of the predictors. The first one is the cumulative inclusion probability of the predictive variables that is computed from the Gibbs sampler as the average value of the latent variable $\gamma_{n,k}$. The resulting quantity indicates the relative frequency with which each regressor is included into the composite model. I also consider two characteristics of the coefficients: posterior mean and posterior standard deviation. Finally, I also explore the posterior distribution of coefficients conditional on them being “large”. I obtain this distribution by conditioning on the latent variable $\gamma_{n,k} = 1$.

Let me first discuss the tables and figures used in this section. Posterior beliefs for the coefficients are represented in Figure 2. Each figure compares box plots of posterior distributions for coefficients in front of a corresponding regressor across countries. For each country, I present two boxes that correspond to the estimation method with and without SUR. Given that the inclusion probabilities are relatively small, the resulting distribution is close to zero. The box plots are complemented by Table 10 that reports posterior beliefs about coefficients. The summary table shows the posterior mean of the coefficient for each country followed by its standard deviation and the probability that the corresponding coefficient is positive. The estimated inclusion probabilities can be found in Table 11. To facilitate the

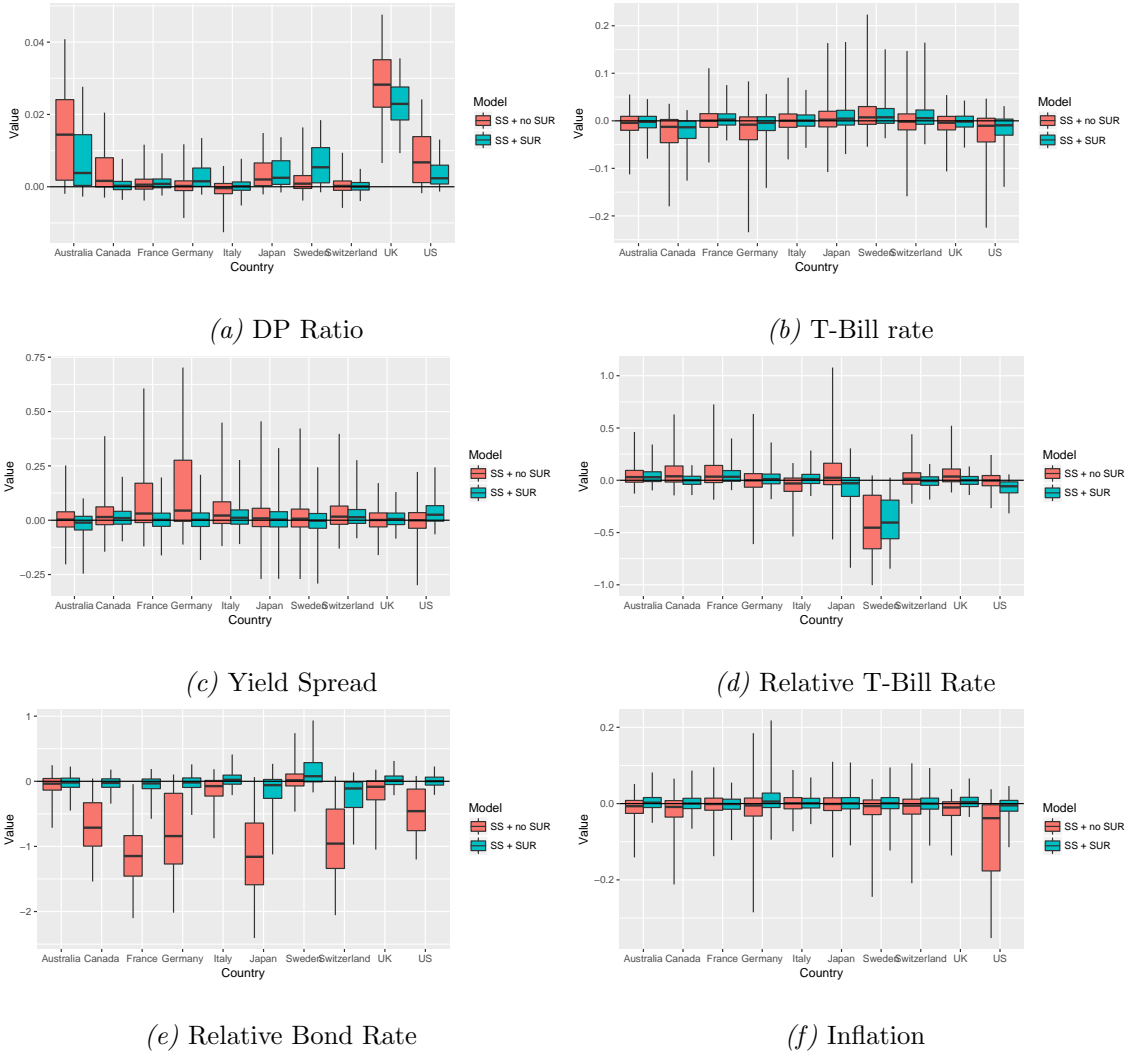


Figure 2: Posterior Distribution of Slope Coefficients (Basic Specification)

This figure displays posterior beliefs for slope coefficients. Each figure compares box plots of posterior distributions for coefficients in front of a corresponding regressor across countries and models. For each country and each regressor, there are two box plots. One box reflects the distribution with SUR while the other corresponds to the distribution obtained from estimating the predictive regression separately for each country. The center line of each box indicates the median of the distribution, and the box and the vertical lines correspond to the 75% and 95% credible sets, respectively. “SS + SUR” refers to an estimation method with spike-and-slab prior and SUR. “SS + No SUR” refers to an estimation method that forces the covariance matrix of errors to be diagonal. The set of predictors comprises the short-term interest rate tb , the yield spread (ys), the short-term interest rate relative to its 12-month moving average (rtb), the long-term government bond yield relative to its 12-month moving average (rbr), annual inflation rate ($infl$), and log dividend-price ratio dp . All predictors are in logs.

country	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>	
Australia							
Top 1	1	0	0	0	0	0	0.19
Top 2	0	0	0	0	0	0	0.12
Top 3	1	0	0	1	0	0	0.06
Canada							
Top 1	0	0	0	0	0	0	0.24
Top 2	0	1	0	0	0	0	0.10
Top 3	0	0	1	0	0	0	0.07
France							
Top 1	0	0	0	0	0	0	0.25
Top 2	0	0	0	1	0	0	0.07
Top 3	0	0	1	0	0	0	0.07
Germany							
Top 1	0	0	0	0	0	0	0.14
Top 2	1	0	0	0	0	0	0.09
Top 3	0	0	0	0	0	1	0.07
Italy							
Top 1	0	0	0	0	0	0	0.26
Top 2	0	0	1	0	0	0	0.08
Top 3	1	0	0	0	0	0	0.07
Japan							
Top 1	0	0	0	0	0	0	0.07
Top 2	1	0	0	0	0	0	0.07
Top 3	0	0	0	1	0	0	0.05
Sweden							
Top 1	1	0	0	1	0	0	0.13
Top 2	1	0	0	1	1	0	0.08
Top 3	0	0	0	1	0	0	0.07
Switzerland							
Top 1	0	0	0	0	0	0	0.18
Top 2	0	0	0	0	1	0	0.12
Top 3	0	0	0	0	0	1	0.05
UK							
Top 1	1	0	0	0	0	0	0.41
Top 2	1	1	0	0	0	0	0.08
Top 3	1	0	0	0	1	0	0.08
US							
Top 1	0	0	0	0	0	0	0.16
Top 2	0	0	0	1	0	0	0.08
Top 3	0	0	1	0	0	0	0.08

Table 9: Models with Highest Posterior Odds

This table reports the three best-performing model specifications for each country, where “1” indicates inclusion and “0” exclusion of the predictive variable. The last column reports the estimated posterior odds assigned to a particular combination of regressors.

comparison with earlier studies, Table 11 also gives the estimate inclusion probability if the correlation of errors is ignored and predictive regressions are estimated separately for each country. The data are relatively uninformative and, thus, the prior inclusion probability has a big influence on the posterior probabilities; hence, it is unproductive to focus on the obtained inclusion probabilities themselves. A natural way of measuring the relevance of a particular predictor variable is to see if the posterior inclusion probability exceeds the prior inclusion probability. In particular, the regressor can be considered to be predictive of future returns if the posterior probability is higher than 50%. Finally, Table 9 presents the three best-performing model specifications with the highest probability of inclusion. The model specifications are defined by inclusion (1) or exclusion (0) of specific predictor variables.

I now turn to results.

First, evidence in favor of predictability is quite low. At most, two predictors are retained as useful in the highest-probability models. Moreover, for six countries out of ten, the model with no regressors has the highest posterior odds. There is a single instance of the inclusion probability being higher than 90% in the dividend-price ratio for the U.K. market. This is also the only instance of the probability that the coefficient is positive being 100%. The second strong predictor is the T-bill rate that can forecast the returns for Sweden with an inclusion probability of 80% and the coefficient being negative with 95% probability. Thus, the predictability seems to be limited to a particular set of countries and a small subset of predictive variables.

The model uncertainty assumption is justified. Table 9 shows that there is little support for a single model that should be used to forecast return. The estimated probability of the “best model” is less than 25% for all countries with the exception of the U.K. Moreover, the second and third best model often feature probabilities close to that of the first model. Overall, this suggests that the data is inconclusive about a particular combination of regressors that should be used to forecast returns.

Next, the importance of interest rate variables and, in particular, the relative bond rate appears to be spurious. As can be seen from Table 11, the inclusion probabilities for the relative bond rate when SUR is ignored are usually revised upwards relative to the prior. This is by largely consistent with earlier results by Schrimpf (2010) who found rbr to be among the best predictors for four out of five countries (France, Germany, Japan, and the U.S.). The predictive power of interest rate variables for international markets was also noted by Hjalmarrsson (2010), Rapach et al. (2005), and Ang and Bekaert (2002). However, after the SUR model is imposed, the inclusion probabilities for the rbr are reconsidered downwards. In fact, the posterior inclusion probability of the relative bond rate is lower than the prior inclusion probability for every country, with the average inclusion probability of only 24%. Inspection of Figure 2 shows that the lower inclusion probabilities are complemented by lower absolute values of the coefficients. Note that the credible sets for most countries either include zero or are centered around zero after the cross-country correlation of errors is incorporated. Moreover, Table 9 shows that the relative bond rate is rarely included in one of the best-performing models.

Other interest rate variables also do not appear as powerful predictors, with some country-specific exceptions. In particular, the returns on the Swedish market are, with high probability, predicted by the relative Treasury bill rate. The average inclusion probabilities for these predictors decline accounting for the correlation of errors. Figure 2 also shows that the posterior means of coefficients tend to become small in absolute value after SUR is incorporated. Moreover, the credible sets typically include zero, with a tight 75% credible interval.

Overall, we have no evidence to support the claim that interest rate variables can forecast returns. The forecasting power of interest rate variables is, thus, an artifact of ignoring the correlation of errors across countries.

Finally, the dividend-price ratio appears to be the only robust predictor with the highest average inclusion probability of 43%. In particular, the inclusion probability is revised

		<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>
Australia	Mean	0.009	-0.005	-0.022	0.052	-0.034	0.003
	SD	(0.010)	(0.029)	(0.080)	(0.106)	(0.155)	(0.030)
	$P_{>0}$	0.83	0.45	0.40	0.69	0.42	0.54
Canada	Mean	0.001	-0.025	0.017	-0.000	-0.035	0.003
	SD	(0.003)	(0.039)	(0.069)	(0.071)	(0.124)	(0.035)
	$P_{>0}$	0.57	0.24	0.59	0.50	0.39	0.51
France	Mean	0.001	0.005	0.005	0.060	-0.064	-0.003
	SD	(0.003)	(0.028)	(0.078)	(0.120)	(0.179)	(0.034)
	$P_{>0}$	0.69	0.56	0.52	0.70	0.37	0.48
Germany	Mean	0.003	-0.013	0.005	0.024	-0.039	0.021
	SD	(0.004)	(0.045)	(0.083)	(0.121)	(0.181)	(0.071)
	$P_{>0}$	0.77	0.41	0.52	0.56	0.43	0.59
Italy	Mean	0.000	0.001	0.027	0.020	0.035	0.002
	SD	(0.003)	(0.028)	(0.089)	(0.100)	(0.146)	(0.029)
	$P_{>0}$	0.54	0.51	0.60	0.57	0.59	0.53
Japan	Mean	0.005	0.017	0.006	-0.109	-0.176	-0.002
	SD	(0.005)	(0.054)	(0.124)	(0.272)	(0.356)	(0.047)
	$P_{>0}$	0.88	0.60	0.52	0.37	0.32	0.50
Sweden	Mean	0.006	0.019	-0.006	-0.388	0.175	-0.001
	SD	(0.006)	(0.046)	(0.109)	(0.243)	(0.290)	(0.045)
	$P_{>0}$	0.88	0.65	0.48	0.05	0.72	0.52
Switzerland	Mean	0.000	0.017	0.031	-0.008	-0.232	-0.001
	SD	(0.002)	(0.050)	(0.085)	(0.081)	(0.307)	(0.044)
	$P_{>0}$	0.53	0.61	0.63	0.45	0.21	0.50
UK	Mean	0.023	-0.002	0.009	-0.002	0.020	0.006
	SD	(0.006)	(0.023)	(0.052)	(0.069)	(0.126)	(0.024)
	$P_{>0}$	1.00	0.46	0.57	0.49	0.56	0.60
US	Mean	0.004	-0.025	0.040	-0.076	0.004	-0.011
	SD	(0.004)	(0.046)	(0.077)	(0.093)	(0.108)	(0.038)
	$P_{>0}$	0.87	0.30	0.70	0.18	0.51	0.40

Table 10: Bayesian Estimates of Slope Parameters (Basic Specification).

This table reports the summary of the coefficient estimates. For each country and regressor, the table reports the posterior mean followed by the posterior standard deviation and the probability that the estimate is positive. Bold numbers denote instances in which the posterior probability of the slope being either positive or negative is less than 5%. The set of predictors comprises the short-term interest rate *tb*, the yield spread (*ys*), the short-term interest rate relative to its 12-month moving average (*rtb*), the long-term government bond yield relative to its 12-month moving average (*rbr*), annual inflation rate (*infl*), and log dividend-price ratio *dp*. All predictors are in logs.

	Spike-and-Slab Prior + SUR						Ignore SUR					
	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>
Australia	0.62	0.21	0.24	0.24	0.20	0.20	0.74	0.24	0.24	0.26	0.24	0.26
Canada	0.23	0.33	0.21	0.16	0.17	0.22	0.40	0.34	0.29	0.32	0.79	0.31
France	0.23	0.19	0.22	0.26	0.22	0.20	0.24	0.24	0.40	0.35	0.93	0.25
Germany	0.38	0.28	0.23	0.23	0.21	0.34	0.27	0.37	0.46	0.34	0.75	0.38
Italy	0.23	0.20	0.25	0.20	0.19	0.19	0.27	0.23	0.32	0.30	0.32	0.21
Japan	0.56	0.31	0.30	0.39	0.37	0.26	0.42	0.29	0.33	0.42	0.85	0.27
Sweden	0.62	0.29	0.27	0.81	0.36	0.25	0.31	0.32	0.31	0.76	0.26	0.30
Switzerland	0.18	0.29	0.24	0.17	0.43	0.25	0.23	0.31	0.30	0.25	0.80	0.32
UK	0.99	0.18	0.17	0.15	0.17	0.17	0.99	0.23	0.20	0.28	0.35	0.28
US	0.48	0.33	0.28	0.29	0.14	0.24	0.62	0.35	0.26	0.21	0.65	0.54
prior mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
average	0.45	0.26	0.24	0.29	0.25	0.23	0.45	0.29	0.31	0.35	0.60	0.31

Table 11: Bayesian Estimates of Inclusion Probabilities (Basic Specification)

This table reports the mean inclusion probability for each regressor. The numbers in bold denote instances in which the inclusion probability is higher than the prior. The country-specific means are followed by the prior inclusion probability. The last row reports the average inclusion probability for each regressor across countries. “Ignore SUR” refers to an estimation method that forces the covariance matrix of errors to be diagonal. The set of predictors comprises the short-term interest rate *tb*, the yield spread (*ys*), the short-term interest rate relative to its 12-month moving average (*rtb*), the long-term government bond yield relative to its 12-month moving average (*rbr*), annual inflation rate (*infl*), and log dividend-price ratio *dp*. All predictors are in logs.

upwards relative to the prior for the U.K., Australia, and (borderline) Sweden and Japan. In addition, the 75% credible sets do not include zero for six countries out of ten as can be seen from Figure 2. For these six countries, the dividend-price ratio is included in at least one of the tops three models. As reported in Table 10, this is the only predictor, where the posterior mean has a consistent sign across countries. That being said, the dividend-price ratio does not necessarily has the highest selection probability for each particular country. For example, for Switzerland the relative bond rate was selected in 42% of cases while the dividend-price ratio has an inclusion probability of only 17%.

Incorporating the SUR model does change the inference for the predictive power of the dividend-price ratio. In particular, the inclusion probabilities for Sweden and Japan are significantly higher after accounting for the correlation of errors. Conversely, the inclusion probability for the U.S. is revised downwards. The change in probability is accompanied by the change in the posterior mean, which increases for Sweden and Japan. Note that in general the posterior mean is revised down for the countries that have an atypically high value of the mean:i.e., Australia, the U.K., and the U.S.

The institutional organization of the financial market in a particular country may be informative about the perspective predictors. For some countries, such as France and Germany, banks play a central role in financial intermediation (so-called “bank-based” financial system), whereas for countries such as the U.S. and the U.K. the stock markets tend to be more important for individual investors due to their connection to the retirement system (so-called, “market-based” financial system).¹⁰ Schrimpf (2010) observes that the dividend-price ratio is more likely to be an important predictor of returns for “market-based” financial systems. My results suggest limited support for this hypothesis as three out of four countries, for which the dividend-price ratio is important are “market-based”: Australia, Sweden and the U.K. On the other hand, France, Germany, and Italy are “bank-based” financial systems and the dividend-price ratio does not predict returns in these countries. However,

¹⁰More information about the difference between “bank-based” and “market-based” financial systems can be found in Allen et al. (2001) and Demirgüç-Kunt and Levine (1999).

the returns for other “market-based” economies in the sample - Canada and Switzerland - are also not predicted by the dividend-price ratio.

2.4.2. Asset Allocation Analysis

This section examines the implications of model uncertainty and SUR for optimal asset allocation. In particular, I consider a Bayesian investor who maximizes their expected utility given a predictive density of future excess returns. I focus on the CER following the prior portfolio choice studies such as Kandel and Stambaugh (1996), Avramov (2002), and Shanken and Tamayo (2012). In particular, I study the certainty equivalent loss of an investor who is forced to hold a suboptimal portfolio and then use this measure to gage the economic significance of the model uncertainty and SUR.

Sensitivity of Asset Allocation to Model Uncertainty and SUR

In this subsection, I focus on the economic implications of model uncertainty and error correlations. Following Penasse (2016) and Shanken and Tamayo (2012), I consider three values for each predictive variable: its historical mean and 1.5 standard deviations above and below the historical mean. Other predictors are fixed at the historical mean.

I start by examining the sensitivity of the portfolio weights to the changes in predictive variables. The resulting portfolio weights are reported in Table 12. The first three columns provide an optimal weight for an investor who takes into account both model uncertainty and SUR for three values of each regressor. The second column corresponds to the optimal weight when all regressors are at their historical mean. In this scenario, the investor has no reason to time the market, so these allocations are identical to those of an investor who does not believe in return predictability. Moreover, they are identical across all regressors and methods and any difference is due to simulation noise. Thus, it is the same for all methods and omitted from further consideration. The following columns report optimal weights for

an investor who ignores either model uncertainty or SUR or both.

The first result is that taking into account model uncertainty leads to an allocation that is significantly less sensitive to the values of the predictive variables. Less sensitive allocation means that incorporating model uncertainty supports lower evidence in favor of predictability; thus, the investor times the market less aggressively. For example, if the dividend-price ratio is “high”, investors who ignore model uncertainty would typically allocate all of their wealth to equity. In contrast, when the dividend-price ratio is “low”, they would keep all their wealth in the risk-free assets. However, an investor who considers model uncertainty would always keep a portfolio that includes both types of assets (the only exception is a British investor who remains very sensitive to the value of the dividend-price ratio). A similar pattern can be observed for other predictive variables.

The influence of SUR on the asset allocation decision is more nuanced and depends on the regressor. More often than not, the investor who discards the correlation of errors is more sensitive to the values of the predictive variables. This is particularly true for the relative bond rate, which is considered to be an important predictor, if SUR is ignored. This reflects the fact that adding SUR, on average, leads to lower values of coefficients and, thus, reduces the incentive to time the market. On the other hand, incorporating the correlation of errors results in more precise estimates, so that the generated returns are less volatile and the investor feels more confident about market timing.

Economic Significance of Model Uncertainty and SUR

To judge the economic significance of SUR and model uncertainty, I explore the utility loss of an investor who is forced to ignore them and, thus, holds a suboptimal portfolio. Utility loss is computed as the loss in an annualized CER and provides an economic metric for gauging the effect of ignoring model uncertainty and SUR. The first three columns of Table 13 list the CERs of investors who hold the optimal portfolio. The subsequent columns display the certainty equivalent losses of investors who allocate wealth based on return-generating

processes that do not account for either SUR or model uncertainty.

Table 13 delivers two interesting observations. First, the costs of ignoring model uncertainty are economically significant. For example, if the dividend-price ratio is “high”, the the loss ranges from 0 to 3.4% per year. Second, the costs of ignoring SUR are, on average, lower than the costs of ignoring model uncertainty. The costs are economically meaningful only when they result in a substantial difference in inference. For example, when the relative bond rate deviates from its historical mean, the losses may be as high as 3.5% per year. These losses indicate that the relative bond rate is considered to be an important predictor if the SUR is ignored; thus, the investors time the market aggressively. On the other hand, this predictive variable appears to be useless if the correlation of errors is accounted for, so that the optimal behavior does not vary that much depending on the value of the regressor. In contrast, in the case of predictors that are usually discarded as worthless, such as the yield spread or the T-bill rate, the costs of discarding SUR are, essentially, zero. Similarly, both with and without SUR the dividend-price ratio is considered as a modestly useful regressor, so the costs are relatively low.

Predictor Values	SS + SUR		Ignore SUR		Ignore SS		Ignore Both		SS + SUR		Ignore SUR		Ignore SS		Ignore Both	
	-1.5	0	1.5	-1.5	1.5	-1.5	1.5	-1.5	1.5	-1.5	0	1.5	-1.5	1.5	-1.5	1.5
	<i>dp</i>															
Australia	0.18	0.35	0.54	0.00	0.69	0.00	1.00	0.00	1.00	0.27	0.35	0.43	0.25	0.44	0.00	0.87
Canada	0.39	0.40	0.41	0.24	0.57	0.00	1.00	0.00	1.00	0.39	0.40	0.40	0.25	0.55	0.31	0.52
France	0.33	0.37	0.39	0.31	0.43	0.00	1.00	0.00	0.80	0.30	0.37	0.44	0.26	0.46	0.00	0.73
Germany	0.27	0.34	0.42	0.33	0.39	0.00	0.92	0.21	0.45	0.32	0.34	0.37	0.35	0.35	0.17	0.52
Italy	0.15	0.14	0.15	0.19	0.11	0.00	0.36	0.25	0.06	0.12	0.14	0.17	0.21	0.11	0.00	0.29
Japan	0.09	0.29	0.50	0.09	0.51	0.00	1.00	0.00	0.86	0.36	0.29	0.24	0.24	0.38	0.45	0.16
Sweden	0.23	0.44	0.65	0.34	0.54	0.00	1.00	0.00	1.00	0.79	0.44	0.08	0.81	0.05	1.00	0.00
Switzerland	0.62	0.61	0.58	0.62	0.62	0.00	1.00	0.32	0.91	0.62	0.61	0.60	0.57	0.68	0.31	0.93
UK	0.00	0.51	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.51	0.51	0.52	0.39	0.63	0.47	0.55
US	0.38	0.60	0.84	0.05	1.00	0.00	1.00	0.00	1.00	0.76	0.60	0.46	0.65	0.61	1.00	0.15
	<i>tb</i>															
Australia	0.36	0.35	0.34	0.41	0.31	0.66	0.07	0.31	0.38	0.40	0.35	0.32	0.42	0.27	0.64	0.06
Canada	0.52	0.40	0.26	0.58	0.24	1.00	0.00	0.73	0.09	0.45	0.40	0.37	1.00	0.00	0.51	0.29
France	0.34	0.37	0.39	0.35	0.37	0.00	0.89	0.00	1.00	0.42	0.37	0.33	1.00	0.00	0.71	0.02
Germany	0.35	0.34	0.32	0.42	0.28	0.30	0.40	0.34	0.34	0.37	0.34	0.32	0.71	0.00	0.55	0.13
Italy	0.13	0.14	0.15	0.14	0.15	0.00	0.31	0.08	0.20	0.13	0.14	0.19	0.23	0.05	0.18	0.10
Japan	0.25	0.29	0.33	0.27	0.34	0.00	0.91	0.00	0.65	0.40	0.29	0.21	0.81	0.00	0.46	0.13
Sweden	0.36	0.44	0.50	0.37	0.51	0.00	0.90	0.00	1.00	0.36	0.44	0.54	0.42	0.43	0.18	0.69
Switzerland	0.55	0.61	0.67	0.62	0.63	0.00	1.00	0.00	1.00	0.78	0.61	0.47	1.00	0.01	1.00	0.00
UK	0.51	0.51	0.51	0.55	0.47	0.71	0.32	0.47	0.54	0.50	0.51	0.54	0.70	0.34	0.53	0.49
US	0.72	0.60	0.50	0.79	0.41	1.00	0.00	1.00	0.00	0.62	0.60	0.62	1.00	0.06	1.00	0.12
	<i>ys</i>															
Australia	0.38	0.35	0.33	0.35	0.37	0.28	0.45	0.03	0.68	0.36	0.35	0.38	0.40	0.30	0.00	0.70
Canada	0.37	0.40	0.42	0.31	0.50	0.47	0.34	0.06	0.77	0.40	0.40	0.40	0.51	0.31	0.12	0.68
France	0.38	0.37	0.36	0.26	0.48	0.06	0.68	0.00	1.00	0.38	0.37	0.36	0.39	0.34	0.94	0.00
Germany	0.35	0.34	0.35	0.21	0.48	0.20	0.50	0.00	0.74	0.32	0.34	0.38	0.39	0.32	0.31	0.35
Italy	0.14	0.14	0.18	0.07	0.22	0.00	0.45	0.00	0.47	0.15	0.14	0.15	0.14	0.14	0.00	0.37
Japan	0.30	0.29	0.29	0.27	0.33	0.32	0.29	0.08	0.52	0.30	0.29	0.28	0.33	0.28	0.71	0.00
Sweden	0.45	0.44	0.43	0.40	0.46	0.38	0.50	0.10	0.76	0.44	0.44	0.43	0.50	0.37	0.75	0.10
Switzerland	0.57	0.61	0.69	0.56	0.69	0.00	1.00	0.00	1.00	0.64	0.61	0.61	0.68	0.56	0.88	0.32
UK	0.49	0.51	0.54	0.53	0.48	0.43	0.60	0.41	0.62	0.48	0.51	0.56	0.63	0.39	0.40	0.62
US	0.53	0.60	0.71	0.63	0.59	1.00	0.15	1.00	0.00	0.65	0.60	0.56	0.99	0.22	0.82	0.41

Table 12: In-Sample Asset Allocation

This table reports the optimal allocation for various prior beliefs and three levels of each predictor. The predictor increases from 1.5 standard deviations below each mean to 1.5 standard deviations above its mean while other predictors are held at their average.

Predictor Values	SS + SUR			Ignore SUR			Ignore SS			Ignore Both			SS + SUR			Ignore SUR			Ignore SS			Ignore Both			
	-1.5	0	1.5	-1.5	1.5	1.5	-1.5	1.5	1.5	-1.5	1.5	1.5	-1.5	0	1.5	-1.5	1.5	1.5	-1.5	1.5	1.5	-1.5	1.5	1.5	
	<i>dp</i>																								
Australia	0.25	0.98	2.33	-0.25	-0.19	-0.25	-1.71	-1.71	-0.25	-1.71	-1.71	-0.25	-1.71	0.58	0.98	1.52	-0.00	-0.00	-0.58	-1.57	-1.57	-0.58	-1.57	-0.58	-2.30
Canada	1.00	1.05	1.11	-0.16	-0.17	-1.00	-2.24	-2.24	-1.00	-2.24	-2.24	-1.00	-2.24	0.99	1.05	1.06	-0.13	-0.14	-0.05	-0.09	-0.09	-0.05	-0.09	-0.05	-2.34
France	1.04	1.32	1.45	-0.00	-0.02	-1.04	-3.70	-3.70	-1.04	-3.70	-3.70	-1.04	-3.70	0.86	1.32	1.86	-0.01	-0.01	-0.86	-0.85	-0.85	-0.86	-0.85	-0.86	-3.06
Germany	0.72	1.19	1.78	-0.04	-0.01	-0.72	-2.58	-2.58	-0.72	-2.58	-2.58	-0.72	-2.58	1.06	1.19	1.36	-0.01	-0.00	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-1.12
Italy	0.30	0.27	0.30	-0.02	-0.02	-0.30	-0.58	-0.58	-0.30	-0.58	-0.58	-0.30	-0.58	0.20	0.27	0.37	-0.11	-0.05	-0.19	-0.20	-0.20	-0.19	-0.20	-0.19	-0.00
Japan	0.07	0.70	2.16	-0.00	-0.00	-0.07	-2.10	-2.10	-0.07	-2.10	-2.10	-0.07	-2.10	1.08	0.70	0.47	-0.12	-0.17	-0.07	-0.05	-0.05	-0.07	-0.05	-0.07	-2.57
Sweden	0.65	2.40	5.16	-0.14	-0.14	-0.65	-1.55	-1.55	-0.65	-1.55	-1.55	-0.65	-1.55	7.68	2.40	0.07	-0.01	-0.01	-0.56	-0.07	-0.07	-0.56	-0.07	-0.56	-0.07
Switzerland	2.29	2.26	2.04	-0.00	-0.01	-2.29	-1.06	-1.06	-2.29	-1.06	-1.06	-2.29	-1.06	2.31	2.26	2.19	-0.02	-0.03	-0.58	-0.62	-0.62	-0.58	-0.62	-0.58	-0.94
UK	0.00	1.82	10.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.82	1.82	1.85	-0.10	-0.09	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-1.63
US	0.80	2.03	3.97	-0.60	-0.14	-0.80	-0.14	-0.14	-0.80	-0.14	-0.14	-0.80	-0.14	3.23	2.03	1.16	-0.07	-0.13	-0.32	-0.54	-0.54	-0.32	-0.54	-0.32	-0.02
	<i>tb</i>																								
Australia	1.06	0.98	0.94	-0.02	-0.01	-0.72	-0.59	-0.59	-0.72	-0.59	-0.59	-0.72	-0.59	1.30	0.98	0.82	-0.00	-0.02	-0.45	-0.55	-0.55	-0.45	-0.55	-0.45	-0.82
Canada	1.77	1.05	0.46	-0.03	-0.00	-1.53	-0.46	-0.46	-1.53	-0.46	-0.46	-1.53	-0.46	1.32	1.05	0.86	-1.95	-0.86	-0.02	-0.04	-0.04	-0.02	-0.04	-0.02	-0.86
France	1.11	1.32	1.49	-0.00	-0.00	-1.11	-2.38	-2.38	-1.11	-2.38	-2.38	-1.11	-2.38	1.69	1.32	1.07	-3.30	-1.07	-0.85	-0.97	-0.97	-0.85	-0.97	-0.85	-1.07
Germany	1.26	1.19	1.01	-0.04	-0.01	-0.03	-0.06	-0.06	-0.03	-0.06	-0.06	-0.03	-0.06	1.39	1.19	1.06	-1.13	-1.06	-0.32	-0.38	-0.38	-0.32	-0.38	-0.32	-1.06
Italy	0.22	0.27	0.32	-0.00	-0.00	-0.22	-0.35	-0.35	-0.22	-0.35	-0.35	-0.22	-0.35	0.22	0.27	0.50	-0.13	-0.26	-0.03	-0.11	-0.11	-0.03	-0.11	-0.03	-0.50
Japan	0.53	0.70	0.92	-0.01	-0.00	-0.53	-2.85	-2.85	-0.53	-2.85	-2.85	-0.53	-2.85	1.38	0.70	0.36	-1.39	-0.36	-0.03	-0.05	-0.05	-0.03	-0.05	-0.03	-0.36
Sweden	1.57	2.40	3.05	-0.00	-0.00	-1.57	-2.02	-2.02	-1.57	-2.02	-2.02	-1.57	-2.02	1.56	2.40	3.55	-0.05	-0.13	-0.37	-0.27	-0.27	-0.37	-0.27	-0.37	-0.01
Switzerland	1.82	2.26	2.70	-0.03	-0.01	-1.82	-0.66	-0.66	-1.82	-0.66	-0.66	-1.82	-0.66	3.67	2.26	1.35	-0.29	-1.35	-0.29	-1.30	-1.30	-0.29	-1.30	-0.29	-1.35
UK	1.84	1.82	1.81	-0.01	-0.01	-0.26	-0.26	-0.26	-0.26	-0.26	-0.26	-0.26	-0.26	1.74	1.82	2.03	-0.28	-0.28	-0.00	-0.02	-0.02	-0.00	-0.02	-0.00	-2.03
US	2.95	2.03	1.38	-0.03	-0.05	-0.43	-1.38	-1.38	-0.43	-1.38	-1.38	-0.43	-1.38	2.14	2.03	2.13	-0.80	-1.72	-0.44	-0.40	-0.40	-0.44	-0.40	-0.44	-1.37
	<i>ys</i>																								
Australia	1.17	0.98	0.87	-0.01	-0.01	-0.08	-0.11	-0.11	-0.08	-0.11	-0.11	-0.08	-0.11	1.02	0.98	1.14	-0.01	-0.05	-1.01	-0.85	-0.85	-1.01	-0.85	-1.01	-0.35
Canada	0.90	1.05	1.18	-0.02	-0.04	-0.06	-0.04	-0.04	-0.06	-0.04	-0.04	-0.06	-0.04	1.03	1.05	1.03	-0.08	-0.05	-0.51	-0.53	-0.53	-0.51	-0.53	-0.51	-0.38
France	1.41	1.32	1.28	-0.15	-0.13	-0.98	-0.96	-0.96	-1.41	-3.95	-3.95	-1.41	-3.95	1.37	1.32	1.23	-0.00	-0.00	-3.02	-1.23	-1.23	-3.02	-1.23	-3.02	-1.23
Germany	1.21	1.19	1.21	-0.20	-0.18	-0.22	-0.22	-0.22	-0.21	-1.57	-1.57	-0.21	-1.57	1.01	1.19	1.48	-0.05	-0.04	-0.00	-0.01	-0.01	-0.00	-0.01	-0.00	-0.30
Italy	0.25	0.27	0.43	-0.06	-0.02	-0.25	-1.03	-1.03	-0.25	-1.17	-1.17	-0.25	-1.17	0.32	0.27	0.30	-0.00	-0.00	-0.32	-0.66	-0.66	-0.32	-0.66	-0.32	-0.16
Japan	0.78	0.70	0.70	-0.01	-0.02	-0.00	-0.00	-0.00	-0.43	-0.45	-0.45	-0.43	-0.45	0.77	0.70	0.66	-0.01	-0.00	-1.37	-0.66	-0.66	-1.37	-0.66	-1.37	-0.66
Sweden	2.51	2.40	2.26	-0.03	-0.01	-0.06	-0.07	-0.07	-1.49	-1.35	-1.35	-1.49	-1.35	2.35	2.40	2.26	-0.05	-0.04	-1.20	-1.32	-1.32	-1.20	-1.32	-1.20	-2.26
Switzerland	1.98	2.26	2.85	-0.00	-0.00	-1.98	-0.59	-0.59	-1.98	-0.59	-0.59	-1.98	-0.59	2.45	2.26	2.26	-0.01	-0.02	-0.36	-0.51	-0.51	-0.36	-0.51	-0.36	-1.01
UK	1.69	1.82	2.05	-0.01	-0.03	-0.03	-0.02	-0.02	-0.05	-0.04	-0.04	-0.05	-0.04	1.59	1.82	2.19	-0.17	-0.20	-0.04	-0.03	-0.03	-0.04	-0.03	-0.04	-2.19
US	1.55	2.03	2.80	-0.06	-0.07	-1.25	-1.73	-1.73	-1.25	-2.80	-2.80	-1.25	-2.80	2.33	2.03	1.73	-0.65	-0.64	-0.17	-0.12	-0.12	-0.17	-0.12	-0.17	-1.73

Table 13: In-Sample Certainty Equivalent Returns

This table reports CERs and losses for various prior beliefs and three levels of each predictor. The predictor increases from 1.5 standard deviations below each mean to 1.5 standard deviations above its mean while other predictors are held at their average.

2.5. Out-of-sample Analysis

In this section, I complement the in-sample results by providing empirical evidence on the out-of-sample (OOS) predictability of the equity premia in international stock markets. I consider the predictability both from an economic and statistical perspective. Statistically, I assess whether the in-sample return predictability documented in the previous section results in meaningful improvement of the forecast based on the out-of-sample R_{OOS}^2 . Economically, I consider portfolio choices of investors with a power utility function and calculate their certainty equivalent returns. Finally, I explore how predictive performance evolves over time.

The main benchmark is a naive forecasting rule based on the average of past returns. The benchmark is motivated by paper of Welch and Goyal (2007), who show that it yields a superior performance relative to more complex predictive models that often feature unstable OOS performance. Similarly, in the context of the international markets, Schrimpf (2010) detects only a limited amount of predictability. Penasse (2016) explores economic evidence of predictability and finds that both the certainty equivalent return (CER) and Sharpe ratios improve significantly relative to the naive benchmark under hierarchical priors and non-negativity constraints on equity premium.

Before proceeding, it is important to caution that the importance of the out-of-sample predictability is a debated topic in the literature. Several theoretical studies discuss the theoretical reasons why the OOS performance may be poor. For example, Inoue and Kilian (2005) argue in favor of in-sample predictability tests because they are more powerful. Cochrane (2008) and Lettau and Van Nieuwerburgh (2007) show that the predictability in the return-generating process can naturally coexist with a complete absence of OOS return predictability. I still find the out-of-sample exercise necessary because international investors care about whether they can exploit the predictive variables to predict the future returns better than the benchmark.

The OOS analysis is based on a sample that spans 20 years from January 1997 through December 2016, so that at least 22 years of data are used for the in-sample training. prediction error. The examination is based upon two schemes. The first scheme, the rolling window approach, fixes the estimation window size and drops distant observations as recent ones are added. The size of the window is set to be 20 years. The second scheme, the recursive window approach, uses all available data. The model is re-estimated monthly.

2.5.1. Statistical Performance of Forecasts

This section explores the statistical properties of the forecast generated by multiple approaches. Throughout the section I focus solely on the posterior mean.

Table 14 reports the OOS R_{OOS}^2 as in Campbell and Thompson (2007). The R_{OOS}^2 measures the reduction in the mean squared forecast error relative to the naive historical average benchmark:

$$R_{OOS,n}^2 = 1 - \frac{\sum_{\tau=t}^T (r_{n,t} - \hat{r}_{n,t})^2}{\sum_{\tau=t}^T (r_{n,t} - \bar{r}_{n,1:t-1})^2},$$

where $r_{n,t}$ is the true return for country n at period t , $\hat{r}_{n,t}$ is the forecast and $\bar{r}_{n,1:t-1}$ is the historical mean based on the data up to period $t - 1$. A positive R^2 implies that the predictive regression produces lower average mean-squared prediction error than the historical average. To compare the performance across models I use the test statistic from Clark and West (2006) and Clark and West (2007).

Several results emerge.

First, the historical average is typically superior to the model-based forecasts. Even after taking into account model uncertainty and SUR, for most countries the OOS R^2 is negative meaning that they under-perform relative to the naive benchmark. Moreover, positive OOS R^2 is usually statistically not different from zero. The highest possible R^2 is less than 2%. Overall, we can spot some evidence in favor of predictability for four countries: Australia, Sweden, the U.K., and the U.S. However, this result is not robust as it depends on the

	Recursive Scheme			Rolling Scheme		
SUR	✓		✓	✓		✓
Spike & Slab	✓	✓		✓	✓	
Australia	-0.26	-0.34	-3.22	-0.41	0.53	-3.99
Canada	-0.59	-1.24	-3.81	-0.53	-1.05	-5.63
France	-0.44	-0.25	-1.79	-0.33	-0.85	-2.72
Germany	-0.35	-0.08	-1.81	-0.79	-0.87	-4.79
Italy	-0.66	-0.53	-4.27	-0.17	-1.19	-1.33
Japan	-0.52	-1.37	-2.33	0.21	-1.89	-4.29
Sweden	0.40	0.20	-2.27	1.30**	0.90*	-2.82
Switzerland	-0.24	-1.61	-4.11	-0.46	-2.83	-7.69
UK	1.54**	1.42**	-0.55	0.33	0.91*	-1.93
US	0.12	0.04	-0.14	0.09	0.60*	0.30
average	-0.10	-0.38	-2.43	-0.08	-0.58	-3.49

Table 14: R_{OOS}^2 (Basic Specification)

This table reports the out-of-sample R^2 in percents. The forecasts are generated either recursively with an expanding window or with a rolling window of 20 years from January 1997 through December 2016. The R_{OOS} is computed for the posterior mean. Clark and West test by Clark and West (2007) and Clark and West (2006) is used to compare the out-of-sample mean squared prediction errors produced by the models relative with those based on the historical mean under one-sided hypothesis. The stars flag levels of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

forecasting scheme and whether or not the correlation of errors is considered.

Second, taking into account model uncertainty offers a huge economic improvement. Incorporating model uncertainty helps to avoid large forecasting mistakes, which leads to the increase of the R_{OOS}^2 by an average of 2-3%. This result is stable across countries and forecasting schemes.

Third, incorporating the cross-correlation of errors does improve the forecast but the improvement is relatively modest. However, country-specific regressions deliver better results for individual countries that are likely to have a relatively strong predictability pattern. In particular, under a rolling scheme the country-specific averaging with spike-and-slab prior out-perform the native benchmark for Australia, Sweden, and the U.K. The regressions with SUR that take into account model uncertainty also deliver positive R_{OOS}^2 , but it is much lower than for country-specific regressions and not statistically significant. The under-performance of SUR is likely to be caused by the fact that given the relatively small sample size of 20 years, the gains of adding SUR are outweighed by the loss of precision due to the increased number of coefficients.

2.5.2. Variations of Forecast Performance

In this section, I focus on the specification with model uncertainty and SUR and discuss how predictability patterns change over time. I consider only the recursive forecasting scheme because the rolling forecasting scheme offers no additional insights. As emphasized by Welch and Goyal (2007), the forecast performance may vary greatly over time. Following Schrimpf (2010), I investigate the time-varying performance using the Net-SSE plots presented in Figure 3. These plots display cumulative squared error relative to the naive model:

$$\sum_{\tau=1}^t (r_{n,t} - \bar{r}_{n,1:t-1})^2 - \sum_{\tau=1}^t (r_{n,t} - \hat{r}_{n,t})^2$$

When the line in the graph has an upward slope, the model outperforms the naive benchmark in terms of squared errors.

Figure 3 shows a significant variation in the degree of OOS predictability over time and across countries. The general result is that the evidence in favor of predictability is rather modest and the episodes of predictability are relatively short as also found by Schrimpf (2010). For many countries, there is only a single period when returns are predictable that took place. For example, in the U.K. this period is responsible for the out-performance of the model relative to the naive benchmark. Thus, even though returns in the U.K. appear to be predictable based on the OOS R^2 , they were actually only predictable for a few years. The fact that the period is similar for many countries implies that the predictability may be associated with some temporary mis-pricing that took place after the dot-com crush. The only exception is the United States where we can observe the upwards trend in the 2010s as well. Thus, the U.S. appears to be the only country with the robust predictability pattern.

2.5.3. Economic Performance: Implications for Asset Allocation

This subsection examines the implications of model uncertainty and SUR for optimal asset allocation.

Table 15 compares the average asset allocation for the out-of-sample analysis under different specifications. Table 16 reports the annualized CERs.

Table 15 shows that consistent with the in-sample results, both incorporating cross-sectional data in the form of SUR and taking into account model uncertainty leads to lower allocation to equity. After both of them are accounted for, investor typically end up with an average allocation very similar to the one produced by following historical means. Another observation is that the investor who uses the rolling scheme is usually less aggressive than his counterpart who uses the recursive forecasting scheme.

The first set of results confirms findings from the previous section. In particular, the evidence in favor of return predictability is weak and the historical mean routinely outperforms all competing model specifications. Taking into account model uncertainty usually offers a huge improvement. On the other hand, the benefits of using SUR are relatively small and

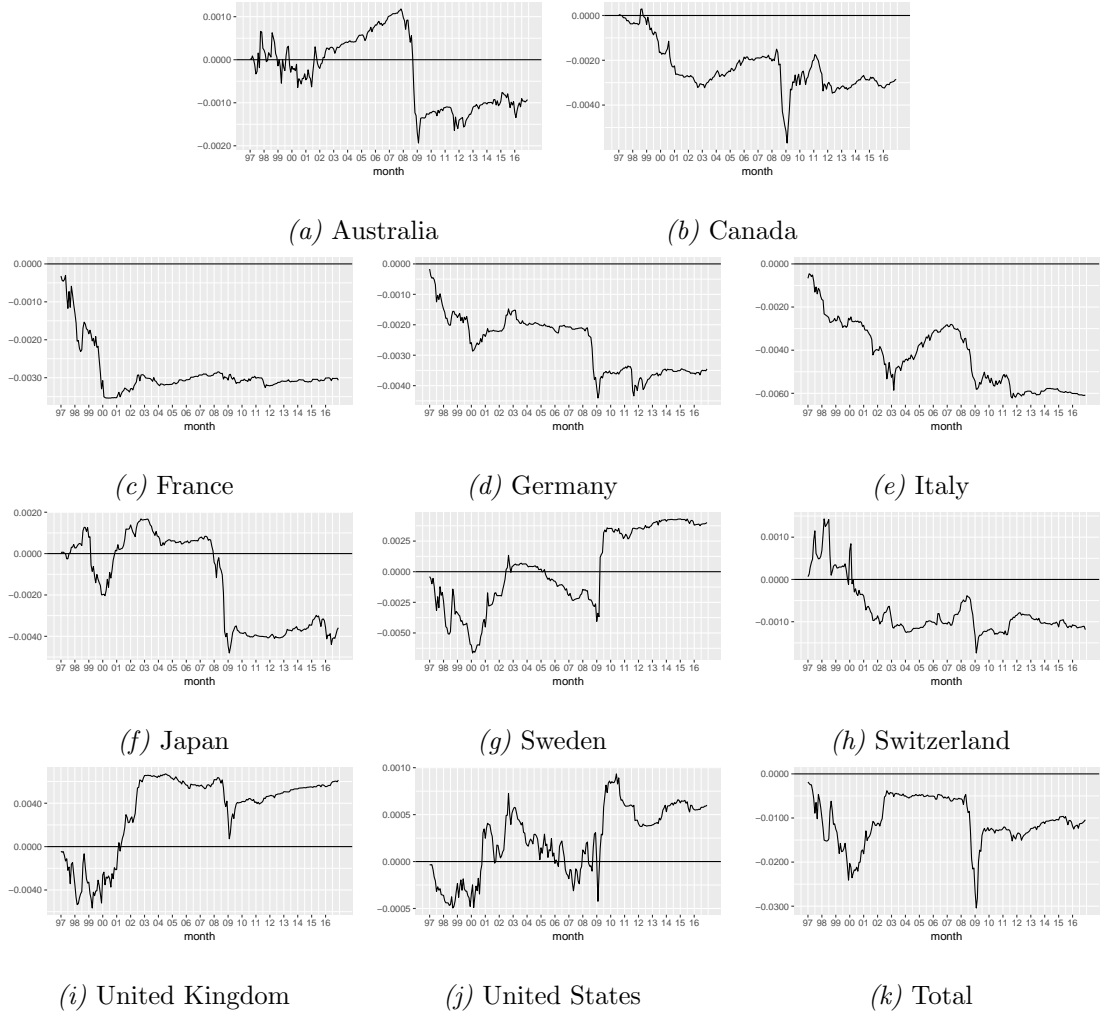


Figure 3: Net-SSE

This figure displays Net-SSE plots of the forecasts based on the composite model with model uncertainty and SUR. Net-SSE is a difference of squared error of the forecast based on historical averages minus the forecast of the conditional model. The forecasts are generated recursively with an expanding window from January 1997 through December 2016.

	Mean	Recursive Scheme			Rolling Scheme		
SUR		✓		✓	✓		✓
Spike & Slab		✓	✓		✓	✓	
Australia	0.33	0.33	0.42	0.50	0.19	0.58	0.61
Canada	0.35	0.46	0.59	0.59	0.33	0.60	0.59
France	0.36	0.34	0.42	0.37	0.23	0.50	0.41
Germany	0.35	0.34	0.42	0.21	0.15	0.42	0.29
Italy	0.16	0.21	0.21	0.26	0.06	0.23	0.32
Japan	0.29	0.28	0.32	0.32	0.08	0.22	0.42
Sweden	0.43	0.39	0.40	0.32	0.23	0.46	0.38
Switzerland	0.63	0.62	0.70	0.29	0.37	0.68	0.29
UK	0.50	0.33	0.41	0.44	0.18	0.52	0.52
US	0.61	0.57	0.64	0.50	0.39	0.64	0.52
average	0.40	0.39	0.45	0.38	0.22	0.49	0.43

Table 15: Out-of-Sample Asset Allocation (Basic Specification)

This table reports average allocation to the market index for an investor with a power utility function with a coefficient of relative risk aversion $A = 5$. The forecasts are generated either recursively or with a rolling window from January 1997 through December 2016. The asset allocation is updated monthly.

	Mean	Recursive Scheme			Rolling Scheme		
SUR		✓		✓	✓		✓
Spike & Slab		✓	✓		✓	✓	
Australia	0.89	0.74	0.86	0.36	0.27	0.69	-0.04
Canada	1.26	0.71	0.29	0.80	0.62	0.73	0.19
France	0.99	0.57	1.20*	0.79	0.56	0.73	-0.13
Germany	0.65	0.11	-0.02	-0.60	0.04	-0.71	-3.54
Italy	0.12	-0.43	-0.23	-1.83	0.23	-0.47	-0.11
Japan	-0.26	-1.06	-1.72	-2.37	-0.26	-1.55	-0.92
Sweden	1.41	1.95***	1.68***	0.45	2.70***	2.81***	0.03
Switzerland	1.05	0.41	-0.19	-1.05	0.19	-1.13	-1.77
UK	0.39	1.08***	1.03***	0.89**	0.34	0.06	-0.01
US	0.91	1.06**	0.69	1.33	1.22**	2.64***	1.79***
average	0.74	0.52	0.36	-0.12	0.59	0.38	-0.45

Table 16: Out-of-Sample Certainty Equivalent Returns (Basic Specification)

This table reports the CERs for an investor with a power utility function with a coefficient of relative risk aversion $A = 5$. The CERs are annualized and reported in percents. The forecasts are generated either recursively or with a rolling window from January 1997 through December 2016. The asset allocation is updated monthly. Delta method as in Shanken and Tamayo (2012) is used to test if the excess CER is bigger than zero. The stars flag levels of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

vary from country to country. Under the rolling scheme incorporating SUR leads to CER losses for most countries.

However, there are some substantial differences in results obtained from Table 14 and Table 16. For example, Table 14 says that the U.S. investor cannot significantly outperform the naive forecast if he uses the recursive forecasting scheme. Someone who focuses just on the posterior mean would, therefore, conclude that returns in the U.S. are unpredictable. However, the results in Table 16 show that the investor can still greatly benefit from using the predictive regression. This confirms the earlier thesis by Barberis (2000) and Kandel and Stambaugh (1996) that even statistically insignificant changes in inference can lead to statistically significant changes in asset allocation. Another explanation for this discrepancy is that the R_{OOS}^2 only focuses on the posterior mean of the predicted equity premium and discards all other moments. On the other hand, the optimal allocation decision is based on the whole distribution and takes into account, in particular, the posterior variance of the predicted equity premium. The variance of returns is often believed to be easier to predict than the posterior mean; therefore, the investor may benefit from using a predictive regression, even if is not optimal in terms of predicting the mean.

2.6. Extensions

In this section I consider to extensions to the basic model presented above: borrowing strength from cross-section and adding stochastic volatility.

2.6.1. Borrowing Strength From Cross-Section

Predictive regression such as in (2.1) may yield imprecise estimates because the return data is very uninformative. A natural way to obtain more precise estimates is to combine country-specific and cross-sectional information, which is known as borrowing strength from the cross-section.

Overall, the literature suggests that the predictability patterns should be different across

countries but offers little guidance about why they differ and how much they differ. For example, consider the coefficient in front of the dividend-price ratio. Even though the model coefficients differ across countries, they describe the same relationships. A priori there is no reason to expect that these coefficients are extremely different. It is thus reasonable to assume that parameter values in one country are informative about parameter values in another country and that the difference between them is just a work of chance. This belief can be expressed mathematically by imposing a hierarchical prior on the parameters. Hierarchical prior means that country-specific parameter values, the first level of the hierarchy, are modeled as draws from the same common distributions, the second level of the hierarchy. The parameters of the common distribution, the hyper-parameters, are then inferred from the data.

In this subsection, I explore if borrowing strength from the cross-section helps to forecast returns. I consider two mechanisms of borrowing strength from the cross-section: imposing hierarchical prior beliefs either on the coefficient values or on the inclusion probabilities. I find that none of these mechanism neither change the key in-sample results nor lead to a significant improvement of the out-of-sample forecasting performance. This chapter thus contradicts the work by Penasse (2016) who argues that imposing hierarchical priors on the coefficients from the dividend-price ratio delivers superior out-of-sample performance in the large majority of countries. I believe that the improved performance is largely driven by the non-negativity constraint on the risk premium imposed by Penasse (2016).

I start by introducing the changes to the methodological framework and proceed to discuss the key results.

Updated Econometric Framework

The system of prior beliefs in (2.2) is now updated as follows:

$$\begin{aligned}
 \beta_{n,k} | \gamma_{n,k}, v_0, \mu_k, \tau_k^2 &\stackrel{ind}{\sim} (1 - \gamma_{n,k}) \mathcal{N}(0, v_0^2 \tau_k^2) + \gamma_{n,k} \mathcal{N}(\mu_k, \tau_k^2) \\
 \gamma_{n,k} | \omega_{n,k} &\stackrel{ind}{\sim} \text{Bernoulli}(\omega_k) \\
 \tau_k^{-2} | a_{0,k}, b_{0,k} &\stackrel{ind}{\sim} \Gamma(a_{0,k}, b_{0,k}) \\
 \omega_k &\stackrel{ind}{\sim} U[0, 1] \\
 \mu | \mu_0, M_0 &\stackrel{ind}{\sim} \mathcal{N}(\mu_0, M_0)
 \end{aligned} \tag{2.6}$$

Prior (2.6) deviates from the prior (2.2) in two ways. First, the “slab” part for the coefficients β is now centered around the common mean (μ) rather than zero. Second, the prior on the inclusion probabilities γ is governed by the common inclusion probability ω , which is not a fixed value as before and is determined by the data.

Note that the prior on the coefficients $\beta_{n,k}$ is not, strictly speaking, a spike-and-slab prior. In particular, if the coefficient is “large” the prior shrinks the posterior values toward the common mean μ rather than toward zero. The common mean μ_k reflects the “average” value of the coefficient and is determined from the data. Thus, this form prior on the coefficients $\beta_{n,k}$ can be thought of as a clustering mechanism with two clusters. A priori each coefficient can with some probability either belong to a cluster of zeros or to a cluster centered around the common mean. I impose a normal prior on the vector of the common means $\mu = [\mu_1, \dots, \mu_K]$. Note that the shrinkage toward the common mean is weak relative to the setup without the spike around zero. To understand why, recall that the weight of the prior common mean is inversely proportional to the prior variance, τ^2 . In turn, the estimated prior variance remains relatively large in this setting because the estimate is based both on the spike and on the slab part of the prior.

The second change relative to the prior in (2.2) is that the prior probability of inclusion

ω_k is the same for all predictive variables of the same type. The probability ω_k is the common probability that a regressor k is helpful in forecasting returns. The prior shrinks the posterior estimates for each probability toward the “common” value. Specifically, amount of shrinkage depends on the likelihood of $\beta_{n,k}$ being close to zero versus the likelihood that it is far from zero. The bigger is the difference and the more information we can infer from the data about the actual value of the coefficient, the less important is the prior probability. In contrast, if the difference is small and the data cannot distinguish between zero and non-zero values well, the posterior estimate of the inclusion probability is strongly shrunk toward the common probability. I impose a uniform prior on the common probability.

The Gibbs sampler used to estimate the model with prior (2.6) as well as the prior specifications can be found in Appendix 2.8.2.

In- and Out-of-Sample Results

In this section I present in- and out-of-sample analysis for the model specification augmented with hierarchical priors. In particular, I show that the key results reported in the previous sections remain intact.

Let us first discuss the posterior beliefs about the inclusion probabilities and coefficient values under hierarchical priors. The posterior inclusion probabilities are reported in Table 17, while the box plots depicting the coefficient values are displayed in Figure 4. Several conclusions emerge.

First, introducing hierarchical prior beliefs on means leads to very modest changes in the inclusion probabilities. In particular, the inclusion probabilities are, on average, slightly revised upwards for the dividend-price ratio relative to the estimates obtained without the hierarchical prior. On the opposite, the inclusion probabilities for other predictive variables get slightly lower. The change in probability, however, rarely exceeds 2%. These changes in inclusion probabilities are accompanied by the small revisions of the posterior mean values. For example, as can be seen from Figure the posterior means are, on average, slightly

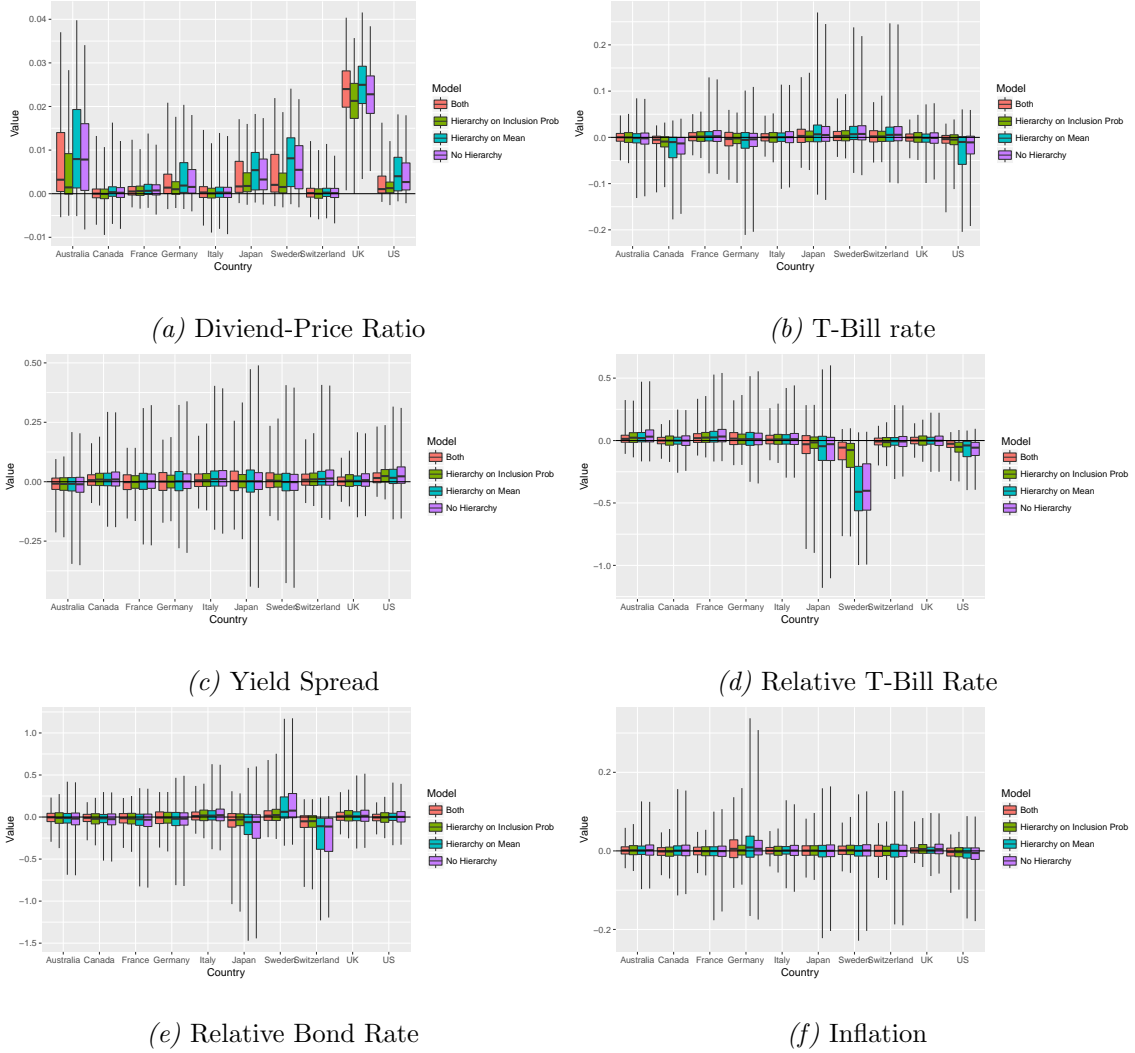


Figure 4: Posterior Distribution of Slope Coefficients (Borrowing from Cross-Section)

This figure displays posterior beliefs for slope coefficients. Each figure compares box plots of posterior distributions for coefficients in front of a corresponding regressor across countries and models. For each country and each regressor, there are four box plots. They correspond to the coefficient values estimated under different priors. “Hierarchy on Mean” and “Hierarchy on Inclusion Prob” refers to specifications that impose either hierarchical prior on mean or inclusion probabilities. “No Hierarchy” refers to the basic specification without hierarchical prior beliefs as given in (2.2). “Both” refers to the specification with hierarchical prior both on the coefficient values and on the inclusion probabilities as given in (2.6). The center line of each box indicates the median of the distribution, the box and the vertical lines correspond to the 75% and 95% credible sets. The set of predictors comprises the short-term interest rate tb , the yield spread (ys), the short-term interest rate relative to its 12-month moving average (rtb), the long-term government bond yield relative to its 12-month moving average (rbr), annual inflation rate ($infl$), and log dividend-price ratio dp . All predictors are in logs.

	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>
	No Hierarchy						Hierarchy on Mean					
Australia	0.62	0.21	0.24	0.24	0.20	0.20	0.68	0.17	0.20	0.21	0.16	0.17
Canada	0.23	0.33	0.21	0.16	0.17	0.22	0.22	0.30	0.18	0.13	0.14	0.18
France	0.23	0.19	0.22	0.26	0.22	0.20	0.23	0.16	0.19	0.22	0.19	0.18
Germany	0.38	0.28	0.23	0.23	0.21	0.34	0.43	0.24	0.20	0.20	0.19	0.30
Italy	0.23	0.20	0.25	0.20	0.19	0.19	0.20	0.17	0.22	0.18	0.17	0.16
Japan	0.56	0.31	0.30	0.39	0.37	0.26	0.49	0.27	0.26	0.38	0.33	0.22
Sweden	0.62	0.29	0.27	0.81	0.36	0.25	0.67	0.26	0.24	0.79	0.32	0.23
Switzerland	0.18	0.29	0.24	0.17	0.43	0.25	0.16	0.25	0.21	0.15	0.39	0.22
UK	0.99	0.18	0.17	0.15	0.17	0.17	1.00	0.15	0.14	0.13	0.14	0.14
US	0.48	0.33	0.28	0.29	0.14	0.24	0.41	0.27	0.25	0.27	0.12	0.20
prior mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
average	0.45	0.26	0.24	0.29	0.25	0.23	0.45	0.22	0.21	0.27	0.21	0.20
	Hierarchy on Inclusion Prob						Both					
Australia	0.40	0.03	0.04	0.05	0.03	0.03	0.46	0.02	0.03	0.03	0.02	0.02
Canada	0.14	0.05	0.04	0.03	0.03	0.03	0.12	0.03	0.02	0.02	0.02	0.02
France	0.15	0.03	0.04	0.06	0.03	0.03	0.14	0.02	0.03	0.03	0.02	0.02
Germany	0.23	0.04	0.04	0.05	0.03	0.05	0.24	0.03	0.03	0.03	0.02	0.04
Italy	0.15	0.03	0.04	0.04	0.03	0.03	0.14	0.02	0.03	0.03	0.02	0.02
Japan	0.32	0.05	0.06	0.10	0.07	0.04	0.33	0.03	0.04	0.07	0.05	0.03
Sweden	0.34	0.04	0.05	0.26	0.05	0.04	0.33	0.02	0.03	0.20	0.03	0.02
Switzerland	0.12	0.04	0.04	0.03	0.08	0.04	0.12	0.03	0.03	0.02	0.05	0.02
United Kingdom	1.00	0.03	0.03	0.03	0.03	0.03	0.98	0.02	0.02	0.02	0.02	0.02
United States	0.21	0.04	0.06	0.07	0.02	0.04	0.22	0.03	0.04	0.05	0.01	0.02
prior mean	0.34	0.11	0.12	0.14	0.12	0.11	0.34	0.10	0.11	0.13	0.10	0.10
average	0.30	0.04	0.04	0.07	0.04	0.04	0.31	0.03	0.03	0.05	0.03	0.02

Table 17: Bayesian Estimates of Inclusion Probabilities (Borrowing from Cross-Section)

This table reports the mean inclusion probability for each regressor. The numbers in bold denote instances in which the inclusion probability is higher than the prior. “Hierarchy on Mean” and “Hierarchy on Inclusion Prob” refers to specifications that impose either hierarchical prior on mean or inclusion probabilities. “No Hierarchy” refers to the basic specification without hierarchical prior beliefs as given in (2.2). “Both” refers to the specification with hierarchical prior both on the coefficient values and on the inclusion probabilities as given in (2.6). The last row reports the average inclusion probability for each regressor across countries. The set of predictors comprises the short-term interest rate tb , the yield spread (ys), the short-term interest rate relative to its 12-month moving average (rtb), the long-term government bond yield relative to its 12-month moving average (rbr), annual inflation rate ($infl$), and log dividend-price ratio dp . All predictors are in logs.

higher for the coefficients in front of the dividend-price ratio if the hierarchical prior on the coefficients is introduced. This is driven by the fact that the prior mean comes into play only if a particular coefficient is “far” from zero, which is rarely the case.

Why is the contribution of the hierarchical prior on the posterior coefficients so small? The key reason is that the common mean is important only if the coefficient is classified as “large”. Most predictive variables are rarely selected and thus do not depend on the common mean. Moreover, in order to make a precise inference about the common mean, multiple coefficients in front of this predictive variables should be considered as “large”. Thus, the data is not very informative about the posterior values of the common means. Another cause is, of course, the limited shrinkage toward the common mean. The amount of shrinkage is inversely proportional to the variance implied by the prior. As the variance τ^2 is “large” by construction of the spike-and-slab prior, the shrinkage towards the common mean is low.

Second, the hierarchical prior beliefs on the inclusion probabilities does significantly change the level selection probabilities. The average inclusion probabilities for most regressors decrease from around 20% to less than 5%, and the average inclusion probability for the dividend-price ratio is revised down from 45% to 30%. Moreover, the posterior inclusion probabilities are very sensitive to the prior beliefs: for example, the inclusion probability for the relative T-bill rate in Sweden declines from 80% to 25% when the hierarchical prior on the inclusion probabilities is introduced. This sensitivity highlights how uninformative the data is about the coefficient values and is consistent with earlier findings of Schrimpf (2010).

The only robust conclusion is that the dividend-price ratio can forecast future returns in the U.K. because its inclusion probability remains higher than 95%. The sensitivity towards the prior shows how uninformative the data is about whether the predictive variables can forecast returns. Many obtained coefficients are almost equally likely to be either “large” or “small”, so the prior belief about which group they belong to is the determining factor.

The conclusions about the relevance of particular predictor variables, however, are not sensitive to the change of prior beliefs. As before, the relevance of particular predictor variables is measured as posterior probability of inclusion exceeding the prior probability of inclusion. Based on this metric, most predictive variables are not informative about the future returns. There is not a single case when another predictive variable becomes relevant after the hierarchical prior is imposed. In particular, the dividend-price ratio appears to be informative about future returns in Australia, the U.K. and, potentially, Japan and Sweden. In addition, the relative T-bill rate appears to be an important prediction for the returns on the Swedish market index.

Overall, hierarchical priors contribute little. Consequently, the asset allocation with and without borrowing strength from the cross-section are essentially identical and the certainty equivalent losses from ignoring the borrowing strength from cross-section are, basically, zero. The certainty equivalent losses and asset allocation decision are not reported to save space.

Let me turn to the out-of-sample results now. The statistical measure of forecast performance relative to the naive benchmark out-of-sample R_{OOS}^2 is reported in Table 18, while the average asset allocation and the certainty equivalent risk-free returns are displayed in Tables 19 and 20.

Both economic and statistical measures show that adding hierarchical prior beliefs do not result in better out-of-sample performance. The R_{OOS}^2 does increase on average but the increase is tiny and statistically insignificant. For the majority of countries the forecasting model still falls short relative to the historical mean.

The asset allocation is essentially identical with and without hierarchical priors. Consequently, the key results are robust to the hierarchical priors. Using hierarchical priors does improve economic measures of the out-of-sample fit for some countries but the improvement is very small. In addition, it varies greatly across countries, so it is impossible to infer which type of hierarchical prior beliefs should be used. As before, investors in Sweden, the

	Recursive Scheme			Rolling Scheme		
	✓	✓	✓	✓	✓	✓
Hierarchy on Mean						
Hierarchy on Inclusion Prob						
Australia	-0.26	0.00	-0.08	0.08	-0.41	0.16
Canada	-0.59	-0.47	-0.23	-0.14	-0.53	-0.43
France	-0.44	-0.36	-0.16	-0.18	-0.33	-0.60
Germany	-0.35	-0.21	-0.06	-0.15	-0.79	-1.07
Italy	-0.66	-0.60	-0.26	-0.23	-0.17	-0.94
Japan	-0.52	-0.56	0.00	-0.08	0.21	-0.27
Sweden	0.40	0.49	0.20	0.18	1.30**	0.75
Switzerland	-0.24	-0.22	0.06	-0.09	-0.46	-1.08
UK	1.54**	1.61**	1.77**	1.54**	0.33	0.85*
US	0.12	0.17	0.18	0.18	0.09	1.19**
average	-0.10	-0.01	0.14	0.11	-0.08	-0.14

Table 18: R_{OOS}^2 (Borrowing From Cross-Section)

This table reports the out-of-sample R^2 in percents. The forecasts are generated either recursively with an expanding window or with a rolling window of 20 years from January 1997 through December 2016. The R_{OOS} is computed for the posterior mean. Clark and West test by Clark and West (2007) and Clark and West (2006) is used to compare the out-of-sample mean squared prediction errors produced by the models relative with those based on the historical mean under one-sided hypothesis. The stars flag levels of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

	Mean	Recursive Scheme		Rolling Scheme	
Hierarchy on Mean		✓	✓	✓	✓
Hierarchy on Incl Prob			✓		✓
Australia	0.33	0.32	0.31	0.31	0.19
Canada	0.35	0.46	0.39	0.39	0.33
France	0.36	0.34	0.35	0.35	0.23
Germany	0.35	0.34	0.35	0.35	0.15
Italy	0.16	0.20	0.19	0.19	0.06
Japan	0.29	0.28	0.27	0.30	0.08
Sweden	0.43	0.39	0.39	0.43	0.23
Switzerland	0.63	0.62	0.62	0.64	0.37
UK	0.50	0.33	0.33	0.32	0.18
US	0.61	0.57	0.57	0.57	0.39
average	0.40	0.39	0.38	0.38	0.22
					0.42
					0.41
					0.41

Table 19: Out-of-Sample Asset Allocation (Borrowing From Cross-Section)

Asset Allocation. This table reports average allocation to the market index for an investor with a power utility function with a coefficient of relative risk aversion $A = 5$. The forecasts are generated either recursively or with a rolling window from January 1997 through December 2016. The asset allocation is updated monthly.

	Mean	Recursive Scheme			Rolling Scheme				
Hierarchy on Mean		✓	✓	✓	✓	✓	✓		
Hierarchy on Incl Prob		✓	✓	✓	✓	✓	✓		
Australia	0.89	0.74	0.87	0.80	0.89	0.27	0.06	0.25	0.29
Canada	1.26	0.71	0.83	1.06	1.17	0.62	0.68	0.94	0.79
France	0.99	0.57	0.66	0.86	0.84	0.56	0.30	0.49	0.55
Germany	0.65	0.11	0.34	0.56	0.41	0.04	-0.73	-0.57	-0.57
Italy	0.12	-0.43	-0.38	-0.09	-0.07	0.23	-0.60	-0.29	-0.33
Japan	-0.26	-1.06	-1.08	-0.38	-0.47	-0.26	-0.44	-0.30	-0.54
Sweden	1.41	1.95***	2.06***	1.67***	1.65***	2.70***	2.61***	2.23***	2.26***
Switzerland	1.05	0.41	0.46	0.90	0.69	0.19	-0.48	-0.17	-0.11
UK	0.39	1.08***	1.14***	1.14***	1.00***	0.34	0.49	0.35	0.43
US	0.91	1.06**	1.10***	1.06***	1.09***	1.22**	2.50***	1.72***	1.77***
average	0.74	0.52	0.60	0.76	0.72	0.59	0.44	0.46	0.45

Table 20: Out-of-Sample Certainty Equivalent Returns (Borrowing From Cross-Section)

This table reports the CERs for an investor with a power utility function with a coefficient of relative risk aversion $A = 5$. The CERs are annualized and reported in percents. The forecasts are generated either recursively with an expanding window from January 1997 through December 2016. The asset allocation is updated monthly. Delta method as in Shanken and Tamayo (2012) is used to test if the excess CER is bigger than zero. The stars flag levels of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

U.K. and the U.S. can increase their risk-free return relative to the naive mean. However, for other countries an investor would still benefit most from using the historical mean and standard deviation to calculate the optimal allocation. As before, even for these countries the episodes of predictability are very short-lived and unstable.

2.6.2. Stochastic Volatility

Returns do not have a constant variance as was previously assumed in this chapter. Instead, they exhibit volatility clustering. Taking into account the volatility dynamics is essential for asset allocation decisions that depends on the perceived predicted variance of the return. Moreover, it may yield more precise estimates of the coefficients and the expected risk premium. Recall that the precision of the Bayesian estimates of the coefficient values β depends on the precision observed in the data. Intuitively, observations related to the periods of low volatility are more informative about the value of the expected risk premium and, therefore, should receive higher weight when forming a posterior estimates. Conversely, observations from high volatility periods should receive lower weight. The volatility dynamics, however, has not been studied in the context of predictive regressions yet. Related literature is limited to a paper by Johansson (2009) who shows that incorporating SV can improve the in-sample forecasts of the capital asset pricing model.

In this subsection, I consider using mean factor stochastic volatility model (SV) to the model in (2.1) and reconsider the in- and out-of-sample results. The key findings reported in the previous sections, however, do not change after the SV is incorporated.

Updated Econometric Framework

I use a variation of the mean factor model by Jacquier et al. (1999), Pitt and Shephard (1999), and Kim et al. (1998). In this model, the innovation is decomposed into two additive components: an idiosyncratic component and an univariate latent factor f . The predictive

regression in (2.1) is modified as follows

$$r_{n,t+1} = \alpha_n + b_n f_t + x'_{n,t} \beta_n + u_{n,t+1}, \quad n = 1, \dots, N; \quad t = 1, \dots, T. \quad (2.7)$$

To ensure that the model is identified I set the loading $b_{US} = 1$. The factor f follows a univariate SV process:

$$\begin{aligned} f_t &= \exp\{h_t/2\}\epsilon_t, \quad t = 1, \dots, T \\ h_{t+1} &= \mu_h + \phi(h_t - \mu) + \eta_t \\ h_1 &\sim \mathcal{N}\left(\mu_h, \frac{\sigma_\eta^2}{1 - \phi^2}\right) \\ \eta_t &\sim \mathcal{N}(0, \sigma_\eta^2) \\ \epsilon_t &\sim \mathcal{N}(0, 1) \end{aligned}$$

I now introduce assumptions regarding the correlation structure. The factor innovations ϵ_t , the volatility innovations η_t , and the innovations errors u_t are all independent of each other. In this specifications the correlation between errors in two countries is driven by their loadings on the common factor f . Therefore, to avoid identification issues I assume that the covariance matrix of the country-specific innovations u_t Σ is diagonal with $\Sigma_{n,n} = \sigma_n^2$.

In- and Out-of-Sample Analysis

In this subsection, I provide more details on the in-sample and out-of-sample results for the stochastic volatility model in (2.7).

2.6.3. In- and Out-of-Sample Results

The posterior inclusion probabilities are reported in Table 21, while the box plots depicting the coefficient values are displayed in Figure 5. Several conclusions emerge.

Incorporating the SV does lead to some changes in the posterior beliefs about inclusion

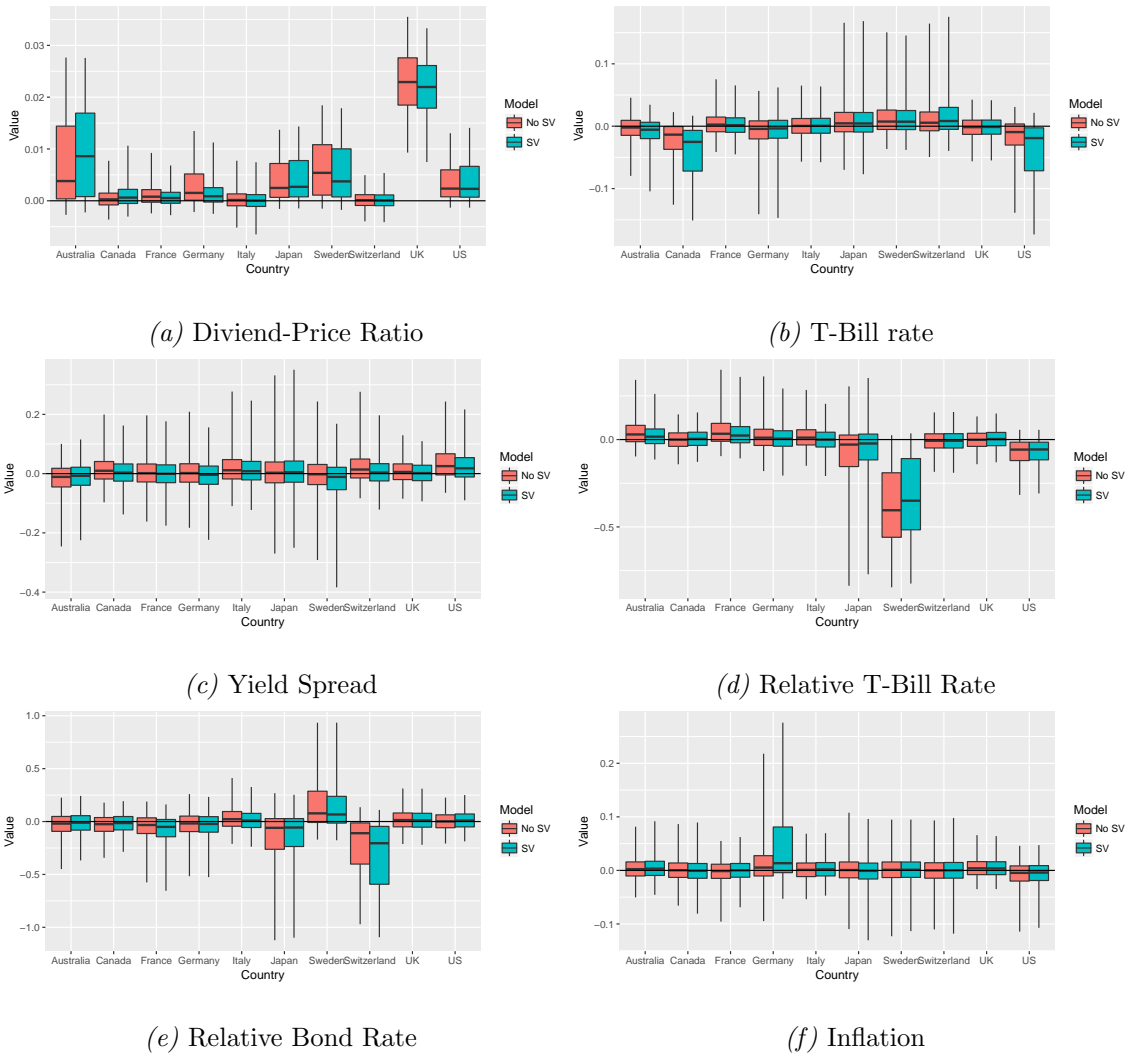


Figure 5: Posterior Distribution of Slope Coefficients (Stochastic Volatility)

This figure displays posterior beliefs for slope coefficients. Each figure compares box plots of posterior distributions for coefficients in front of a corresponding regressor across countries and models. For each country and each regressor, there are two box plots. “SV” refers to specifications that includes stochastic volatility as in (2.7). “No SV” refers to the basic specifications without stochastic volatility as in (2.1). The center line of each box indicates the median of the distribution, the box and the vertical lines correspond to the 75% and 95% credible sets. The set of predictors comprises the short-term interest rate tb , the yield spread (ys), the short-term interest rate relative to its 12-month moving average (rtb), the long-term government bond yield relative to its 12-month moving average (rbr), annual inflation rate ($infl$), and log dividend-price ratio dp . All predictors are in logs.

probabilities. Most changes, however, do not exceed 10%.

In particular, the posterior inclusion probabilities for the dividend-price ratio are revised downwards relative to the specification without the SV. Most notably, the inclusion probability for Sweden declines from 68% to 54%. However, even after the downward revision of inclusion probabilities, the dividend-price ratio still appears to be the most important predictive variable. The coefficient values as displayed in Figure 5 also largely remain the same with and without SV.

There are two notable changes. First, in the U.S. the inclusion probabilities for the dividend-price ratio and the T-bill rate are revised upwards, so that the resulting probability exceeds the prior. Second, in Switzerland the inclusion probability for the relative bond rate increases from 43% to 54%. These changes are reflected by the increased values of the corresponding coefficients.

Overall, the changes are associated with transition from SUR to country-specific regression. They are caused by the fact that the correlation structure is less complex under the SV model because it is determined solely by the factor loadings. For example, recall from Table 11 that without SUR the dividend-price ratio does not appear to be important for Sweden.

I now proceed to the out-of-sample results. The statistical measure of forecast performance relative to the naive benchmark out-of-sample R^2_{OOS} is reported in Table 22, while the average asset allocation and the certainty equivalent risk-free returns are displayed in Tables 23 and 24.

Both economic and statistical measures show that incorporating SV leads to worse performance. The OOS R^2 declines almost uniformly across countries. The poor forecasting properties of the posterior mean are probably explained by the estimation noise as the model with the SV has more parameters as it includes latent volatilities and factor values for each time period.

	No SV						SV					
	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>	<i>dp</i>	<i>tb</i>	<i>ys</i>	<i>rtb</i>	<i>rbr</i>	<i>infl</i>
Australia	0.62	0.21	0.24	0.24	0.20	0.20	0.60	0.23	0.22	0.20	0.19	0.21
Canada	0.23	0.33	0.21	0.16	0.17	0.22	0.31	0.49	0.21	0.16	0.16	0.24
France	0.23	0.19	0.22	0.26	0.22	0.20	0.20	0.19	0.23	0.23	0.26	0.19
Germany	0.38	0.28	0.23	0.23	0.21	0.34	0.33	0.28	0.24	0.22	0.22	0.43
Italy	0.23	0.20	0.25	0.20	0.19	0.19	0.21	0.20	0.24	0.20	0.18	0.19
Japan	0.56	0.31	0.30	0.39	0.37	0.26	0.57	0.30	0.30	0.37	0.36	0.27
Sweden	0.62	0.29	0.27	0.81	0.36	0.25	0.60	0.30	0.30	0.76	0.35	0.26
Switzerland	0.18	0.29	0.24	0.17	0.43	0.25	0.20	0.33	0.21	0.18	0.54	0.25
UK	0.99	0.18	0.17	0.15	0.17	0.17	1.00	0.18	0.15	0.15	0.17	0.16
US	0.48	0.33	0.28	0.29	0.14	0.24	0.60	0.51	0.23	0.29	0.14	0.25
prior mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
average	0.45	0.26	0.24	0.29	0.25	0.23	0.46	0.30	0.23	0.27	0.26	0.25

Table 21: Bayesian Estimates of Inclusion Probabilities (Stochastic Volatility)

This table reports the mean inclusion probability for each regressor under the model specification with stochastic volatility. “SV” refers to specifications that includes stochastic volatility as in (2.7). “No SV” refers to the basic specifications without stochastic volatility as in (2.1). The numbers in bold denote instances in which the inclusion probability is higher than the prior. The last row reports the average inclusion probability for each regressor across countries. The set of predictors comprises the short-term interest rate tb , the yield spread (ys), the short-term interest rate relative to its 12-month moving average (rtb), the long-term government bond yield relative to its 12-month moving average (rbr), annual inflation rate ($infl$), and log dividend-price ratio dp . All predictors are in logs.

The asset allocations decisions are sensitive to the predicted variance, so incorporating SV is expected to influence the asset allocation decisions. In particular, investors who incorporates SUR tend to allocate more of his funds in equity, timing the market more aggressively during the low volatility periods. Consequently, the allocation also tends to vary more across time periods. However, incorporating SV typically results in CER losses relative to the investors who ignore it. These utility losses are mostly driven by the periods of low returns that are usually associated with high volatility episodes. The investors who incorporate SV are usually more heavily invested in equity, so they are hit harder by the market crashes.

2.7. Conclusion

An international investor faces a high degree of uncertainty when deciding which predictive variables to include in the regression to forecast the equity premium. Ignoring model uncer-

	Recursive Scheme		Rolling Scheme	
	Ignore SV	SV	Ignore SV	SV
Australia	-0.26	-0.54	-0.41	-0.71
Canada	-0.59	-0.23	-0.53	-0.11
France	-0.44	-0.87	-0.33	-1.15
Germany	-0.35	-0.23	-0.79	-1.13
Italy	-0.66	-1.76	-0.17	-1.39
Japan	-0.52	-0.54	0.21	-0.44
Sweden	0.40	0.05	1.30**	0.07
Switzerland	-0.24	-0.88	-0.46	-1.51
UK	1.54**	1.02	0.33	0.47
US	0.12	0.15	0.09	1.64**
average	-0.10	-0.46	-0.08	-0.42

Table 22: R_{OOS}^2 (Stochastic Volatility)

This table reports the out-of-sample R^2 in basis points. The forecasts are generated either recursively with an expanding window of with a rolling window of 20 years from January 1997 to December 2016. The R_{OOS} is computed for the posterior mean. R_{OOS}^2 is reported in percentage (multiplied by 100). The stars flag levels of significance for the Clark and West test by Clark and West (2007) and Clark and West (2006). If a p-value is less than 0.01 it is flagged with three stars (***). If a p-value is less than 0.05 it is flagged with two stars (**). If a p-value is less than 0.01 it is flagged with three stars (***).

tainty may lead investors to make poor asset allocation decisions. This chapter examines the portfolio choices of a Bayesian investor who accounts for model uncertainty.

I find that incorporating model uncertainty leads to more conservative asset allocation, which is less sensitive to predictive variables. The resulting optimal portfolio weights deliver superior OOS performance when compared with models that ignore model uncertainty. The improvement of the portfolio weights is driven both by the better estimate of the expected equity premium and its variance. However, in the majority of countries, an investor does not perceive utility gains relative to the naive forecast based on the sample means. More broadly, the findings in this chapter suggest that the international comparisons are essential to understand the nature of the equity premium. In particular, I find that the U.S. is not representative of other countries as they are characterized by a strong predictability not observed for other capital markets.

	Mean	Recursive Scheme		Rolling Scheme	
		Ignore SV	SV	Ignore SV	SV
Australia	0.33	0.33	0.83	0.19	0.88
Canada	0.35	0.46	0.95	0.33	0.97
France	0.36	0.34	0.89	0.23	0.97
Germany	0.35	0.34	0.97	0.15	1.00
Italy	0.16	0.21	0.35	0.06	0.51
Japan	0.29	0.28	0.56	0.08	0.51
Sweden	0.43	0.39	0.74	0.23	0.88
Switzerland	0.63	0.62	0.99	0.37	1.00
UK	0.50	0.33	0.67	0.18	0.80
US	0.61	0.57	1.00	0.39	0.99
average	0.40	0.39	0.79	0.22	0.85

Table 23: Out-of-Sample Asset Allocation (Stochastic Volatility)

This table reports average allocation to the market index for an investor with a power utility function with a coefficient of relative risk aversion 5. The forecasts are generated either recursively or with a rolling window from January 1997 to December 2016. The asset allocation is updated monthly.

	Mean	Recursive Scheme		Rolling Scheme	
		Ignore SV	SV	Ignore SV	SV
Australia	0.89	0.74	0.79	0.27	0.20
Canada	1.26	0.71	0.29	0.62	1.93**
France	0.99	0.57	-0.71	0.56	-3.03
Germany	0.65	0.11	-7.39	0.04	-5.78
Italy	0.12	-0.43	1.85***	0.23	-3.50
Japan	-0.26	-1.06	-3.74	-0.26	-1.22
Sweden	1.41	1.95***	-2.70	2.70***	-0.91
Switzerland	1.05	0.41	-5.10	0.19	0.03
UK	0.39	1.08***	-3.07	0.34	0.67
US	0.91	1.06**	0.97	1.22**	0.13
average	0.74	0.52	-1.88	0.59	-1.15

Table 24: Out-of-Sample Certainty Equivalent Returns (Stochastic Volatility)

This table reports the CERs for an investor with a power utility function with a coefficient of relative risk aversion 5. The CERs are annualized and reported in basis points. The forecasts are generated recursively with an expanding window from January 1997 to December 2016. The asset allocation is updated monthly. The stars flag levels of significance. If a p-value is less than 0.01 it is flagged with three stars (***). If a p-value is less than 0.05 it is flagged with two stars (**). If a p-value is less than 0.10 it is flagged with one star (*). The p-values are calculated based on the delta method as in Shanken and Tamayo (2012).

This chapter suggests several avenues for future research. For example, the framework developed here can be extended to study the predictors of asset returns in other panels. It can be applied to the cross-section of the U.S. stocks to determine the most valuable predictors of returns.

2.8. Appendix

2.8.1. Estimation Procedure

Likelihood

It is helpful to-rewrite the system of equations (2.1) in stacked form

$$\begin{bmatrix} r_{1,t+1} \\ r_{2,t+1} \\ \vdots \\ r_{N,t+1} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} x'_{1,t} & 0 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & x'_{N,t} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} u_{1,t+1} \\ \vdots \\ u_{N,t+1} \end{bmatrix}$$

or

$$r_{t+1} = \alpha + X_t \beta + u_{t+1}, \quad t = 1, \dots, T; \quad u_{t+1} \sim \mathcal{N}_N(0, \Omega^{-1}). \quad (2.8)$$

Let $D = \{r_1, \dots, r_T\}$ denote all available data on returns. The conditional likelihood of the model (2.8) is

$$p(D|\beta, \Omega, X) \propto |\Omega|^{T/2} \exp \left[-\frac{1}{2} \sum_{t=1}^T (r_t - X_t \beta)' \Omega (r_t - X_t \beta) \right]. \quad (2.9)$$

Prior Beliefs

The prior beliefs on the coefficients (2.2) can be conveniently re-written in the matrix form

$$\beta|\gamma, v_0, \tau^2 \sim \mathcal{N}(0, B_0(\gamma, \tau^2, v_0)),$$

where $B_0(\gamma, \tau^2, v_0)$ is an $NK \times NK$ diagonal matrix with each diagonal element corresponding to the prior variance: $\{B_0\}_{i,i} = (v_0 + \gamma_{n,k} * (1 - v_0)) \tau_k^2$. Thus, the variance of

the coefficient $\beta_{n,k}$ is $v_0\tau_k^2$ if it is close to zero ($\gamma_{n,k} = 0$) and τ_k^2 otherwise.

I follow Ishwaran et al. (2005) to set shape and scale parameters for the prior distribution of variance τ_k^2 and small variance $v_0 = 0.005$. Ishwaran et al. (2005) use a shape parameter 5 and a scale parameter 50 when working with standardized and centered regressors. I adjust this prior, so that it corresponds to the scale of regressors and returns. In particular, I use the training sample from the period 1971-1974 consisting of data for eight countries: Australia, Canada, France, Japan, Germany, Sweden, the U.K. and the U.S. I find the average standard deviation for each regressor and returns and obtain the new scale parameter:

$$a_{0,k} = 5 \tag{2.10}$$

$$b_{0,k} = 50 \times \left(\frac{1}{\sqrt{T}} \frac{SD(\text{returns})}{SD(\text{regressor } k)} \right)^2 \tag{2.11}$$

In my training sample $SD(\text{returns}) = 0.06$ and $SD(\text{regressor}) = (0.4, 0.035, 0.015, 0.010, 0.006, 0.033)$.

I use the full length of the sample, so $\sqrt{T} = \sqrt{503}$. When working with a rolling window forecast I use $\sqrt{T} = \sqrt{240}$ that corresponds to 20 years of monthly data.

Posterior and Gibbs Sampler

The prior beliefs can be written as a prior density $p(\Psi)$, where $\Psi = [\beta, \Omega, \tau, \gamma, \omega]$ includes all parameters to be estimated. The posterior density of the parameters is calculated following the Bayes rule:

$$p(\Psi|D, X) \propto p(D|\beta, \Omega, X)p(\Psi)$$

The Gibbs sampler consists of five blocks. I generate 150,000 draws, the first 50,000 draws are discarded and the remaining 100,000 draws are used for the inference. I evaluate sample convergence using the visual inspection and the test by Geweke et al. (1991) that compares mean in the first and in the last part of the sample. I initialize the sampler with the OLS estimates based on single-country regressions.

Below I detail the conditional posteriors for various parameters and latent variables. At each step it is implicit that I condition on fixed values of other variables based on the previous simulation.

0. Set starting values

Select a starting values for the covariance matrix $\Sigma^{(i-1)}$, vector of variances around the common mean $\tau^{2(i-1)}$, and vector of selection indicators $\gamma^{(i-1)}$.

1. **Sampling β**

Construct a prior variance $B_0^{(i-1)}$ using $\gamma^{(i-1)}$ and $\tau^{2(i-1)}$. Draw $\beta^{(i)} \sim \mathcal{N}(\bar{\beta}^{(i)}, B^{(i)})$ where

$$B^{(i)} = \left[B_0^{-1(i-1)} + \sum_{t=1}^T X_t' \Sigma^{-1(i-1)} X_t \right]^{-1}$$

$$\bar{\beta}^{(i)} = B^{(i)} \left[\sum_{t=1}^T X_t' \Sigma^{-1(i-1)} (r_t - \alpha^{(i-1)}) \right]$$

2. **Sampling α**

Draw $\alpha^{(i)} \sim \mathcal{N}(\bar{\alpha}^{(i)}, V^{(i)})$ where

$$V^{(i)} = \frac{1}{T} \Sigma^{(i-1)}$$

$$\bar{\alpha}^{(i)} = V^{(i)} \left[\sum_{t=1}^T i_N' \Sigma^{-1(i-1)} (r_t - X_t \beta^{(i)}) \right],$$

where i_N is a unit vector of length N .

3. **Sampling Σ**

The inverse of the covariance matrix follows a Wishart distribution. In particular, $\Sigma^{-1(i)} \sim \mathcal{W}_N(\rho_T, R_T^{(i)})$ where

$$R_T^{(i)} = \left[\sum_{t=1}^T (y_t - \alpha^{(i)} - X_t \beta^{(i)}) (y_t - \alpha^{(i)} - X_t \beta^{(i)})' \right]^{-1}$$

$$\rho_T = T$$

4. **Sampling τ**

Draw $\tau_k^{-2(i)} \stackrel{ind}{\sim} \text{Gamma}(a_k^{(i)}, b_k^{(i)})$ where

$$\begin{aligned} a_k^{(i)} &= a_{0,k} + \frac{1}{2} \\ b_k^{(i)} &= b_{0,k} + \sum_{k=1}^K \frac{1}{2} \frac{\beta_{n,k}^{2(i)}}{v_0 + \gamma_{n,k}(1 - v_0)} \end{aligned}$$

5. **Sampling γ**

Draw $\gamma_{n,k}^{(i)} \stackrel{ind}{\sim} \text{Bernoulli}\left(\frac{\omega_{1,k}^{(i)}}{\omega_{1,k}^{(i)} + \omega_{2,k}^{(i)}}\right)$ where

$$\begin{aligned} \omega_{1,k}^{(i)} &= \pi_{n,k} \exp\left\{-\frac{\beta_{n,k}^{2(i)}}{2\tau_k^{2(i)}}\right\} \\ \omega_{1,k}^{(i)} &= (1 - \pi_{n,k})v_0^{-1/2} \exp\left\{-\frac{\beta_{n,k}^{2(i)}}{2v_0\tau_k^{2(i)}}\right\} \end{aligned}$$

6. This completes one iteration of the sampler. Repeat steps 1-5 I times

7. Discard the first B draws for burn in.

2.8.2. *Estimation Procedure when Borrowing Strength from Cross-Section*

It is helpful to introduce some notation. Let H be a $K \times NK$ matrix that reflects the hierarchical structure and consists of N $K \times K$ stacked identity matrices I_K : $H = [I_K, \dots, I_K]'$. The matrix maps the $K \times 1$ vector of common means μ into an $NK \times 1$ vector of coefficients β . A matrix $G = \text{diag}(\gamma_{n,k})$ is an $NK \times NK$ diagonal matrix with the selection vector γ on the main diagonal. The matrix is used to select coefficients from the coefficient vector β that are far from zero.

The hierarchical beliefs on the coefficients (2.6) can be conveniently re-written in the matrix

form as follows:

$$\beta|\gamma, v_0, \tau^2, \mu \sim \mathcal{N}(\beta_0(\gamma, \mu), B_0(\gamma, \tau^2, v_0)),$$

where $\beta_0(\gamma, \mu) = GH\mu$ is the vector of prior mean, so that the mean for a particular $\beta_{n,k}$ is μ_k if $\gamma_{n,k} = 1$ and zero otherwise. As before, $B_0(\gamma, \tau^2, v_0)$ is a $NK \times NK$ diagonal prior covariance matrix with each diagonal element corresponding to the prior variance of the coefficient $\beta_{n,k}$: $(v_0 + \gamma_{n,k} * (1 - v_0)) \tau_k^2$.

The parameters $a_{0,k}$, $b_{0,k}$ and v_0 are set as described in the appendix 2.8.1. I use an informative albeit diffuse prior on the common mean μ . An informative prior is required in order to update draws when none of the relevant variables is selected, e.g., all $\gamma_{n,k} = 0$ for a certain k . I set $\mu_{0,k} = 0$ for all k . The prior variance M_0 is a diagonal matrix with each diagonal element $M_{0,\{k,k\}} = \frac{b_{0,k}}{a_{0,k}-1}$.

I now introduce the Gibbs sampler. The procedure to sample intercepts α , covariance matrix Σ and the vector of large variances τ^2 are the same as before without the hierarchical prior. The conditional posterior distribution for the coefficient vector β differs slightly because the prior mean is equal to μ rather than zero if the coefficients that are “large”. Similarly, sampling the vector of inclusion indicators γ is adjusted, so that the prior mean is the draw of the average inclusion probability ω rather than the set of fixed prior means. Finally, two more steps are added to the sampler: sampling the vector of common means μ and the vector of the common inclusion probabilities ω .

0. Set starting values

Select a starting values for the covariance matrix $\Sigma^{(i-1)}$, vector of common means $\mu^{(i-1)}$, vector of variances around the common mean $\tau^{2(i-1)}$, and vector of selection indicators $\gamma^{(i-1)}$.

1. **Sampling** β

Construct a prior variance $B_0^{(i-1)}$ and prior mean $\beta_0^{(i-1)}$ using $\mu^{(i-1)}$, $\gamma^{(i-1)}$ and $\tau^{2(i-1)}$.

Draw $\beta^{(i)} \sim \mathcal{N}(\bar{\beta}^{(i)}, B^{(i)})$ where

$$B^{(i)} = \left[B_0^{-1(i-1)} + \sum_{t=1}^T X_t' \Omega^{(i-1)} X_t \right]^{-1}$$

$$\bar{\beta}^{(i)} = B^{(i)} \left[B_0^{-1} \beta_0 + \sum_{t=1}^T X_t' \Omega^{(i-1)} (r_t - \alpha^{(i-1)}) \right]$$

2. Sampling α

Draw $\alpha^{(i)} \sim \mathcal{N}(\bar{\alpha}^{(i)}, V^{(i)})$ where

$$V^{(i)} = \frac{1}{T} \Sigma^{(i-1)}$$

$$\bar{\alpha}^{(i)} = V^{(i)} \left[\sum_{t=1}^T i_N' \Sigma^{-1(i-1)} (r_t - X_t \beta^{(i)}) \right],$$

3. Sampling Ω

The inverse of the covariance matrix follows a Wishart distribution. In particular,

$\Sigma^{-1(i)} \sim \mathcal{W}_N(\rho_T, R_T^{(i)})$ where

$$R_T^{(i)} = \left[\sum_{t=1}^T (y_t - \alpha^{(i)} - X_t \beta^{(i)}) (y_t - \alpha^{(i)} - X_t \beta^{(i)})' \right]^{-1}$$

$$\rho_T = T$$

4. Sampling μ

Note that only selected coefficients should be used to sample μ . Draw $\mu^{(i)} \sim \mathcal{N}(\bar{\mu}^{(i)}, M^{(i)})$

where

$$M^{(i)} = \left[M_0^{-1} + G^{(i-1)'} H' B^{-1(i-1)} H G^{(i-1)} \right]^{-1}$$

$$\bar{\mu}^{(i)} = M^{(i)} \left[M_0^{-1} \mu_0 + G^{(i-1)'} H' B^{-1(i-1)} \beta^{(i-1)} H G^{(i-1)} \right]$$

5. Sampling τ

Draw $\tau_k^{-2(i)} \stackrel{ind}{\sim} \text{Gamma}(a_k^{(i)}, b_k^{(i)})$ where

$$a_k^{(i)} = a_{0,k} + \frac{1}{2}$$

$$b_k^{(i)} = b_{0,k} + \sum_{k=1}^K \frac{1}{2} \frac{(\beta_{n,k}^2 - \gamma_{n,k} \mu_k)^2}{v_0 + (1 - \gamma_{n,k})}$$

6. Sampling γ

Draw $\gamma_{n,k}^{(i)} \stackrel{ind}{\sim} \text{Bernoulli}\left(\frac{\omega_{1,k}^{(i)}}{\omega_{1,k}^{(i)} + \omega_{2,k}^{(i)}}\right)$ where

$$\omega_{1,k}^{(i)} = \omega_k \exp\left\{-\frac{(\beta_{n,k}^{(i)} - \mu_k^{(i)})^2}{2\tau_k^{2(i)}}\right\}$$

$$\omega_{1,k}^{(i)} = (1 - \omega_k)v_0^{-1/2} \exp\left\{-\frac{(\beta_{n,k}^{(i)})^2}{2v_0\tau_k^{2(i-1)}}\right\}$$

7. Sampling ω

Draw $\omega_k^{(i)} \stackrel{ind}{\sim} \text{Beta}(1 + \#\{n : \gamma_{n,k} = 1\}, 1 + \#\{n : \gamma_{n,k} = 0\})$.

8. This completes one iteration of the sampler. Repeat steps 1-7 I times

9. Discard the first B draws for burn in.

2.8.3. Estimation Procedure under Stochastic Volatility

The model in (2.7) is completed with priors in 2.2, and priors on the SV process. The priors on the SV process are the same as in Kim et al. (1998). Let $\phi = 2\phi^* - 1$, then

$$\sigma_\eta^{-2} \sim \text{Gamma}(a_{0,\eta}, b_{0,\eta})$$

$$\phi^* \sim \text{Beta}(\phi_{0,1}, \phi_{0,2}) \mid |\phi| < 1$$

I use $a_{0,\eta} = 2.5$, $b_{0,\eta} = 0.025$, $\phi_{0,1} = 20$, $\phi_{0,2} = 1.5$.

I impose uninformative priors on the intercepts α , variances σ_n^2 and the mean of the volatility process μ_h .

I now introduce the Gibbs sampler.

Below I detail the conditional posteriors for various parameters and latent variables. At each step it is implicit that I condition on fixed values of other variables based on the previous simulation.

0. Set starting values

Select a starting values for the covariance matrix $\Sigma^{(i-1)}$, vector of variances around the common mean $\tau^{2(i-1)}$, and vector of selection indicators $\gamma^{(i-1)}$.

1. **Sampling β**

Construct a prior variance $B_0^{(i-1)}$ using $\gamma^{(i-1)}$ and $\tau^{2(i-1)}$. Draw $\beta^{(i)} \sim \mathcal{N}(\bar{\beta}^{(i)}, B^{(i)})$ where

$$B^{(i)} = \left[B_0^{-1(i-1)} + \sum_{t=1}^T X_t' \Sigma^{-1(i-1)} X_t \right]^{-1}$$

$$\bar{\beta}^{(i)} = B^{(i)} \left[\sum_{t=1}^T X_t' \Sigma^{-1(i-1)} \left(r_t - \alpha^{(i-1)} - f_t^{(i-1)'} b^{(i-1)} \right) \right]$$

2. **Sampling α**

Draw $\alpha^{(i)} \sim \mathcal{N}(\bar{\alpha}^{(i)}, V^{(i)})$ where

$$V^{(i)} = \frac{1}{T} \Sigma^{(i-1)}$$

$$\bar{\alpha}^{(i)} = V^{(i)} \left[\sum_{t=1}^T i_N' \Sigma^{-1(i-1)} \left(r_t - X_t \beta^{(i)} - f_t^{(i-1)'} b^{(i-1)} \right) \right],$$

where i_N is a unit vector of length N .

3. **Sampling Σ**

The covariance matrix Σ is diagonal with diagonal elements following inverse Gamma distribution. They can be sampled independently of each other. In particular, $\sigma_n^{-2(i)} \sim \text{Gamma}\left(a_{1,n}, b_{1,n}^{(i)}\right)$ where

$$\begin{aligned} a_{1,\eta} &= a_{0,n} + \frac{T}{2} \\ b_{1,\eta}^{(i)} &= b_{0,n} + \frac{1}{2} \sum_{t=1}^T \left(r_{n,t} - \alpha_n^{(i)} - x'_{n,t} \beta_n^{(i)} - b_n^{(i-1)} f_t^{(i-1)} \right)^2 \end{aligned}$$

4. Sampling τ

Draw $\tau_k^{-2(i)} \stackrel{\text{ind}}{\sim} \text{Gamma}\left(a_k^{(i)}, b_k^{(i)}\right)$ where

$$\begin{aligned} a_k^{(i)} &= a_{0,k} + \frac{1}{2} \\ b_k^{(i)} &= b_{0,k} + \sum_{k=1}^K \frac{1}{2} \frac{\beta_{n,k}^{2(i)}}{v_0 + \gamma_{n,k}(1 - v_0)} \end{aligned}$$

5. Sampling γ

Draw $\gamma_{n,k}^{(i)} \stackrel{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{\omega_{1,k}^{(i)}}{\omega_{1,k}^{(i)} + \omega_{2,k}^{(i)}}\right)$ where

$$\begin{aligned} \omega_{1,k}^{(i)} &= \pi_{n,k} \exp\left\{-\frac{\beta_{n,k}^{2(i)}}{2\tau_k^{2(i)}}\right\} \\ \omega_{1,k}^{(i)} &= (1 - \pi_{n,k}) v_0^{-1/2} \exp\left\{-\frac{\beta_{n,k}^{2(i)}}{2v_0\tau_k^{2(i)}}\right\} \end{aligned}$$

6. Sampling f

Draw $f_t \stackrel{ind}{\sim} \mathcal{N}(\bar{f}_t, \sigma_{f,t}^2)$ where

$$\begin{aligned}\bar{f}_t &= \sigma_{f,t}^2 \left[\sum_{t=1}^T b^{(i-1)'\Sigma^{-1(i)}} \left(r_t - \alpha^{(i-1)} - X_t \beta^{(i)} \right) \right] \\ \sigma_{f,t}^2 &= \left[b^{(i-1)'\Sigma^{-1(i)}} b^{(i-1)} + \exp\{h_t^{(i-1)}\} \right]^{-1}\end{aligned}$$

7. Sampling b

Let $b = [b_{-US}, b_{US}]$. Similarly, define $X_{t,-US}$, $r_{t,-US}$, α_{-US} , β_{-US} and Σ_{-US} . Draw $b_{-US}^{(i)} \sim \mathcal{N}(\bar{b}^{(i)}, V^{(i)})$ where

$$\begin{aligned}V^{(i)} &= \left[\sum_{t=1}^T X'_{t,-US} \Sigma_{-US}^{-1(i-1)} f_t \right]^{-1} \\ \bar{\beta}^{(i)} &= B^{(i)} \left[\sum_{t=1}^T f_t \Sigma_{-US}^{-1(i-1)} \left(r_{t,-US} - \alpha_{-US}^{(i-1)} - X_{t,-US} \beta_{-US}^{(i)} \right) \right]\end{aligned}$$

8. **Sampling** $\mu_h, \sigma_\eta^2, h, \phi$ as in Kim (1994)

9. This completes one iteration of the sampler. Repeat steps 1-8 I times

10. Discard the first B draws for burn in.

BIBLIOGRAPHY

- F. AitSahlia, T. Doellman, and S. Sardarli. Mean-variance spanning tests: The fiduciary case in 401 (k) plans. 2017.
- F. Allen, S. Morris, and A. Postlewaite. Finite bubbles with short sale constraints and asymmetric information. *Journal of Economic Theory*, 61(2):206–229, 1993.
- F. Allen, D. Gale, et al. *Comparative financial systems: a survey*. Citeseer, 2001.
- D. W. Andrews and P. J. Barwick. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, 80(6):2805–2826, 2012.
- D. W. Andrews and P. Guggenberger. Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. *Journal of Econometrics*, 152(1):19–27, 2009.
- D. W. Andrews and G. Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.
- A. Ang and G. Bekaert. International asset allocation with regime shifts. *Review of Financial studies*, 15(4):1137–1187, 2002.
- A. Ang and G. Bekaert. Stock return predictability: Is it there? *The Review of Financial Studies*, 20(3):651–707, 2006.
- A. Ang, R. J. Hodrick, Y. Xing, and X. Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.
- D. Avramov. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458, 2002.
- N. Barberis. Investing for the long run when returns are predictable. *The Journal of Finance*, 55(1):225–264, 2000.
- A. Beber and M. Pagano. Short-selling bans around the world: Evidence from the 2007–09 crisis. *The Journal of Finance*, 68(1):343–381, 2013.
- V. Binsbergen, H. Jules, and R. S. Koijen. Predictive regressions: A present-value approach. *The Journal of Finance*, 65(4):1439–1471, 2010.
- P. Bossaerts and P. Hillion. Implementing statistical criteria to select return forecasting models: what do we learn? *The Review of Financial Studies*, 12(2):405–428, 1999.
- M. W. Brandt and K. Q. Wang. Time-varying risk aversion and unexpected inflation. *Journal of Monetary Economics*, 50(7):1457–1498, 2003.

- J. Y. Campbell. Stock returns and the term structure. *Journal of financial economics*, 18 (2):373–399, 1987.
- J. Y. Campbell and R. J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228, 1988.
- J. Y. Campbell and S. B. Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531, 2007.
- J. Y. Campbell and T. Vuolteenaho. Inflation illusion and stock prices. *American Economic Review*, 94(2):19–23, 2004.
- J. Y. Campbell and M. Yogo. Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60, 2006.
- I. A. Canay. El inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics*, 156(2):408–425, 2010.
- M. M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52 (1):57–82, 1997.
- J. Chen, H. Hong, and J. C. Stein. Breadth of ownership and stock returns. *Journal of financial Economics*, 66(2):171–205, 2002.
- S.-S. Chen. Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, 33(2):211–223, 2009.
- S. Chib and E. Greenberg. Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2):183–206, 1994.
- T. E. Clark and K. D. West. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1-2):155–186, 2006.
- T. E. Clark and K. D. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, 138(1):291–311, 2007.
- J. H. Cochrane. The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575, 2008.
- J. H. Cochrane. Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108, 2011.
- K. M. Cremers. Stock return predictability: A bayesian model selection perspective. *The Review of Financial Studies*, 15(4):1223–1249, 2002.
- A. Demirgüç-Kunt and R. Levine. Bank-based and market-based financial systems: Cross-country comparisons. 1999.

- F. A. DeRoos and T. E. Nijman. Testing for mean-variance spanning: a survey. *Journal of empirical finance*, 8(2):111–155, 2001.
- D. W. Diamond and R. E. Verrecchia. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, 18(2):277–311, 1987.
- E. Dimson, P. Marsh, and M. Staunton. The worldwide equity premium: a smaller puzzle. In *Handbook of the equity risk premium*, pages 467–514. Elsevier, 2008.
- D. Duffie, N. Garleanu, and L. H. Pedersen. Securities lending, shorting, and pricing. *Journal of Financial Economics*, 66(2-3):307–339, 2002.
- A. Dugalic and D. T. Patino. Short-sale constraints and the market portfolio. 2017.
- E. J. Elton, M. J. Gruber, and C. R. Blake. The adequacy of investment choices offered by 401 (k) plans. *Journal of Public Economics*, 90(6):1299–1314, 2006.
- A. Estrella and G. A. Hardouvelis. The term structure as a predictor of real economic activity. *The journal of Finance*, 46(2):555–576, 1991.
- E. F. Fama and K. R. French. Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25, 1988.
- E. F. Fama and K. R. French. Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, 25(1):23–49, 1989.
- E. F. Fama and K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465, 1992.
- E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.
- E. F. Fama and K. R. French. Dissecting anomalies. *The Journal of Finance*, 63(4):1653–1678, 2008.
- E. F. Fama and K. R. French. Choosing factors. *Fama-Miller Working Paper*, 2015a.
- E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015b.
- E. F. Fama and K. R. French. Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29(1):69–103, 2016.
- E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, pages 607–636, 1973.
- W. E. Ferson, S. Sarkissian, and T. T. Simin. Spurious regressions in financial economics? *The Journal of Finance*, 58(4):1393–1413, 2003.

- A. Frazzini and L. H. Pedersen. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.
- E. I. George and R. E. McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.
- E. I. George and R. E. McCulloch. Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373, 1997.
- J. Geweke et al. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA, 1991.
- M. R. Gibbons, S. A. Ross, and J. Shanken. A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152, 1989.
- P. Giot and M. Petitjean. International stock return predictability: Statistical evidence and economic significance. 2006.
- C. R. Harvey and Y. Liu. Lucky factors. *Available at SSRN 2528780*, 2015.
- E. Hjalmarsson. Predicting global stock returns. 2010.
- R. J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies*, 5(3):357–386, 1992.
- H. Hong and J. C. Stein. Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, 16(2):487–525, 2003.
- K. Hou, C. Xue, and L. Zhang. A comparison of new factor models. Technical report, National Bureau of Economic Research, 2014a.
- K. Hou, C. Xue, and L. Zhang. Digesting anomalies: An investment approach. *Review of Financial Studies*, page hhu068, 2014b.
- K. Hou, C. Xue, and L. Zhang. A comparison of new factor models. 2017.
- A. Inoue and L. Kilian. In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402, 2005.
- H. Ishwaran, J. S. Rao, et al. Spike and slab variable selection: frequentist and bayesian strategies. *The Annals of Statistics*, 33(2):730–773, 2005.
- E. Jacquier, N. G. Polson, P. Rossi, et al. *Stochastic volatility: Univariate and multivariate extensions*. CIRANO, 1999.
- R. Jarrow. Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices. *The Journal of Finance*, 35(5):1105–1113, 1980.

- A. C. Johansson. Stochastic volatility and time-varying country risk in emerging markets. *The European Journal of Finance*, 15(3):337–363, 2009.
- S. Kandel and R. F. Stambaugh. On the predictability of stock returns: an asset-allocation perspective. *The Journal of Finance*, 51(2):385–424, 1996.
- J.-Y. Kim. Bayesian asymptotic theory in a time series model with a possible nonstationary process. *Econometric Theory*, pages 764–773, 1994.
- S. Kim, N. Shephard, S. Chib, et al. Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies*, 65(3):361–393, 1998.
- D. A. Kodde and F. C. Palm. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica: journal of the Econometric Society*, pages 1243–1248, 1986.
- R. S. Koijen, S. Van Nieuwerburgh, et al. Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3(1):467–491, 2011.
- A. Kudo. A multivariate analogue of the one-sided test. *Biometrika*, 50(3/4):403–418, 1963.
- M. Lettau and S. Van Nieuwerburgh. Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence. *The Review of Financial Studies*, 21(4):1607–1652, 2007.
- K. Li, A. Sarkar, and Z. Wang. Diversification benefits of emerging markets subject to portfolio constraints. *Journal of Empirical Finance*, 10(1):57–80, 2003.
- J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, pages 13–37, 1965.
- E. M. Miller. Risk, uncertainty, and divergence of opinion. *The Journal of finance*, 32(4):1151–1168, 1977.
- T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.
- S. Nagel. Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics*, 78(2):277–309, 2005.
- C. J. Neely and P. Weller. Predictability in international asset returns: A reexamination. *Journal of Financial and Quantitative Analysis*, 35(4):601–620, 2000.
- R. Novy-Marx and M. Velikov. A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147, 2015.
- E. Ofek and M. Richardson. Dotcom mania: The rise and fall of internet stock prices. *The Journal of Finance*, 58(3):1113–1137, 2003.

- B. S. Paye and A. Timmermann. Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315, 2006.
- J. Penasse. Return predictability: Learning from the cross-section. *Available at SSRN 2575725*, 2016.
- M. Pitt and N. Shephard. Time varying covariances: a factor stochastic volatility approach. *Bayesian statistics*, 6:547–570, 1999.
- D. E. Rapach, M. E. Wohar, and J. Rangvid. Macro variables and international stock return predictability. *International journal of forecasting*, 21(1):137–166, 2005.
- J. P. Romano and A. M. Shaikh. Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, 138(9):2786–2807, 2008.
- J. P. Romano, A. M. Shaikh, and M. Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002, 2014.
- M. S. Rozeff. Dividend yields are equity risk premiums. 1984.
- M. Rubinstein. Great moments in financial economics: Iii. short-sales and stock prices. 2004.
- P. A. Saffi and K. Sigurdsson. Price efficiency and short selling. *Review of Financial Studies*, page hhq124, 2010.
- A. Schrimpf. International stock return predictability under model uncertainty. *Journal of International Money and Finance*, 29(7):1256–1282, 2010.
- E. Sentana. The econometrics of mean-variance efficiency tests: a survey. *The Econometrics Journal*, 12(3):C65–C101, 2009.
- J. Shanken and A. Tamayo. Payout yield, risk, and mispricing: A bayesian analysis. *Journal of Financial Economics*, 105(1):131–152, 2012.
- W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.
- W. F. Sharpe. Capital asset prices with and without negative holdings. *The Journal of Finance*, 46(2):489–509, 1991.
- M. J. Silvapulle and P. K. Sen. *Constrained statistical inference: Order, inequality, and shape constraints*, volume 912. John Wiley & Sons, 2011.
- R. Sloan. Do stock prices fully reflect information in accruals and cash flows about future earnings?(digest summary). *Accounting review*, 71(3):289–315, 1996.

- J. H. Stock and M. W. Watson. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829, 2003.
- N. Tang, O. S. Mitchell, G. R. Mottola, and S. P. Utkus. The efficiency of sponsor and participant portfolio choices in 401 (k) plans. *Journal of Public Economics*, 94(11):1073–1085, 2010.
- W. Torous, R. Valkanov, and S. Yan. On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, 77(4):937–966, 2004.
- J. A. Wachter and M. Warusawitharana. Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148(2):162–178, 2009.
- J. A. Wachter and M. Warusawitharana. What is the chance that the equity premium varies over time? evidence from regressions on the dividend-price ratio. *Journal of Econometrics*, 186(1):74–93, 2015.
- I. Welch and A. Goyal. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508, 2007.
- A. Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298):348–368, 1962.