

Return Predictability: Learning from the Cross-Section *

Julien Penasse[†]

Abstract

This paper develops an estimation framework in which the true parameters of international return processes share a common distribution. The model (i) makes efficient use of the cross-sectional correlation in the residuals, (ii) incorporates cross-sectional information in the estimation process, and (iii) introduces economic constraints on equity premium forecasts. The effect on estimation precision is remarkably strong and manifests itself both in- and out-of-sample. Once cross-sectional information is accounted for, the international evidence of return predictability appears much less heterogeneous than previously reported. The United States stands out as the exception rather than the rule in having both an unusually large long-term equity premium and an unusually strong return predictability.

Keywords: Stock return predictability, Bayesian panel VAR, Asset allocation.

JEL: C22, C23, G11, G12, G15.

*This version: March 30, 2015. I am thankful for comments from Christophe Boucher, Edouard Challe, Guillaume Chevillon, Stéphane Chrétien, Pierre Collin-Dufresne, Joost Driessen, Rik Frehen, Robert Hodrick, Michael Johannes, Frank de Jong, Bryan Kelly, Malik Kerkour, Tim Kroencke, Hao Liang, Lorenzo Naranjo, Michaela Pagel, Benjamin Pognard, Patrice Poncet, Olivier Scaillet, as well as seminar participants at Columbia Business School, BI Norwegian Business School, Luxembourg School of Finance, Paris Dauphine, Banque de France, EPFL, Cambridge University, Ecole Polytechnique, ESSEC Business School, THEMA, Caisse des Dépôts et Consignations, the Econometric Society 2014 European Winter Meeting in Madrid, the RES 2015 meeting in London and the EEA 2014 meeting in Toulouse. The support of the Institut Louis Bachelier-Collège de France Research Initiative “Long-Term Asset Allocation,” together with its partners CNP Assurances and Caisse des Dépôts et Consignations is gratefully acknowledged. Part of this paper circulated previously as “International Return Predictability and the Term Structure of Risk.” All remaining errors are my own.

[†]Tilburg University and Ecole Polytechnique, Département d’Economie, 91128 Palaiseau Cedex, France. E-mail: julien.penasse@gmail.com. Phone: +33 6 81 20 41 54.

1 Introduction

There is widespread evidence that a number of macro-financial variables can predict stock returns.¹ However, in spite of this vast literature, there is still considerable uncertainty on the magnitude and pervasiveness of return predictability. As stressed by [Pastor and Stambaugh \(2012\)](#), even after observing more than two centuries of US data, “investors do not know the values of the parameters of the return-generating process, especially the parameters related to the conditional expected return.” International evidence of return predictability is even less conclusive. If anything, the prevailing view seems to be that “return predictability is neither a uniform, nor a universal feature across international capital markets” ([Schrimpf, 2010](#)).

However, an important source of information has been neglected in previous literature: cross-sectional evidence. Many studies, obviously, document evidence of return predictability outside the United States,² but most of these studies consider each country in isolation. The approach developed in this paper — learning from the cross-section³ — considers them jointly.

Three mechanisms can improve estimation precision by exploiting cross-country information. Consider a world with two countries:

$$\begin{aligned}r_{t+1}^{US} &= \theta^{US} + \beta^{US} x_t^{US} + u_{t+1}^{US} \\r_{t+1}^{UK} &= \theta^{UK} + \beta^{UK} x_t^{UK} + u_{t+1}^{UK}\end{aligned}$$

where, in each country, excess returns are predictable by a variable x_t , e.g. the domestic dividend-price ratio. The first mechanism is to treat international data as a set of seemingly unrelated equations ([Zellner, 1962](#)). International stock returns are correlated, a detail that is ignored in single-country regressions. The OLS estimator of the parameter of interest, say $\hat{\beta}^{US}$, is not efficient, because it assumes that the residual covariance is diagonal, i.e., $cov(u_{t+1}^{US}, u_{t+1}^{UK}) = 0$. An improved mechanism should exploit the covariance

¹See, for example, [Fama and Schwert \(1977\)](#); [Rozeff \(1984\)](#); [Keim and Stambaugh \(1986\)](#); [Campbell \(1987\)](#); [Campbell and Shiller \(1988\)](#); [Fama and French \(1989\)](#); [Ferson and Harvey \(1991\)](#); [Campbell and Shiller \(1991\)](#); [Hodrick \(1992\)](#).

²Recent studies include [Campbell \(2003\)](#), [Ang and Bekaert \(2007\)](#), [Schrimpf \(2010\)](#), [Hjalmarsson \(2010\)](#) and [Rapach et al. \(2005, 2013\)](#).

³[Jones and Shanken \(2005\)](#) coined the term for the purpose of evaluating mutual fund performance.

of the residuals.

The second mechanism is to treat the parameters of each country’s return process as random variables. For example, suppose that I estimate $\hat{\beta}^{US}$ to be close to zero, but at the same time, I find a strong relation between the dividend-price ratio and future returns in the United Kingdom. The previous literature considers these two elements of evidence as distinct facts.⁴ In the framework developed in this paper, I will assume that coefficients are drawn from a common normal distribution. This evidence would conduct me to revise my prior about return predictability in both countries. The reason for doing so originates from [de Finetti’s \(1964\)](#) exchangeability assumption, further developed by [Lindley and Smith \(1972\)](#) for linear regression models: When there is not enough information to allow for precise estimation of an individual effect, it is natural to assume that the difference with other individual effects is the work of chance.

To better understand the concept of exchangeability, consider the return process in a hypothetical country, Zembla.⁵ We have histories of stock returns and dividend-price ratios for other countries but not Zembla. In the absence of any relevant information about Zembla’s return process, our best guess would be $\beta^Z \sim \mathcal{N}(\bar{\beta}, \sigma_{\bar{\beta}})$, where $\bar{\beta}$ and $\sigma_{\bar{\beta}}$ are the hyperparameters characterizing the “population” of international return generating processes. If we *do* have histories of stock returns and dividend-price ratios for Zembla, should we overlook the information about $\bar{\beta}$ and $\sigma_{\bar{\beta}}$? The single-country estimate of $\hat{\beta}^Z$ may be imprecise, perhaps less precise than the common mean $\bar{\beta}$. In that case, intuition suggests a better estimate of β^Z would put considerable weight to the common mean. In contrast, if the single-country estimate were more precise, it would be given more weight.

The third mechanism originates in the Bayesian paradigm that prior knowledge about model parameters should be used in the estimation process. Assume for example that for the sample of Zembla’s returns, stocks underperformed bills, so that excess stock returns are negative on average. It seems likely that a Zemblian investor would follow the obser-

⁴[Ang and Bekaert \(2007\)](#), [Hjalmarsson \(2010\)](#) and [Rapach et al. \(2013\)](#) consider pooled regressions, i.e. take the polar viewpoint that the slope parameters are identical across countries. One exception is [Westerlund and Narayan \(2014\)](#), who construct a test of return predictability that relaxes the assumption of slope homogeneity.

⁵ Old Zembla’s fields where my gray stubble grows,
And slaves make hay between my mouth and nose.

(From “Pale Fire” by Vladimir Nabokov.)

vation in, e.g., [Merton \(1980\)](#) that the expected market risk premium should be positive, and adjust accordingly her estimation of the domestic equity premium. I incorporate this idea in my approach. Specifically, because, in the asset pricing literature, researchers tend to disagree on the interpretation of return predictability, I consider two forms of equity premium (EP) restrictions. The weak one only imposes that equity premium is non-negative in the long run, leaving room for behavioral interpretations of return predictability in terms of temporary mispricing (see, e.g., [Shiller, 2014](#)). The strong one further requires that the equity premium is always non-negative, in the line of [Pettenuzzo et al. \(2014\)](#).⁶

These are the three main ingredients of the Bayesian approach that I pursue in this paper. This approach builds on previous research that considers predictability from a Bayesian investment perspective, pioneered by the work of [Kandel and Stambaugh \(1996\)](#).⁷ Accounting for cross-sectional information is the main way in which I depart from previous research. For example, a growing number of papers study time-variation in predictability (e.g., [Henkel et al. \(2011\)](#); [Pettenuzzo and Timmermann \(2011\)](#); [Pastor and Stambaugh \(2009, 2012\)](#); [Dangl and Halling \(2012\)](#); [Johannes et al. \(2014\)](#)). In contrast, a Bayesian analysis of variation across countries seems long overdue. In that aspect, this paper can also be seen as a generalization of the pooled approaches to return predictability of [Ang and Bekaert \(2007\)](#) and [Hjalmarsson \(2010\)](#).

My study also builds on [Chib and Greenberg \(1995a\)](#), who introduce an exchangeable prior in the seemingly unrelated regression model of [Zellner \(1962\)](#). In their empirical application, [Chib and Greenberg \(1995a\)](#) consider a dynamic panel but implicitly assume that the initial values of the predictor are fixed and hence provide no additional information. In contrast, in my setting, the regressors are not pre-determined, so that additional information can be extracted if the later are stationary as in [Stambaugh \(1999\)](#).

The main empirical question I am trying to address here is whether return predictability really differs across countries, and if so, how much. I reexamine the international evidence of stock return predictability using a large data set of fifteen countries, concen-

⁶Along the same line, [Shanken and Tamayo \(2012\)](#) consider if, in light of their difference in prior beliefs, Eugene Fama and Richard Thaler would invest and time the market differently, after observing the same evidence on return predictability.

⁷[Wachter \(2010\)](#) surveys this literature on Bayesian portfolio choice.

trating on one of the most popular predictors of stock returns, namely the dividend-price ratio.

My main empirical findings can be summarized as follows. Cross-sectional learning substantially alters coefficient estimates. The effect on parameter precision is unambiguously strong. This gain in precision manifests itself both in and out-of-sample, where forecasts based on the exchangeable prior typically outperform forecasts based on models that ignore the cross-section. From an asset allocation perspective, the cost of ignoring cross-sectional information varies dramatically across countries, from roughly zero in the United States to about 150 basis points per year. Estimation risk is also substantially mitigated so that stocks look typically safer in the long run than when an uninformative prior is assumed.

Importantly, the international heterogeneity is substantially blurred, in contrast to previous literature. The United States stands out as the exception rather than the rule in having both an unusually large long-term equity premium and an unusually strong return predictability. Within the framework of this paper, a US investor would invest substantially more in stocks, and time the market much more aggressively, than in any other country. In the remaining countries, whether stocks emerge as predictable depends crucially on the theoretical interpretation of return predictability. When the constraint that risk premia cannot go negative is imposed, there is little evidence of stocks return predictability outside the US, the UK, and Japan. Additionally, since this restriction is poorly supported by the data, the reader is left with two mutually exclusive conclusions. Either (i) predictability may be associated to some form of temporary mispricing so that returns can sometimes be predictably negative, or (ii) expected returns may not fluctuate at all.

The rest of this paper is organized as follows. Section 2 introduces the model with learning from the cross-section and provides an overview of the estimation procedure. Section 3 discusses the empirical results, both in- and out-of-sample and illustrates the consequence of learning on asset allocation. Section 4 concludes.

2 Methodology

2.1 A model of international return predictability

It is common in the return predictability literature (e.g. [Kandel and Stambaugh, 1996](#); [Stambaugh, 1999](#); [Barberis, 2000](#); [Wachter and Warusawitharana, 2009](#)) to use vector auto regressions (VAR) to capture the relation between asset returns and predictor variables. I follow this literature and assume that the returns of stocks in excess of the risk-free rate $r_{i,t}$ is a linear function of lagged predictors such as the dividend-price-ratio. To facilitate comparison with previous studies, I will concentrate on a single regressor, $x_{i,t}$, that follows an AR(1) process. The model takes the form:

$$r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1} \quad (1)$$

$$x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1} \quad (2)$$

where the innovations $\epsilon_{i,t+1} = (u_{i,t+1}, v_{i,t+1})'$ are normal, i.i.d. across t and cross-sectionally correlated with covariance Σ .

This process characterizes excess return dynamics in $i = 1, 2, \dots, N$ countries over time $t = 0, 1, \dots, T$. Taking expectation of Equation (1), we see that $\theta_i + \beta_i x_{i,t}$ is the conditional equity premium (EP) for country i . If β_i differs from zero, then the EP varies over time in country i .

Most of the existing literature considers predictive regressions in isolation (or focuses on a single country). Alternatively, several recent papers have considered pooled regression by assuming that the slope coefficients β_i are equal across countries ([Ang and Bekaert, 2007](#); [Hjalmarsson, 2010](#); [Rapach et al., 2013](#)). Intuitively, the first approach ignores the meaningful information contained in the cross-section, while the second makes the strong assumption that the data-generating processes are similar across countries. This paper takes another direction and considers Equations (1) - (2) as a system of seemingly unrelated regressions (SUR). If the innovations are correlated, it is well known that one can obtain more efficient estimates of the model by considering them jointly ([Zellner, 1962](#)).

Some care must be taken regarding the structure of the covariance matrix Σ , which

directly enters into the SUR estimates of the model parameters. The vector of innovations ϵ_{t+1} is of dimension $2N$ and therefore requires to handle $2N(2N + 1)/2$ covariance terms. Most of these terms are likely to be redundant and imprecisely estimated. To circumvent this problem, I assume the following factor structure for the innovations:

$$u_{i,t+1} = \delta_i^u u_{N,t+1} + \tilde{u}_{i,t+1} \quad (3a)$$

$$v_{i,t+1} = \delta_i^v u_{N,t+1} + \tilde{v}_{i,t+1} \quad (3b)$$

where $\delta_N^u \equiv 1$, $\tilde{u}_{N,t+1} \equiv 0$, and

$$\begin{pmatrix} \tilde{u}_{i,t+1} \\ \tilde{v}_{i,t+1} \end{pmatrix} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \tilde{\Sigma}_i) \text{ for } i = 1 \dots N - 1 \text{ and} \quad (4a)$$

$$\tilde{v}_{N,t+1} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \tilde{\sigma}_{N,\tilde{v}}^2) \quad (4b)$$

Finally, I assume that

$$u_{N,t+1} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{N,u}^2). \quad (5)$$

Equations (3) through (5) describe an econometric model where the innovations in the first $N - 1$ countries are subject to common shocks corresponding to stock return innovations in country N . This specification allows for country-specific variances and lets the correlation between the unexpected returns $u_{i,t}$ and the innovation in the predictor $v_{i,t}$ vary across countries. For a country $i < N$, the correlation between returns and the predictor is given by the covariance matrix $\tilde{\Sigma}_i$, while for $i = N$ it follows from the factor structure. Stacking the slope coefficients δ_i^u and δ_i^v in a vector δ and the idiosyncratic variances in a block diagonal matrix $\tilde{\Sigma}$, we can recover Σ as

$$\Sigma = \sigma_{N,u}^2 \delta \delta' + \tilde{\Sigma}. \quad (6)$$

Note that if the panel data set is not balanced, it is convenient to let N be the country with the longest history (the United States in this paper), so that $\sigma_{N,u}^2$ can be estimated with greater precision.

So far, I have discussed little about the distribution of the initial values $x_{i,0}$. A

possibility is to treat them as fixed constants and therefore assume that they convey no information about the parameters of the data-generating process. [Stambaugh \(1999\)](#) shows that ignoring the stochastic nature of the initial conditions can have unexpectedly deep consequences for both Bayesian estimation and portfolio choice.⁸ Fortunately, it is possible to construct a likelihood function that specifically accounts for the initial conditions, provided $\rho_i \in (-1, 1)$ for all i . This assumption is ubiquitous in the literature. In particular, present-value models that impose transversality require that the dividend yield must be stationary (although empirically the coefficient is often very close to one). It is helpful to stack the predictors across countries and rewrite Equation (2) as a stationary vector autoregression (VAR) of order one,

$$x_t = \alpha + \rho x_t + v_{t+1}, \quad (7)$$

where α is a vector of intercepts and ρ is a diagonal matrix of slopes. When $\rho_i \in (-1, 1)$ for all i , we see from (7) that the vector of predictors has an unconditional mean of

$$\mu_x = (I_N - \rho)^{-1} \alpha \quad (8)$$

and a variance of V_x satisfying

$$\text{vec}(V_x) = (I_{N^2} - (\rho \otimes \rho))^{-1} \text{vec}(\Sigma_x) \quad (9)$$

where Σ_x is the submatrix of Σ containing covariances for the vector x_t .⁹ I next discuss how this information can be used in the Bayesian estimation of the model.

2.2 Prior beliefs

When estimating Equations (1) - (2) in different countries, the econometrician will typically conclude that the data-generating processes differ significantly across countries (see, e.g., [Ang and Bekaert, 2007](#); [Hjalmarsson, 2010](#); [Rapach et al., 2013](#)). The key premise of

⁸See also [Wachter and Warusawitharana \(2009, 2014\)](#) in the Bayesian portfolio choice literature and [Anderson and Hsiao \(1982\)](#) for a discussion in the frequentist setting.

⁹[Stambaugh \(1999\)](#) derives a similar expression for a single country VAR with multiple predictors. Equation (7) describes a multi-country VAR with a single predictor, which unconditional moments happen to have the same form.

this paper is that these differences may simply be the effect of chance. In this subsection, I introduce two classes of prior beliefs that correspond to this idea.

2.2.1 Exchangeability

Economic theory offers little guidance as to why predictability should be significant in some countries and not in others. As argued by [Hjalmarsson \(2010\)](#), countries that share many common characteristics should be more likely to exhibit similar predictability patterns than those that do not. [Menzly et al. \(2004\)](#) propose an external habit persistence model close to [Campbell and Cochrane \(1999\)](#), that generates cross-sectional heterogeneity in return predictability. This heterogeneity follows from the difference in asset cash flows' exposure to aggregate consumption. Although their model studies predictability across industries, the argument can be extended to international markets, viewing each country as an individual asset ([Hjalmarsson, 2010](#)). However [Menzly et al.](#) suggest that the heterogeneity is likely to be small. Other plausible sources of international heterogeneity are heterogeneous preferences (e.g. risk aversion), institutional differences, and measurement errors. Particular to the dividend-price ratio, [Engsted and Pedersen \(2010\)](#) and [Rangvid et al. \(2014\)](#) attribute international heterogeneity to differences in share repurchases and dividend smoothing.¹⁰

If the previous literature suggests that international return processes are likely to be different, it leaves us with the empirical question of how much they differ. In this light, a reasonable assumption may be to treat each country's vector of coefficients $\zeta_i = (\theta_i, \beta_i, \alpha_i, \rho_i)'$ as a realization of a random variable:

$$\zeta_i = \bar{\zeta} + \eta_i. \tag{10}$$

Such a prior is often denoted as *exchangeable* ([de Finetti, 1964](#)) and, provided individual coefficients are not too dispersed, causes the Bayesian estimates to be shrunk towards the common mean vector $\bar{\zeta} = (\bar{\theta}, \bar{\beta}, \bar{\alpha}, \bar{\rho})$. It is convenient to assume that the dispersion terms η_i are normally distributed. However, for the equity premium process to be stationary, it is necessary to truncate the distribution for the autoregressive parameters ρ_i to lie

¹⁰See, e.g., [Fama and French \(2001\)](#) for US evidence and [Renneboog and Trojanowski \(2007\)](#); [Von Eije and Megginson \(2008\)](#); [Andres et al. \(2009\)](#) for European evidence.

between -1 and 1 . Formally, I assume that, for each country i ,

$$\eta_i \sim \mathcal{N}(0, \Delta) \tag{11}$$

when $\rho_i \in (-1, 1)$. That is, this prior puts a zero probability mass to $|\rho_i| > 0$.

For the remaining parameters of the model, namely δ_i , $\tilde{\Sigma}_i$, $\tilde{\sigma}_{N,\bar{v}}^2$, $\sigma_{N,u}^2$, $\bar{\zeta}$ and Δ , I choose a prior that is uninformative in the sense of [Jeffreys \(1961\)](#) (see Appendix A.1).

It would be of interest to refine the exchangeability assumption further by enriching the right-hand side by country-specific characteristics or to identify clusters of countries sharing similar characteristics. Such analysis would typically require a larger number of countries. For example, [Canova \(2004\)](#) applies a similar framework to identify convergence clubs in income per capita growth but uses a dataset of 144 European regions. Alternatively, one could increase N by studying return predictability at the individual stock or industry levels.¹¹ I leave this possibility to further research and concentrate on international evidence in order to facilitate comparison with previous research.

Notice that this prior nests as special cases the traditional approach, which treats each data-generating process independently, and the pooled approach where cross-country heterogeneity is assumed to be negligible. The former arises when one assumes that individual parameters have little in common and ignores the cross-sectional correlation of the innovations (i.e., when $\Delta^{-1} = 0$ and when the off-diagonal terms of Σ are dogmatically set to zero for $i \neq j$). The latter corresponds to the assumption that $\zeta_i = \bar{\zeta}$.

2.2.2 Equity premium restrictions

A complementary way to mitigate estimation errors is to incorporate economic constraints on the model parameters. In this paper, I impose sign restrictions on excess return forecasts.

Most models that successfully account for return predictability propose mechanisms where risk or risk premia vary over time.¹² Quoting [Pettenuzzo et al. \(2014\)](#), “it is

¹¹[Cosemans et al. \(2012\)](#) propose a hierarchical approach to estimate individual stock betas but consider economically motivated priors, instead of treating individual betas as random draws from the cross-sectional mean.

¹²Such mechanisms include time-varying relative risk aversion (e.g. [Campbell and Cochrane, 1999](#)), time-varying aggregate consumption risk ([Bansal and Yaron, 2004](#)) and time-varying consumption disas-

difficult to imagine an equilibrium setting where risk-averse investors would hold stocks if their expected compensations were negative.” Building on earlier work by [Campbell and Thompson \(2008\)](#), [Pettenuzzo et al.](#) suggest to restrict the predictive regression by constraining the parameters θ_i and β_i so that the forecasts $\theta_i + \beta_i x_{i,t}$ are non-negative at all points in time.

Although a negative equity premium could theoretically arise if stocks hedge against other risk factors ([Boudoukh et al., 1997](#); [Pettenuzzo et al., 2014](#)), negative return forecasts are generally interpreted as evidence of mispricing (e.g., reflecting behavioral biases).¹³ In 1991, Eugene Fama noted that “there is no evidence that low D/P signals bursting bubbles, that is, negative expected stock returns” ([Fama, 1991](#)). Since then, the advocates of a purely rational view of asset price fluctuations have remained unconvinced of the existence of fads or bubbles (see, e.g. [Schwert, 2003](#); [Pastor and Veronesi, 2006](#)).

The previous literature has indeed provided mixed evidence that excess stock returns are predictably negative. For example, [Kothari and Shanken \(1997\)](#) show that an investor who assigns a 50% prior probability that expected returns are never negative comes away with a posterior probability of only 8% for the period 1926-1991, but conclude that expected returns are never negative in the postwar period. [Eleswarapu and Thompson \(2007\)](#) propose a bootstrap test that relies on out-of-sample forecasts, and reject the null that excess return forecasts are always non-negative. [Driesprong et al. \(2008\)](#) find that oil prices frequently forecast negative stock returns and disprove any interpretation in terms of risk premia.¹⁴

Confronted with this lack of decisive evidence, it is interesting to consider weak and strong forms of sign restriction. Under the assumption that the predictor $x_{i,t}$ is stationary with mean μ_{x_i} , the excess return series is also stationary with unconditional mean

$$\mu_{r_i} = E(\theta_i + \beta_i x_{i,t} + u_{i,t} | \Psi) = \theta_i + \beta_i \mu_{x_i}.$$

ters ([Gabaix, 2012](#)).

¹³In any consumption based asset pricing model, the risk premium must be non-negative provided the conditional covariance between aggregate consumption and aggregate wealth is positive. If the risk premium can plausibly be negative for some assets, which returns are negatively correlated with consumption, it is unlikely to hold for the stock market as a whole.

¹⁴In a similar vein, [Penasse and Renneboog \(2014\)](#) show that trading volume negatively predicts changes in fine art auction prices, and argue that this predictability cannot be related to time-varying risk or risk aversion.

When excess returns are not predictable $\beta_i = 0$, so that the equity premium is simply θ_i . The weak restriction imposes that the unconditional equity premium is non-negative in all countries. Given stock return high volatility, it is not unlikely that stocks persistently underperform safe assets, so that average excess returns happen to be negative, particularly in some countries with short histories. Formally, the prior in (10) is modified to

$$\zeta_i = \bar{\zeta} + \eta_i, \quad \rho_i \in (-1, 1), \quad \theta_i, \beta_i \in W_{\zeta_i}, \quad i = 1, \dots, N, \quad (12)$$

where W_{ζ_i} is the set such that

$$W_{\zeta_i} = \{\theta_i + \beta_i \mu_{x_i} \geq 0\}. \quad (13)$$

This restriction essentially says that excess return forecasts cannot be negative in population, and therefore does not rule out “mispricing.” The strong form, in contrast, encompasses the weak restriction and the one proposed by [Pettenuzzo et al. \(2014\)](#) that excess return forecasts are always non-negative. In that case θ_i and β_i must belong to the set

$$S_{\zeta_i} = \{\theta_i + \beta_i \mu_{x_i} \geq 0; \theta_i + \beta_i \min(x_i) \geq 0\}, \quad (14)$$

where $x_i = \{x_{i,0}, \dots, x_{i,T}\}$.

2.3 Bayesian estimation

I next describe how the model is estimated. The model takes the form of a three-stage hierarchy. The first stage corresponds to the likelihood of the data conditioned on country-level parameters. The second stage corresponds to the distribution of these parameters. The third stage corresponds to the distribution of the common means of the parameters. This setup allows to specify uninformative priors in the third stage of the hierarchy, and thus it also lets the data “speak” about country heterogeneity, given the assumed likelihood. In this subsection, I provide a brief overview of the estimation methodology. Additional details are given in Appendix A.

The investor begins with the prior beliefs described in the previous subsection. These beliefs can be written as a prior density $p(\Psi)$, where $\Psi = (\zeta, \delta_i, \tilde{\Sigma}_i, \tilde{\sigma}_{N,\bar{v}}^2, \sigma_{N,u}^2, \bar{\zeta}, \Delta)_{i=1}^{N-1}$.

Let $D \equiv \{r_1, \dots, r_T, x_0, x_1, \dots, x_T\}$ represent the data available to the investor at time T .

The posterior density of the parameter Ψ is computed as

$$p(\Psi|D) \propto L(D|\zeta, \Sigma)p(\Psi) \quad (15)$$

where $p(\Psi)$ denotes the prior density of the parameters, and $L(D|\zeta, \Sigma)$ is the likelihood function for the seemingly unrelated regression model.

It is helpful to rewrite the system (1) - (2) in stacked form:

$$\begin{bmatrix} r_{1,t+1} \\ x_{1,t+1} \\ \vdots \\ r_{N,t+1} \\ x_{N,t+1} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,t} & 0 & \cdots & & 0 \\ 0 & & 1 & x_{1,t} & & \\ \vdots & & & \ddots & & \\ & & & & 1 & x_{N,t} \\ 0 & & & & & 1 & x_{N,t} \end{bmatrix} \begin{bmatrix} (\theta_1, \beta_1)' \\ (\alpha_1, \rho_1)' \\ \vdots \\ (\theta_N, \beta_N)' \\ (\alpha_N, \rho_N)' \end{bmatrix} + \begin{bmatrix} u_{1,t+1} \\ v_{1,t+1} \\ \vdots \\ u_{N,t+1} \\ v_{N,t+1} \end{bmatrix}$$

or

$$y_{t+1} = X_t \zeta + \epsilon_{t+1}, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma) \quad (16)$$

The likelihood function is obtained by extending the setup described in [Chib and Greenberg \(1995a\)](#) to a time series setting in which the regressor is not pre-determined:

$$L(D|\zeta, \Sigma) \propto |\Sigma|^{-T/2} \exp \left[-\frac{1}{2} \sum_{t=1}^T (y_{t+1} - X_t \zeta)' \Sigma^{-1} (y_{t+1} - X_t \zeta) \right] \quad (17a)$$

$$\times \mathcal{N}(\mu_x, V_x). \quad (17b)$$

The first block of Equation (17) describes the likelihood function for the observations $1, \dots, T$ and treats the first observations, described by the vector x_0 , as a constant. The second block corresponds to the likelihood of the first vector of observations. It is common to refer to Equation (17) as the true likelihood and to Equation (17a) as the conditional likelihood (see [Stambaugh, 1999](#); [Wachter and Warusawitharana, 2009, 2014](#)).

The posterior distribution of the model parameters can then be obtained by integrating out the hyperparameters from the joint posterior density (15), which can be done with a Gibbs sampler (see, e.g. [Chib and Greenberg, 1995a](#); [Hsiao et al., 1998](#)). Gibbs sampling

is an iterative Markov Chain Monte Carlo (MCMC) procedure for obtaining a sequence of observations that are approximated from a specified multivariate probability distribution. Starting from some arbitrary initial values of the parameters, it samples successively from the posterior distribution of each parameter, conditional on the values of the other parameters sampled in the latest iteration. For this posterior density, I use a four-block Gibbs sampler as in [Chib and Greenberg \(1995a\)](#). The algorithm is extended to account for the stochastic nature of the initial observation, the economic restrictions on the model parameters, and the factor structure of the innovations (see Appendix A.2). The first two blocks correspond to the individual parameters ζ_i and to the parameters of the covariance matrix Σ . The last two blocks correspond to the meta distribution of the individual parameters, $\bar{\zeta}$ and Δ . In particular, the conditional distribution for the country-specific parameters ζ is given by

$$\zeta|D, \Psi_{-\beta} \sim \mathcal{N}(m_\zeta, V_\zeta) \times \mathcal{N}(\mu_x, V_x), \quad \rho_i \in (-1, 1), \quad i = 1, \dots, N \quad (18)$$

where

$$m_\zeta = V_\zeta \left(\Delta_N^{-1} A_0 \bar{\zeta} + \sum_{t=1}^T X_t' \Sigma^{-1} y_t \right) \quad (19)$$

$$V_\zeta = \left(\Delta_N^{-1} + \sum_{t=1}^T X_t' \Sigma^{-1} X_t \right)^{-1}. \quad (20)$$

and $\Delta_N^{-1} = I_N \otimes \Delta^{-1}$.¹⁵

Equations (19) and (20) provide insight on how cross-country information is accounted for in the posterior distribution and, hence, it is worth discussing them in detail. Note, first, that the Bayes estimator of ζ differs from the classical sampling estimator. In the classical setting it makes no sense to estimate the individual parameters, which are treated as random variables.¹⁶ Conditional on $\rho_i \in (-1, 1)$ for all i , the latter is the SUR estimator, given by

$$\left(\sum_{t=1}^T X_t' \hat{\Sigma}^{-1} X_t \right)^{-1} \left(\sum_{t=1}^T X_t' \hat{\Sigma}^{-1} y_t \right),$$

¹⁵The distribution is conditional to the remaining parameters, so that for example $\Psi_{-\beta} = (\Sigma, \bar{\zeta}, \Delta_N)$.

¹⁶In contrast, the frequentist estimator of the common mean vector $\bar{\zeta}$ is identical to the Bayesian estimator, conditional on Δ_N and Σ , see [Hsiao and Pesaran \(2008\)](#).

with $\hat{\Sigma}$ being a consistent estimate of Σ . The conditional Bayes estimator, m_ζ , is a weighted average of the SUR estimate and the common mean vector $\bar{\zeta}$. The weights are respectively proportional to Δ_N^{-1} and $\sum_{t=1}^T X_t' \Sigma^{-1}$. In other words, the Bayes estimator gives weights to both time-series and cross-sectional evidence, and uses weights that correspond to the precision of the time-series and cross-sectional information. As a result, the Bayes estimator shrinks the estimates of country coefficients to $\bar{\zeta}$. When the time dimension increases, more information about individual coefficients becomes available, which marginalizes cross-sectional information, so that the Bayes estimates gradually converges to the SUR estimates (see [Hsiao and Pesaran, 2008](#)).

This weighted average form is reminiscent of the common Bayes estimator with an informative prior, which is a weighted average of the OLS estimate and the prior mean. Conditional on $\bar{\zeta}$, Δ_N^{-1} , and x_0 , the prior for ζ is indeed

$$\zeta | \bar{\zeta}, \Delta_N, x_0 \sim \mathcal{N}(\bar{\zeta}, \Delta_N), \quad \rho_i \in (-1, 1), i = 1, \dots, N.$$

Conceptually, $\bar{\zeta}$ can be seen as an informative prior with precision Δ_N^{-1} , which information proceeds from the data. As noted in [Jones and Shanken \(2005\)](#), ignoring this information is tantamount to specifying a joint prior distribution in which the beliefs about ζ_i are mutually independent.

Finally, equity premium restrictions can be imposed in the first block of the Gibbs sampler. In order to draw from the restricted posterior distribution, I use the Metropolis-Hastings algorithm (see [Chib and Greenberg, 1995b](#); [Griffiths, 2003](#); [Johannes and Polson, 2003](#)). The reader is referred to Appendix A.2 for details.

2.4 Predictive distribution and portfolio choice

Following previous portfolio choice studies, (see for example [Kandel and Stambaugh, 1996](#); [Shanken and Tamayo, 2012](#); [Wachter and Warusawitharana, 2014](#)), I consider an individual investor with power utility and coefficient of relative risk aversion A . This investor maximizes at time T

$$E \left[\frac{W_{T+1}^{1-A}}{1-A} \middle| D \right] \tag{21}$$

for $A = 5$, where W_{T+1} is her time $T + 1$ wealth as she invests a fraction $0 \leq \omega \leq 1$ of her wealth in stocks:¹⁷

$$W_{T+1} = W_T(\omega \exp\{r_{i,T+1} + r\} + (1 - \omega) \exp\{r\}). \quad (22)$$

In the above, $r_{i,T+1}$ represents the excess stock return in country i , and r is the risk-free rate, which I assume is constant and equal to 120 bp per quarter throughout.¹⁸

The expectation (21) is taken with respect to the predictive distribution

$$p(r_{i,T+1}|D) = \int p(r_{i,T+1}|\Psi)p(\Psi|D)d\Psi. \quad (23)$$

The predictive distribution can be seen as a mixture of distributions, each conditioned to the set of parameter values Ψ and integrated over the probability distribution of these parameters. The maximization problem of the investor is solved numerically as

$$\omega^* \approx \arg \max_{\omega} \frac{1}{L} \sum_{l=1}^L \left(\frac{[\omega \exp\{r_{i,T+1}^{(l)} + r\} + (1 - \omega) \exp\{r\}]^{1-A}}{1 - A} \right) \quad (24)$$

where $r_{i,T+1}^{(l)}$, $l = 1, \dots, L$ are returns drawn from the MCMC output using the predictive distribution (23).¹⁹

For any portfolio weight ω , the certainty equivalent return (CER) in excess of the risk-free rate solves

$$\frac{\exp\{(CER + r)\}^{1-A}}{1 - A} = E \left[\frac{(\omega \exp\{r_{i,T+1} + r\} + (1 - \omega) \exp\{r\})^{1-A}}{1 - A} \middle| D \right] \quad (25)$$

The CER of a risky portfolio represents the excess return earned with certainty that would provide the investor with the same utility as the expected utility derived from the risky portfolio. The CER gives a convenient metric for the utility (expressed in monetary terms) perceived by our investor when she holds a given portfolio.

¹⁷More precisely, I restrict the allocation to the interval $0 \leq \omega \leq 0.999$ to avoid bankruptcy concerns (see [Kandel and Stambaugh, 1996](#)).

¹⁸Observe that, from (21) and (22), the optimal allocation to stocks does not depend on the level of the risk-free rate, although the level of utility does.

¹⁹As suggested in [Shanken and Tamayo \(2012\)](#), I slightly amend this procedure to incorporate antithetic sampling (see [Bauwens et al., 2000](#), pp. 75-76) to improve computational efficiency.

This metric is used twice in this paper. In Section 3.5, I measure the economic significance of prior beliefs by taking the difference between an optimal portfolio’s CER and the CER obtained from an alternative portfolio, suboptimal from the subjective point of view of an hypothetical investor. This difference measures the loss that this investor would face if forced to hold the suboptimal portfolio. In Section 3.6, it is used in a similar way to quantify the performance achieved by out-of-sample strategies.

3 Empirical Results

3.1 Data

My data set consists of quarterly data for fifteen OECD countries: Australia (AUS), Belgium (BEL), Canada (CAN), Denmark (DNK), France (FRA), Germany (DEU), Italy (ITA), Japan (JPN), the Netherlands (NLD), Norway (NOR), Spain (ESP), Sweden (SWE), Switzerland (CHE), the United Kingdom (GBR) and the United States (USA). Due to data availability, sample periods differ between countries. The US data covers 1952:Q1 to 2013:Q1. For the remaining countries, the series span 1971:Q1 to 2013:Q1, except Denmark (1972:Q1), the Netherlands (1986:Q1), Norway (1979:Q1), Spain (1978:Q2) and Switzerland (1974:Q1). More countries could be included, but the benefit of cross-sectional learning may be outweighed by country-specific estimation risk by using data available for shorter time periods.

I concentrate on one of the most popular predictors of stock returns, the dividend-price ratio (Rozeff, 1984; Fama and French, 1988; Campbell and Shiller, 1988), which is related to future stock returns via the present-value identity. My US stock data is the Standard & Poor’s 500 value-weighted index, downloaded from the CRSP. International equity prices and total returns are retrieved from Morgan Stanley Capital International’s (MSCI) database. I compute excess return by subtracting the continuously compounded 3-month short-term interest rate. The latter is downloaded from the OECD statistics or FRED (USA), and the IMF International Financial Statistics depending on availability. Following the convention, I compute the dividend-price ratio as the twelve-month moving sum of dividends minus the log of stock prices. Descriptive statistics for the two variables are available in the Internet Appendix that accompanies this paper.

3.2 Regression results

Posterior beliefs about the predictability coefficient β_i for each of the fifteen countries are represented in Figure 1. The figure shows boxplots of the posterior distribution under the single-country benchmark and with the exchangeable prior under weak and strong equity premium restrictions. The center line of each boxplot indicates the median of the distribution, the box and the vertical lines include 75% and 99% of the observations, respectively. Tables 1 and 2 supplement the boxplots by providing estimates for the remaining key parameters.²⁰

[Insert Figure 1 and Tables 1 and 2 about here]

The top panel of Figure 1 represents the estimates under the traditional framework that ignores cross-country information (see, for example [Stambaugh, 1999](#)). The median coefficients are mostly positive (except Germany and Italy). A Bayesian investor would conclude that the predictive coefficient is significantly different from zero in Australia and the United Kingdom, in line with the frequentist results in, e.g., [Hjalmarsson \(2010\)](#). The posterior distributions differ greatly across countries, as previously documented in the literature, and it seems difficult to explain why predictability is so strong in some countries and so weak in others.

Is there enough evidence to support this large heterogeneity? Panel (b) of Figure 1 presents the posterior distribution with the exchangeable prior and weak economic restrictions. For comparison purposes, the classical SUR estimates are also reported (denoted by diamonds). The cross-country heterogeneity documented in Panel (a) is clearly blurred, due to the cross-sectional correlation of the innovations (as can be seen from the SUR estimates), and also due to the exchangeable prior, which further shrinks the coefficients to the international mean.²¹ This suggests that there are, by large, enough similarities to make a meaningful use of cross-country information. The gain in precision is visually striking. For example, the 99% credible sets are typically twice tighter than in the benchmark case that imposes an uninformative prior. Countries with limited data (e.g. the Netherlands) and countries where the benchmark estimates are statistically fragile (e.g.

²⁰Results for the other models and data sets are given in the Internet Appendix.

²¹The weak economic restrictions have little impact on the predictability coefficients, as can be verified in Table 2 of the Internet Appendix.

Germany and Norway) gain more from cross-sectional learning. The mean predictive coefficient is no longer negative in Italy and is more than halved in the United Kingdom. For all countries, at least 75% of the posterior distribution is positive, which provides more comfort in favor of international return predictability than in the benchmark case. The UK, the US, and (to a lesser extent) Japan seem to stand out from the whole sample of countries by exhibiting a much stronger evidence of predictability (with posterior probabilities of a positive coefficient larger than 99%).

Observe also that the posterior distribution of the US slope coefficient is barely affected by exchangeability, reflecting the fact that it is readily estimated quite precisely with US data alone. Recall that the latter spans a longer period of time than for the remaining countries; hence, we should not be surprised by the robustness of US coefficients.

Bayesian estimates of the β_i , when the prior is modified to include strong equity premium constraints, can be seen on Panels (c) of Figure 1. The gain in precision and the shrinkage of individual coefficients toward the common means is further amplified (note the change in scale for the vertical axis). As noted by [Pettenuzzo et al. \(2014\)](#), the constraint has to hold at each point in time (and, in my case, for each country simultaneously). Therefore the number of constraints grows in proportion to sample size. This large set of potentially binding constraints pins down the coefficients of the predictive regression, and thus increases precision.²²

The evidence of return predictability is considerably weaker, however. The predictability coefficients are weighted downward in all countries and are all centered around zero, except again in the US and the UK (and to some extent, Japan). The downward effect is quantitatively large. Even in the United States, a one standard deviation increase in the dividend-price ratio raises next-quarter returns by 0.51% (0.53%) in the benchmark (exchangeable) case. Once the strong equity premium restrictions are imposed, the economic effect falls down to 0.34%. The reason for this shift is that the constraint places an upper bound on the magnitude of return predictability. For example, when the dividend-price ratio is very low, a positive β_i would predict low returns, absent of strong EP restriction. Thus, β_i cannot be too large so that more mass is given to draws where the predictive

²²A important difference is that economic constraints increase precision by shrinking the set of admissible coefficients, while the exchangeable prior use cross-country information to increase precision.

coefficient is low, including draws where it is negative.

We also learn from Table 1 that the log-dividend price ratio is more persistent when the equity premium constraint is entertained. [Stambaugh \(1999\)](#) shows that when the correlation between innovations in returns and innovations in the predictive variable is negative, underestimating the persistence ρ_i leads to overestimating the evidence for predictability β_i .²³ The Stambaugh correlation is typically strongly negative for the dividend price ratio.²⁴ The strong EP restriction produces draws of β_i that are typically smaller, which happen to correspond to draws where ρ_i is relatively large.

To further understand how similar are the parameters across countries, information on the posterior distribution of the common mean hyperparameters is provided in Table 2. When the weak economic restriction is imposed, the common mean for the predictability coefficient is clearly positive. Its posterior distribution is centered at 0.019 with a 99% credible set of [0.002, 0.037]. Consistent with the discussion on individual country coefficients, the strong economic restriction weakens the economic significance of return predictability. The distribution is now centered at 0.005 with a 99% credible set of [-0.007, 0.017], and the posterior probability that the coefficient is negative (not shown) is 12%.

[Insert Table 2 about here]

3.3 The long-run equity premium

In this subsection, I discuss how the exchangeable prior and the economic restriction influence posterior beliefs about the unconditional or long-run equity premium (EP). As discussed in Section 2.2.2, when the dividend-price ratio is stationary, so is the equity premium. The unconditional EP writes $\mu_{r_i} = E(\theta_i + \beta_i x_{i,t} + u_{i,t} | \Psi) = \theta_i + \beta_i \mu_{x_i}$ and is

²³[Stambaugh \(1999\)](#) shows that the relation between β and ρ is nearly linear under various specifications and proves that the relation is given by

$$E(\beta_i | D) - \hat{\beta}_i \approx E\left(\frac{\sigma_{i,uv}}{\sigma_v^2} \middle| D\right) [E(\rho_i | D) - \hat{\rho}_i] \quad (26)$$

where $\hat{\beta}_i$ and $\hat{\rho}$ are the OLS estimates of β_i and ρ_i , and $\sigma_{i,uv}$ and σ_v^2 are elements of the covariance matrix Σ . $\sigma_{i,uv}$ is often referred to as the Stambaugh correlation.

²⁴The posterior mean for the Stambaugh correlation ranges from -0.63 to -0.96 across countries (-0.95 for the United States). It is almost identical across specifications.

represented on Figure 2 for the same prior combinations as before. The top panel displays boxplots of the posterior distribution for the uninformative prior. The equity premium is often computed as the sample average of excess returns, and the latter is also reported for comparison purpose.

We see on Figure 2 that, in the benchmark case, the long-run equity premium is typically unstable. Multiplied by 400 to get annualized values, the average posterior ranges from -1.97% (Italy) to 8.3% (US).²⁵ With the exception of Australia, the US, and the UK, the 99% credible sets are very large. This occurs because the long-run EP is a nonlinear function of the model parameters and can therefore be more unstable than the latter.²⁶ The consequence is that a typical investor would allow a significant probability to cases where the long-run equity premium is negative (the posterior mean is even negative for Italy, as is the sample average of excess returns).

[Insert Figure 2 about here]

Panel (b) of Figure 2 represents the unconditional equity premia for an investor with exchangeable prior who explicitly excludes the possibility that the unconditional EP is negative. The posteriors are all above zero (by construction) but are also much less dispersed with posterior means ranging from 1.76% to 8.17% (again for Italy and the United States, see Table 1). Recall that the weak economic restriction is imposed to hold simultaneously for all countries. As a result, the distributions are not truncated at zeros, reflecting the fact that the constraint rarely binds for the same country. Rather, the posterior distributions are clearly asymmetric toward zero in the countries where the constraint binds the most such as Italy, Germany, or the Netherlands.

Once the strong economic restriction is imposed, we see on Panel (c) that the long-run EP is higher on average. The posterior mean ranges from 3.38% to 7.73% . Comfortingly, [Dimson et al. \(2008\)](#) report a similar range of 2.8% to 7.1% for the same countries but

²⁵The values for the remaining countries are reported in Table 2 of the Internet Appendix.

²⁶In a similar exercise, [Wachter and Warusawitharana \(2014\)](#) study the impact on the long-run equity premium of the assumption regarding the initial value of the predictor. They find that using the true likelihood for the data successfully pins down the equity premium, which appears remarkably stable across specifications (they only consider US data). In unreported results with the conditional likelihood, I indeed found that the equity premium is even more dispersed than the one shown on Panel (a), which relies on the true likelihood.

computed on a century of data.²⁷

3.4 Test for negative expected excess returns

This paper proposes two forms of economic restrictions on the equity premium forecasts. Both restrictions impose that the unconditional equity premium cannot be negative. The weak form leaves room for the possibility that the equity premium is sometimes negative, while the strong form additionally imposes that it must always be non-negative. In this subsection, I test if the strong restriction is supported by the data. As argued earlier, the literature has provided mixed evidence on the non-negativity of excess return forecasts. Since most of this literature focuses only on US data, it is worth considering this hypothesis within the framework of this paper.

In the Bayesian model comparison paradigm, testing the non-negativity of equity premium forecasts corresponds to selecting the model favored by posterior odds. Let \mathcal{M}_w represent the weak equity premium restriction and \mathcal{M}_s represent the strong form. The weak form should be preferred over the strong form if the posterior odds of model \mathcal{M}_w over model \mathcal{M}_s , $\frac{p(\mathcal{M}_w|D)}{p(\mathcal{M}_s|D)}$, is greater than one. In general, the posterior odds ratio equals the prior odds ratio multiplied by the Bayes factor, which is the ratio of the marginal likelihoods of each model. The marginal likelihood can be difficult to compute in high-dimensional models like the present one.²⁸ In our case, fortunately, the weak form restriction nests the strong one. Hence, it is possible to compute $p(\mathcal{M}_s|D)$ directly from the MCMC output by simply keeping records of how many draws among model \mathcal{M}_w satisfy the constraints of model \mathcal{M}_s (see, for example, [Koop, 2003](#)).

The posterior probabilities that returns are non-negative as well as information on the share of forecasts that are negative among countries are reported in Table 3. The posterior probabilities range from 1% (Japan) to 43% (Switzerland), which implies that the posterior odds are all above one. This suggests that one should reject the strong restriction in favor of the weak one, for all countries. However, the evidence is far from

²⁷The ranking differ substantially, however. For example, Italy is found to be one of the most profitable markets of the century with an average excess return of 6.6%. This further confirms that long-run equity premia is essentially unstable, and justifies the use of economically motivated priors for the purpose of inference.

²⁸see e.g. [Chib \(1995\)](#); [Raftery \(1996\)](#).

decisive. The posterior probability is below 10% for five countries (Japan, Spain, the UK, France and Canada). For the US, it is an unimpressive 13%, corresponding to posterior odds of 6.5.

[Insert Table 3 about here]

It is instructive to study how often predicted excess returns tend to be negative. We see in Table 3 that the share of negative forecasts varies substantially across countries. In the United States, forecasts are on average negative for about 10 quarters (i.e., only 4% of the total forecasts), mostly corresponding to the years 1999-2001. At the other end of the spectrum, about half of the forecasts of Japan excess returns are negative. In the remaining countries, roughly 20% of the forecasts are negative. The United States stands out as having the lowest share of negative forecasts, and yet a relatively large posterior odds against a model where returns are never negative. This reflects the fact that the average forecast is typically higher in the US, which has the highest unconditional equity premium (see Table 1), and, at the same time, that the model parameters are estimated with more precision than in the remaining countries.

In summary, in spite of a clear gain in posterior precision, there is no decisive evidence that excess returns are predictably negative. The data provides some comfort to the view that returns can sometimes be negative in some countries, but critics may still legitimately argue that this rejection is simply the effect of chance, or e.g., non-linearity in the predictive relation. In any case, a somehow uncomfortable conclusion remains. As noted in Section 3.2, the predictive coefficients are rarely significant when the strong restriction is imposed. This means that, under the paradigm that excess return predictability reflects fluctuations in risk premia, there is actually little evidence that risk premia fluctuate at all. Put differently, the main conclusion is that, if excess returns are predictable, predictability may be associated to some form of temporary mispricing.

3.5 Asset allocation analysis

I have discussed above the statistical significance of return predictability. Since the economic significance of return predictability is also of interest, I next examine the implications of various prior beliefs for optimal asset allocation. To do so, I take the point of

view of a Bayesian investor maximizing her expected utility with respect to her predictive probability distribution. Forcing this investor to hold a suboptimal portfolio (with respect to her own belief) is a way to measure the economic significance of a prior. Following [Kandel and Stambaugh \(1996\)](#), the certainty equivalent loss our investor receives if forced to hold this portfolio is an adequate metric to measure the impact of this prior, through the lens of asset allocation. This metric has been extensively used in previous Bayesian finance applications with informative priors (e.g. [Avramov, 2002](#); [Handa and Tiwari, 2006](#); [Shanken and Tamayo, 2012](#)).

To understand how prior beliefs affect asset allocation, I begin with a combination of the exchangeable prior with the weak (strong) economic restriction and alternatively ignore each prior. The resulting portfolio weights are reported in Table 4. The top panel studies allocation with respect to the weak prior that only restricts the unconditional equity premium while the bottom panel presents the result for the strong EP restriction. Following [Shanken and Tamayo \(2012\)](#), I consider values of the dividend price ratio at its historical mean and 1.5 standard deviations above or below the mean. The first line of each panel gives which value of the predictor is considered. The first three columns report the allocation to stocks of the optimal portfolio, and the next columns show the allocations for the various “suboptimal” portfolios associated to an investor who ignores exchangeability, the economic restriction, or both. Symmetrically, Table 4 reports the optimal certainty equivalent return (CER) in excess of the risk free rate associated with the optimal allocation and the CER losses for the suboptimal portfolios.

[Insert Tables 4 and 5 about here]

I start with the discussion of the results in regards to the weak EP restriction. The second column of Table 4 corresponds to the optimal asset allocation when the dividend-price ratio is at its historical average. When this is the case, the investor has no reason to time the market, and hence these allocations are identical to those of an investor who believes that $\beta_i = 0$. We can assess the economic strength of return predictability by comparing how the optimal allocation to stocks varies with the level of the predictor. We see that the typical investor would time the market quite aggressively. Expected returns rise with the log dividend-price ratio, and so do the typical allocation to stocks. For

example, an Australian investor would respectively allocate 13%, 29%, and 44% to stocks as the predictor increases in 1.5 standard deviation increments. These allocations are quite representative of the remaining countries, with the notable exception of the United States. In the case of the latter, the equity premium is sensibly higher (as can be seen in Figure 2), and a US investor would want to respectively invest 28%, 84%, and 100% of her wealth in stocks.

The optimal CERs corresponding to these portfolios are reported in the first three columns of Table 5. The CERs are annualized and expressed in basis points (bps). For example, a Spanish investor would choose to invest all her wealth to the risk-free rate when the predictor is 1.5 standard deviations below its mean. This investor would naturally earn a CER of zero as reported in the fourth column. Returning to our Australian investor, we see that she gets CERs of 15, 79 and 190 bps, respectively. For a US investor, the gains are dramatically higher: 47, 411, and 1043 bps, respectively. This reflects the higher US equity premium and larger allocations to stocks.

When considering the suboptimal portfolio weights, observe that the effects of the priors vary considerably across countries. A US investor would barely amend her allocation if she were to ignore either the weak restriction or the information from the cross-section. This point echoes the finding in Section 3.2 that the US predictability coefficient was quite unaffected by exchangeability. Consequently, the utility costs for holding a suboptimal portfolio are negligible (between 0 and 8 bps per year). In other countries, however, the effect of either exchangeability or the weak economic restriction can be substantial. For example, a Japanese investor ignoring exchangeability would allocate about 19% of her wealth to stock when the predictor is low, and zero with any other prior. This investor would be willing to pay 135 bps to return to the optimal allocation. As for the weak restriction alone, the costs are mostly negligible. The main exception is Italy where the equity premium would generally be negative absent any economic constraint. Hence, for Italy, the weak restriction typically binds and is worth 34 bps in monetary terms.

In some case, the economic restriction and the exchangeable prior play a similar role in asset allocation so that the cost of ignoring either assumption is small. When the dividend-price ratio is high, the Australian investor would change her allocation to stocks from 44% to 53% (37%) when overlooking exchangeability (EP restriction) with a marginal

CER loss of 6 (5) bps. However, if she were to ignore both assumptions, her allocation would surge to 83%, with an associated utility loss of 147 bps. On average, the cost of ignoring both restrictions is slightly higher than 11 bps when the dividend price ratio is low, and 41 bps when it is high.

Panel (b) of Table 4 reports the allocation with respect to the strong economic restriction. In most cases, an investor with such beliefs will be more skeptical about return predictability and hence her stock position will be less aggressive. For instance, the allocation of the Australian investor barely changes from 33% to 42% as the level of the predictor increases. The main exceptions are the UK and the US, which is consistent with the results of Section 3.2 that the predictive coefficients remain statistically significant in these countries after imposing the strong economic restriction.

Because the strong restriction tends to significantly pin down the predictive coefficient, we expect the exchangeable prior to have a fairly limited impact. Indeed, this is confirmed in most countries. In some cases such as in France or Japan, ignoring exchangeability leads an investor to not time the market at all. The costs associated to overlooking exchangeability are overall small (2 to 6 bps on average, depending on the level of the predictor).

In contrast, the effect of the strong EP restriction is substantial. It concentrates on the low dividend scenario. The utility costs are large, ranging from 42 to 127 bps. For example, the Australian investor who would normally invest a third of her wealth in stocks would reduce her allocation to 7%. The reason for this is that the strong restriction essentially binds when the dividend-price ratio is low, forcing the equity premium to remain non-negative. This is consistent with the observation that the strong EP restriction is poorly supported by the data (see Section 3.4).

To summarize, the exchangeable prior can have a substantial impact on portfolio choice, albeit not systematically. The weak economic restriction has a modest effect and is sometimes redundant for an investor with an exchangeable prior. In contrast, the strong restriction forces the investor to be more cautious, and the loss of ignoring this constraint can be sizable.

3.6 Out-of-sample analysis

I argue in this paper that cross-sectional information and economic restrictions can be exploited to improve estimation precision. Under the relatively mild prior that countries share some similarities, the results indicate that international predictability is much less heterogeneous than would indicate the estimation of individual predictive regression. In this section, I evaluate the out-of-sample performance of these priors.

Goyal and Welch (2008) recently show that a simple forecasting rule based on the historical average of past returns outperformed most predictors that had been suggested by the literature. In turn, a number of papers have proposed more sophisticated models that can improve the statistical and economic value of out-of-sample forecasts. A common argument is that estimation on short samples, combined with model uncertainty and parameter instability, render conventional predictive regression forecasts unreliable. Hence, imposing reasonable constraints on the forecasts or coefficient estimates can reduce forecast errors. As previously noted, Campbell and Thompson (2008) and Pettenuzzo et al. (2014) impose that the equity risk premium cannot be negative and find that it improves both statistical and economic measures of out-of-sample performance.²⁹

From a forecasting point of view, the exchangeable prior can be seen as a different way to restrict coefficient estimates. Pooling models have been successfully used for the purpose of GDP growth forecasting (see, e.g. Mittnik, 1990; Zellner and Hong, 1989; Hoogstrate et al., 2000). Hjalmarsson (2010) finds that equity premium forecasts based on the fixed effect estimator often outperform those based on the time-series estimates. This paper expands this idea by allowing partial pooling and by incorporating risk premium restrictions in the Bayesian estimation.

Before proceeding, it is important to caution that even wrong restrictions can lead to superior out-of-sample forecasts. In the mean squared forecast error (MSFE) sense,

²⁹Several other studies study economically motivated constraints. For example, Pastor and Stambaugh (2009, 2012) impose that the sign of the correlation between shocks to unexpected and expected returns is negative. Wachter and Warusawitharana (2009, 2014) develop a class of informative priors that assign a low probability to high R^2 in the predictive regression. Many other approaches have been shown to successfully improve out-of-sample forecasts of equity returns, including forecast combinations and Bayesian model averaging (see, e.g., Cremers, 2002; Avramov, 2002; Rapach et al., 2009; Schrimpf, 2010; Dangl and Halling, 2012), factor models (e.g. Ludvigson and Ng, 2007; Neely et al., 2014; Kelly and Pruitt, 2013) and regime and time-varying coefficient models (e.g. Pesaran and Timmermann, 2002; Paye and Timmermann, 2006; Rapach, 2006; Henkel et al., 2011; Pettenuzzo and Timmermann, 2011; Dangl and Halling, 2012; Johannes et al., 2014).

this occurs when the bias resulting from imposing false restrictions is outweighed by the reduction of the estimator’s variance due to the restriction (see e.g. Hoogstrate et al., 2000). More generally, out-of-sample analysis compares forecasts, not models. For the latter (e.g., to assess the relevance of the strong economic restriction) full sample analysis such as the one conducted in Section 3.4 should be preferred to out-of-sample exercises (Diebold, 2012).³⁰

Following Hjalmarsson (2010), I exclude countries with less than 40 years of data. I use a forecasting period of 20 years, which leaves minimally 20 years of in-sample training period for each country. The forecasting exercise occurs recursively on an expanding window. To save on computation time, I update the posterior distributions every year (rather than quarterly).

3.6.1 Forecast performance

I evaluate out-of-sample forecasting ability by comparing forecasts based on predictive regressions and forecasts that are based on historical average of country excess returns. For each country, I compare historical average forecasts to forecasts generated from a competing predictive regression model. The historical average forecast corresponds to the constant expected excess return model, i.e., $\hat{r}_{i,t+1} = \hat{\theta}_{i,t}$, while the predictive regressions forecasts are obtained as $\hat{r}_{i,t+1} = \hat{\theta}_{i,t} + \hat{\beta}'_{i,t} z_{i,t}$. Forming forecasts in this manner simulates the situation of an investor in real time.

Unlike the rest of this paper, the approach I use to statistically evaluate forecasts is purely frequentist in nature and mainly relies on the out-of-sample R^2 statistic introduced by Campbell and Thompson (2008), R^2_{OS} . The R^2_{OS} measures the proportional reduction in MSFE for the competing model relative to the historical average benchmark over the out-of-sample period \underline{t}, \dots, T :

$$R^2_{OS} = 1 - \frac{\left(\sum_{t=\underline{t}}^T r_{i,t} - \hat{r}_{i,t}\right)^2}{\left(\sum_{t=\underline{t}}^T r_t^i - \bar{r}_{i,t}\right)^2} \quad (27)$$

³⁰Likewise, Cochrane (2008) and Lettau and Van Nieuwerburgh (2008) show that genuine predictability in the data-generating process can naturally coexist with a complete absence of out-of-sample return predictability.

where $\hat{r}_{i,t}$ and $\bar{r}_{i,t}$ are respectively forecasts obtained by predictive regression and forecasts based on the historical mean. A positive R_{OS}^2 implies that the predictive regression has lower average mean-squared prediction error than the historical average return.

I also compute the [Clark and West \(2007\)](#) *MSFE-adjusted* statistic to test the null that the historical average MSFE is less than or equal to the predictive regression MSFE, against the alternative hypothesis that the historical average MSFE is greater than the predictive regression MSFE. The [Clark and West \(2007\)](#) test is designed to compare forecasts from nested models. To compare the performances of the uninformative and exchangeable priors, which are not nested, I use the [Diebold and Mariano \(1995\)](#) and [West \(2006\)](#) (*DMW*) statistic.³¹

3.6.2 Economic performance

The approach I just described only measures the statistical performance of excess return forecasts based on the various priors discussed in this paper. It is also interesting to ask if these forecasts could have generated meaningful out-of-sample gains to our Bayesian investor. Just as I have evaluated the in-sample economic significance of the priors in the previous section, it is quite natural to assess out-of-sample performance using the certainty equivalent (excess) return metric (CER).

Within the framework of Section 2.4, I compare the portfolio choices of an investor who actively attempts to time the market to an investor who dogmatically believes that $\beta_i = 0$ and whose predictive distribution only depends on the history of past returns.³² For each investor, it is straightforward to compute her out-of-sample portfolio weights and hence her realized utility.

3.6.3 Results

Table 6 reports the out-of-sample R^2 and CER for the uninformative and exchangeable priors after imposing either weak or strong economic restrictions. The ‘ Δ ’ columns report the difference between either metric so that a positive value indicates an improvement

³¹Note that comparing differences in MSFE is not strictly identical to comparing out-of-sample R^2 because the latter is based on a *ratio* of MSFE. Therefore, it is possible that a statistically superior model (according to the *DMW* statistic) obtains a lower R_{OS}^2 than a statistically inferior model.

³²See, e.g. [Barberis \(2000, pp. 233-236\)](#).

from the non-informative prior.

I start with Panel A, which gives the R_{OS}^2 for the two models, as well as the difference between the two R_{OS}^2 . The table also indicates if the reported results are statistically significant.

[Insert Table 6 about here]

The first striking result is the poor out-of-sample predictive power of the dividend-price ratio. The weak EP-constrained forecasts are on average negative, respectively -1.01% (-0.26%) for the uninformative (exchangeable) priors. The OLS forecasts (not shown) are even worse, being negative for ten out of eleven countries. This poor predictive power is consistent with the previous literature and motivated more sophisticated forecasting models.

Observe that the exchangeable prior offers a first improvement, typically reducing large forecasting mistakes. In particular for the United States, the increase in R_{OS}^2 is a significant 4.64% over the uninformative prior (-7.24%), although the forecasts still fail to beat the historical benchmark.

We learn from the right columns of Panel A that the strong EP constraint sharpens forecasts further, the average R_{OS}^2 increasing by 0.23%. The constraint does a better job than the exchangeable prior in avoiding large mistakes (the lowest R_{OS}^2 is still for the US but is now -1.29%). This result is in line with [Pettenuzzo et al. \(2014\)](#) and hence indicates that their finding is robust to international data. From a forecasting viewpoint, we also see that exchangeability and the strong EP restriction complement each other. The fifth column of Panel A indicates that forecasts using the exchangeable prior jointly with the strong economic restriction systematically beat the historical average. The average R_{OS}^2 increases by 0.52% (albeit not systematically).

The results expressed in CER terms and reported in Panel B confirm that the exchangeable prior leads to substantially better forecast performance. Notice that both the uninformative and the exchangeable priors achieve a performance superior to the historical mean. The average CER is 25 bps per year for the uninformative prior and 47 bps for the exchangeable prior. The strong economic restriction improves the average performance, which is 56 bps for the uninformative prior and 61 bps for the exchangeable prior.

Consistent with earlier discussion, the outperformance is systematic once the strong EP restriction is imposed.

In summary, the exchangeable prior and the strong EP restriction help avoid large forecasting mistakes and improve out-of-sample forecasts in the large majority of cases. Returns are systematically predictable when the two are combined. This is remarkable because this exercise is performed on a period (1993-2013), in which the previous literature finds little predictability. For example, [Pettenuzzo et al. \(2014\)](#), document that EP constrained forecasts of US equity returns outperform the historical average during the 1947-2010 period but underperform during the 1979-2010 period. Similarly, the time-varying parameter model of [Dangl and Halling \(2012\)](#) fails to consistently beat the historical average during the 1988-2002 period. This is also true for studies that consider international evidence (see e.g. [Henkel et al., 2011](#); [Schrimpf, 2010](#); [Hjalmarsson, 2010](#)).

3.7 The term structure of risk

Whether stocks are safer in the long run has received considerable attention in the recent literature, particularly following the thought-provoking paper of [Pastor and Stambaugh \(2012\)](#) who claim that stocks may be riskier in the long run.³³ The Bayesian literature, in particular, emphasizes that uncertainty about expected returns is proportionally more important when investment horizon increases. In this final section, I extend the previous research by studying the effect of exchangeability and equity premium restrictions on the term structure of stock volatility. Interest centers on the following quantity:

$$\sigma_i^2(k|D) = \frac{1}{k} \text{Var}(r_{t \rightarrow t+k}^i | D) \quad (28)$$

where $r_{t \rightarrow t+k}^i = \sum_{j=1}^k r_{t+j}$ is the cumulated k -quarter ahead excess return. The predictive system (1)-(2) constitutes a restricted VAR that is commonly studied in the literature.

³³A partial list includes [Siegel \(1998\)](#); [Stambaugh \(1999\)](#); [Barberis \(2000\)](#); [Campbell and Viceira \(2002, 2005\)](#); [Bec and Gollier \(2009\)](#); [Jondeau and Rockinger \(2010\)](#); [Favero and Tamoni \(2010\)](#); [Diris \(2011\)](#); [Pettenuzzo and Timmermann \(2011\)](#); [Cales et al. \(2013\)](#); [Hovenaars et al. \(2014\)](#), and [Johannes et al. \(2014\)](#).

To facilitate comparison with previous works, it is useful to rewrite it as:

$$\begin{pmatrix} r_{i,t+1} \\ x_{i,t+1} \end{pmatrix} = \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} + \begin{pmatrix} 0 & \beta_i \\ 0 & \rho_i \end{pmatrix} \begin{pmatrix} r_{i,t} \\ x_{i,t} \end{pmatrix} + \begin{pmatrix} u_{i,t+1} \\ v_{i,t+1} \end{pmatrix} \quad (29)$$

or more compactly

$$z_t^i = \Phi_0^i + \Phi_1^i z_{t-1}^i + v_t^i \quad (30)$$

where $v_t^i \sim \mathcal{N}(0, \Sigma_i)$.³⁴ Campbell and Viceira (2002, 2005) show that the VAR framework provides a natural setup to compute multi-period moments. They find that the conditional variance of stock return grows more slowly with the investment horizon, which is consistent with the “conventional wisdom” popularized by Siegel (1998) that stocks are safer in the long run. To see such a case, note that under the assumption that Σ_i is constant over time, the conditional k -quarter variance is given by:³⁵

$$\sigma_i^2(k|D, \Psi_i) = \frac{1}{k} M \text{Var} (z_{t+1}^i + z_{t+2}^i + \dots + z_{t+k}^i) M' \quad (31)$$

where the vector $M = (1, 0)'$ extracts stock returns from the vector z_t^i ; Ψ_i is the set of parameters for country i and

$$\begin{aligned} \text{Var} (z_{t+1}^i + z_{t+2}^i + \dots + z_{t+k}^i) &= \Sigma_i + (I + \Phi_1^i) \Sigma_i (I + \Phi_1^i)' \\ &+ (I + \Phi_1^i + \Phi_1^i \Phi_1^i) \Sigma_i (I + \Phi_1^i + \Phi_1^i \Phi_1^i)' \\ &+ \dots \\ &+ (I + \Phi_1^i + \dots + (\Phi_1^i)^{k-1}) \Sigma_i (I + \Phi_1^i + \dots + (\Phi_1^i)^{k-1})' \end{aligned} \quad (32)$$

Equation (31) describes the term structure of risk faced by an investor who understands that a fraction of stock returns is predictable. Perhaps counterintuitively, although returns are predictable, volatility does not necessarily decrease with horizon. Consider the two-period variance:

$$\frac{1}{2} \text{Var}(r_{t \rightarrow t+2}) = \frac{1}{2} \text{Var}(r_{t+1}) + \frac{1}{2} \text{Var}(r_{t+2}) + \text{cov}(r_{t+1}, r_{t+2})$$

³⁴Therefore, Σ_i are block diagonal elements of the $2N \times 2N$ matrix Σ .

³⁵The reader is referred to the Internet Appendix for more details.

Absent return predictability, returns will not be autocorrelated, and the two-period variance will be identical to the sum of single-period variance, which I assume equal. If returns are predictable, the covariance term may be positive or negative so that volatility may increase or decrease with the investment horizon. More generally, return dynamics will be captured by the VAR. An investor who neglects predictability and therefore consider Equation (30) assuming that the coefficients of Φ_1^i are zero will face a flat term structure of volatility.

Importantly, Equation (31) provides the term structure of risk conditional on a given set of parameters. Since a typical investor would be uncertain about the true values of these parameters, a Bayesian investor would rather be interested in the predictive variance (Pastor and Stambaugh, 2012):

$$\sigma_i^2(k|D) = E(\sigma_i^2(k|D, \Psi_i)|D) + Var(E(r_{t \rightarrow t+k}^i | \Psi_i, D)|D) \quad (33)$$

Put differently, Equation (33) accounts for parameter uncertainty while (31) does not. The second term of Equation (33) is the variance of the conditional mean and adds positively to the expected conditional variance. Thus, the predictive variance will generally be larger than the conditional variance studied by Campbell and Viceira (2002, 2005).

Figures 3 and 4 plot the predictive volatility for each of the fifteen countries with investment horizon increasing from one quarter to 15 years for the three specifications introduced earlier.

The curves with continuous lines correspond to the traditional case with a non-informative prior. Observe that for most countries, the predictive volatility is larger at higher horizons.³⁶ In this simple setup with a single predictor and where the VAR is restricted, mean reversion of returns is the byproduct of the correlation between innovations in returns and innovations in the predictive variable (the Stambaugh correlation $\sigma_{i,uv}$) and the predictive slope β_i (see Campbell and Viceira, 2005). As stressed earlier, the correlation is strongly negative and the predictive slope is typically positive, inducing mean reversion in returns. This pertains to the conditional variance (the first term in

³⁶Notice also that the single-period volatility differs in this benchmark case. In the more general framework, the volatility is estimated from the factor structure given by Equations (3)-(5), and therefore uses an estimate of US stock volatility, which is computed from a longer period.

Equation (33)). However, once parameter uncertainty is accounted for, the mean reversion is more than compensated by uncertainty about expected returns (the second term in Equation (33)). The effect becomes stronger with investment horizon, increasing the slope of the term structure. This is typically the case for countries where a long history of data is not available such as the Netherlands, or where the predictive slope is estimated with low precision such as Switzerland. For those countries, the predictive volatility is typically J-shaped, and the 15-year predictive volatility can be substantially larger than the one-quarter volatility (e.g., up to about 10% for Italy). These results are largely consistent with previous literature (see [Bec and Gollier \(2009\)](#) for French data and [Jondeau and Rockinger \(2010\)](#) for a larger set of European countries). In contrast, the predictive volatility is clearly downward sloping for the United States, as is typically the case in the literature with uninformative priors ([Stambaugh, 1999](#); [Barberis, 2000](#)).

[Insert Figures 3 and 4 about here]

The predictive volatilities with the exchangeable prior are represented by dotted lines. We see that the curves are now downward sloping for most countries, which reflects two effects. First, the slope coefficients are partially shrunk toward the common mean with increasing mean reversions in countries where it was initially weak or negative. For example, this is the case for Italy and Germany. Second, parameter uncertainty is clearly mitigated, reducing volatility at long horizons. The consequences are of first order for most countries with the effect being a flattening of the term structures of risk and, on average, modestly decreasing by 1%-3%. Two notable exceptions are Japan, the US, and the UK where stocks remain manifestly safer in the long run.

Finally, the predictive volatilities with the exchangeable prior and economic constraints are represented by dashed lines. I showed in Section 3.2 that imposing economic constraints tends to reduce the economic significance of predictability. The consequence, which is visible in Figures 3 and 4, is a clear flattening of the term structure of risk. Again notable exceptions are the US and the UK.

4 Conclusion

The previous literature has produced a number of comprehensive studies of international return predictability. The common thread of this literature is that the return processes differ substantially across countries. This paper takes as a premise that this heterogeneity may be the effect of chance. It also argues that these previous studies, by treating each country separately, have neglected important information about predictability as a whole.

I develop an estimation framework which assumes that the true parameters of international return processes share a common normal distribution. By treating the evidence of return predictability jointly, the Bayesian investor makes efficient use of the cross-sectional correlation of the data and learns about the common means of the parameters. The investor can then use both country-specific and international estimates and weight them according to their respective precision. The resulting model nests as special case the classical approach that considers individual countries separately.

My empirical results suggest international heterogeneity is much smaller than previously reported, and as a result, I find that investment decisions differ little across countries when priors are exchangeable. The United States seems to be the exception rather than the rule, because it is characterized by both a high equity premium and a high degree of stock return predictability.

Because countries differ less than previously reported, the typical investor learns substantially from the cross-section. Estimates based on the exchangeable prior deliver superior out-of-sample performance in a large majority of countries. The benefit is even systematic when equity premium forecasts are constrained to be positive.

Appendix

A Bayesian framework

A.1 Likelihood and prior beliefs

It is convenient to rewrite Equation (16) as

$$y = X\zeta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma \otimes I_T) \quad (\text{A.1})$$

where $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_N)$, $y = (y'_1, y'_2, \dots, y'_N)'$, $\epsilon = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_N)'$, y_i and ϵ_i are $2T \times 1$ vector of the left-hand side variables and innovation terms for country i . Finally, $X = \text{diag}(X_1, X_2, \dots, X_N)$, $X_i = (\iota_T, (x_{i,1}, x_{i,2}, \dots, x_{i,T})')$ where ι_T denotes a vector of ones.

Equation (10), which corresponds to the second stage of the hierarchy, can be rewritten as

$$\zeta = A\bar{\zeta} + \eta, \quad \rho_i \in (-1, 1), \quad i = 1, \dots, N-1 \quad (\text{A.2})$$

with

$$\eta \sim \mathcal{N}(0, \Delta_\zeta), \quad \Delta_\zeta^{-1} = I_N \otimes \Delta^{-1} \quad (\text{A.3})$$

and where $A = (I_G, I_G, \dots, I_G)'$ maps the $G \times 1$ vector of common coefficients $\bar{\zeta}$ to the $NG \times 1$ vector ζ . $G = 4$ is the number of coefficients per country and \otimes denotes the Kronecker product. Equations (A.2) - (A.3) say that the individual coefficients $\zeta_i = (\theta_i, \beta_i, \alpha_i, \rho_i)'$ follow a normal distribution with $G \times G$ covariance matrix Δ .

The likelihood function is now written as follows:

$$p(D|\zeta, \Sigma) \propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} (y - X\zeta)' (\Sigma^{-1} \otimes I_T) (y - X\zeta) \right\} \times \mathcal{N}(\mu_x, V_x), \quad (\text{A.4})$$

where μ_x , and V_x are given in Equations (8) and (9).

Further, I assume the following priors for the model parameters:

$$\begin{aligned}
p(\zeta|\bar{\zeta}, \Delta_\zeta) &\propto |\Delta_\zeta|^{-N/2} \exp \left\{ -\frac{1}{2}(\zeta - A_0\bar{\zeta})' \Delta_\zeta^{-1} (\zeta - A_0\bar{\zeta}) \right\}, \quad \rho_i \in (-1, 1), \quad i = 1, \dots, N-1 \\
p(\delta_i, \tilde{\Sigma}_i) &\propto |\tilde{\Sigma}_i|^{-3/2}, \quad i = 1, \dots, N-1 \\
p(\delta_N^v, \sigma_{N,\bar{v}}^2) &\propto 1/\sigma_{N,\bar{v}}^2 \\
p(\sigma_{N,u}^2) &\propto 1/\sigma_{N,u}^2 \\
p(\bar{\zeta}, \Delta_\zeta^{-1}) &\propto |\Delta_\zeta^{-1}|^{-(G+1)/2}.
\end{aligned}$$

The prior on ζ reflects the assumption that individual coefficients are drawn from a common distribution. For the remaining parameters, the priors are uninformative in the sense of [Jeffreys \(1961\)](#). I assume that the parameter vectors are mutually independent. The joint posterior for $\Psi = (\zeta, \delta_i, \tilde{\Sigma}_i, \tilde{\sigma}_{N,\bar{v}}^2, \sigma_{N,u}^2, \bar{\zeta}, \Delta)_{i=1}^{N-1}$ is thus³⁷

$$p(\Psi|D) \propto p(D|\zeta, \Sigma)p(\zeta|\bar{\zeta}, \Delta_\zeta^{-1})p(\bar{\zeta})p(\Sigma)p(\Delta_\zeta^{-1}) \quad (\text{A.5})$$

where

$$p(\Sigma) \propto \prod_{i=1}^{N-1} p(\delta_i, \tilde{\Sigma}_i)p(\delta_N^v, \sigma_{N,\bar{v}}^2)p(\sigma_{N,u}^2). \quad (\text{A.6})$$

A.2 Sampling from the posterior

The Gibbs sampler consists of four blocks, which I detail below.³⁸ I initialize the Gibbs sampler with SUR estimates (or least squares estimates constrained to fit either of the equity premium restrictions). I start the sampler using 2,000 draws that I discard; I use the subsequent sample of 10,000 draws for the purpose of inference. When an equity premium constraint is entertained, the draws are strongly autocorrelated and keeping every tenth draw appears to be optimal for storage considerations. The convergence is supported by visual inspection of the posterior draws and by the MCMC diagnostics of [Raftery and Lewis \(1995, 1992b,a\)](#), [Geweke \(1992\)](#) numerical standard errors and relative numerical efficiency estimates and the Geweke chi-squared test comparing the means from

³⁷Note that it is equivalent to estimate Δ or Δ_ζ , since $\Delta_\zeta^{-1} = I_N \otimes \Delta^{-1}$.

³⁸In this appendix, I describe the approach for a balanced panel of countries for simplicity. The extension to the unbalanced case is straightforward (see [Schmidt \(1977\)](#) and footnote 40).

the first and last part of the sample.³⁹

1. Conditional posterior for ζ

Viewing the joint posterior in Equation (A.5) as a function of only ζ yields the following conditional posterior for ζ (conditional on $\rho_i \in (-1, 1)$ for all i and on the initial observations):

$$\begin{aligned} p(\zeta|D, \Psi_{-\zeta}, x_0) &\propto p(D|\zeta, \Sigma)p(\zeta|\bar{\zeta}, \Delta_\zeta^{-1}) \\ &\propto \exp\left\{-\frac{1}{2}(y - X\zeta)'(\Sigma^{-1} \otimes I_T)(y - X\zeta)\right\} \\ &\quad \exp\left\{-\frac{1}{2}(\zeta - A_0\bar{\zeta})'\Delta_\zeta^{-1}(\zeta - A_0\bar{\zeta})\right\} \end{aligned}$$

Note that

$$(y - X\zeta)'(\Sigma^{-1} \otimes I_T)(y - X\zeta) = (\zeta - \hat{\zeta})'X'(\Sigma^{-1} \otimes I_T)(\zeta - \hat{\zeta})X + \text{terms independent of } \zeta$$

where $\hat{\zeta} = [X'(\Sigma^{-1} \otimes I_T)X]^{-1}X'(\Sigma^{-1} \otimes I_T)y$.

The conditional posterior for ζ is thus proportional to the terms in the exponents,

$$(\zeta - \hat{\zeta})'X'(\Sigma^{-1} \otimes I_T)(\zeta - \hat{\zeta})X + (\zeta - A_0\bar{\zeta})'\Delta_\zeta^{-1}(\zeta - A_0\bar{\zeta}).$$

This expression is similar to the standard multivariate distribution with an informative prior about ζ (see e.g. [Koop et al. \(2007\)](#) p. 108-110). It can be rewritten as

$$(\hat{\zeta} - A_0\bar{\zeta})'X'(\Sigma^{-1} \otimes I_T)XV_\zeta\Delta_\zeta^{-1}(\hat{\zeta} - A_0\bar{\zeta}) + (\zeta - m_\zeta)'V_\zeta^{-1}(\zeta - m_\zeta)$$

where

$$\begin{aligned} V_\zeta &= (X'(\Sigma^{-1} \otimes I_T)X + \Delta_\zeta^{-1})^{-1} \\ m_\zeta &= V_\zeta(X'(\Sigma^{-1} \otimes I_T)y + \Delta_\zeta^{-1}A_0\bar{\zeta}) \end{aligned}$$

³⁹These convergence tools are implemented in Matlab Econometrics Toolbox, written by James P. LeSage (see www.spatial-econometrics.com).

ζ only enters through the term $(\zeta - m_\zeta)'V_\zeta^{-1}(\zeta - m_\zeta)$, we can thus write

$$p(\zeta|D, \Psi_{-\zeta}, x_0) \propto \exp \left\{ -\frac{1}{2}(\zeta - m_\zeta)'V_\zeta^{-1}(\zeta - m_\zeta) \right\}.$$

This is the kernel of a normal density and therefore ζ obeys a multivariate normal distribution with mean m_ζ and variance V_ζ . In order to account for the initial conditions, it is necessary to multiply this expression by the distribution for the vector x_0 , which yields the following posterior distribution for ζ :

$$\zeta|D, \Psi_{-\zeta} \sim \mathcal{N}(m_\zeta, V_\zeta) \times \mathcal{N}(\mu_x, V_x), \quad \rho_i \in (-1, 1), \quad i = 1, \dots, N \quad (\text{A.7})$$

This expression does not take the form of a standard density function because of the terms in the likelihood involving x_0 . Therefore, I use the Metropolis-Hastings algorithm (independence chain sampling, see [Chib and Greenberg, 1995b](#)) to sample from the posterior. The proposal density for ζ is normal with mean m_ζ and variance V_ζ .

When an equity premium constraint is entertained, the prior on ζ is defined as

$$p(\zeta|\bar{\zeta}, \Delta_\zeta) \propto |\Delta_\zeta|^{-N/2} \exp \left\{ -\frac{1}{2}(\zeta - A_0\bar{\zeta})'\Delta_\zeta^{-1}(\zeta - A_0\bar{\zeta}) \right\},$$

$$\rho_i \in (-1, 1), \quad \theta_i, \beta_i \in E_{\zeta_i}, \quad i = 1, \dots, N$$

where E_{ζ_i} is either

$$W_{\zeta_i} = \{\theta_i + \beta_i \mu_{x_i} \geq 0\}$$

or

$$S_{\zeta_i} = \{\theta_i + \beta_i \mu_{x_i} \geq 0; \theta_i + \beta_i \min(x_{i,t}) \geq 0\}.$$

The posterior is therefore modified to

$$\zeta|\Sigma, \bar{\zeta}, \Delta_\zeta \sim \mathcal{N}(m_\zeta, V_\zeta) \times \mathcal{N}(\mu_x, V_x), \quad \rho_i \in (-1, 1), \quad \theta_i, \beta_i \in E_{\zeta_i}, \quad i = 1, \dots, N. \quad (\text{A.8})$$

It is again necessary to use the Metropolis-Hastings algorithm to draw from the posterior, although in this case with many binding constraints, I rely on random walk sampling (see [Griffiths, 2003](#)).

2. Conditional posterior for Σ

To draw from the posterior distribution for Σ , I treat the residuals as an auxiliary model, so that standard results for the multivariate regression model apply. Define the $T \times 2$ matrices $e_i = (\epsilon'_{i,1}, \dots, \epsilon'_{i,T})'$ and $\tilde{e}_i = (\tilde{\epsilon}'_{i,1}, \dots, \tilde{\epsilon}'_{i,T})'$, and the $T \times 1$ vector $u_N = (u_{N,1}, \dots, u_{N,T})'$. Equation (3) can be rewritten as

$$e_i = u_N \delta'_i + \tilde{e}_i \text{ for } i < N.$$

From results in e.g. Zellner (1971), the posterior distributions for $\tilde{\Sigma}_i$ and δ_i are

$$\begin{aligned} \tilde{\Sigma}_i | D, \Psi_{-\tilde{\Sigma}_i}, x_0 &\sim \text{iWishart}(T-2, S_i) \\ \delta_i | D, \Psi_{-\delta_i}, x_0 &\sim \mathcal{N}(\hat{\delta}_i, \tilde{\Sigma}_i \otimes (u'_N u_N)^{-1}) \end{aligned}$$

where $S_i = (e_i - u_N \hat{\delta}_i)'(e_i - u_N \hat{\delta}_i)$ and $\hat{\delta}_i = (u'_N u_N)^{-1} u'_N e_i$.⁴⁰

Similarly define $e_N = (v_{N,1}, \dots, v_{N,T})$ and $\tilde{e}_N = (\tilde{v}_{N,1}, \dots, \tilde{v}_{N,T})$ so that $e_N = u_N \delta_N^v + \tilde{e}_N$, then

$$\begin{aligned} \sigma_{N,\tilde{v}}^2 | D, \Psi_{-\sigma_{N,\tilde{v}}^2}, x_0 &\sim \text{iGamma}\left(\frac{T-1}{2}, \frac{s_{N,\tilde{v}}}{2}\right) \\ \delta_N^v | D, \Psi_{-\delta_N^v}, x_0 &\sim \mathcal{N}(\hat{\delta}_N^v, \sigma_{N,\tilde{v}}^2 (u'_N u_N)^{-1}) \end{aligned}$$

with $s_{N,\tilde{v}} = (e_i - u_N \hat{\delta}_N^v)'(e_i - u_N \hat{\delta}_N^v)$ and $\hat{\delta}_N^v = (u'_N u_N)^{-1} u'_N e_N$.

Finally

$$\sigma_{N,u}^2 | D, \Psi_{-\sigma_{N,u}^2}, x_0 \sim \text{iGamma}\left(\frac{T-1}{2}, \frac{s_{N,u}}{2}\right)$$

with $s_{N,u} = \sum_{t=1}^T (u_{N,t} - \bar{u}_N)^2$ and $\bar{u}_N = T^{-1} \sum_{t=1}^T u_{N,t}$.

The above distributions condition on the initial condition. To integrate over the distribution for x_0 , I use the same Metropolis-Hastings step as previously for ζ , using the above distributions as candidates.

⁴⁰A benefit of this approach is that the factor δ_i and the variance term $\tilde{\Sigma}_i$ can be estimated using all data available for country i . In contrast, the previous literature (e.g. Chib and Greenberg (1995b)) typically relies on a more general prior for Σ , which entails that the posterior must be computed from the overlapping observations and therefore necessitates to drop a significant part of the data.

3. Conditional Posterior for $\bar{\zeta}$

Using the fact that $\zeta = A_0\bar{\zeta} + \eta$, with a diffuse prior, it can be verified (see, e.g., [Smith, 1973](#)) that the conditional posterior is normal with mean

$$m_{\bar{\zeta}} = V_{\bar{\zeta}} (A_0' \Delta_{\zeta}^{-1} \zeta)$$

and covariance

$$V_{\bar{\zeta}} = (A_0' \Delta_{\zeta}^{-1} A_0)^{-1}.$$

4. Conditional Posterior for Δ_{ζ}^{-1} and Δ^{-1}

From Equation (A.5) and using the definition $\Delta_{\zeta}^{-1} = I_N \otimes \Delta^{-1}$, the conditional posterior density for the dispersion parameters is given by:

$$\begin{aligned} p(\Delta_{\zeta}^{-1} | \zeta, \bar{\zeta}, D) &\propto p(\zeta | \bar{\zeta}, \Delta_{\zeta}^{-1}) p(\Delta_{\zeta}^{-1}) \propto \prod_{i=1}^N p(\zeta_i | \bar{\zeta}, \Delta^{-1}) p(\Delta^{-1}) \\ &\propto |\Delta^{-1}|^{N/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\zeta_i - \bar{\zeta})' \Delta^{-1} (\zeta_i - \bar{\zeta}) \right\} \times |\Delta^{-1}|^{-(G+1)/2} \\ &\propto |\Delta^{-1}|^{(N-G)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\sum_{i=1}^N (\zeta_i - \bar{\zeta})' (\zeta_i - \bar{\zeta}) \right] \Delta^{-1} \right\} \end{aligned}$$

Therefore $\Delta^{-1} | D, \Psi_{-\Delta^{-1}} \sim \text{Wishart} \left(\left[\sum_{i=1}^N (\zeta_i - \bar{\zeta})' (\zeta_i - \bar{\zeta}) \right], N \right)$.

References

- Anderson, Theodore Wilbur, and Cheng Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47–82.
- Andres, Christian, André Betzer, Marc Goergen, and Luc Renneboog, 2009, Dividend policy of German firms, *Journal of Empirical Finance* 16, 175–187.
- Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.
- Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.
- Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.
- Bauwens, Luc, Michel Lubrano, and Jean-Francois Richard, 2000, *Bayesian Inference in Dynamic Econometric Models* (Oxford University Press, Oxford).
- Bec, Frédérique, and Christian Gollier, 2009, Assets returns volatility and investment horizon: the French case, *CESifo Working Paper* 2622.
- Boudoukh, Jacob, Matthew Richardson, and Robert F Whitelaw, 1997, Nonlinearities the relation premium and between the the equity risk term structure, *Management Science* 43, 371–385.
- Cales, Ludovic, Eric Jondeau, and Michael Rockinger, 2013, Long-term portfolio management with a structural macroeconomic model, *Swiss Finance Institute Research Paper* 13.
- Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.
- Campbell, John Y., 2003, Consumption-based asset pricing, in George M. Constantinides, Milton Harris, and Rene M. Stulz, eds., *Handbook of the Economics of Finance*, 803–888 (North-Holland, Amsterdam).
- Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.
- Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.
- Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rate movements: A bird’s eye view, *The Review of Economic* 58, 495–514.
- Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.
- Campbell, John Y., and Luis M. Viceira, 2002, *Strategic Asset Allocation* (Oxford University Press, Oxford, UK).
- Campbell, John Y., and Luis M. Viceira, 2004, *Long-horizon mean-variance analysis: A user guide*, Manuscript, Harvard University, Cambridge, MA.

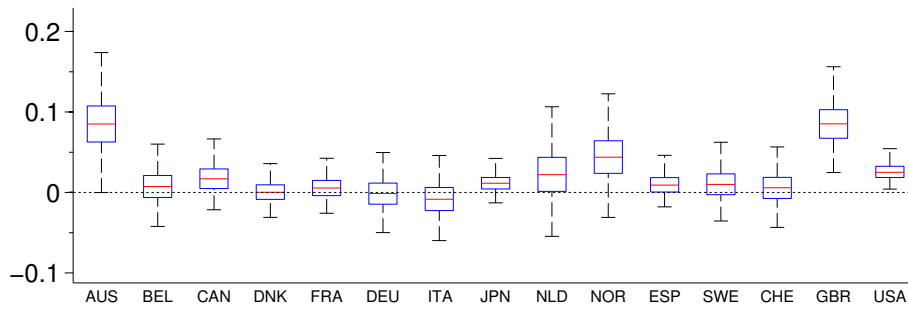
- Campbell, John Y., and Luis M Viceira, 2005, The term structure of the risk-return tradeoff, *Financial Analysts Journal* 61, 34–44.
- Canova, Fabio, 2004, Testing for convergence clubs in income per capita: A predictive density approach, *International Economic Review* 45, 49–77.
- Chib, Siddhartha, 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association* 90, 1313–1321.
- Chib, Siddhartha, and Edward Greenberg, 1995a, Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models, *Journal of Econometrics* 68, 339–360.
- Chib, Siddhartha, and Edward Greenberg, 1995b, Understanding the metropolis-hastings algorithm, *American Statistician* 49, 327–335.
- Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.
- Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.
- Cosemans, Mathijs, Rik Frehen, Peter C. Schotman, and Rob Bauer, 2012, *Estimating Security Betas Using Prior Information Based on Firm Fundamentals*, Working paper, Rotterdam School of Management, Erasmus University.
- Cremers, K.J. Martijn, 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.
- Dangl, Thomas, and Michael Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157–181.
- de Finetti, Bruno, 1964, Foresight: Its logical laws in subjective sources, in Henry E. Kyburg, and Howard E. Smokler, eds., *Studies in Subjective Probability*, 93–158 (Wiley, New York).
- Diebold, Francis X., 2012, *Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests*, Working paper 18391, National Bureau of Economic Research.
- Diebold, Francis X., and Robert S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253–263.
- Dimson, Elroy, Paul Marsh, and Mike Staunton, 2008, The worldwide equity premium: a smaller puzzle, in Rajnish Mehra, ed., *Handbook of the Equity Risk Premium*, 467–514 (Elsevier).
- Diris, Bart F., 2011, *Model uncertainty for long-term investors*, Working paper, Erasmus University Rotterdam.
- Driesprong, Gerben, Ben Jacobsen, and Benjamin Maat, 2008, Striking oil: Another puzzle?, *Journal of Financial Economics* 89, 307–327.
- Eleswarapu, Venkat R., and Rex Thompson, 2007, Testing for negative expected market return premia, *Journal of Banking & Finance* 31, 1755–1770.
- Engsted, Tom, and Thomas Q. Pedersen, 2010, The dividend-price ratio does predict dividend growth: International evidence, *Journal of Empirical Finance* 17, 585–605.
- Fama, Eugene F., 1991, Efficient Capital Markets: II, *Journal of Finance* 46, 1575–1617.

- Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.
- Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.
- Fama, Eugene F., and Kenneth R. French, 2001, Disappearing dividends: changing firm characteristics or lower propensity to pay?, *Journal of Financial Economics* 60, 3–43.
- Fama, Eugene F., and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.
- Favero, Carlo A., and Andrea Tamoni, 2010, Demographics and US stock market fluctuations, *CESifo Economic Studies* 57, 25–43.
- Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.
- Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.
- Geweke, John, 1992, Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., *Bayesian Statistics, Volume 4*, 169–193 (Oxford University Press).
- Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.
- Griffiths, William E., 2003, Bayesian inference in the seemingly unrelated regressions model, in David Giles, ed., *Computer-Aided Econometrics* (CRC Press).
- Handa, Puneet, and Ashish Tiwari, 2006, Does stock return predictability imply improved asset allocation and performance? Evidence from the U.S. stock market, *Journal of Business* 79, 2423–2468.
- Henkel, Sam James, J. Spencer Martin, and Federico Nardari, 2011, Time-varying short-horizon predictability, *Journal of Financial Economics* 99, 560–580.
- Hjalmarsson, Erik, 2010, Predicting global stock returns, *Journal of Financial and Quantitative Analysis* 45, 49–80.
- Hodrick, RJ, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial studies* 5, 357–386.
- Hoevenaars, Roy P. P. M., Roderick D. J. Molenaar, Peter C. Schotman, and Tom B. M. Steenkamp, 2014, Strategic asset allocation for long-term investors: Parameter uncertainty and prior information, *Journal of Applied Econometrics* 29, 353–376.
- Hoogstrate, Andre, J. Franz C. Palm, and Gerard A. Pfann, 2000, Pooling in dynamic panel-data models: An application to forecasting GDP growth rates, *Journal of Business & Economic Statistics* 18, 274–283.
- Hsiao, Cheng, and M. Hashem Pesaran, 2008, Random coefficient panel data models, in László Mátyás, and Patrick Sevestre, eds., *The Econometrics of Panel Data*, volume 1233, third edition (Springer).
- Hsiao, Cheng, M. Hashem Pesaran, and A. Kamil Tahmiscioglu, 1998, Bayes estimation of short-run coefficients in dynamic panel data models, in Cheng Hsiao, Kajal Lahiri, Lung Fei Lee, and M. Hashem Pesaran, eds., *Analysis of Panels and Limited Dependent*

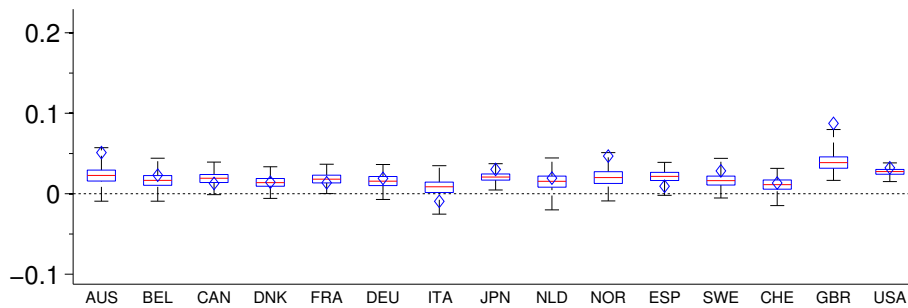
- Variables: A Volume in Honour of G. S. Maddala*, number 96, chapter 11, 268–296 (Cambridge University Press, Cambridge).
- Jeffreys, H, 1961, *The Theory of Probability* (Oxford University Press, Clarendon).
- Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential Learning, Predictability, and Optimal Portfolio Returns, *Journal of Finance* 69, 611–644.
- Johannes, Michael, and Nicholas Polson, 2003, MCMC methods for continuous-time financial econometrics, in Yacine Ait-Sahalia, and Lars Hansen, eds., *Handbook of Financial Econometrics* (Elsevier, North-Holland).
- Jondeau, Eric, and Michael Rockinger, 2010, *Portfolio allocation for European markets with predictability and parameter uncertainty*, Working paper, University of Geneva.
- Jones, Christopher S., and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507–552.
- Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385–424.
- Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.
- Kelly, Bryan, and Seth Pruitt, 2013, Market Expectations in the Cross Section of Present Values, *Journal of Finance* 68, 1721–1756.
- Koop, Gary, 2003, *Bayesian Econometrics* (Wiley, New York).
- Koop, Gary, Dale J. Poirier, and Justin L. Tobias, 2007, *Bayesian Econometric Methods* (Cambridge University Press).
- Kothari, Smitu P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.
- Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, *Review of Financial Studies* 21, 1607–1652.
- Lindley, Dennis V., and Adrian F. M. Smith, 1972, Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society* 34, 1–41.
- Ludvigson, Sydney C., and Serena Ng, 2007, The empirical risk-return relation: A factor analysis approach, *Journal of Financial Economics* 83, 171–222.
- Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding predictability, *Journal of Political Economy* 112, 1–47.
- Merton, Robert C., 1980, On estimating the expected return on the market, *Journal of Financial Economics* 8, 323–361.
- Mittnik, S, 1990, Macroeconomic forecasting using pooled international data, *Journal of Business & Economic Statistics* 8, 205–208.
- Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou, 2014, Forecasting the equity risk premium: the role of technical indicators, *Management Science* 60, 1772–1791.
- Pastor, Lubos, and Robert F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* LXIV.
- Pastor, Lubos, and Robert F. Stambaugh, 2012, Are Stocks Really Less Volatile in the Long Run?, *Journal of Finance* 67, 431–478.

- Pastor, Lubos, and Pietro Veronesi, 2006, Was there a Nasdaq bubble in the late 1990s?, *Journal of Financial Economics* 81, 61–100.
- Paye, Bradley S., and Allan G. Timmermann, 2006, Instability of return prediction models, *Journal of Empirical Finance* 13, 274–315.
- Penasse, Julien, and Luc Renneboog, 2014, *Bubbles and Trading Frenzies: Evidence from the Art Market*, Working paper, Tilburg University.
- Pesaran, M. Hashem, and Allan G. Timmermann, 2002, Market timing and return prediction under model instability, *Journal of Empirical Finance* 9, 495–510.
- Pettenuzzo, Davide, and Allan G. Timmermann, 2011, Predictability of stock returns and asset allocation under structural breaks, *Journal of Econometrics* 164, 60–78.
- Pettenuzzo, Davide, Allan G. Timmermann, and Rossen Valkanov, 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* 1–37.
- Raftery, Adrian E., 1996, Hypothesis testing and model selection, in Walter R. Gilks, Sylvia Richardson, and David Spiegelhalter, eds., *Markov chain Monte Carlo in practice*, chapter 10 (Chapman and Hall/CRC).
- Raftery, Adrian E., and Steven Lewis, 1992a, How many iterations in the Gibbs sampler, *Bayesian statistics* 4, 763–773.
- Raftery, Adrian E., and Steven M. Lewis, 1992b, One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo, *Statistical Science* 7, 493–497.
- Raftery, Adrian E., and Steven M. Lewis, 1995, The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms, in Walter R. Gilks, and David J. Spiegelhalter, eds., *Practical Markov Chain Monte Carlo*, 115–130 (Chapman and Hall).
- Rangvid, Jesper, M Schmeling, and A Schrimpf, 2014, Dividend predictability around the world, *Journal of Financial and Quantitative Analysis* (forthcoming).
- Rapach, David E., 2006, Structural breaks and predictive regression models of aggregate U.S. stock returns, *Journal of Financial Econometrics* 4, 238–274.
- Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2009, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.
- Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2013, International stock return predictability: What is the role of the United States?, *Journal of Finance* 68, 1633–1662.
- Rapach, David E., Mark E. Wohar, and Jesper Rangvid, 2005, Macro variables and international stock return predictability, *International Journal of Forecasting* 21, 137–166.
- Renneboog, Luc, and Grzegorz Trojanowski, 2007, Control structures and payout policy, *Managerial Finance* 33, 43–64.
- Rozeff, Michael, 1984, Dividend yields are equity risk premiums, *Journal of Portfolio management* 68–75.
- Schmidt, Peter, 1977, Estimation of seemingly unrelated regressions with unequal numbers of observations, *Journal of Econometrics* 5, 365–377.
- Schrimpf, Andreas, 2010, International stock return predictability under model uncer-

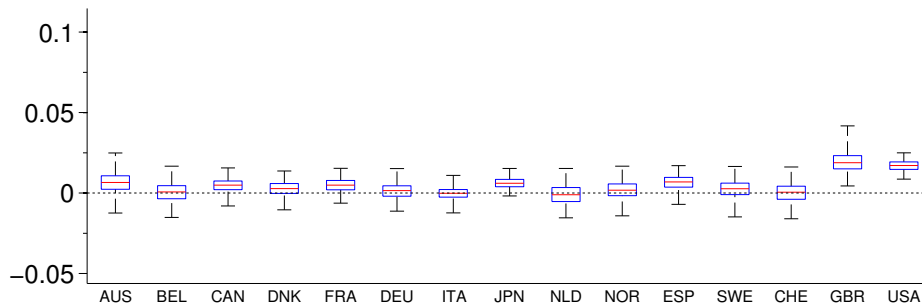
- tainty, *Journal of International Money and Finance* 29, 1256–1282.
- Schwert, G. William, 2003, Anomalies and market efficiency, in George Constantinides, Milton Harris, and René Stulz, eds., *Handbook of the Economics of Finance*, chapter 15, 937–972 (Elsevier).
- Shanken, Jay, and Ane Tamayo, 2012, Payout yield, risk, and mispricing: A Bayesian analysis, *Journal of Financial Economics* 105, 131–152.
- Shiller, Robert J., 2014, Speculative asset prices, *American Economic Review* 104, 1486–1517.
- Siegel, Jeremy J., 1998, *Stocks for the Long Run* (McGraw Hil, New York).
- Smith, Adrian F. M., 1973, A general Bayesian linear model, *Journal of the Royal Statistical Society. Series B (Methodological)* 35, 67–75.
- Stambaugh, Robert F., 1999, Predictive Regressions, *Journal of Financial Economics* 54, 375–421.
- Von Eije, Henk, and William L. Megginson, 2008, Dividends and share repurchases in the European Union, *Journal of Financial Economics* 89, 347–374.
- Wachter, Jessica a., 2010, Asset Allocation, *Annual Review of Financial Economics* 2, 175–206.
- Wachter, Jessica A., and Missaka Warusawitharana, 2009, Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market?, *Journal of Econometrics* 148, 162–178.
- Wachter, Jessica A., and Missaka Warusawitharana, 2014, What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio, *Journal of Econometrics* (forthcoming).
- West, Kenneth D., 2006, Forecast Evaluation, in Graham Elliott, Clive W. J. Granger, and Allan Timmermann, eds., *Handbook of Economic Forecasting, Volume 1*, volume 1, chapter 3, 99–134 (Elsevier, Amsterdam).
- Westerlund, J., and P. Narayan, 2014, A Random Coefficient Approach to the Predictability of Stock Returns in Panels, *Journal of Financial Econometrics* 0, 1–60.
- Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* 57, 348–368.
- Zellner, Arnold, 1971, *An introduction to Bayesian inference in econometrics* (Wiley, New York).
- Zellner, Arnold, and Chansik Hong, 1989, Forecasting international growth rates using Bayesian shrinkage and other procedures, *Journal of Econometrics* 40, 183–202.



(a) Uninformative prior



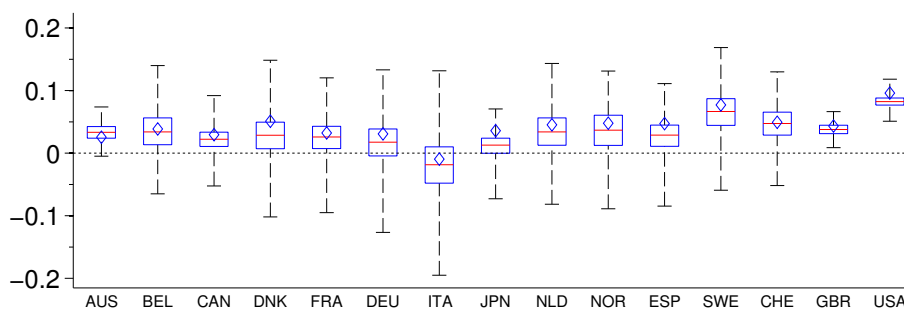
(b) Exchangeable prior and weak economic restriction



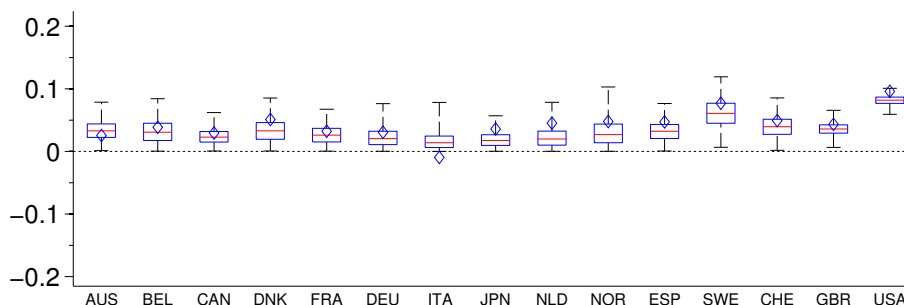
(c) Exchangeable prior and strong economic restriction

Figure 1: Posterior distribution for the dividend-price ratio coefficient

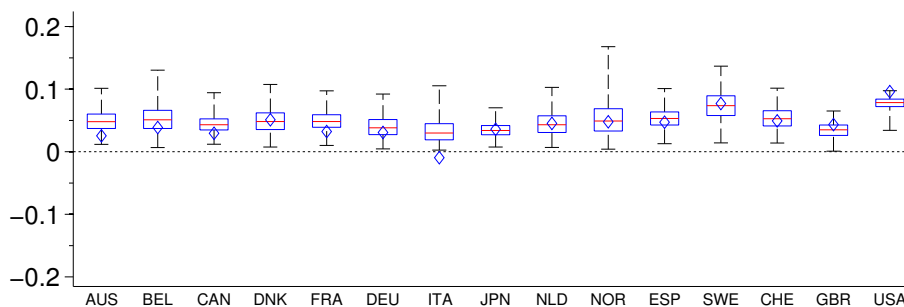
This figure depicts the posterior distributions of the slope coefficients β_i in a regression of excess returns on the log dividend-price ratio $r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1}$ where the latter follows an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$ for each of the fifteen countries i . For a given country, the red bar represents the median of the posterior distribution, the box corresponds to the first and third quartiles, and the whiskers give the 99% credible set. The diamonds in Panel (b) indicate the frequentist SUR estimates.



(a) Uninformative prior



(b) Exchangeable prior and weak economic restriction



(c) Exchangeable prior and strong economic restriction

Figure 2: The long-run equity premium

This figure depicts the posterior distributions of the unconditional equity premium $E(\theta_i + \beta_i x_{i,t} + u_{i,t+1}) = \theta_i + \beta_i \mu_{x_i}$, where x_i is the log-dividend-price ratio in country i . Unconditional means are annualized and continuously compounded. For a given country, the red bar represents the median of the posterior distribution, the box corresponds to the first and third quartiles, and the whiskers give the 99% credible set. The diamonds indicate the sample average of the (continuously compounded) excess return.

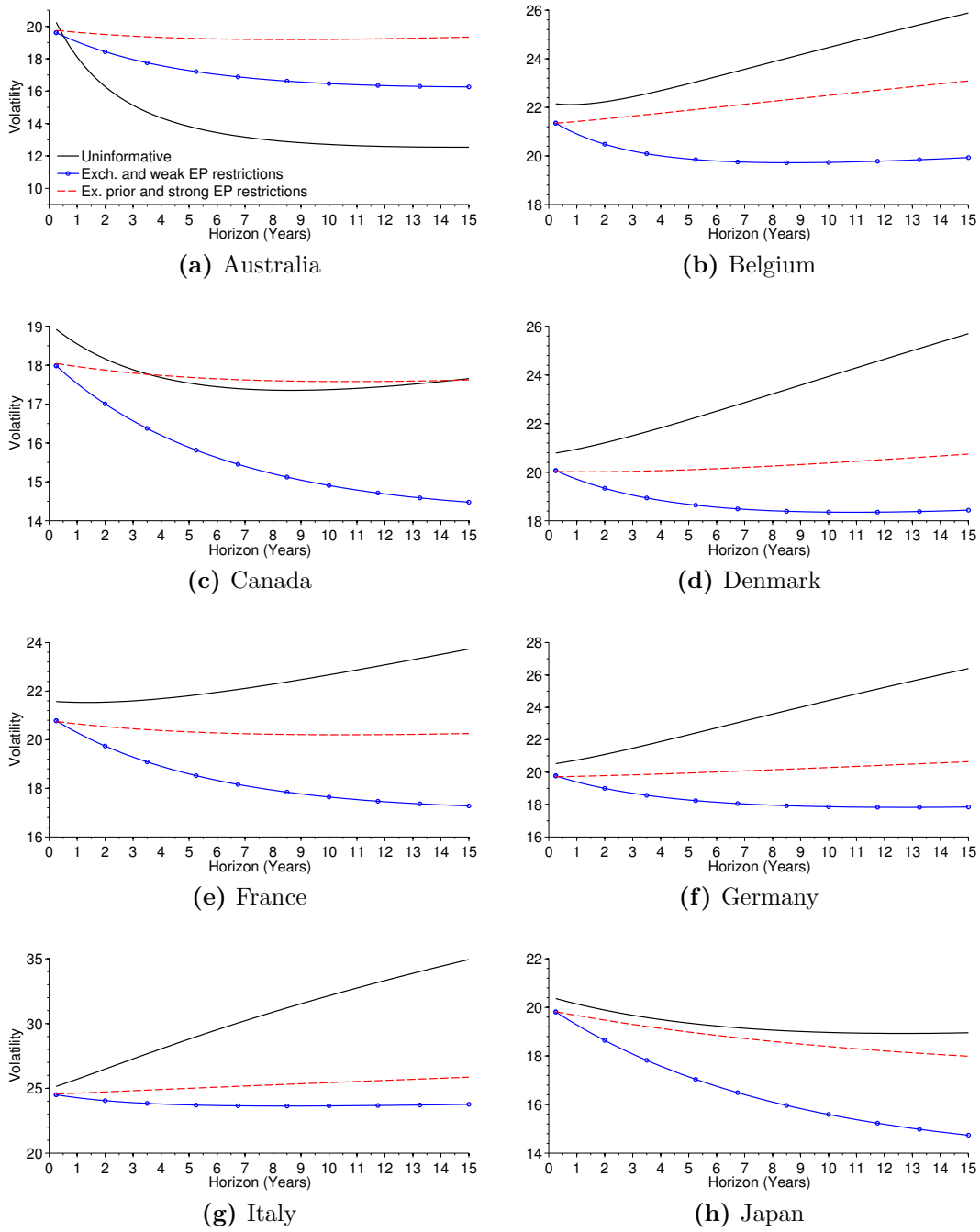


Figure 3: Term Structure of Risk

This figure shows the term structure of annualized predictive excess return volatilities for each country and for the three models. The straight lines use only country-level information. The dotted lines correspond to the exchangeable prior and constrain the unconditional equity premium to be positive. The dashed lines feature cross-sectional learning and constrain the posterior distribution of the equity premium to be non-negative.

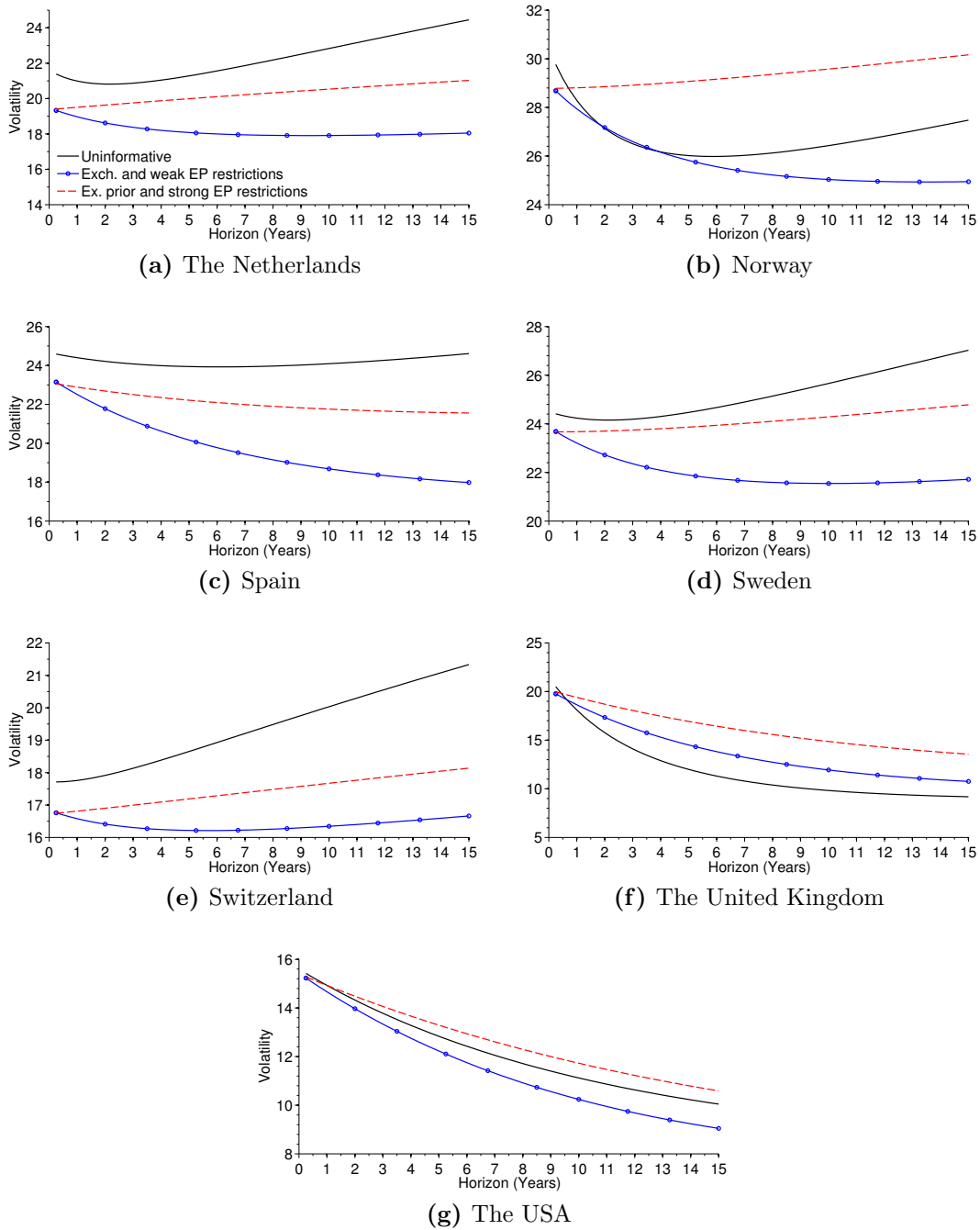


Figure 4: Term Structure of Risk (Continued)

This figure shows the term structure of annualized predictive excess return volatilities for each country and for the three models. The straight lines use only country-level information. The dotted lines correspond to the exchangeable prior and constrain the unconditional equity premium to be positive. The dashed lines feature cross-sectional learning and constrain the posterior distribution of the equity premium to be non-negative.

Table 1: Bayesian estimates for the country-level parameters

Panel A: Weak economic restriction									
	β_i	s.d.	$p < 0$	ρ_i	s.d.	μ_{x_i}	s.d.	μ_{r_i}	s.d.
Australia	0.023	0.011	0.02	0.934	0.016	-3.27	0.1	3.36	1.55
Belgium	0.017	0.01	0.04	0.912	0.019	-3.15	0.13	3.21	1.87
Canada	0.019	0.008	0.01	0.949	0.009	-3.61	0.11	2.39	1.22
Denmark	0.014	0.007	0.03	0.944	0.013	-3.76	0.2	3.37	1.83
France	0.018	0.007	0	0.95	0.01	-3.51	0.15	2.69	1.5
Germany	0.016	0.009	0.04	0.93	0.013	-3.62	0.11	2.29	1.56
Italy	0.008	0.011	0.21	0.919	0.016	-3.63	0.11	1.76	1.5
Japan	0.021	0.006	0	0.969	0.008	-4.14	0.24	1.91	1.22
Netherlands	0.015	0.011	0.09	0.914	0.019	-3.34	0.11	2.29	1.62
Norway	0.02	0.011	0.03	0.927	0.017	-3.46	0.19	3.11	2.21
Spain	0.021	0.008	0.01	0.958	0.01	-2.97	0.2	3.25	1.6
Sweden	0.017	0.009	0.02	0.932	0.014	-3.54	0.15	6.12	2.35
Switzerland	0.011	0.009	0.11	0.93	0.017	-3.76	0.14	3.97	1.74
United Kingdom	0.04	0.011	0	0.959	0.014	-3.28	0.16	3.59	1.05
United States	0.027	0.004	0	0.977	0.005	-3.56	0.12	8.17	0.77
Panel B: Strong economic restriction									
	β_i	s.d.	$p < 0$	ρ_i	s.d.	μ_{x_i}	s.d.	μ_{r_i}	s.d.
Australia	0.007	0.007	0.15	0.952	0.012	-3.39	0.14	4.96	1.77
Belgium	0.001	0.006	0.46	0.931	0.016	-3.24	0.16	5.3	2.31
Canada	0.005	0.004	0.14	0.965	0.006	-3.79	0.14	4.42	1.41
Denmark	0.003	0.005	0.27	0.956	0.012	-3.89	0.27	4.97	1.97
France	0.005	0.004	0.12	0.962	0.008	-3.72	0.18	4.95	1.62
Germany	0.001	0.005	0.38	0.946	0.011	-3.73	0.12	4.03	1.78
Italy	0	0.004	0.52	0.935	0.013	-3.72	0.15	3.38	2
Japan	0.006	0.003	0.02	0.981	0.006	-4.61	0.3	3.47	1.17
Netherlands	-0.001	0.006	0.57	0.931	0.016	-3.45	0.14	4.49	1.93
Norway	0.002	0.006	0.37	0.946	0.015	-3.62	0.26	5.33	2.86
Spain	0.007	0.005	0.09	0.972	0.008	-3.28	0.25	5.35	1.66
Sweden	0.002	0.006	0.31	0.947	0.012	-3.64	0.18	7.36	2.4
Switzerland	0	0.006	0.46	0.944	0.013	-3.84	0.17	5.39	1.78
United Kingdom	0.019	0.007	0	0.979	0.009	-3.65	0.3	3.4	1.29
United States	0.017	0.003	0	0.986	0.005	-3.9	0.23	7.73	1.05

This table reports posterior means and standard deviations for the regression of excess returns on the log dividend-price ratio $r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1}$ where the log dividend-price ratio follows an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$. The weak economic restriction imposes the unconditional forecasts of the equity premium to be non-negative for all countries. The strong economic restriction additionally imposes the same constraint on all forecasts. μ_{x_i} and μ_{r_i} (in annualized, percentage terms) are the unconditional means for the dividend-price ratio and the continuously compounded excess return, respectively. $p < 0$ is the probability that β_i is smaller than zero. The longest data is for the US and spans 1952:Q1 to 2013:Q1, and the shortest spans 1986:Q1 to 2013:Q1.

Table 2: Bayesian estimates for the common mean hyperparameters

	Weak economic restriction				Strong economic restriction			
	Mean	s.d.	99% credible set		Mean	s.d.	99% credible set	
$\bar{\theta}$	0.075	0.02	0.017	0.132	0.031	0.013	-0.004	0.068
$\bar{\beta}$	0.019	0.006	0.002	0.037	0.005	0.004	-0.007	0.017
$\bar{\alpha}$	-0.207	0.034	-0.305	-0.128	-0.161	0.028	-0.24	-0.087
$\bar{\rho}$	0.94	0.01	0.91	0.965	0.956	0.008	0.932	0.978

This table reports Bayesian estimates for the vector of common means across countries $\bar{\zeta} = (\bar{\theta}, \bar{\beta}, \bar{\alpha}, \bar{\rho})$. The weak economic restriction requires the unconditional forecasts of the equity premium to be non-negative for all countries. The strong economic restriction additionally imposes the same constraint on all forecasts.

Table 3: Posterior probability that expected returns are non-negative

	Posterior probability of non-neg. expected returns	Share of negative expected returns		
		Mean	75% credible set	
Australia	0.31	0.12	0	0.2
Belgium	0.29	0.15	0	0.25
Canada	0.08	0.23	0.08	0.35
Denmark	0.23	0.22	0.01	0.42
France	0.08	0.3	0.1	0.5
Germany	0.17	0.25	0.06	0.43
Italy	0.22	0.21	0.01	0.38
Japan	0.01	0.51	0.47	0.58
Netherlands	0.24	0.23	0.02	0.41
Norway	0.17	0.32	0.07	0.51
Spain	0.06	0.38	0.2	0.54
Sweden	0.32	0.1	0	0.18
Switzerland	0.43	0.09	0	0.14
United Kingdom	0.07	0.24	0.1	0.36
United States	0.13	0.04	0.01	0.07

This table reports the posterior probability that excess return forecasts are non-negative and gives statistics about the distribution of negative forecasts. The forecasts are based on the model with an exchangeable prior and where the coefficients are constrained so that the long-term equity premium is non-negative (i.e., weak economic restriction). Posterior probabilities are computed from the MCMC output as the percentage of draws where the coefficients additionally satisfy the condition that excess return forecasts are non-negative (i.e., strong economic restriction). The right panel reports the average share of negative forecasts and the 75% credible set.

Table 4: Influence of prior beliefs on asset allocation

Predictor	Optimal weight (%)			Exchangeability			Weight (%) ignoring Economic restriction			Both		
	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5
Panel A: Weak economic restriction												
Australia	13	29	44	26	39	53	7	23	37	0	28	83
Belgium	8	25	44	19	27	35	3	19	35	17	24	32
Canada	0	24	51	9	26	43	0	18	42	0	21	43
Denmark	2	25	48	25	28	30	0	19	41	22	23	24
France	0	21	49	14	24	34	0	17	45	12	20	28
Germany	2	20	38	24	25	26	0	12	29	20	19	17
Italy	9	16	22	14	18	22	0	1	7	11	5	0
Japan	0	9	44	19	22	24	0	6	41	0	9	28
Netherlands	5	20	34	12	30	47	0	7	19	6	24	42
Norway	3	14	26	13	23	33	0	8	19	0	16	39
Spain	0	18	47	5	19	34	0	14	44	5	17	28
Sweden	13	30	47	22	32	43	11	27	43	22	31	40
Switzerland	18	36	52	33	41	49	12	28	45	31	40	47
UK	0	28	63	8	40	72	0	25	59	0	27	96
USA	28	84	100	31	78	100	17	77	100	25	79	100
Panel B: Strong economic restriction												
Australia	33	38	42	40	43	47	7	23	39	0	28	83
Belgium	33	33	34	31	34	36	3	19	36	17	24	32
Canada	33	40	46	34	33	34	0	18	43	0	21	43
Denmark	31	35	40	33	33	34	0	19	43	22	23	24
France	28	35	42	30	31	32	0	17	45	12	20	28
Germany	30	31	32	32	31	31	0	12	29	20	19	17
Italy	21	21	21	22	23	25	0	1	7	11	5	0
Japan	20	31	42	28	28	28	0	6	42	0	9	28
Netherlands	34	33	33	35	38	41	0	7	20	6	24	42
Norway	22	23	24	26	29	31	0	8	19	0	16	39
Spain	23	33	42	24	26	29	0	13	45	5	17	28
Sweden	34	36	38	35	37	38	11	27	43	22	31	40
Switzerland	49	48	48	49	48	47	12	29	45	31	40	47
UK	24	41	58	31	47	63	0	24	60	0	27	96
USA	63	96	100	55	83	100	17	76	100	25	79	100

This table reports the optimal allocation to stock for various priors and three levels of the predictor. The investor has power utility with a relative risk aversion of 5. The weights are evaluated as the predictor incrementally increases from 1.5 standard deviations below its mean to 1.5 standard deviations above. The optimal weight column gives the allocation for a combination of exchangeable prior and weak (strong) economic restriction. The next columns correspond to an allocation of an investor who alternatively ignores exchangeability, the economic restriction, or both. The weak restriction imposes that the unconditional equity premium is non-negative and is analyzed in panel A. The strong restriction additionally imposes that all forecasts of excess returns are non-negative and is studied in panel B.

Table 5: Influence of prior beliefs on certainty equivalent returns (CER)

	Optimal CER (bps)			Exchangeability			CER Loss of ignoring Economic restriction			Both		
	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5
Panel A: Weak economic restriction												
Predictor	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5
Australia	15	79	190	18	11	6	3	3	5	15	0	147
Belgium	6	75	217	15	0	8	3	4	8	9	0	17
Canada	0	48	203	8	0	5	0	3	5	0	1	4
Denmark	0	63	234	55	1	34	0	4	5	41	0	60
France	0	49	257	37	1	24	0	2	2	32	0	45
Germany	0	40	143	47	2	15	0	6	9	30	0	43
Italy	13	37	75	4	1	0	13	33	34	1	18	75
Japan	0	9	195	135	15	39	0	1	1	0	0	25
Netherlands	2	35	109	5	10	15	2	15	21	0	2	5
Norway	2	43	141	21	14	11	2	9	11	2	0	36
Spain	0	42	298	20	0	21	0	2	1	18	0	45
Sweden	26	127	305	10	1	2	1	1	2	10	0	6
Switzerland	24	88	191	15	2	1	3	4	4	12	1	2
UK	0	74	381	17	14	8	0	1	1	0	0	107
USA	47	411	1043	0	2	0	8	3	0	1	2	0
Panel B: Strong economic restriction												
Predictor	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5	-1.5	0	1.5
Australia	106	138	174	5	3	2	63	21	1	106	9	169
Belgium	122	125	131	0	0	0	102	22	1	29	8	1
Canada	90	127	169	0	3	11	90	37	1	90	27	1
Denmark	95	125	159	0	1	4	95	26	1	8	15	26
France	81	130	189	0	2	11	81	35	1	25	23	20
Germany	85	92	101	0	0	0	85	33	1	10	15	22
Italy	68	66	65	0	1	2	68	60	28	15	40	65
Japan	42	96	173	5	1	19	42	64	0	42	46	19
Netherlands	111	105	100	0	3	6	111	66	16	74	7	8
Norway	98	108	118	5	7	10	98	46	5	98	11	45
Spain	70	139	230	0	6	22	70	48	1	44	31	23
Sweden	161	185	209	0	0	0	74	12	3	22	3	1
Switzerland	163	161	164	0	0	0	92	25	0	20	5	0
UK	59	168	331	5	3	3	59	27	1	59	19	167
USA	233	550	961	4	10	0	127	24	0	87	23	0

This table reports optimal certainty equivalent (excess) returns (CER) and losses (in basis points per year) when the investor is forced to hold a “suboptimal” portfolio based on an alternative prior. The investor has power utility with a relative risk aversion of 5, and the corresponding allocations to stocks are reported in Table 4. The CER and losses are evaluated as the predictor incrementally increases from 1.5 standard deviations below its mean to 1.5 standard deviations above. The first three columns give the CER with respect to an exchangeable prior with weak (strong) economic restriction. The next columns correspond to an allocation of an investor who alternatively ignores exchangeability, the economic restriction, or both. The weak restriction imposes that the unconditional equity premium is non-negative and is analyzed in panel A. The strong restriction additionally imposes that all forecasts of excess returns are non-negative and is studied in panel B.

Table 6: Out-of-sample results

Country	Weak economic restriction			Strong economic restriction		
	Uninform. (1)	Exchang. (2)	Δ (3)	Uninform. (4)	Exchang. (5)	Δ (6)
Panel A: Out-of-sample forecast performance						
Australia	1.47	3.05*	1.58	-1.05	1.24	2.29
Belgium	-2.02	-0.85	1.17**	-0.19	0.36	0.54
Canada	-0.7	-0.21	0.49	2.63***	1.79**	-0.84
Denmark	-1.94	-3.11	-1.18	1.89***	1.25***	-0.64
France	-1.69	-1.76	-0.07	0.45	0.32	-0.12
Germany	-0.81	-0.41	0.4	1.18*	1.00*	-0.18
Italy	0.94	-0.43	-1.36	0.78	0.7	-0.08
Japan	2.37**	2.26*	-0.11	1.13***	1.47*	0.34
Sweden	-5.73	-3.06	2.68***	0.72*	0.44	-0.28
UK	4.19*	4.26**	0.06	1.1	3.64*	2.54*
USA	-7.24	-2.6	4.64***	-1.29	0.85	2.14**
Average	-1.01	-0.26	0.75	0.23	1.19	0.52
Panel B: Out-of-sample economic performance						
Australia	62	84	22	53	60	7
Belgium	106	170	64	80	15	-65
Canada	-29	-10	19	65	66	1
Denmark	2	-37	-39	69	84	15
France	-13	1	14	26	26	0
Germany	5	18	13	54	41	-13
Italy	12	-12	-24	15	5	-10
Japan	146	137	-9	111	60	-51
Sweden	-48	0	48	22	44	22
UK	72	63	-9	57	75	18
USA	-44	103	147	69	191	122
Average	25	47	22	56	61	4

Panel A reports the [Campbell and Thompson \(2008\)](#) out-of-sample R_{OS}^2 (in percent), measuring the proportional reduction in mean-squared forecast error (MSFE) for each model relative to the historical average benchmark. Panel B gives certainty equivalent returns (again relative to the historical average benchmark) for asset allocations based on recursive forecasts of excess returns for an investor with power utility with a relative risk aversion of 5. All forecasts are performed on the out-of-sample period 1993-2013. The first (last) three columns present out-of-sample results with weak (strong) equity premium constraints on the coefficients. In each case, the Uninform. (Exchang.) column corresponds to an uninformative (exchangeable) prior. ‘ Δ ’ columns indicate the difference between both prior so that a positive value indicates a better forecast when using the exchangeable prior. The forecasts in Panel A (forecasting gains between the uninformative and exchangeable priors) are also evaluated according to [Clark and West \(2007\)](#) *MSFE-adjusted* (respectively [Diebold and Mariano \(1995\)](#) and [West \(2006\)](#)) statistic (see Section 3.6).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Internet Appendix to

**Return Predictability:
Learning from the Cross-Section**

The term structure of risk

Conditional k -period Moments

This section details how to extract moments of real returns from the excess real returns in the VAR(1) model. The reader is referred to [Campbell and Viceira \(2004\)](#) for additional details. I drop the index i for convenience. First, I derive a set of equations relating z_{t+k} to its current value z_t plus a weighted sum of interim shocks:

$$\begin{aligned}
 z_{t+1} &= \Phi_0 + \Phi_1 z_t + v_{t+1} \\
 z_{t+2} &= \Phi_0 + \Phi_1 z_{t+1} + v_{t+2} \\
 &= \Phi_0 + \Phi_1 \Phi_0 + \Phi_1 \Phi_1 z_t + \Phi_1 v_{t+1} + v_{t+2} \\
 z_{t+k} &= \Phi_0 + \Phi_1 \Phi_0 + \Phi_1^2 \Phi_0 + \dots + \Phi_1^{k-1} \Phi_0 + \Phi_1^k z_t \\
 &= +\Phi_1^{k-1} v_{t+1} + \Phi_1^{k-2} v_{t+2} + \dots + \Phi_1 v_{t+k-1} + v_{t+k}
 \end{aligned}$$

Taking the sum and reordering terms yields:

$$\begin{aligned}
 z_{t+1} + \dots + z_{t+k} &= [k + (k-1)\Phi_1 + (k-2)\Phi_1^2 + \dots + \Phi_1^{k-1}] \Phi_0 \\
 &+ (\Phi_1^k + \Phi_1^{k-1} + \dots + \Phi_1) z_t \\
 &+ (1 + \Phi_1 + \dots + \Phi_1^{k-1}) v_{t+1} \\
 &+ (1 + \Phi_1 + \dots + \Phi_1^{k-2}) v_{t+2} \\
 &+ \dots \\
 &+ (1 + \Phi_1) v_{t+k-1} + v_{t+k}
 \end{aligned}$$

Or more compactly :

$$z_{t+1} + z_{t+2} + \dots + z_{t+k} = \left[\sum_{n=0}^{k-1} (k-n) \Phi_1^n \right] \Phi_0 + \left[\sum_{m=1}^k \Phi_1^m \right] z_t + \sum_{q=1}^k \left[\sum_{p=0}^{k-q} \Phi_1^p \right] v_{t+q}$$

I am now able to compute conditional variance:

$$\begin{aligned} \text{Var}(z_{t+1} + z_{t+2} + \dots + z_{t+k}) &= \text{Var}\left(\left[\sum_{n=0}^{k-1} (k-n) \Phi_1^n\right] \Phi_0 + \left[\sum_{m=1}^k \Phi_1^m\right] z_t + \sum_{q=1}^k \left[\sum_{p=0}^{k-q} \Phi_1^p v_{t+q}\right]\right) \\ &= \text{Var}\left(\sum_{q=1}^k \left[\sum_{p=0}^{k-q} \Phi_1^p v_{t+q}\right]\right) \end{aligned}$$

as all other terms are constant or already known at time t . Expanding this expression yields:

$$\begin{aligned} \text{Var}(z_{t+1} + z_{t+2} + \dots + z_{t+k}) &= \Sigma_i + (I + \Phi_1) \Sigma_i (I + \Phi_1)' \\ &\quad + (I + \Phi_1 + \Phi_1 \Phi_1) \Sigma_i (I + \Phi_1 + \Phi_1 \Phi_1)' \\ &\quad + \dots \\ &\quad + (I + \Phi_1 + \dots + \Phi_1^{k-1}) \Sigma_i (I + \Phi_1 + \dots + \Phi_1^{k-1})'. \end{aligned}$$

I am only interested in extracting conditional moments per period from the portion of the VAR that contains returns, which I extract as follows:

$$E(r_{t \rightarrow t+k} | \Psi_i, z_t) = \frac{1}{k} M E(z_{t+1} + z_{t+2} + \dots + z_{t+k}). \quad (\text{A-1})$$

$$\text{Var}(r_{t \rightarrow t+k} | \Psi_i) = \frac{1}{k} M \text{Var}(z_{t+1} + z_{t+2} + \dots + z_{t+k}) M'. \quad (\text{A-2})$$

with $M = (1, 0)'$.

Predictive variance

The predictive variance is computed from the output of the MCMC. For each draw $\Psi^{(d)}$ ($d=1 \dots L$) from the posterior density $p(\Psi|D)$, I compute the conditional means and variance as described above

$$E_{i,d}(k) = E\left(r_{t \rightarrow t+k} \middle| \Psi_i^{(d)}, z_t\right) \quad (\text{A-3})$$

$$V_{i,d}(k) = \text{Var}\left(r_{t \rightarrow t+k} \middle| \Psi_i^{(d)}\right). \quad (\text{A-4})$$

Observe that the conditional mean must be measured for a given z_t . Following [Hoevenaars et al. \(2014\)](#), I choose to set it at the unconditional mean $z_t = (I - \Phi_1)^{-1} \Phi_0$. It is then straightforward to compute the predictive variance as

$$\bar{V}_i(k) = \frac{1}{L} \sum_d V_{i,d}(k) + \frac{1}{L} \sum_d (E_{i,d}(k) - \bar{E}_i(k))^2 \quad (\text{A-5})$$

where $\bar{E}_i(k) = \frac{1}{L} \sum_d E_{i,d}(k)$.

References

- Campbell, John Y., and Luis M. Viceira, 2004, *Long-horizon mean-variance analysis: A user guide*, Manuscript, Harvard University, Cambridge, MA.
- Hoevenaars, Roy P. P. M., Roderick D. J. Molenaar, Peter C. Schotman, and Tom B. M. Steenkamp, 2014, Strategic asset allocation for long-term investors: Parameter uncertainty and prior information, *Journal of Applied Econometrics* 29, 353–376.

Table 1: Summary Statistics

		Excess return		Log div.-price ratio	
		Mean	s.d.	Mean	s.d.
Australia	Q1:1971 - Q1:2013	2.54	20.02	-3.23	0.11
Belgium	Q1:1971 - Q1:2013	3.87	21.58	-3.10	0.21
Canada	Q1:1971 - Q1:2013	2.95	18.45	-3.61	0.18
Denmark	Q1:1972 - Q1:2013	5.09	20.17	-3.82	0.28
France	Q1:1971 - Q1:2013	3.23	21.00	-3.54	0.27
Germany	Q1:1971 - Q1:2013	3.07	20.04	-3.66	0.19
Italy	Q1:1971 - Q1:2013	-0.95	24.48	-3.58	0.22
Japan	Q1:1971 - Q1:2013	3.58	19.93	-4.37	0.28
Netherlands	Q1:1986 - Q1:2013	4.54	20.50	-3.40	0.15
Norway	Q1:1979 - Q1:2013	4.79	29.08	-3.59	0.20
Spain	Q2:1978 - Q1:2013	4.72	23.03	-3.10	0.31
Sweden	Q1:1971 - Q1:2013	7.70	23.82	-3.61	0.23
Switzerland	Q1:1974 - Q1:2013	4.94	17.19	-3.85	0.18
United Kingdom	Q1:1971 - Q1:2013	4.34	20.47	-3.28	0.14
United States	Q1:1953 - Q1:2013	9.60	15.26	-3.51	0.20

This table reports the mean and standard deviations (in parenthesis) for log excess stock returns (in percent per year) and the log dividend-price ratio.

Table 2: Bayesian estimates for the country-level parameters: alternative priors I

Panel A: Uninformative prior									
	β_i	s.d.	$p < 0$	ρ_i	s.d.	μ_{x_i}	s.d.	μ_{r_i}	s.d.
Australia	0.085	0.033	0.01	0.866	0.037	-3.24	0.08	3.33	1.46
Belgium	0.008	0.02	0.36	0.904	0.031	-3.12	0.17	3.49	3.55
Canada	0.018	0.018	0.17	0.951	0.02	-3.58	0.22	2.19	2.3
Denmark	0.001	0.013	0.49	0.957	0.019	-3.74	0.39	2.75	3.96
France	0.006	0.014	0.35	0.964	0.016	-3.46	0.35	2.44	3.39
Germany	-0.001	0.019	0.53	0.941	0.023	-3.6	0.21	1.61	4.1
Italy	-0.008	0.021	0.66	0.923	0.027	-3.53	0.2	-1.97	5.2
Japan	0.012	0.011	0.13	0.978	0.011	-4.04	0.52	1.06	2.34
Netherlands	0.023	0.032	0.24	0.886	0.041	-3.38	0.15	3.38	3.91
Norway	0.044	0.03	0.07	0.885	0.039	-3.52	0.18	3.57	3.85
Spain	0.01	0.013	0.24	0.971	0.014	-2.88	0.47	2.64	3.16
Sweden	0.011	0.019	0.3	0.933	0.025	-3.55	0.25	6.54	3.91
Switzerland	0.006	0.019	0.39	0.915	0.03	-3.81	0.17	4.64	3.1
United Kingdom	0.086	0.026	0	0.912	0.029	-3.28	0.12	3.78	1.06
United States	0.026	0.01	0	0.979	0.01	-3.47	0.3	8.28	1.08
Panel B: Exchangeable prior (without economic restriction)									
	β_i	s.d.	$p < 0$	ρ_i	s.d.	μ_{x_i}	s.d.	μ_{r_i}	s.d.
Australia	0.022	0.011	0.01	0.935	0.015	-3.22	0.11	2.67	1.64
Belgium	0.015	0.01	0.08	0.914	0.019	-3.11	0.14	2.05	2.4
Canada	0.019	0.007	0.01	0.949	0.009	-3.55	0.11	1.82	1.25
Denmark	0.014	0.007	0.04	0.944	0.013	-3.72	0.19	2.42	2.13
France	0.019	0.007	0	0.949	0.01	-3.46	0.15	2.04	1.72
Germany	0.015	0.009	0.04	0.93	0.013	-3.57	0.11	0.99	1.95
Italy	0.007	0.01	0.23	0.917	0.015	-3.51	0.13	-2.45	2.96
Japan	0.021	0.006	0	0.968	0.008	-4.13	0.23	1.27	1.5
Netherlands	0.013	0.011	0.1	0.915	0.018	-3.27	0.11	0.13	2.29
Norway	0.019	0.011	0.04	0.925	0.018	-3.39	0.18	0.77	3.18
Spain	0.022	0.009	0.01	0.957	0.01	-2.92	0.21	2.49	1.82
Sweden	0.016	0.009	0.03	0.932	0.014	-3.49	0.15	5.52	2.24
Switzerland	0.011	0.009	0.11	0.93	0.017	-3.74	0.14	3.14	1.86
United Kingdom	0.041	0.012	0	0.957	0.014	-3.24	0.17	3.63	1.11
United States	0.029	0.004	0	0.975	0.005	-3.47	0.11	8.3	0.74

This table reports posterior means and standard deviation for the regression of excess returns on the log dividend-price ratio $r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1}$ where the log dividend-price ratio follows an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$. μ_{x_i} and μ_{r_i} (in annualized, percentage terms) are the unconditional means for the dividend-price ratio and the continuously compounded excess return, respectively. $p < 0$ is the probability that β_i is smaller than zero. The longest data is for the US and spans 1952:Q1 to 2013:Q1, and the shortest spans 1986:Q1 to 2013:Q1.

Table 3: Bayesian estimates for the country-level parameters: alternative priors II

Panel A: Uninformative prior, weak economic restriction									
	β_i	s.d.	$p < 0$	ρ_i	s.d.	μ_{x_i}	s.d.	μ_{r_i}	s.d.
Australia	0.021	0.034	0.27	0.865	0.038	-3.25	0.07	6.23	2.76
Belgium	0.008	0.02	0.34	0.903	0.031	-3.15	0.16	4.36	2.95
Canada	0.014	0.017	0.22	0.949	0.02	-3.62	0.16	3.01	2.04
Denmark	0.001	0.013	0.46	0.956	0.019	-3.83	0.32	4.09	2.92
France	0.007	0.013	0.3	0.963	0.016	-3.55	0.28	3.53	2.48
Germany	0.001	0.019	0.49	0.939	0.023	-3.68	0.16	3.42	2.95
Italy	0.004	0.021	0.43	0.922	0.027	-3.66	0.18	2.89	2.89
Japan	0.001	0.011	0.46	0.977	0.011	-4.27	0.38	2.98	2.81
Netherlands	0.022	0.031	0.24	0.883	0.04	-3.41	0.12	4.63	2.99
Norway	0.019	0.03	0.26	0.882	0.039	-3.56	0.15	6.26	3.97
Spain	0.012	0.013	0.18	0.97	0.014	-2.99	0.37	3.58	2.24
Sweden	0.011	0.019	0.29	0.933	0.025	-3.57	0.21	6.88	3.26
Switzerland	0.006	0.019	0.36	0.914	0.03	-3.82	0.15	5.06	2.59
United Kingdom	0.041	0.026	0.06	0.912	0.029	-3.28	0.12	6.68	1.93
United States	0.023	0.01	0	0.978	0.01	-3.46	0.3	8.41	0.89
Panel B: Uninformative prior, strong economic restriction									
	β_i	s.d.	$p < 0$	ρ_i	s.d.	μ_{x_i}	s.d.	μ_{r_i}	s.d.
Australia	0.005	0.018	0.38	0.88	0.024	-3.27	0.07	7.27	2.67
Belgium	0.002	0.009	0.41	0.909	0.025	-3.19	0.15	5.71	2.62
Canada	0	0.007	0.48	0.964	0.009	-3.72	0.18	4.28	2.12
Denmark	0.001	0.006	0.47	0.957	0.016	-3.89	0.31	5.04	2.49
France	0.001	0.006	0.45	0.97	0.01	-3.66	0.28	4.74	2.37
Germany	0	0.007	0.52	0.941	0.015	-3.73	0.14	4.51	2.31
Italy	0.002	0.006	0.42	0.925	0.018	-3.72	0.14	4.2	2.31
Japan	0	0.005	0.49	0.979	0.008	-4.36	0.36	3.81	2.17
Netherlands	0.004	0.013	0.38	0.902	0.028	-3.46	0.13	6.37	2.95
Norway	0.004	0.013	0.38	0.897	0.028	-3.61	0.15	8.29	3.77
Spain	0.002	0.006	0.39	0.98	0.008	-3.16	0.44	4.83	2.44
Sweden	0.001	0.01	0.44	0.943	0.018	-3.62	0.24	7.99	3.28
Switzerland	-0.001	0.01	0.51	0.921	0.026	-3.84	0.16	5.91	2.5
United Kingdom	0.021	0.016	0.11	0.933	0.019	-3.33	0.13	7.55	2
United States	0.014	0.006	0.01	0.987	0.006	-3.53	0.4	8.4	0.95

This table reports posterior means and standard deviation for the regression of excess returns on the log dividend-price ratio $r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1}$ where the log dividend-price ratio follows an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$. Weak economic restrictions impose the unconditional forecasts of the equity premium to be non-negative for all countries. Strong economic restrictions additionally impose the same restriction on all forecasts. μ_{x_i} and μ_{r_i} (in annualized, percentage terms) are the unconditional means for the dividend-price ratio and the continuously compounded excess return, respectively. $p < 0$ is the probability that β_i is smaller than zero. The longest data is for the US and spans 1952:Q1 to 2013:Q1, and the shortest spans 1986:Q1 to 2013:Q1.