LSF Research Working Paper Series

N°. 16-03

Date: February 2016

Title: Return Predictability: Learning from the Cross-Section

Author(s)*: Julien Penasse

Abstract : Long histories of returns are needed but often lacking when estimating the equity premium. This paper studies stock return predictability from the perspective of a Bayesian investor who has access to international data. Learning across countries arises whenever this investor believes that international return processes share a common distribution. The model allows for samples of different lengths and introduces economic constraints on equity premium forecasts. Empirically, estimates are more reliable, an effect that manifests itself both in- and out-of-sample. International predictability also appears much less heterogeneous than previously reported.

***Corresponding Author's Address:** Tel: +352 46 66 44 58 24; Fax : +352 46 66 44 6835
E-mail address: julien.penasse@uni.lu

The LSF Research Working Paper Series is available online:
http://wwwen.uni.lu/recherche/fdef/luxembourg_school_of_finance_research_in_finance/working_papers

For editorial correspondence, please contact:
sophie.lux@uni.lu

University of Luxembourg
Faculty of Law, Economics and Finance
Luxembourg School of Finance
4 Rue Albert Borschette
L-1246 Luxembourg

Return Predictability: Learning from the Cross-Section*

Julien Penasse†

Abstract

Long histories of returns are needed but often lacking when estimating the equity premium. This paper studies stock return predictability from the perspective of a Bayesian investor who has access to international data. Learning across countries arises whenever this investor believes that international return processes share a common distribution. The model allows for samples of different lengths and introduces economic constraints on equity premium forecasts. Empirically, estimates are more reliable, an effect that manifests itself both in- and out-of-sample. International predictability also appears much less heterogeneous than previously reported.

*This version: February 2, 2016. In preparing this draft I have benefited from discussions with Michael Brennan, Bob Hodrick, Michael Johannes, Rajnish Mehra, Lorenzo Naranjo, Bradley Paye, Patrice Poncet, Davide Pettenuzzo, Richard Priestley, Barbara Rossi, Adrien Verdelhan, and comments from Christophe Boucher, Francesca Carrieri, Edouard Challe, Guillaume Chevillon, Stéphane Chrétien, Pierre Collin-Dufresne, Joost Driessen, Rik Frehen, Mike Gallmeyer, Frank de Jong, Bryan Kelly, Malik Kerkour, Tim Kroencke, Hao Liang, Benjamin Poignard, Kalle Rinne, Olivier Scaillet and Denitsa Stefanova, as well as seminar participants at Columbia Business School, BI Norwegian Business School, Luxembourg School of Finance, Paris Dauphine, Banque de France, Université d'Orléans, EPFL, Cambridge University, Ecole Polytechnique, ESSEC Business School, THEMA, Caisse des Dépôts et Consignations, the 2015 EFA meeting, the 2015 SoFiE conference, the 2015 NFA meeting, the Econometric Society 2014 European Winter Meeting, the RES 2015 meeting and the EEA 2014 meeting. 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.

†Luxembourg School of Finance, 4 rue Albert Borschette, 1246 Luxembourg, Luxembourg. E-mail: julien.penasse@uni.lu.

A fundamental quantity for economists and practitioners is the equity premium, the expected return on stocks in excess of the risk-free rate. In spite of vast literature, economists and investors still debate on how much the equity premium varies over time, i.e. whether aggregate returns are forecastable (Cochrane, 2011). For example, Pastor and Stambaugh (2012) argue that even after observing over 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. For example, Ang and Bekaert (2007) find that "none of the [return predictability] patterns in other countries resembles the US pattern." This paper incorporates three mechanisms to estimate the equity premium in predictive regressions: seemingly unrelated regressions (SUR), exchangeability and equity premium restrictions. These mechanisms exploit international data to improve estimation precision — hence the term "learning from the cross-section".[1]

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 main parameters of interest are the slopes β^{US} and β^{UK}. Previous literature has shown that even statistically insignificant estimates can produce large differences in asset allocations (Kandel and Stambaugh, 1996; Barberis, 2000).

The first mechanism treats 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 slope, say $\hat{\beta}^{US}$, is not efficient, because it treats the residual covariance matrix as diagonal, i.e., $cov(u_{t+1}^{US}, u_{t+1}^{UK}) = 0$. An improved mechanism should exploit the covariance of the residuals.

The second mechanism treats the parameters of each country's return process as

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

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 predictive relation in the United Kingdom. Prior literature considers these two elements of evidence as distinct facts.[2] In the framework developed in this paper, I 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.[3] Suppose histories of stock returns and dividend-price ratios are available for other countries but not for 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. On the other hand, if data *is* available 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 should be given more weight.

The third mechanism originates in the Bayesian paradigm that prior knowledge should be used in the estimation process. Assume for example that in 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 observation that the expected

[2] 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.

[3] Old Zembla's fields where my gray stubble grows,
And slaves make hay between my mouth and nose.

(From "Pale Fire" by Vladimir Nabokov.)

market risk premium should be positive (e.g. Merton (1980)). The investor would thus adjust accordingly their estimation of the domestic equity premium. I consider two forms of equity premium (EP) restrictions. The weak one requires the equity premium to be non-negative in the long run, leaving room for interpretations of return predictability in terms of temporary mispricing. The strong one further imposes that the equity premium is always non-negative.[4]

This paper contributes to a growing amount of literature that studies predictability from a Bayesian investment perspective.[5] Accounting for cross-sectional information is the main way in which this paper departs from that literature. While a number of papers study *time-variation* in predictability,[6] an analysis of variation *across countries* seems long overdue. In this respect, this paper can also be seen as a generalization of the pooled approaches to return predictability of Ang and Bekaert (2007) and Hjalmarsson (2010). On an intuitive level, this paper is also related to Kelly and Pruitt (2013, 2015), who propose a latent factor model that efficiently combines multiple predictors to forecast a single equity index. The mechanism presented herein is different because it uses the cross-section to learn about individual model parameters, in a context where individual data is scarce. In this dimension, this paper complements prior contributions that show how to estimate parameters with time series of different lengths (Stambaugh, 1997; Singleton, 2006; Lynch and Wachter, 2013).

Empirically, the main objective is to understand how much return processes differ across countries. I reexamine the international evidence of stock return predictability using a large data set of fifteen countries. I find that learning from the cross section increases precision and can substantially change the strength of return predictability. This gain in precision manifests itself both in- and out-of-sample, where forecasts based

[4]This paper is not the first to study sign restriction in the equity premium forecasts. Campbell and Thompson (2008) truncate the forecasts at zero and Pettenuzzo et al. (2014) introduce this idea in a Bayesian setup. Along 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.

[5]Wachter (2010) surveys this literature on Bayesian portfolio choice.

[6]See e.g., Henkel et al. (2011); Pettenuzzo and Timmermann (2011); Pastor and Stambaugh (2009, 2012); Dangl and Halling (2012); Johannes et al. (2014)

on the exchangeable prior routinely outperform forecasts based on models ignoring the cross-section. From an asset allocation perspective, the model produces weights that are reasonable and consistent across countries. The cost of ignoring cross-sectional information and equity premium restrictions is generally substantial. Estimation risk is also vastly mitigated so that stocks typically look safer in the long run than when an uninformative prior is assumed.

Importantly, international heterogeneity weakens considerably in contrast to previous literature. As presaged in Ang and Bekaert (2007), the main exception is the US, where the equity premium is unusually large and where return predictability is unusually strong. 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 critically on the theoretical interpretation of return predictability. Outside the US, the UK, and Japan, the evidence of predictability disappears when risk premia are constrained to be non-negative. Since this restriction is poorly supported by the data, the reader is left with two mutually exclusive conclusions; either predictability may be associated to some form of temporary mispricing and returns can sometimes be predictably negative, or expected returns may not fluctuate at all.

The remainder of this paper is organized as follows: Section I introduces the model which involves learning from the cross-section and provides an overview of the estimation procedure; Section II discusses the empirical results and illustrates the consequence of learning on asset allocation; Section III concludes.

I. Methodology

A. 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, $r_{i,t}$, are a linear function of a single predictor.[7] To ease comparison with earlier studies, I concentrate on the dividend-price ratio, $x_{i,t}$, which follows an AR(1) process. The model takes the form of a panel VAR:

$$r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1} \tag{1}$$

$$x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}. \tag{2}$$

This process characterizes excess return dynamics in $i = 1, 2, \ldots N$ countries over time $t = 0, 1, \ldots T$. 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 a $2N \times 2N$ covariance matrix, Σ. 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 regressions by assuming that the slope coefficients, β_i, are equal across countries (Ang and Bekaert, 2007; Hjalmarsson, 2010; Rapach et al., 2013). The first approach ignores cross-sectional information, 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 more efficient estimates can be obtained by considering them jointly (Zellner, 1962).

I now introduce assumptions regarding the covariance matrix, Σ. The conditional distribution of the country parameters is centered at the SUR estimator, which weights observations by Σ. The vector of cross-country innovations is of dimension $2N$ and therefore requires to handle $2N(2N + 1)/2$ covariance terms. Most of these terms are

[7]Of course, this framework could be extended to multiple predictors and assets. Such an extension is of particular interest in the context of strategic asset allocation (see e.g. Hoevenaars et al. (2014); Diris et al. (2014)). This is left for further research.

likely to be redundant and imprecisely estimated. Using an unconstrained covariance matrix matters little in-sample, but can result in poor out-of-sample performance. To circumvent this issue, I assume the following structure for the innovations:

$$
\begin{aligned}
u_{i,t+1} &= \delta_i^u u_{N,t+1} + \tilde{u}_{i,t+1} &\quad (3a)\\
v_{i,t+1} &= \delta_i^v u_{N,t+1} + \tilde{v}_{i,t+1} &\quad (3b)
\end{aligned}
$$

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

$$
\begin{aligned}
\begin{pmatrix} \tilde{u}_{i,t+1} \\ \tilde{v}_{i,t+1} \end{pmatrix} &\overset{i.i.d.}{\sim} \mathcal{N}(0, \tilde{\Sigma}_i) \text{ for } i = 1 \ldots N-1 \text{ and} &\quad (4a)\\
\tilde{v}_{N,t+1} &\overset{i.i.d.}{\sim} \mathcal{N}(0, \tilde{\sigma}^2_{N,\tilde{v}}) &\quad (4b)
\end{aligned}
$$

Finally, I assume that

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

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 allows the correlation between the unexpected returns, $u_{i,t}$, and the innovation in the predictor, $v_{i,t}$, to 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, $\tilde{\Sigma}$, Σ is recovered as

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

When sample lengths differ across countries, it is convenient to let N be the country with the longest history (the United States in this paper). This allows $\sigma^2_{N,u}$ to be estimated with greater precision.

B. Prior beliefs

When estimating Equations (1) - (2) in different countries, the econometrician will often conclude that the data-generating processes differ substantially across countries (see, e.g., Ang and Bekaert, 2007; Hjalmarsson, 2010; Rapach et al., 2013). The key premise of this paper is that these differences may simply be the effect of chance. In this subsection, I introduce two types of beliefs corresponding to this idea.

B.1. Exchangeability

Economic theory offers little guidance as to why predictability should be significant in some countries and not in others. The evidence of return predictability has led to the development of asset pricing models incorporating time-varying relative risk aversion (Campbell and Cochrane, 1999) or time-varying risks (Bansal and Yaron, 2004; Wachter and Warusawitharana, 2009; Gabaix, 2012). For countries to differ, risk sharing must be imperfect and/or deep parameters (e.g. preferences, elasticities) must vary across countries. I am not aware of an attempt to quantify such cross-country differences.[8] In the (reduced form) context of this paper, a reasonable assumption is 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:[9]

$$\zeta_i = \bar{\zeta} + \eta_i. \tag{7}$$

Such a prior is often denoted as *exchangeable* (de Finetti, 1964). This prior specification causes the Bayesian estimates to shrink towards the common mean vector $\bar{\zeta} = (\bar{\theta}, \bar{\beta}, \bar{\alpha}, \bar{\rho})$. Specifically, posterior precision (reciprocal of variance) is the sum of the cross-country and country-level precisions, while the posterior mean is a precision-weighted average of the common mean and the country-level estimate.

It is convenient to assume that the dispersion terms, η_i, are normally distributed.

[8] A growing literature relies on consumption based asset pricing models to study international financial markets, see e.g. Backus et al. (2001); Brandt and Santa-Clara (2002); Lustig and Verdelhan (2007); Colacito and Croce (2011).

[9] This sort of econometric model is often referred to as a random coefficients model (e.g., Swamy 1970).

However for the equity premium process to be stationary, it is necessary to truncate the distribution for the autoregressive parameters, ρ_i, to lie between -1 and 1. Formally, I assume that, for each country i, $\eta_i \sim \mathcal{N}(0, \Delta)$, when $\rho_i \in (-1, 1)$. Hence this prior puts a zero probability mass to $|\rho_i| > 0$. For the remaining parameters of the model, namely δ_i, $\tilde{\Sigma}_i$, $\tilde{\sigma}^2_{N,\tilde{v}}$, $\sigma^2_{N,u}$, $\bar{\zeta}$ and Δ, I choose priors uninformative in the sense of Jeffreys (1961) (see Appendix A).

One could 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 require a larger number of countries. Alternatively, one could increase N by studying return predictability at the individual stock or industry levels.[10] I leave this possibility for further research and concentrate on international evidence to facilitate comparison with the previous literature.

This prior combines the traditional approach (which treats each data-generating process independently) and the pooled approach (where cross-country heterogeneity is assumed to be negligible) as special cases. The former arises when $\Delta^{-1} = 0$ and when the off-diagonal terms of Σ are dogmatically set to zero for $i \neq j$). As noted in Jones and Shanken (2005), this amounts to specifying a prior in which the beliefs about predictability are independent across countries. The latter corresponds to the assumption that $\zeta_i = \bar{\zeta}$.

B.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. Although a negative equity premium could theoretically arise if stocks hedge against other risk factors, negative return forecasts are generally interpreted as evidence of mispricing (e.g., reflecting behavioral biases). For instance, Eugene Fama noted in 1991 that "there is no evidence that low D/P signals bursting bubbles, that is, negative

[10] 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.

expected stock returns" (Fama, 1991).

Prior literature has provided mixed evidence that excess stock returns are predictably negative. For example, Kothari and Shanken (1997) show how 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; however, they conclude that expected returns are never negative in the postwar period. Eleswarapu and Thompson (2007) propose a bootstrap test which 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 often forecast negative stock returns and disprove any interpretation in terms of risk premia.[11]

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, excess return series are stationary with unconditional mean

$$\mu_{r_i} = E(\theta_i + \beta_i x_{i,t} + u_{i,t}|\Psi) = \theta_i + \beta_i \mu_{x_i}, \tag{8}$$

where μ_{x_i} is the unconditional mean of the dividend-price ratio, $\mu_{x_i} = \frac{\alpha_i}{1-\rho_i}$. When excess returns are not predictable $\beta_i = 0$ and the equity premium is equal to θ_i. The weak restriction imposes that the unconditional equity premium is non-negative in all countries. Formally, the prior in (7) is modified to

$$\zeta_i = \bar{\zeta} + \eta_i, \quad \rho_i \in (-1, 1),\ \theta_i, \beta_i \in W_{\zeta_i},\ i = 1, \ldots, N, \tag{9}$$

where W_{ζ_i} is the set such that $W_{\zeta_i} = \{\theta_i + \beta_i \mu_{x_i} \geq 0\}$.

This restriction requires excess return forecasts to be non-negative in population, and so 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

[11]Negative excess return forecasts have also been documented in other markets, such as the dry bulk shipping industry (Greenwood and Hanson, 2015) and the fine art market (Penasse and Renneboog, 2015).

are always non-negative. In this case θ_i and β_i must belong to the set $S_{\zeta_i} = \{\theta_i + \beta_i \mu_{x_i} \geq 0; \min(\theta_i + \beta_i x_i) \geq 0\}$, where $x_i = \{x_{i,0}, \ldots, x_{i,T}\}$.

C. Bayesian estimation

I next describe how the model is estimated, with additional details in Appendix A. 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, given the common mean hyperparameters. The third stage corresponds to the distribution of these hyperparameters. This last stage is absent in non-hierarchical models, where second-stage parameters are estimated given a set of fixed prior beliefs.[12] Here, in contrast to these models, second-stage prior beliefs are updated in the third stage. This setup allows specification of uninformative priors in the third stage of the hierarchy, and thus to let the data "speak" about country heterogeneity.

Care must be taken when the predictor is persistent and when the predictor and return innovations are correlated. From the frequentist point of view, predictive regressions generally produce biased slope estimates (Stambaugh, 1986, 1999).[13] In contrast, from the Bayesian point of view, the posterior obtained from an uninformative prior is centered around the OLS estimate. Hence, the frequentist view is that OLS estimates should be corrected, while the Bayesian view is that they need not be. Recent research, however, shows that this disagreement depends on how initial conditions are modeled. Treating the initial value of the predictor as a fixed constant amounts to assuming that it conveys no information about the parameters of the data-generating process. This is often labeled as the conditional likelihood approach in the literature. Alternatively, it is possible to construct an 'exact' likelihood function incorporating the stochastic nature of the initial condition (Box and Tiao, 1973). Stambaugh (1999) and Marcet and Jaroci (2014) find

[12]This is by far the most common setup in the literature. Models that instead use hierarchical models include Wachter and Warusawitharana (2009, 2015), Pastor and Stambaugh (2001) and Pettenuzzo and Timmermann (2011).

[13]Several paper study the Stambaugh bias in the frequentist domain, e.g. Lewellen (2004); Campbell and Yogo (2006); Chiquoine and Hjalmarsson (2009).

that when the initial value is treated as a fixed constant, frequentist and Bayesian estimates can differ substantially, but that the difference narrows when it is treated as a random variable. In this light, I follow the exact likelihood approach throughout (see Appendix A). I empirically evaluate the effect of following the exact likelihood approach versus the conditional likelihood approach in Section II.B.

Posterior distributions do not have analytical closed forms in this model. Closed form solutions for each parameter are only available conditional on the remaining parameters of the model. For example, the common mean parameter posterior is a precision-weighted average of individual country parameters. At the same time, this common mean serves as prior for the individual coefficients. Thus both parameters depend on each other. We can nevertheless compute the posterior distributions with a Gibbs sampler (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 based on Chib and Greenberg (1995a). The algorithm is augmented with Metropolis-Hastings steps when needed (see Chib and Greenberg, 1995b; Griffiths, 2003; Johannes and Polson, 2003 and Appendix B). 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.

II. Empirical Results

A. Data

My data set consists of quarterly data for fifteen OECD countries with at least 25 years of data: 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).

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). The dividend-price ratio is directly related to future stock returns via the present-value identity, which makes it a natural choice to forecast returns. Previous research has also argued that dividends could be useful to estimate the unconditional equity premium (e.g., Fama and French (2002); Donaldson et al. (2010)), which is another purpose of the present paper. The US stock data is the Standard & Poor's 500 value-weighted index, downloaded from 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 from the total equity return (in logs). The former is downloaded from the OECD statistics or FRED (USA), and the IMF International Financial Statistics depending on availability. Following convention, I compute the log dividend-price ratio as the twelve-month moving sum of dividends minus the log of stock prices.

B. Posterior distributions

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 correspond to the 75% and 99% credible sets, respectively. Since the model has several features that may affect posterior beliefs simultaneously, Table I supplements the boxplots by providing posterior means (the Bayesian "estimates") for several variants of the model, as well as OLS estimates. Table I also indicates posterior standard deviations (equivalent to conventional standard errors) in parentheses. The last row of Table I gives the standard deviation of the slopes across countries, normalized by the standard deviation of OLS coefficients. Thus, for a given model, a value below unity indicates coefficients that are less dispersed than OLS coefficients.

The top panel of Figure 1 gives the beliefs based on uninformative priors (column (2) in Table I). The posterior means are mostly positive (except Germany and Italy). A Bayesian investor would conclude that the predictive coefficient is significantly different from zero in the US, 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 shows the beliefs based on the exchangeable prior, seemingly unrelated regressions (SUR), and weak economic restrictions (column (6) in Table I). For comparison purposes, the classical SUR estimates are also reported (denoted by diamonds). The cross-country heterogeneity documented in Panel (a) clearly weakens. Table I shows that the cross-country dispersion drops to 0.338 with the exchangeable prior alone, which means the slope coefficients are less dispersed by about two thirds. Dispersion falls below 0.278 when the SUR model is introduced, which suggests that exchangeability explains most of the reduction in cross-country dispersion. This is unsurprising since the prior shrinks posteriors toward the cross-sectional average. Comparing columns (5) and (6), we also see that the weak economic restriction has a negligible impact on the slope coefficients, which is also expected.

The increase in precision is visually striking. The 99% credible sets are typically twice tighter than in the uninformative prior. We see in Table I that this gain comes both from the exchangeable prior and the SUR model. Countries with limited data (e.g. the Netherlands) and countries where the benchmark estimates are imprecise (e.g. Australia, Germany) gain more from cross-sectional learning. The mean predictive coefficient is no longer negative for Italy and is more than halved for 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 uninformative prior 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 predictability (with posterior probabilities of a positive coefficient larger than 99%).[14]

Bayesian estimates of the predictive slopes with strong equity premium constraints can be seen on Panel (c) of Figure 1. Precision increases again and individual coefficients further shrink toward the common means (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 this case, for each country simultaneously). Therefore the number of constraints grows in proportion to the size of the sample. This large set of potentially binding constraints pins down the coefficients of the predictive regression, and thus increases precision.[15]

The evidence of return predictability is considerably weaker, however. 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). This occurs because too large of a predictive slope violates the strong EP restriction. Hence more probability mass is given to draws where the predictive coefficient is low, including draws where it is negative.

Table I also gives the impact of correctly treating the initial observation of the dividend-price ratio as random (the exact likelihood approach) against the more common conditional likelihood approach. Stambaugh (1999) shows that in the context of predictive regressions, this choice can have unexpectedly deep consequences for both Bayesian estimation and portfolio choice.[16] The first column shows OLS estimates, which are equivalent to posterior means with conditional likelihoods and uninformative priors, and

[14]This result persists if the analysis is restricted to a balanced panel instead.

[15]A 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.

[16]See also Wachter and Warusawitharana (2009, 2015).

the second column gives the posterior means with exact likelihoods. While precision increases substantially, predictive slopes change little on average. Additionally, we see that there is little convergence of the country parameters. It is of interest to compare columns (4) and (5), which give the effect of the exact likelihood in the SUR model with an exchangeable prior. The coefficients and standard errors are almost identical, indicating that these two key features marginalize the role of the likelihood specification.

Overall these results suggest that the exchangeable prior is quite informative. This means there are, by large, enough similarities to make a meaningful use of cross-country information. To further understand how similar the parameters are across countries, Table II gives the posterior beliefs of the common mean hyperparameters, which also serve as prior beliefs in the estimation of individual coefficients. When the weak economic restriction is imposed, the common mean for the predictability coefficient is clearly positive. Its posterior distribution is centered at 0.020 with a 99% credible set of $[0.000, 0.041]$. 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 $[-0.006, 0.016]$, and the posterior probability that the coefficient is negative (not shown) is 11.3%. Also evident from Table II is that the log-dividend price ratio is more persistent when the equity premium constraint is entertained. Stambaugh (1999) shows that if the correlation between innovations in returns and innovations in the predictive variable is negative, underestimating the persistence, ρ_i, leads to overestimating the predictive slopes, β_i. This correlation is always strongly negative for the dividend-price ratio. 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.

C. The long-run equity premium

In this section, I discuss how the exchangeable prior and the economic restriction influence posterior beliefs about the unconditional or long-run equity premium (EP). The unconditional EP is a nonlinear function of the model parameters, which is given by Equation

(8). The posterior distributions implied by the same prior combinations as before are plotted in Figure 2. All numbers are annualized. The equity premium is traditionally computed as the sample average of excess returns, and the latter is also reported in blue diamonds for comparison purpose. In the same fashion as in Table I, Table III reports the posterior means and standard deviations for several variants of the model.

The top panel of Figure 2 displays boxplots of the posterior distribution for the uninformative prior. We observe a large heterogeneity across countries, which parallels the heterogeneity in sample averages of excess returns.[17] Column (2) of Table III shows that posterior means range from -1.8% (Italy) to 8.2% (US). Except for 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.[18] Hence, a Bayesian investor would give a large 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).

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

[17] Also noted is that the posterior means are consistently below the sample average. As noted in Avdis and Wachter (2015), the difference between the sample average of excess returns and the model-implied long-run equity premium can be expressed as (country indices dropped for convenience and use hats to denote posterior means):

$$\frac{1}{T}\sum_{t=1}^{T} r_t - \hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\theta} + \hat{\beta} x_t + \hat{u}_t\right) - \hat{\theta} - \hat{\beta}\hat{\mu}_x. \tag{10}$$

In general the shocks to excess returns $\hat{u}_t$ will not sum up to zero. We can rewrite this difference as:

$$\frac{1}{T}\sum_{t=1}^{T} r_t - \hat{\mu}_r = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t + \hat{\beta}\left(\bar{x} - \hat{\mu}_x\right), \tag{11}$$

which is the sum of unexpected return shocks over the sample period (first term) and a term reflecting whether the long-term dividend-price ratio is estimated to be above or below its sample average, $\bar{x}$ (second term). The long-run dividend-price ratio is generally close to the sample average, so that the second term is negligible. Hence shocks to excess returns must have been positive on average.

[18] In a similar exercise, Wachter and Warusawitharana (2009) and Avdis and Wachter (2015) study the impact on the long-run equity premium of the assumption regarding the initial value of the predictor. They find that using the exact 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 find that the equity premium is even more dispersed than the one shown on Panel (a), which relies on the exact likelihood.

negative. The posteriors are all above zero (by construction) but are also much less dispersed. Posterior means range from 1.8% to 8.2% (again for Italy and the United States, see column (6) of Table III). Recall that the weak economic restriction is imposed to hold simultaneously for all countries. As a result, the distributions are not truncated at zero, indicating 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, Japan and 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 3.2% to 7.9%. Coincidentally, Dimson et al. (2008) report a similar range of 2.8% to 7.1% for the same countries but computed on a century of data.[19]

D. 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 lets the equity premium be negative sometimes, while the strong form imposes that it must always be non-negative. In this section, I test whether the strong restriction is supported by the data. As noted 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 unity. The Bayes rule implies that the posterior odds ratio equal the prior odds multiplied by the ratio of marginal likelihoods.

[19]The ranking differ, 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 inference purposes.

The marginal likelihood can be difficult to compute in high-dimensional models like the present one (see e.g. Chib 1995). Fortunately, in our case, 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 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 of non-negative expected returns and information on the share of negative forecasts across countries are reported in Table IV. The posterior probabilities range from 2% (Japan) to 42% (Switzerland), which corresponds to posterior odds above unity in all countries. In the Bayesian sense, one should reject the strong restriction in favor of the weak one but the evidence is far from decisive. The posterior probability is below 10% for only three countries (Japan, Spain and the UK). For the US, it is an unimpressive 10.4%, corresponding to posterior odds of 8.6.

It is instructive to study how often predicted excess returns are negative. Table IV shows that the share of negative forecasts varies considerably across countries. In the United States, only 5.4% of the forecasts are negative (i.e. forecasts are negative for about 13 quarters, mostly corresponding to the years 1999-2001). At the other end of the spectrum, about half of Japan excess return forecasts are negative. In the remaining countries, roughly 20% of the forecasts are negative. The United States has the lowest share of negative forecasts, and yet a relatively large posterior odds against a model where returns are never negative. This stems from the fact that the average forecast is generally higher in the US, which has the highest unconditional equity premium (see Table III), 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. However, a somewhat uncomfortable conclusion remains. As noted in Section II.B, the predictive coefficients are rarely significant when the strong restriction is imposed. Thus under the paradigm that excess return predictability reflects fluctuations in (positive) risk premia, there is 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.

E. Asset allocation analysis

The previous sections have discussed the statistical significance of return predictability. I now examine the implications of various prior beliefs for optimal asset allocation. To do so, I take the point of view of a Bayesian investor with a given predictive distribution who maximizes his expected utility. Forcing this investor to hold a suboptimal portfolio is a way to gauge 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.

As in previous portfolio choice studies (for example Kandel and Stambaugh, 1996; Avramov, 2002; Shanken and Tamayo, 2012), I consider an individual investor with power utility and coefficient of relative risk aversion, A. This investor maximizes at time T

$$E\left[\left.\frac{W_{T+1}^{1-A}}{1-A}\right|D\right] \tag{12}$$

for $A = 5$, where W_{T+1} is his time $T+1$ wealth as he invests a fraction $0 \leq \omega \leq 1$ of his wealth in stocks.[20] Without loss of generality, I normalize initial wealth to unity, $W_T = 1$, so that his terminal wealth is

$$W_{T+1} = \omega \exp\{r_{i,T+1} + r\} + (1-\omega)\exp\{r\}. \tag{13}$$

In the above, $r_{i,T+1}$ is the excess stock return in country i, and r is the risk-free rate, which I assume is constant and equal to 120 basis points (bps) per quarter throughout.[21]

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

[21] Observe that, from (12) and (13), the optimal allocation to stocks does not depend on the level of the risk-free rate, although the level of utility does.

The expectation (12) is taken with respect to the predictive distribution

$$p(r_{i,T+1}|D) = \int p(r_{i,T+1}|\Psi)p(\Psi|D)d\Psi. \tag{14}$$

This 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

$$\omega^* \approx \arg\max_{\omega} \frac{1}{L}\sum_{l=1}^{L}\left(\frac{\left[\omega \exp\{r_{i,T+1}^{(l)} + r\} + (1-\omega)\exp\{r\}\right]^{1-A}}{1-A}\right) \tag{15}$$

where $r_{i,T+1}^{(l)}$, $l = 1, \ldots, L$ are returns drawn from the MCMC output using the predictive distribution (14).[22]

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

$$\frac{\exp\{(CER + r)\}^{1-A}}{1-A} = E\left[\frac{(\omega \exp\{r_{i,T+1} + r\} + (1-\omega)\exp\{r\})^{1-A}}{1-A}\middle| D\right]. \tag{16}$$

The CER of a risky portfolio is 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 the investor when he holds a given portfolio.

To understand how prior beliefs affect asset allocation, I first combine the exchangeable prior with the weak then strong economic restriction and alternatively ignore each belief. The resulting portfolio weights are reported in Table V. 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. I follow Shanken and Tamayo (2012) and consider values of the dividend-price ratio at

[22] As 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.

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. The next columns show the allocations for the various suboptimal portfolios associated with an investor who ignores exchangeability, the economic restriction, or both. Symmetrically, Table V reports the optimal excess CERs associated with the optimal allocation and the CER losses for the suboptimal portfolios.

I first discuss the results for the weak EP restriction. The second column of Table V corresponds to the optimal asset allocation when the dividend-price ratio is at its unconditional mean. 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. It is apparent that the typical investor would time the market aggressively. Expected returns rise with the log dividend-price ratio, and so does the typical allocation to stocks. For example, an Australian investor would respectively allocate 11%, 28%, and 45% 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. For the latter, the equity premium is sensibly higher (as can be seen in Figure 2). A US investor would want to respectively invest 25%, 84%, and 100% of his wealth in stocks.

Panel (b) of Table V shows that investors who believe that the equity premium is always positive will be more skeptical about return predictability. Hence their allocation to stocks will be less sensitive to the level of the dividend-price ratio. For instance, the Australian investor barely changes his allocation from 34% to 41% as the level of the predictor increases. The main exceptions are the UK and the US, in line with the results of Section II.B that predictability persists after imposing the strong economic restriction.

CERs corresponding to optimal portfolios are reported in the first three columns of Table VI. The CERs are annualized and expressed in basis points over the risk-free rate. For example, a Spanish investor would choose to invest all his wealth to the risk-free rate when the predictor is 1.5 standard deviations below its mean. This investor would thus earn a zero CER as reported in the fourth column. Returning to the Australian investor, we see they get CERs of 12, 76 and 192 bps, respectively. For a US investor, the gains are dramatically higher: 36, 408, and 1060 bps, respectively. This reflects the higher US equity premium and larger allocations to stocks as the dividend-price ratio increases.

The last three columns of Table VI give the CER loss of an investor who is forced to hold the portfolio corresponding to the uninformative prior. The results indicate that ignoring exchangeability and the equity premium restrictions correspond to potentially large economic losses. For example, a German investor with an uninformative prior tends to ignore return predictability, so that their allocation is largely independent from the level of the dividend-price ratio. When the dividend-price ratio is high, Table V shows that this investor with weak (strong) restriction would suffer a certainty equivalent loss of 47 (21) bps. These are large numbers, having in mind the total CER from investing in the stock market, which is 148 (100) bps. Of course, the utility costs vary considerably across countries and configurations. They are stronger in countries where predictive slopes differ the most from the cross-section (e.g. Australia, Italy, the UK). The costs are mostly smaller for the US, which enjoys the longest data sample.

F. *Long-term volatility*

So far I have limited my analysis to short horizons, but intuition suggests that parameter uncertainty or 'estimation risk' becomes proportionally more important when investment horizon increases. Previous research has provided support to this intuition (Stambaugh, 1999; Barberis, 2000). More broadly, the asset allocation literature has examined the riskiness of stocks over long horizons. This question has received growing attention following the thought-provoking paper of Pastor and Stambaugh (2012) who claim that stocks may be riskier in the long run.[23]

This section performs the same analysis as Pastor and Stambaugh (2012) for the

[23] See, e.g. Hoevenaars et al. (2014), Johannes et al. (2014), Carvalho et al. (2015) and Avramov et al. (2015).

model specifications. Interest centers on the following quantity:

$$\sigma_i^2(k|D) = \frac{1}{k} Var(r^i_{t\to t+k}|D) \tag{17}$$

where $r^i_{t\to t+k} = \sum_{j=1}^{k} r_{t+j}$ is the cumulated k-quarter ahead excess return in country i. Pastor and Stambaugh (2012) show that Equation (17) can be decomposed as

$$\sigma_i^2(k|D) = E\left(\sigma_i^2(k|D,\Psi_i)\middle|D\right) + Var\left(E(r^i_{t\to t+k}|\Psi_i, D)\middle|D\right). \tag{18}$$

The first term on the right hand side of the above equation is the conditional variance, which depends on the estimated model parameters and has been studied in Campbell and Viceira (2002). The second term is the variance of the conditional mean; this is the estimation risk term which adds positively to the expected conditional variance.

Figure 3 plots 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 solid lines correspond to the uninformative prior specification. For most countries, predictive volatility is increasing for long horizons.[24] In predictive regressions, conditional variance is the byproduct of the correlation between innovations in returns and innovations in the predictive variable and the predictive slope β_i (see Campbell and Viceira, 2005). The predictive variable is strongly negative, while the predictive slope is typically positive, so that the conditional variance is downward sloping. The fact that predictive volatility increases for long horizons reflects estimation risk. Estimation risk becomes stronger with investment horizon, increasing the slope of the term structure of volatility. This risk is particularly prevalent in countries with shorter data (e.g. the Netherlands), or where the predictive slope is estimated with low precision (e.g. Switzerland).

Predictive volatilities with the exchangeable prior and weak EP restriction are repre-

[24]Notice 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.

sented by dotted lines. With the weak restriction, the curves become downward sloping for most countries, which reflects the partial shrinkage of slope coefficients toward the common mean and the much lower estimation risk. We have seen in Section II.B that strong restriction considerably weakens posterior beliefs in return predictability. This is mirrored in the predictive volatilities (dashed lines), which flatten significantly. Overall these results show how beliefs about long-horizon volatility are sensitive to prior specification, and give another indication of the cost of uninformative beliefs.

G. Out-of-sample analysis

This final section evaluates the out-of-sample performance of these priors. To this end, I compare the portfolio choices of an investor who actively attempts to time the market and an investor who dogmatically believes $\beta_i = 0$ and whose predictive distribution only depends on the history of past returns. This benchmark is motivated by Goyal and Welch (2008), who show that a simple forecasting rule based on the historical average of past returns outperforms most predictors suggested by the prior literature. Note that from the forecasting point of view, the exchangeable prior can be seen as a different way to restrict coefficient estimates. For example, pooling models have been used to forecast GDP growth (e.g. Mittnik, 1990; Zellner and Hong, 1989). Hjalmarsson (2010) find 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.[25]

For most countries, the series span about 40 years. The out-of-sample analysis thus starts in 1993 and ends in 2013. This leaves minimally 20 years of in-sample training

[25]Several 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, 2015) develop a class of informative priors which 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. Paye and Timmermann, 2006; Henkel et al., 2011; Pettenuzzo and Timmermann, 2011; Dangl and Halling, 2012; Johannes et al., 2014).

period for eleven countries with at least 40 years of data. Each year, I estimate the model using data available up to that year and compute the predictive distribution of returns.[26] I then calculate the optimal stock allocation for an investor with power utility, as described in Section II.E. Performance is assessed using the certainty equivalent returns (CER) over the risk-free rate, as well as Sharpe ratios.

Table VII reports excess CERs and Sharpe ratios for the historical average benchmark and for the three specifications considered earlier in this paper. Both metrics are annualized, i.e. the CERs are multiplied by 40000 to express them in basis points and the Sharpe ratios are multiplied by 2. Following Shanken and Tamayo (2012), I use asymptotic standard errors to assess statistical significance. Standard errors for the excess CERs are calculated with the Delta method (see Shanken and Tamayo 2012). For the Sharpe ratios, I use the approach described in Jobson and Korkie (1981). In both cases values that are 1.645 standard deviation above zero are reported in Table VII with stars.

The results highlight the poor performance of uninformative priors, although in some countries the historical average performs worse. In the United States, uninformative priors deliver a performance about 70 bps below the historical average, which is in line with Goyal and Welch (2008) and the subsequent literature. The exchangeable prior delivers superior performances, either with the weak or strong restriction. On average, excess CERs double with the weak restriction and increase even more with the strong restriction. Sharpe ratios also double on average with the exchangeable prior, the weak restriction performing slightly better in that case.

Overall the exchangeable prior improves out-of-sample forecasts in the large majority of countries. Both versions of the model outperform the historical average almost systematically. The outperformance of the model over the historical average is interesting because this exercise is performed on a period (1993-2013) in which the prior literature finds little evidence of international predictability (see e.g. Henkel et al., 2011; Schrimpf, 2010; Hjalmarsson, 2010).

[26] I update the model every year to save on computation time, although the data is sampled quarterly.

III. Conclusion

Regressions of excess stock returns on predictors such as the dividend-price ratio yield estimates that are highly imprecise. Imprecise estimates may lead researchers to conclude that return processes vary considerably across countries and may lead practitioners to make poor asset allocation decisions. This paper studies the portfolio choices of an investor who assumes that international return processes are exchangeable. This Bayesian investor makes use of the cross-sectional correlation of the data and learns about the common means of the parameters. The resulting model nests, as special cases, the traditional approach which considers individual countries separately and the alternative approach that pools information across countries.

International return processes are found to be much more alike than previously reported and, consequently, that investment decisions differ little across countries. The resulting portfolio weights deliver superior out-of-sample performance in the large majority of countries. More broadly, this study supports the idea that international comparisons are worthwhile and can improve model estimation and comparison. In particular, my results suggest that the US equity premium is not representative of industrialized countries, because it is characterized by both a high equity premium and a strong predictability of stock returns.

This paper suggests several avenues for future research. First, the framework developed here can be extended to study the drivers of asset returns in more disaggregated panels. For example, there is a growing interest in linking return predictability with stock characteristics (e.g. Cochrane (2011)). Furthermore, many leading asset pricing models suggest that "latent" and hard to detect processes drive asset price fluctuations. Exchangeability is a natural way to estimate such processes. Important literature also suggests that learning can explain many asset pricing puzzles, including predictability (see Pastor and Veronesi (2009) for a review). The equilibrium implications of learning from other countries remain unexplored and are left for further research.

Appendix A. Likelihood and prior beliefs

The investor begins with the prior beliefs described in Section I.B. These beliefs can be written as a prior density $p(\Psi)$, where $\Psi = (\zeta, \delta_i, \tilde{\Sigma}_i, \tilde{\sigma}^2_{N,\tilde{v}}, \sigma^2_{N,u}, \bar{\zeta}, \Delta)_{i=1}^{N-1}$. Let $D \equiv \{r_1, \ldots, r_T, x_0, x_1, \ldots, x_T\}$ represent the data available to the investor at time T.

The posterior density of the parameter Ψ is computed as

$$p(\Psi|D) \propto L(D|\zeta, \Sigma)p(\Psi) \tag{A1}$$

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

$$y_{t+1} = X_t \zeta + \epsilon_{t+1}, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma) \tag{A2}$$

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:[27]

$$\begin{aligned} L(D|\zeta, \Sigma) \quad \propto \quad & |\Sigma|^{-T/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_{t+1} - X_t\zeta)'\Sigma^{-1}(y_{t+1} - X_t\zeta)\right] && \text{(A3a)} \\ & \times \mathcal{N}(\mu_x, V_x). && \text{(A3b)} \end{aligned}$$

The first block of Equation (A3) describes the likelihood function for the observations $1, \ldots, T$ and treats the first observations, described by the vector x_0, as a constant. The

[27]See also Hsiao and Pesaran (2008) for a textbook treatment.

second block corresponds to the likelihood of the first vector of observations. In this setting, this distribution has an unconditional mean of[28]

$$\mu_x = (I_N - \rho)^{-1}\,\alpha \tag{A4}$$

and a variance V_x satisfying

$$\mathrm{vec}(V_x) = (I_{N^2} - (\rho \otimes \rho))^{-1}\,\mathrm{vec}(\Sigma_x) \tag{A5}$$

where α is a vector of intercepts and ρ is a diagonal matrix of slopes in Equation (2), and where Σ_x is the submatrix of Σ containing covariances for the vector x_t.

It is common to refer to Equation (A3) as the exact likelihood and to Equation (A3a) as the conditional likelihood (see Stambaugh, 1999; Wachter and Warusawitharana, 2009, 2015).

It is convenient to rewrite Equation (A2) as

$$y = X\zeta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma \otimes I_T) \tag{A6}$$

where $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_N)$, $y = (y_1', y_2', \ldots, y_N')'$, $\epsilon = (\epsilon_1', \epsilon_2', \ldots, \epsilon_N')'$, y_i and ϵ_i are $2T \times 1$ vector of the left-hand side variables and innovation terms for country i. Finally, $X = \mathrm{diag}(X_1, X_2, \ldots, X_N)$, $X_i = (\iota_T, (x_{i,1}, x_{i,2}, \ldots, x_{i,T})')$ where ι_T denotes a vector of ones.

Equation (7), which corresponds to the second stage of the hierarchy, can be rewritten as

$$\zeta = A\bar{\zeta} + \eta, \quad \rho_i \in (-1, 1),\ i = 1, \ldots, N \tag{A7}$$

with

$$\eta \sim \mathcal{N}(0, \Delta_\zeta),\ \Delta_\zeta^{-1} = I_N \otimes \Delta^{-1} \tag{A8}$$

and where $A = (I_G, I_G, \cdots I_G)'$ 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

[28] See, e.g., Hamilton (1994).

the Kronecker product. Equations (A7) - (A8) 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:

$$p(D|\zeta, \Sigma) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}\left(y - X\zeta\right)' \left(\Sigma^{-1} \otimes I_T\right) \left(y - X\zeta\right)\right\} \times \mathcal{N}(\mu_x, V_x), \tag{A9}$$

where μ_x, and V_x are given in Equations (A4) and (A5).

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

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

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 $\Psi = (\zeta, \delta_i, \tilde{\Sigma}_i, \tilde{\sigma}_{N,\tilde{v}}^2, \sigma_{N,u}^2, \bar{\zeta}, \Delta)_{i=1}^{N-1}$ is thus[29]

$$p(\Psi|D) \propto p(D|\zeta, \Sigma)p(\zeta|\bar{\zeta}, \Delta_\zeta^{-1})p(\bar{\zeta})p(\Sigma)p(\Delta_\zeta^{-1}) \tag{A10}$$

where

$$p(\Sigma) \propto \prod_{i=1}^{N-1} p(\delta_i, \tilde{\Sigma}_i)p(\delta_N^v, \sigma_{N,\tilde{v}}^2)p(\sigma_{N,u}^2). \tag{A11}$$

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

Appendix B. Sampling from the posterior

The Gibbs sampler consists of four blocks, which I detail below.[30] 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 20,000 draws that I discard; I then generate a subsequent sample of 100,000 draws from which every tenth draw is retained 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 the first and last part of the sample.[31]

1. Conditional posterior for ζ

Viewing the joint posterior in Equation (A10) 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):

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

Note that

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

[30] In 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 32).

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

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

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

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

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) which can be rewritten as

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

where

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

Since 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. 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:

$$\zeta | D, \Psi_{-\zeta} \sim \mathcal{N}(m_\zeta, V_\zeta) \times \mathcal{N}(\mu_x, V_x), \quad \rho_i \in (-1, 1),\ i = 1, \ldots, N. \tag{B1}$$

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

$$\begin{aligned} p(\zeta|\bar{\zeta}, \Delta_\zeta) &\propto |\Delta_\zeta|^{-N/2} \exp\left\{-\frac{1}{2}(\zeta - A_0\bar{\zeta})'\Delta_\zeta^{-1}(\zeta - A_0\bar{\zeta})\right\}, \\ & \quad \rho_i \in (-1,1),\ \theta_i, \beta_i \in E_{\zeta_i},\ i = 1, \ldots, N \end{aligned}$$

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

$$\zeta|\Sigma, \bar{\zeta}, \Delta_\zeta \sim \mathcal{N}(m_\zeta, V_\zeta) \times \mathcal{N}(\mu_x, V_x), \quad \rho_i \in (-1,1),\ \theta_i, \beta_i \in E_{\zeta_i},\ i = 1, \ldots, N. \tag{B2}$$

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. Defining the $T \times 2$ matrices $e_i = (\epsilon'_{i,1}, \ldots, \epsilon'_{T,1})'$ and $\tilde{e}_i = (\tilde{\epsilon}'_{i,1}, \ldots, \tilde{\epsilon}'_{T,1})'$, and the $T \times 1$ vector $u_N = (u_{N,1}, \ldots, u_{N,T})'$, 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 $\hat{\delta}_i = (u_N' u_N)^{-1} u_N' e_i$.[32]

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

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

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

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

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.

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

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

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

and covariance

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

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

From Equation (A10) and using the definition $\Delta_\zeta^{-1} = I_N \otimes \Delta^{-1}$, the conditional posterior

[32] A 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 earlier 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.

density for the dispersion parameters is given by:

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

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

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Avdis, Efstathios, and Jessica A. Wachter, 2015, Maximum likelihood estimation of the equity premium, Working paper, NBER Working Paper No. w19684.

Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Avramov, Doron, Scott Cederburg, and Katarina Lucivjansk, 2015, Are stocks riskier over the long run? Taking cues from economic theory.

Backus, David K, Silverio Foresi, and Chris I Telmer, 2001, Affine Term Structure Models and the Forward Premium Anomaly, *The Journal of Finance* 56, 279–304.

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).

Box, George E. P., and George C. Tiao, 1973, *Bayesian inference in statistical analysis* (Addison-Wesley, Reading, MA).

Brandt, Michael W., and Pedro Santa-Clara, 2002, Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* 63, 161–210.

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 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, 2005, The term structure of the risk-return tradeoff, *Financial Analysts Journal* 61, 34–44.

Campbell, John Y., and M Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.

Carvalho, Carlos M., Hedibert F. Lopes, and Robert E. McCulloch, 2015, On the Long Run Volatility of Stocks.

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.

Chiquoine, Benjamin, and Erik Hjalmarsson, 2009, Jackknifing stock return predictions, *Journal of Empirical Finance* 16, 793–803.

Cochrane, John H., 2011, Presidential address: discount rates, *Journal of Finance* 66, 1047–1108.

Colacito, Riccardo, and Mariano M. Croce, 2011, Risks for the Long Run and the Real Exchange Rate, *Journal of Political Economy* 119, 153–181.

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).

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, Franz Palm, and Peter Schotman, 2014, Long-term strategic asset allocation: an out-of-sample evaluation, *Management Science* 61, 2185 – 2202.

Donaldson, R. Glen, Mark J. Kamstra, and Lisa A. Kramer, 2010, Estimating the Equity Premium, *Journal of Financial and Quantitative Analysis* 45, 813–846.

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.

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, 2002, The Equity Premium, *Journal of Finance* 57, 637 – 659.

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.

Greenwood, Robin, and Samuel G Hanson, 2015, Waves in Ship Prices and Investment, *Quarterly Journal of Economics* 130, 55–109.

Griffiths, William E., 2003, Bayesian inference in the seemingly unrelated regressions model, in David Giles, ed., *Computer-Aided Econometrics* (CRC Press).

Hamilton, J. D., 1994, *Time series analysis* (Princeton University Press, Princeton, NJ).

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.

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.

Hsiao, Cheng, and M. Hashem Pesaran, 2008, Random coefficient panel data models, in Lászlo 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, Clarenden).

Jobson, JD, and BM Korkie, 1981, Performance hypothesis testing with the Sharpe and Treynor measures, *Journal of Finance* 36, 889–908.

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).

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.

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Kelly, Bryan, and Seth Pruitt, 2015, The three-pass regression filter: A new approach to forecasting using many predictors, *Journal of Econometrics* 186, 294–316.

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.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

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.

Lustig, Hanno N., and Adrien Verdelhan, 2007, The Cross Section of Foreign Currency Risk Premia and Growth Risk Consumption, *American Economic Review* 97, 89–117.

Lynch, Anthony W., and Jessica A. Wachter, 2013, Using Samples of Unequal Length in Generalized Method of Moments Estimation, *Journal of Financial and Quantitative Analysis* 48, 277–307.

Marcet, Albert, and Marek Jaroci, 2014, Contrasting Bayesian and Frequentist Approaches to Autoregressions: the Role of the Initial Condition, Working paper, Barcelona GSE Working Paper Series No. 776.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *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, 2001, The equity premium and structural breaks, *Journal of Finance* 56, 1207–1239.

Pastor, Lubos, and Robert F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583–1628.

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, 2009, Learning in Financial Markets, *Annual Review of Financial Economics* 1, 361–381.

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, 2015, *Bubbles and Trading Frenzies: Evidence from the Art Market*, Working paper, Tilburg University.

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., 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).

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.

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 uncertainty, *Journal of International Money and Finance* 29, 1256–1282.

Shanken, Jay, and Ane Tamayo, 2012, Payout yield, risk, and mispricing: A Bayesian analysis, *Journal of Financial Economics* 105, 131–152.

Singleton, Kenneth J., 2006, Inference with unequal-length samples, in *Empirical dynamic asset pricing: model specification and econometric assessment*, 88–93 (Princeton University Press).

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., 1986, *Bias in regressions with lagged stochastic regressors*, Working paper, Center for Research in Security Prices, Graduate School of Business, University of Chicago.

Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.

Stambaugh, Robert F., 1999, Predictive Regressions, *Journal of Financial Economics* 54, 375–421.

Swamy, Paravastu A. V. B., 1970, Efficient inference in a random coefficient regression model, *Econometrica* 38, 311–323.

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.

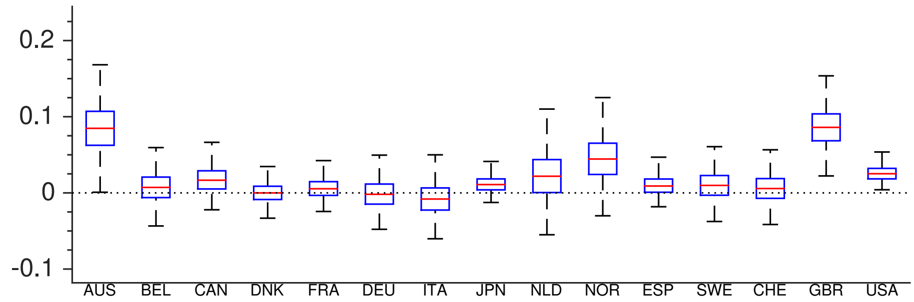
Wachter, Jessica A., and Missaka Warusawitharana, 2015, What is the Chance that the Equity Premium Varies over Time ? Evidence from Predictive Regressions, *Journal of Econometrics* 186, 74–93.

Westerlund, Joakim, and Paresh Narayan, 2014, A Random Coefficient Approach to the Predictability of Stock Returns in Panels, *Journal of Financial Econometrics* (Forthcoming).

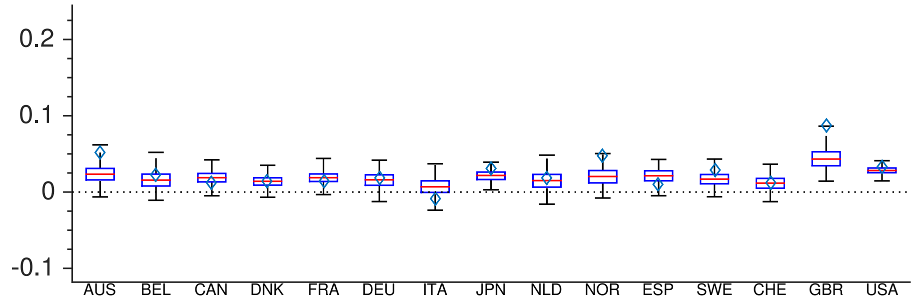
Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* 57, 348–368.

Zellner, Arnold, 1971, *An introduction to Bayesian inference in econometrics* (Wiley, New York).

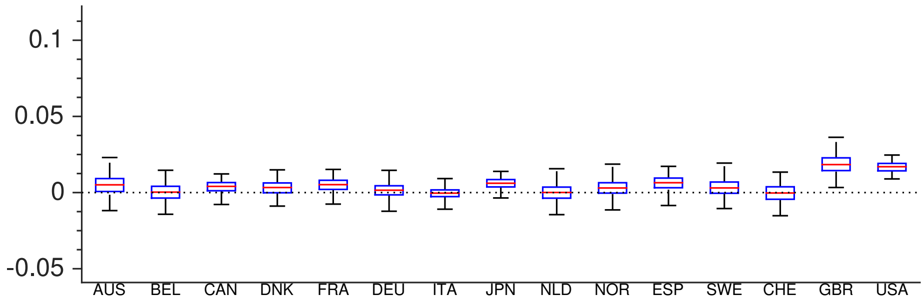
Zellner, Arnold, and Chansik Hong, 1989, Forecasting international growth rates using Bayesian shrinkage and other procedures, *Journal of Econometrics* 40, 183–202.



(a) Uninformative prior

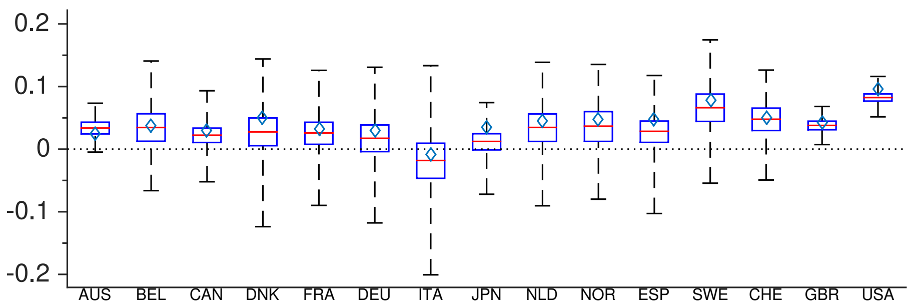


(b) Exchangeable prior and weak economic restriction

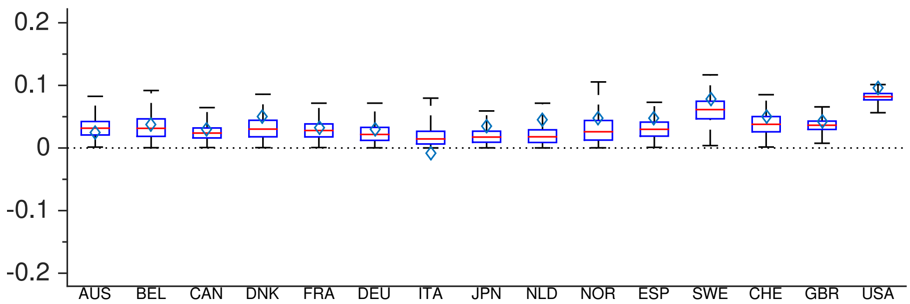


(c) Exchangeable prior and strong economic restriction

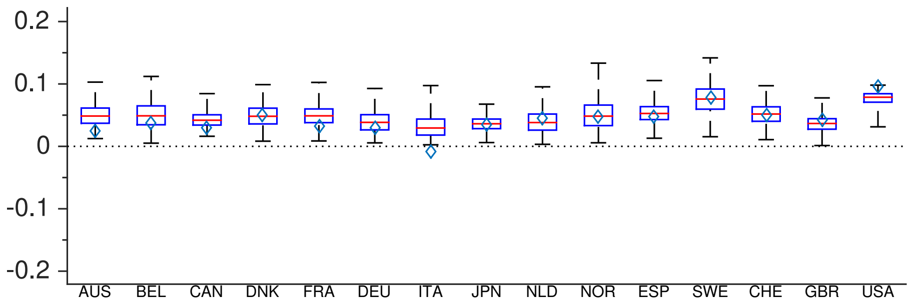
Figure 1: Posterior distribution for the dividend-price ratio coefficient
This figure depicts the posterior distributions of the slope coefficients β_i in a regression of excess returns on the log dividend-price ratio $r_{i,t+1} = \theta_i + \beta_i x_{i,t} + u_{i,t+1}$ where the latter follows an AR(1) 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 center line 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



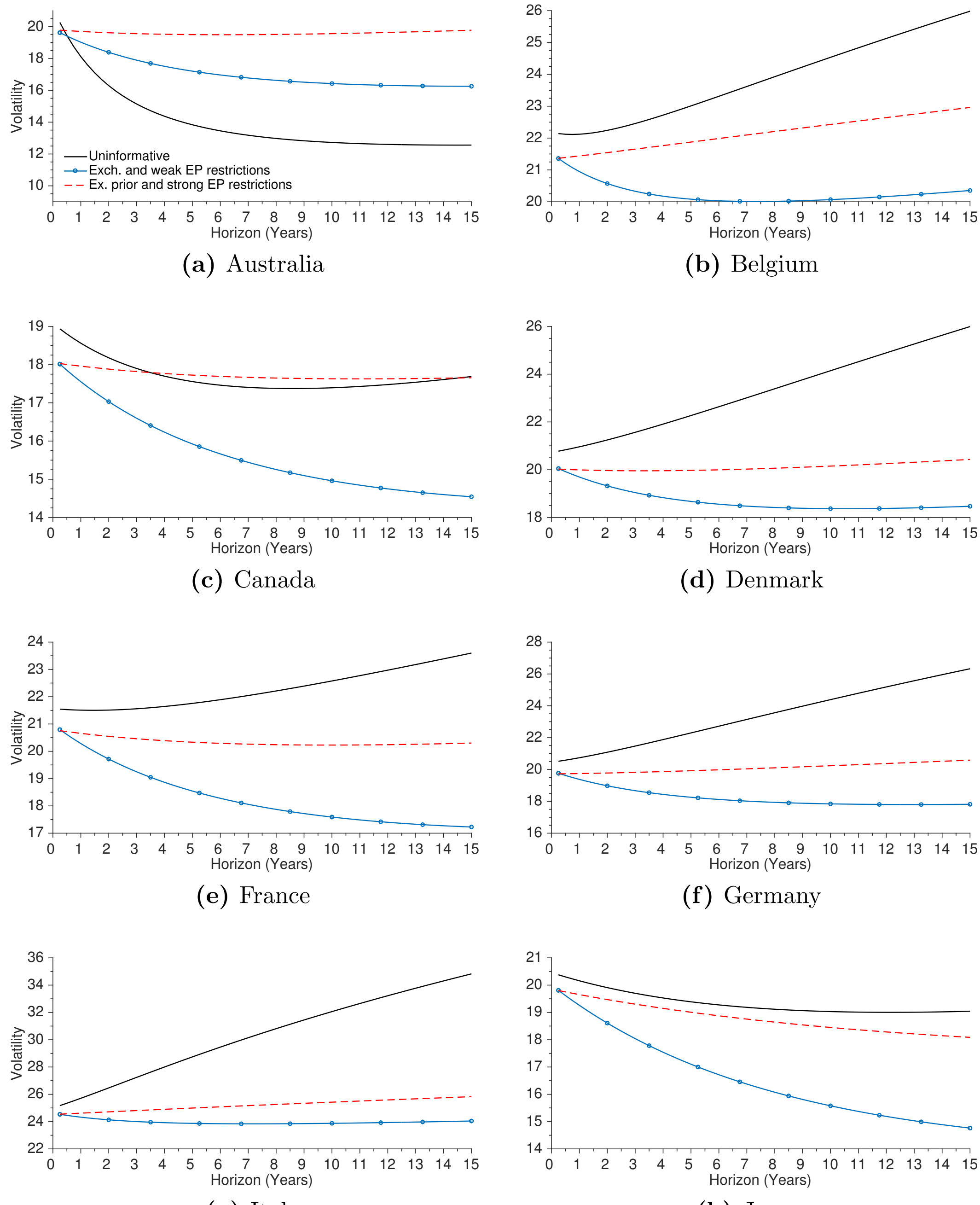
(b) Exchangeable prior and weak economic restriction



(c) Exchangeable prior and strong economic restriction

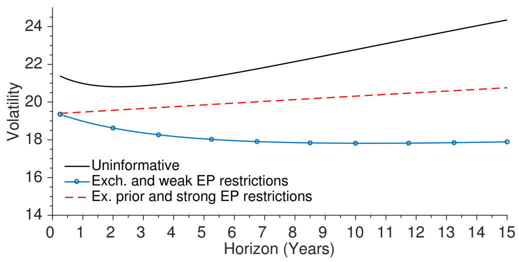
Figure 2: The long-run equity premium

This figure depicts the posterior distributions of the unconditional equity premium $E(\theta_i + \beta_i x_{i,t} + u_{i,t+1}) = \theta_i + \beta_i \frac{\alpha_i}{1-\rho_i}$. x_i is the log-dividend-price ratio in country i and is assumed to follow an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$. Unconditional means are annualized and continuously compounded. For a given country, the center line 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.

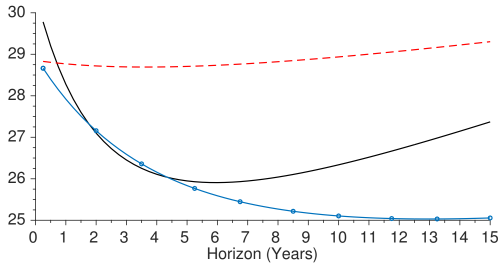


(a) Australia (b) Belgium (c) Canada (d) Denmark (e) France (f) Germany (g) Italy (h) Japan

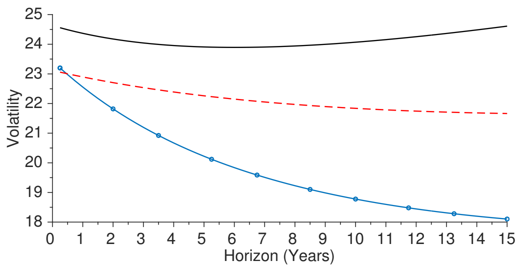
Figure 3: The term structure of risk



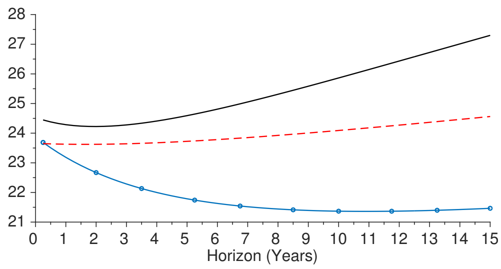
(a) The Netherlands



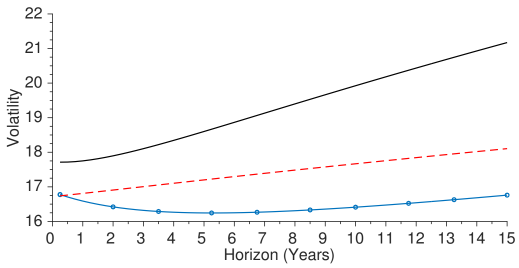
(b) Norway



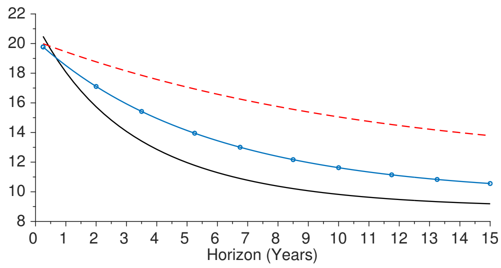
(c) Spain



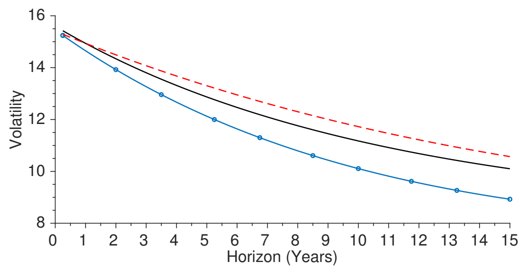
(d) Sweden



(e) Switzerland



(f) The United Kingdom



(g) The USA

Figure 3: The 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 solid 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 I: Bayesian estimates for the country-level slope parameters

	OLS	(2)	(3)	(4)	(5)	(6)	(7)
SUR		×	×	✓	✓	✓	✓
Exchangeable prior		×	✓	✓	✓	✓	✓
Exact likelihood		✓	×	×	✓	✓	✓
Equity premium restr.		×	×	×	×	Weak	Strong
Australia	**0.088**	**0.085**	**0.019**	**0.024**	**0.023**	**0.024**	0.005
	(0.042)	(0.033)	(0.012)	(0.010)	(0.011)	(0.012)	(0.007)
Belgium	0.004	0.007	0.003	0.015	0.016	0.016	0.000
	(0.023)	(0.020)	(0.012)	(0.010)	(0.011)	(0.012)	(0.006)
Canada	0.012	0.017	**0.015**	**0.019**	**0.019**	**0.019**	0.004
	(0.026)	(0.018)	(0.008)	(0.008)	(0.008)	(0.009)	(0.004)
Denmark	0.004	0.000	0.006	**0.014**	**0.014**	**0.014**	0.003
	(0.017)	(0.013)	(0.009)	(0.008)	(0.008)	(0.007)	(0.005)
France	0.000	0.006	0.013	**0.019**	**0.019**	**0.019**	0.005
	(0.017)	(0.013)	(0.008)	(0.008)	(0.007)	(0.008)	(0.004)
Germany	-0.004	-0.001	0.004	**0.015**	**0.015**	0.016	0.002
	(0.025)	(0.019)	(0.010)	(0.009)	(0.009)	(0.010)	(0.005)
Italy	-0.011	-0.008	-0.001	0.007	0.008	0.007	-0.000
	(0.022)	(0.022)	(0.012)	(0.011)	(0.011)	(0.011)	(0.004)
Japan	0.023	0.011	**0.018**	**0.023**	**0.021**	**0.021**	**0.006**
	(0.017)	(0.010)	(0.007)	(0.007)	(0.006)	(0.007)	(0.004)
Netherlands	0.021	0.022	0.003	0.014	0.014	0.015	0.000
	(0.039)	(0.032)	(0.013)	(0.011)	(0.011)	(0.012)	(0.006)
Norway	**0.047**	0.045	0.010	**0.019**	**0.019**	**0.020**	0.003
	(0.026)	(0.030)	(0.012)	(0.011)	(0.011)	(0.012)	(0.006)
Spain	0.008	0.010	**0.019**	**0.023**	**0.022**	**0.021**	0.006
	(0.017)	(0.013)	(0.009)	(0.009)	(0.009)	(0.010)	(0.005)
Sweden	0.008	0.010	0.006	**0.016**	**0.017**	**0.017**	0.003
	(0.020)	(0.019)	(0.010)	(0.009)	(0.009)	(0.009)	(0.006)
Switzerland	0.004	0.006	0.003	0.011	0.011	0.011	-0.000
	(0.027)	(0.019)	(0.010)	(0.009)	(0.009)	(0.010)	(0.006)
United Kingdom	**0.082**	**0.086**	**0.034**	**0.041**	**0.041**	**0.044**	**0.019**
	(0.033)	(0.026)	(0.012)	(0.012)	(0.013)	(0.014)	(0.006)
United States	0.027	**0.026**	**0.027**	**0.030**	**0.029**	**0.028**	**0.017**
	(0.020)	(0.010)	(0.009)	(0.005)	(0.004)	(0.005)	(0.003)
Dispersion	1.000	0.977	0.338	0.278	0.274	0.290	0.193

This table reports posterior means and standard deviations for the slope 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 log dividend-price ratio follows an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$. OLS estimates are also reported in the first column. 'SUR' indicates that the Seemingly Unrelated Regression model of Zellner (1962) is used. 'Exchangeable prior' means that prior beliefs about country-level parameters are given as in Section I.B.1. 'Exact likelihood' indicates that the predictors' initial values are treated as random in the estimation process, as in Stambaugh (1999). 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. Bold numbers denote instances in which the posterior probability of a negative slope is less than 0.05. The longest data is for the US and spans 1952:Q1 to 2013:Q1, and the shortest spans 1986:Q1 to 2013:Q1.

Table II: 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	
$\bar{\theta}$	0.076	0.024	0.016	0.143	0.030	0.012	−0.002	0.066
$\bar{\beta}$	0.020	0.007	0.000	0.041	0.005	0.004	−0.006	0.016
$\bar{\alpha}$	−0.212	0.036	−0.317	−0.124	−0.166	0.030	−0.254	−0.092
$\bar{\rho}$	0.939	0.011	0.908	0.966	0.954	0.009	0.929	0.976

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 III: Bayesian estimates for the country-level equity premia

	Sample	(2)	(3)	(4)	(5)	(6)	(7)
SUR		×	×	✓	✓	✓	✓
Exchangeable prior		×	✓	✓	✓	✓	✓
Exact likelihood		✓	×	×	✓	✓	✓
Equity premium restr.		×	×	×	×	Weak	Strong
Australia	0.026	0.008	0.007	0.005	0.006	0.008	0.013
	(0.031)	(0.015)	(0.021)	(0.016)	(0.016)	(0.016)	(0.018)
Belgium	0.036	0.009	0.010	0.006	0.005	0.008	0.013
	(0.033)	(0.036)	(0.043)	(0.024)	(0.024)	(0.020)	(0.022)
Canada	0.029	0.006	0.007	0.004	0.004	0.006	0.011
	(0.028)	(0.023)	(0.019)	(0.013)	(0.013)	(0.012)	(0.013)
Denmark	0.049	0.007	0.010	0.007	0.006	0.008	0.012
	(0.031)	(0.041)	(0.032)	(0.022)	(0.021)	(0.019)	(0.018)
France	0.031	0.006	0.008	0.005	0.005	0.007	0.012
	(0.032)	(0.034)	(0.052)	(0.018)	(0.017)	(0.015)	(0.017)
Germany	0.028	0.004	0.007	0.003	0.002	0.006	0.010
	(0.031)	(0.039)	(0.029)	(0.020)	(0.019)	(0.015)	(0.018)
Italy	-0.009	-0.004	-0.001	-0.005	-0.006	0.004	0.008
	(0.038)	(0.051)	(0.039)	(0.031)	(0.029)	(0.016)	(0.020)
Japan	0.034	0.003	0.005	0.004	0.003	0.005	0.009
	(0.031)	(0.023)	(0.018)	(0.015)	(0.017)	(0.013)	(0.012)
Netherlands	0.047	0.009	0.011	0.001	0.001	0.005	0.010
	(0.039)	(0.038)	(0.036)	(0.023)	(0.023)	(0.015)	(0.019)
Norway	0.043	0.009	0.010	0.003	0.002	0.007	0.013
	(0.050)	(0.039)	(0.041)	(0.033)	(0.032)	(0.023)	(0.025)
Spain	0.046	0.007	0.012	0.007	0.006	0.008	0.014
	(0.039)	(0.034)	(0.490)	(0.028)	(0.018)	(0.016)	(0.016)
Sweden	0.075	0.017	0.018	0.015	0.014	0.015	0.019
	(0.037)	(0.039)	(0.039)	(0.024)	(0.026)	(0.022)	(0.024)
Switzerland	0.052	0.012	0.012	0.009	0.008	0.009	0.013
	(0.028)	(0.030)	(0.078)	(0.019)	(0.019)	(0.018)	(0.017)
United Kingdom	0.043	0.009	0.009	0.010	0.009	0.009	0.009
	(0.032)	(0.011)	(2.973)	(0.026)	(0.012)	(0.010)	(0.015)
United States	0.096	0.021	0.020	0.020	0.021	0.021	0.020
	(0.020)	(0.011)	(5.937)	(0.008)	(0.008)	(0.008)	(0.012)
Dispersion	1.000	0.244	0.213	0.245	0.253	0.182	0.140

This table reports posterior means and standard deviations for the unconditional equity premium $E(\theta_i + \beta_i x_{i,t} + u_{i,t+1}) = \theta_i + \beta_i \frac{\alpha_i}{1-\rho_i}$. x_i is the log-dividend-price ratio in country i and is assumed to follow an AR(1) process, $x_{i,t+1} = \alpha_i + \rho_i x_{i,t} + v_{i,t+1}$. 'SUR' indicates that the Seemingly Unrelated Regression model of Zellner (1962) is used. 'Exchangeable prior' means that prior beliefs about country-level parameters are centered on the international average. 'Exact likelihood' indicates that the predictors' initial values are treated as random in the estimation process, as in Stambaugh (1999). 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. The longest data is for the US and spans 1952:Q1 to 2013:Q1, and the shortest spans 1986:Q1 to 2013:Q1.

Table IV: 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	30.0	13.8	0.0	22.0
Belgium	34.2	15.4	0.0	25.6
Canada	10.1	23.1	6.5	34.5
Denmark	21.3	23.9	1.8	46.3
France	10.5	29.4	10.1	48.8
Germany	18.3	24.1	4.8	40.5
Italy	24.9	19.8	0.6	36.9
Japan	2.0	50.3	45.8	58.3
Netherlands	23.4	24.8	1.9	41.7
Norway	17.6	33.2	6.6	51.5
Spain	8.2	37.1	19.4	54.0
Sweden	31.6	11.2	0.0	18.5
Switzerland	41.7	11.0	0.0	19.2
United Kingdom	7.7	27.0	12.2	42.9
United States	10.4	5.4	2.1	7.5

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 V: Influence of prior beliefs on asset allocation

	Optimal weight (%)			Weight (%) ignoring Exchangeability			Economic restriction			Both		
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	11	28	45	26	40	52	7	23	38	0	28	83
Belgium	8	25	43	19	28	35	3	19	36	17	24	31
Canada	0	24	49	8	26	42	0	17	43	0	22	43
Denmark	1	24	47	24	27	30	0	19	42	22	23	23
France	0	22	50	13	24	34	0	17	45	12	21	29
Germany	2	20	39	23	25	26	0	12	29	20	18	17
Italy	10	16	22	15	18	21	0	2	8	11	5	0
Japan	0	10	46	19	22	24	0	6	41	0	10	28
Netherlands	3	18	32	12	30	46	0	8	21	7	24	42
Norway	2	14	25	13	23	33	0	8	19	0	16	40
Spain	0	17	47	4	19	33	0	14	44	5	17	29
Sweden	12	29	45	22	32	42	10	27	43	22	31	40
Switzerland	17	34	50	33	42	50	12	29	44	31	40	49
United Kingdom	0	27	66	8	41	73	0	24	61	0	27	96
United States	26	82	100	31	77	100	17	77	100	25	76	100
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	34	36	40	41	45	49	7	23	38	0	28	83
Belgium	32	32	32	31	33	36	3	19	36	17	24	31
Canada	33	37	43	33	34	34	0	17	43	0	21	43
Denmark	30	35	40	33	33	34	0	19	42	22	23	23
France	27	35	43	29	31	32	0	17	45	12	21	29
Germany	29	30	32	32	32	31	0	12	29	20	18	17
Italy	21	20	20	22	23	25	0	2	8	11	5	0
Japan	22	32	42	28	28	28	0	6	41	0	10	29
Netherlands	31	31	31	36	39	42	0	8	21	7	24	42
Norway	20	22	24	26	29	31	0	8	19	0	16	40
Spain	24	32	41	24	26	28	0	14	44	5	17	29
Sweden	34	37	41	36	36	38	10	27	43	22	31	41
Switzerland	47	48	46	49	48	48	12	29	44	32	40	48
United Kingdom	25	40	56	31	48	63	0	24	61	0	28	97
United States	62	96	100	56	85	100	17	77	100	24	75	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 VI: Influence of prior beliefs on certainty equivalent returns (CER)

	Optimal CER (bps)			CER Loss of ignoring								
				Exchangeability			Economic restriction			Both		
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	12	76	192	−24	−15	−5	−2	−4	−4	−12	−2	−141
Belgium	8	76	212	−15	−2	−7	−4	−6	−7	−9	−2	−17
Canada	0	48	196	−7	−1	−3	−0	−4	−2	−0	−1	−2
Denmark	0	60	224	−54	−2	−30	0	−4	−2	−45	−2	−60
France	0	52	272	−36	0	−29	−0	−2	−5	−31	0	−50
Germany	0	40	148	−43	−1	−18	0	−6	−10	−30	1	−47
Italy	16	40	76	−4	−1	−1	−16	−33	−30	−1	−20	−76
Japan	0	8	208	−136	−12	−48	−0	−0	−2	−0	1	−30
Netherlands	0	28	96	−7	−11	−19	0	−8	−10	−0	−2	−10
Norway	0	40	136	−25	−20	−14	0	−7	−9	0	−2	−45
Spain	0	40	296	−17	1	−24	−0	−1	−2	−17	1	−42
Sweden	20	120	300	−10	0	−2	0	0	−2	−11	1	−6
Switzerland	20	80	184	−18	−3	−1	−2	−0	−4	−14	−1	−1
United Kingdom	0	72	428	−24	−18	−5	−0	1	−3	−0	2	−88
United States	36	400	1052	−0	−1	0	−3	−1	0	2	−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	112	132	160	−6	−7	−5	−72	−18	1	−112	−6	−174
Belgium	116	120	120	−0	−2	−2	−99	−21	−2	−26	−9	−1
Canada	88	116	148	0	−2	−6	−88	−34	0	−87	−22	0
Denmark	92	124	164	−2	0	−4	−92	−25	−0	−8	−15	−30
France	80	132	192	0	−3	−11	−80	−37	0	−24	−23	−19
Germany	80	92	100	1	−1	1	−80	−34	1	−7	−15	−21
Italy	68	64	60	−1	−1	−2	−68	−53	−20	−17	−37	−60
Japan	44	100	172	−4	−3	−18	−44	−66	1	−44	−49	−16
Netherlands	88	92	88	−1	−8	−10	−88	−52	−7	−53	−6	−11
Norway	88	104	124	−8	−8	−11	−88	−42	−7	−88	−9	−49
Spain	72	136	224	0	−4	−25	−72	−45	−2	−46	−30	−21
Sweden	160	192	232	−1	0	−2	−80	−15	−0	−19	−4	−0
Switzerland	156	156	152	−1	−1	−2	−86	−26	−2	−17	−5	−2
United Kingdom	60	160	312	−4	−4	−4	−60	−22	−1	−60	−14	−171
United States	228	540	940	−4	−6	−1	−121	−23	−1	−85	−23	−1

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 V. 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 VII: Out-of-sample results

	Excess CERs				Sharpe ratios			
	Hist.	Uninf.	Exchang. prior		Hist.	Uninf.	Exchang. prior	
	average	prior	Weak	Strong	average	prior	Weak	Strong
Australia	62	152	143	130	0.27*	0.40*	0.42*	0.35*
Belgium	−38	31	137	−3	0.14	0.16	0.37*	0.23*
Canada	77	14	69*	151	0.26*	0.12	0.44*	0.37*
Denmark	133	65	95*	221	0.34*	0.29*	0.47*	0.48*
France	30	−2	30	66	0.16	0.06	0.16	0.24*
Germany	19	−35	39	52	0.15	−0.05	0.18	0.22*
Italy	−19	−24	−37	−12	−0.05	−0.10	0.02	0.08
Japan	−166	−15	−39	−113	−0.06	0.00	0.01	0.01
Sweden	149	81	150	189	0.34*	0.26*	0.37*	0.39*
United Kingdom	65	105	127*	141	0.24*	0.43*	0.51*	0.38*
United States	201	132	294*	385	0.49*	0.39*	0.65*	0.59*

This table gives excess CERs and Sharpe ratios for out-of-sample asset allocations. The analysis is performed during the period 1993-2013 for all countries with data in 1973. Allocations between stock and a risk-free asset are calculated each year based the predictive distribution of returns, for an investor with power utility with a relative risk aversion of 5. All values are annualized. The historical average columns correspond to an investor who ignores return predictability and allocates his wealth to stock based on the sample average of past returns. The next columns in both panels correspond to the three priors combination used throughout, i.e. uninformative prior and exchangeable with weak (strong) equity premium restrictions. Stars indicate values that are 1.645 standard errors above zero.