# International return predictability and the term structure of risk

Julien Pénasse*

**Abstract**

This paper explores return predictability within a panel VAR setup. Using quarterly data from 1971 to 2013, I show that stock and bond are actually predictable in 10 countries with well-developed capital markets. While it is generally admitted that stocks are safer on the long run, I obtain a rather flat term structure of risk. Therefore the "conventional wisdom" on long-term investing may not be robust. Using bootstrapped confidence intervals of the conditional moments, I show that stocks are significantly riskier than bonds or bills at any horizon. Long-term correlations across asset classes are very low, even at a ten-year horizon. This result is robust to alternative specifications and data. Overall, while I observe predictability with more precision than in studies focusing on a single country, the consequence of predictability is surprisingly weak.

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

*Keywords:* Cross-sectional dependence; Panel VAR; Pooled regression; Predictive regression; Stock return predictability; Vector autoregression.

*First version :* September 2013

*This version :* December 2013

## 1 Introduction

There is a widespread belief that stocks are safer in the long run. This belief is supported by evidence that hinge on the premise that returns are predictable. Predictability of stocks and bonds has progressively become a stylized fact of the asset pricing literature, see, e.g. Rozeff

*I am grateful to Benjamin Poignard (Institut Louis Bachelier) for exceptional research assistance. 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. I thank the participants of the Research Initiative seminars for helpful comments and suggestions, especially Isabelle Laudier, Didier Janci, Stéphane Gallon, Martial Lasfargues and François Dezorme.

(1984), Fama and French (1988), Campbell (1988), Campbell and Shiller (1991), Pesaran and Timmermann (1995). At the same time, vector autoregressive (VAR) models have been presented as a parsimonious and robust framework capable of integrating mean-reversion in returns (Campbell, 1991; Kandel and Stambaugh, 1989). Evidence of predictability remains mixed, however. Most studies assess predictability over a large timespan but are based on individual time-series regressions, and are subsequently limited to focusing on a single country (usually, the United States). As stressed by Pastor and Stambaugh (2012), even after observing more than two centuries of US data, "investors do not know the values of the parameters of the return-generating process, especially the parameters related to the conditional expected return". Moreover, employing 19th century returns may be useful to historians, but eventually be harmful to practitioners if it leads to overestimating returns predictability. Financial markets have probably gained efficiency in the last century, and inflation risk, which is an important component of long-term risk, has arguably decreased. Empirical support for this idea is provided by Paye and Timmermann (2006), which emphasizes the presence of breaks in a wide range of models and countries. Meantime, international markets have become increasingly tied together. Finally, more synchronous global business cycles, combined with regulatory convergence, have encouraged increasingly homogeneous behavior across and within financial markets.

This convergence requires a departure from traditional time series models. Panel techniques allow to gain statistical precision and minimize the risks associated with forecasting in presence of structural breaks. This is the approach I follow in this paper. I extend the VAR framework of Campbell and Viceira (2002) to a panel of the ten largest developed economies for the period 1971 to 2013. I show that the term structures of stock and bond risk are largely flat on a ten-year horizon. Stocks are riskier than bonds or bills at any horizon, while the risk of the latter more than double in ten years due to inflation risk. These results are robust to several alternative specifications and to variations in data. Finally, the correlation between stocks and both bonds and bills is remarkably low, even at a ten-year horizon, contrasting with previous evidence. Yet both stock and bond returns are found to be predictable, as discussed below.

This paper relates both to the literature on returns predictability and long-term asset allocation. Surprisingly, studies using pooled regressions have been relatively rare. Polk, Thompson, and Vuolteenaho (2006) analyzes the predictive ability of the cross-sectional price of risk using an international panel of 22 countries. Ang and Bekaert (2007) develop a panel extension of Hodrick (1992) and study long horizon predictability for the US, UK, Germany and France. Hjalmarsson (2010) provides a more comprehensive empirical investigation of stock return predictability, but considers neither long-term predictability nor bond returns. In this paper, I show that the GMM-corrected standard error developed in Ang and Bekaert (2007) may underestimate the impact of cross-sectional correlation and therefore overestimate predictability. For example, the bias can be has high as one-third in a panel of four countries.

I therefore rely on bootstrapped standard errors and still provide evidence of predictability. While Ang and Bekaert stress the importance of short-term yield when forecasting stock returns, only the term spread appears as a robust predictor in this setup. The term spread is also able to predict bond excess returns up to one year. I also evaluate the out-of-sample predictability of stock and bond returns. Predictions based on pooled coefficients allow an almost systematic gain in out-of-sample predictions, even when imposing economic restrictions on the return forecasts.

The rest of the paper is organized as follows. Section 2 presents the VAR model, data and provides preliminary statistics of international term structure of volatility. The term structures of risk and conditional correlations among asset classes are compared in Section 3. Section 4 presents the results concerning long-term and out-of-sample predictability. Section 5 concludes.

# 2 The Term Structure of Volatility

## 2.1 Unconditional Volatility

I present on Figure 1 the sample historical volatility of cumulated returns $\sigma_k = \sqrt{\text{var}(r_{t,t+k})/k}$ over the period 1971-2013 for ten of the most industrialized countries, for investment horizons up to 10 years. I use quarterly stock returns obtained from MSCI. Cash and bond returns are obtained from either IMF or OECD. These returns are deflated using consumer inflation rates (OECD)[1]. What is generally expected is that stocks are safer on the long run, so that volatilities should slope downward. This is what we obtain here for several countries, with the notable exception of the United States, the United Kingdom, and Japan. In fact, only Canada and Australia strictly obey such a pattern. For most countries, volatilities seem to follow a hump-shaped pattern, where risk peaks before decreasing consistently with the mean reverting returns hypothesis. The structure of cash and bond risks is much less ambiguous. With the exceptions of Italy and Germany, bond risk seems to be largely flat. Finally cash returns are unambiguously mean averting, reflecting significant inflation risk.

[**Insert Figure 1 about here**]

The flat shape of US stock volatility is striking, in particular when compared to the classical results of Siegel (1998). The reason for this difference is that Siegel use almost two centuries of data where I use only forty-two years. I reproduce in Figure 2 evidence similar to Siegel's by using returns from the S&P Composite data of Robert Shiller's Irrational Exuberance[2]. Stocks are now found to be mean reverting (again after a peak at one year) when using the full series. However, on the subsample 1971-2013 the results are similar, whether we use S&P

[1] I give a more extensive account of data in Section 2.3.
[2] http://aida.econ.yale.edu/ shiller/data.htm.

or MSCI series. This illustrates the likely structural breaks during long sample periods, which further complicate statistical analysis of mean reversion.

**[Insert Figure 2 about here]**

The evidence discussed here relies on nonparametric estimates of volatility which are close to the *price* mean reversion literature. This literature typically rely on variance-ratios[3] to detect mean reversion. Our purpose, however, is to study *returns* mean reversion, which can be obtained when stock returns are negatively autocorrelated. This is precisely what we captured here, but this approach neglects the possibility that other variables may also predict returns. Capturing a richer set of relations requires to build an econometric model, which I present in Section 2.2.

## 2.2 VAR Methodology

A likely consequence of return predictability is the downward slope of the term structure of volatility. A common approach, introduced by Kandel and Stambaugh (1989) and Campbell (1991), is to model an arbitrary set of traded assets and state variables under a vector autoregressive setup. This framework is generally considered to be well-suited to evaluate investment horizon effects, as the conditional moments can be computed from the VAR parameters. A classical result is the mean reversion of stock returns: the annualized volatility of stock returns is lower over long horizons than over short horizons (e.g. Campbell and Viceira 2002).

This evidence has been criticized by an important "Bayesian" literature. The core argument is that parameters uncertainty is neglected when computing long run variances (Barberis, 2000). More recently, Pastor and Stambaugh (2012) have argued that traditional approaches not only ignored estimation risk but also made the implausible assumption of perfect predictors. Estimating a predictive model

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

and assuming that coefficients as given and known by investors is therefore a strong assumption. It is tantamount to assuming that coefficients are perfectly known and that Equation (1) *is* the true model, so that $z_t$ will deliver a perfect proxy for the conditional expected return. Pastor and Stambaugh's critic is related to model instability and uncertainty (e.g. Avramov 2002; Paye and Timmermann 2006).

The "Bayesian" literature has emerged from the debate over predictability, in part because of limited samples. By exploiting the cross sections of predictive regressions, the estimation of pooled versions of (1) should greatly attenuate estimation risk. For example, I show in Section 4.4 that pooled regressions consistently produce much more stable coefficients than their time

[3]See, e.g. Lo and MacKinlay (1988); Poterba and Summers (1988).

series counterparts. In this paper I do not directly address estimation risk but instead produce confidence interval of conditional moments.

I follow the setup described in Campbell and Viceira (2002) and Campbell, Chan, and Viceira (2003). The predictive system is expressed as a VAR(1):

$$\mathbf{z}_t^i = \Phi_0^i + \Phi_1^i \mathbf{z}_{t-1}^i + \mathbf{v}_t^i \tag{2}$$

where

$$\mathbf{z}_t^i = \begin{pmatrix} r_{0,t}^i \\ \mathbf{x}_t^i \\ \mathbf{s}_t^i \end{pmatrix}$$

is a $(K-1) \times 1$ vector, with $r_{0,t}^i$ the real short-term interest rate, $\mathbf{x}_t^i$ the $n \times 1$ vector of log excess returns and $\mathbf{s}_t^i$ the $(K-n-2) \times 1$ vector of variables useful for predicting returns, namely the nominal short rate, the log dividend-price ratio and the term spread[4]. $n$ is the number of assets in excess of the real short-term interest rate. $\Phi_0^i$ is the $(K-1) \times 1$ vector of intercepts and $\Phi_1^i$ is the $(K-1) \times (K-1)$ matrix of slope coefficients. I assume that the VAR process is stationary, i.e. the eigenvalues of $\Phi_1^i$ lie inside the unit circle in absolute value. Finally, $\mathbf{v}_t^i$ is the $(K-1) \times 1$ vector of innovations in asset returns and return forecasting variables, which is assumed to be *i.i.d.* normally distributed :

$$\mathbf{v}_t^i \sim N(0, \mathbf{\Sigma_v^i}), \tag{3}$$

where $\mathbf{\Sigma_v^i}$ is the $(K-1) \times (K-1)$ covariance matrix.

Under the assumption that $\mathbf{\Sigma_v}$ is constant over time, the conditional $k$-quarter variance is given by[5]:

$$\frac{1}{k} V_t \begin{pmatrix} r_{0,t+k}^{(k)} \\ r_{e,t+k}^{(k)} \\ r_{b,t+k}^{(k)} \end{pmatrix} = \frac{1}{k} \mathbf{M} \, V_t \left( \mathbf{z_{t+1}} + \mathbf{z_{t+2}} + ... + \mathbf{z_{t+k}} \right) \, \mathbf{M}' \tag{4}$$

[4]See, for example, Rozeff (1984), Fama and French (1988) and Campbell (1988) for evidence that the dividend yield can predict equity risk premium. The short-term interest rate have been identified as a predictor of stock excess returns in studies including Campbell (1987), Hodrick (1992), Ang and Bekaert (2007) and Lioui and Poncet (2012). Finally, several studies observed similar evidence for the yield spread (Campbell, 1987; Fama and French, 1989; Keim and Stambaugh, 1986). Bond returns predictability, on the other hand, is much less controversial, e.g. Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005; Fama and Bliss, 1987; Lioui and Poncet, 2012.

[5]Dropping the index $i$ for presentation purpose.

where $\mathbf{M}$ is a $(K-1) \times (K-1)$ selection matrix and

$$\begin{aligned} V_t\left(\mathbf{z_{t+1}} + \mathbf{z_{t+2}} + ... + \mathbf{z_{t+k}}\right) = {} & \Sigma_v + \left(\mathbf{I} + \Phi_1\right) \Sigma_v \left(\mathbf{I} + \Phi_1\right)' \\ & + \left(\mathbf{I} + \Phi_1 + \Phi_1\Phi_1\right) \Sigma_v \left(\mathbf{I} + \Phi_1 + \Phi_1\Phi_1\right)' \\ & + ... \\ & + \left(\mathbf{I} + \Phi_1 + ... + \Phi_1^{k-1}\right) \Sigma_v \left(\mathbf{I} + \Phi_1 + ... + \Phi_1^{k-1}\right)' \end{aligned} \tag{5}$$

I use Equations (4) and (5) to compute conditional volatilities and correlation. A bootstrap procedure is followed to compute the confidence intervals of these conditional moments (see Section 3.2). Pastor and Stambaugh (2012) argues that such an approach neglects prediction risk, by overlooking uncertainty about the current expected return and about parameter values. They elegantly show that such risk should be compounded with horizon so that the variance relevant to investors would dramatically increase with horizon. I choose to focus on the true conditional volatility, which compares more favorably to its unconditional counterpart.

## 2.3 Data

I consider ten OECD countries: Australia, Belgium, Canada, France, Germany, Italy, Japan, Sweden, United Kingdom and the United States. In order to assess return predictability and conditional moments, I need to compute excess returns for bills, stocks and bonds, as well as usual predictors, i.e. 3-month log yields, log dividend-price ratios and term spreads for each country. To do so I follow Campbell and Viceira (2002) and Jondeau and Rockinger (2010). The sample starts in 1970Q1 and ends in 2013Q1 for all ten countries. For these countries (see table 1), I am able to collect the relevant series for the full timespan. Inflation rates are obtained from OECD quarterly series. Monthly short-term interest rates are downloaded from the OECD statistics or the IMF International Financial Statistics (for Australia, Italy, Sweden and USA) depending on availability. End-of-quarter values from this monthly series are retained to get quarterly observations ($Y_t$). $y_t$ denotes the 3-month log yield: $y_t = log(1 + Y_t)/4$. $r_{0,t}$ is the real ex post short-term rate, i.e. the difference between $y_t$ and the log inflation rate.

I construct log long bond returns ($r_{b,t}$) and term spreads ($spr_t$) from the IMF monthly series of 10-year Government bond yields. Again, I retain end-of-quarter values to get quarterly series. Bond returns are constructed using the duration loglinear approximation described in Chapter 10 of Campbell, Lo, and MacKinlay (1997):

$$r_{n,t+1} \approx D_{n,t} y_{n,t} - \left(D_{n,t} - 1\right) y_{n-1,t+1}$$

where $n$ is the bond maturity and $D_{n,t}$ is the bond duration. $y_{n,t}$ is the log bond yield

$y_{n,t} = log(1 + Y_{n,t})$. I calculate duration at time $t$ as

$$D_{n,t} \approx \frac{1 - (1 + Y_{n,t})^{-n}}{1 - (1 + Y_{n,t})^{-1}}$$

where $n$ is set to 10 years. I also approximate $y_{n-1,t+1}$ by $y_{n,t+1}$. The log excess return on bond is then $x_{b,t} = r_{b,t} - r_{0,t}$. The term spread $spr$ is the defined as the difference between the long and short interest rate yields.

Equity prices and total returns (in local currency) come from Morgan Stanley Capital International (MSCI) database. The quarterly log real return on the stock index is denoted $r_{e,t}$ and defined as the difference between the log return on equities and the log inflation rate. The log excess return on stocks follows: $x_{e,t} = r_{e,t} - r_{0,t}$. As argued in Bec and Gollier (2009) I choose to work with data excluding tax credits. Following convention, a smoothed dividend series (a sum of dividends from month $t-11$ through month t) is used to compute the dividend yield. The log dividend-price ratio, *ldp* obtains as the log dividend yield less the log price index.

Table 1 gives the first and second moments of the data for each country. Except for the log dividend-price ratio, the sample statistics are in annualized, percentage units. The mean log returns are adjusted by adding one-half of their variance so that they reflect mean gross returns. Excess returns are remarkably homogenous across countries, with a few exceptions (Japan for short-term interest rate, Italy and Sweden for stocks). The equity premium roughly lies between 4 and 6% for most countries during this sample period, as reported in Ang and Bekaert (2007). The bond premium ranges approximately between 2 and 2.5% for all countries except Japan, which low dividend yield and high term spread seem to reflect the peculiar economic environment of the last decades. While most studies of stock riskiness have focused on the United States, it is important to notice that the latter have the lowest excess return volatility at 16.5%, the sample unweighted average being 20.6%. The standard deviations of the six series are quite similar across countries and roughly follows the limited discrepancies of sample averages.

[**Insert Table 1 about here**]

Levin, Lin, and Chu (2002) (LLC) panel unit root tests are also reported in Table 1. LLC tests are a panel generalization of ADF tests and therefore are performed under the null of a unit root. This test increases the power of the conventional ADF test by restricting the autoregressive coefficient to be identical across units. The number of lags is selected in order to minimize the Akaike information criterion. As suggested by Levin, Lin, and Chu, this version of the test is performed on demeaned series, in order to mitigate the impact of cross-sectional dependence. The unit root null is strongly rejected for all variables except $r_0$, where I must accept the absence of stationarity at the 10% confidence level.

## 2.4 Conditional Volatilities

Before turning to a model in which coefficients are similar across countries, it is instructive to reproduce the results of Section 2.1 but exploiting the VAR structure. We now consider the situation of an investor who understands that some fraction of each asset unconditional volatility is actually predictable time-variation in the return. Therefore this portion of the volatility does not count as risk. The long-term risks of asset returns, which dynamics are captured by the VAR, may however increase or decline as the holding period increases. By contrast, if an investor neglects predictability, and therefore estimate Equation 2 assuming that the coefficients of $\Phi_1^i$ are zero, will face a flat term structure of volatility.

Figure 3 plots percent annualized standard deviations of real returns for investment horizons up to 10 years. While it is common to plot standard deviations on very long horizons (e.g. up to 25 years in Bec and Gollier (2009), 50 years in Hoevenaars et al. 2013), I consider a ten-year horizon, having in mind the absence of significant long-term predictability (see Section 4). Overall these conditional volatilities are remarkably consistent with their unconditional counterparts presented in Figure 1. This confirms that VAR is a reasonable approximation for representing the properties of long-term returns, as originally pointed by Kandel and Stambaugh (1988). The evidence of mean reverting returns is weaker, however. Two countries now seem to present significant mean reverting (the UK and Australia). I find mean aversion in two others (France and Italy), while for the remaining countries we may hesitate between hump-shaped and simply flat volatilities. Bond volatilities now seem to peak after one or two years before slightly diminishing. Again, I only find unambiguous mean aversion in the risk of cash investment.

[**Insert Figure 3 about here**]

More importantly, it must be stressed that given the limited sample size, comparing volatilities across countries is a delicate exercise. Bec and Gollier (2009) shows when focusing on a single country, one cannot plausibly distinguish the risk of bonds and stocks after a few quarters. This implies that I cannot either exclude that the shape of the term structure of risk is significantly different from one country to another. I will therefore have to rely on pooled estimates to improve estimation precision. This is the purpose of Section 3, where I also extend Bec and Gollier (2009)'s methodology to panel so as to assess the hierarchy of risks.

# 3 Panel VAR Evidence

## 3.1 Panel VAR Estimation

I now turn to my main result which is an assessment of the term structure of risk under the assumption that coefficients are constant across countries. Estimating the system (2) with pooled coefficients and a country fixed effect is more complex than in studies such as Ang and Bekaert (2007) because (2) includes lagged dependent variable as one of the regressors. With a fixed effect, I can rewrite the error term as

$$\mathbf{v}_t^i = \eta^i + \mathbf{u}_t \tag{6}$$

where $\eta^i$ and $\mathbf{u}_t$ have zero mean, finite variance and are independent of each other. Both $\mathbf{z}_t^i$ and $\mathbf{z}_{t-1}^i$ are functions of the fixed effect $\eta^i$ so that the regressor $\mathbf{z}_{t-1}^i$ in Equation (2) is correlated with the error term. A common approach is to estimate a least square dummy variable (LSDV) model. LSDV generally performs well when $T$ is large since the bias is $O(1/T)$ (Nickell, 1981), but Judson and Owen (1999) showed that the bias could remain problematic in the presence of persistent regressors. A solution is to employ a GMM approach, e.g. Arellano and Bond (1991), which is consistent even when $T$ is small. However this approach is not recommended when $T$ is greater than the number of units because the number of moment conditions would be large relative to the sample size.

Does the inclusion of a fixed effect matters? The failure to include dummy variables when they are needed makes OLS inconsistent, but I show in Section 3.3 that both OLS and LSDV estimators produce close results. Moreover Equation (5) do not depend on $\Phi_0$, so that the inclusion of a fixed effect only affects conditional moments through its potential impact on $\Phi_1$ and $\Sigma_v$. This impact seems negligible for estimation purposes, but large distortions appeared in the bootstrap procedure when using LSDV. I therefore assume $\eta^i = 0$ for all $i$ and estimate pooled coefficients of (2) using OLS.

The results are reported in Table 2, together with bootstrapped standard errors. Each column corresponds to a component of the VAR and its regressors. The first three columns correspond to the traditional assets available to a representative long-term investor: the real short-term rate and excess stock and bond returns. The last three rows correspond to the state variable or predictors: the short-term yield, the log dividend-price ratio and the term spread. The real short-term interest rate is predicted by all variables but equities. Stock excess returns exhibit a statistically significant autocorrelation, albeit the magnitude of this autocorrelation (0.109) is not uncommon in the literature. They are also significantly predicted by real short-term rate, lagged bond excess returns and the log dividend-price ratio, an improvement when compared to the predictive regressions presented before. The real short-term interest rate appears to have greater predictive power than its nominal counterpart, both for stocks and bonds. This is not a complete surprise, as the dynamics of inflation may have played a role in explaining returns in the last decades. Bond excess returns are predicted by the term spread and also present a significant autocorrelation. The last three columns reveal that the state variables are highly persistent, especially the short-term yield and the log dividend-price ratio as frequently observed in the literature. The VAR satisfies the stability condition, the largest eigenvalues of $\Phi_1$ lie inside the unit circle in absolute terms, with a value of 0.988.

**[Insert Table 2 about here]**

Overall, the pattern of statistical predictability is larger than in typical studies focusing on one countries (e.g. Bec and Gollier, 2009; Campbell and Viceira, 2002; Hoevenaars et al., 2013; Jondeau and Rockinger, 2010). The economic significance of the point estimates is however in line or smaller than what is typically observed in time series regressions. The gain in statistical predictability comes from lower standard errors thanks to the size of the sample. It is unlikely to be caused by biased point or standard errors estimates. Pooled regressions with a common intercept do not suffer from Stambaugh (1999) small sample bias (Hjalmarsson, 2008) and as pointed earlier, the OLS coefficients are very close to estimates obtained by LSDV. Besides, the bootstrapped standard errors are robust to cross-sectional correlation and heteroscedasticity.

Table 2 also reports the pooled cross-correlations of residuals. The diagonal terms are annualized standard deviations while the off-diagonal entries are correlations between variables, regardless of their country. These average cross-correlations are remarkably comparable to their country-specific counterpart, see e.g. Campbell and Viceira (2002). Shocks to the dividend-price ratio and stocks are highly negatively correlated, as are shocks on the short-term yield and bond excess returns. The combination of a negative contemporaneous correlation and a positive forecast coefficient induces mean-reversion (Campbell and Viceira, 2005), therefore stocks and bonds will be mean reverting. Notice the low correlations between excess bond and stock return residuals, and between bond excess returns and the real short-term rate residuals (3.35% and 5.73% respectively).

Table 3 presents the results of Brown-Forsythe test of homogeneity of variances between countries, for each of the VAR components. The null of variance homogeneity is largely rejected for all variables, therefore conditional moments must be interpreted as averaged values. I address this heterogeneity by reporting bootstrapped confidence intervals of conditional moments. In my procedure, residuals are resampled in a way that preserves both cross-sectional correlation and heteroscedasticity (see Section 3.2). Table 3 also reports the results of a Roy (1957) and Zellner (1962) test of poolability, which null hypothesis is full poolability, consistently with the assumption of common intercepts across countries. The null is rejected for all series but stocks. Again, this result does not invalidate our analysis and the VAR coefficients can meaningfully be interpreted as average relations between variables.

**[Insert Table 3 about here]**

## 3.2 Conditional Moments

As discussed above, I use Equations (4) and (5) to calculate the conditional $k$-period variance-covariance matrix of real returns. Annualized standard deviations are obtained by dividing conditional standard deviation of cumulative returns by the square root of the horizon.

Since neither the homogeneity nor the constant variance hypotheses strictly hold across countries, I would like to assess the variability of these average relations in a manner that would respect the characteristics of the data. A natural way to measure this variability is to compute confidence bands of the assets conditional moments. Moreover, since homogeneity *is* accepted for stocks, confidence bands can be viewed as an (unconventional) way to capture parameter uncertainty. If pooling allows for an increase in estimation precision, this should reflect in narrow confidence bands. We may also want to check whether the difference between the term structure of e.g. stocks and bonds is statistically significant.

I adapt the parametric residual bootstrap method described in Bec and Gollier (2009) to panel data. Under the assumption that the cross-section of residuals $V_t = (\mathbf{v}_t^1 \dots \mathbf{v}_t^N)$ is *i.i.d* from an unknown distribution and fixing the initial condition on $(\mathbf{z}_t^1 \dots \mathbf{z}_t^N)$, the bootstrap distribution of the VAR coefficients, and by extension of conditional moments, can be computed by simulation. Random draws are made from the cross-section of estimated residuals $\hat{V}_t$ and then $N$ simulated series are calculated by recursion given model (2). For each time $t$ the residuals are contemporaneous to one another, which means that cross-sectional correlation is preserved. Moreover, residuals are assigned to each country; for example, residuals for the United States will always be drawn from the size $T$ sample of United States residuals. This sampling strategy therefore preserves heterogeneity in covariance across countries. For each simulation, pooled coefficients $\hat{\Phi}_0^s$, $\hat{\Phi}_1^s$ and $\hat{\Sigma}_v^s$ are estimated by OLS and the corresponding conditional moments are computed. 10,000 panels are simulated with the same dimensions $(N \times T)$ as the original sample. Bootstrapped standard errors of the coefficients $\Phi_0^s$ and $\Phi_1^s$ are obtained as the standard deviations of $\hat{\Phi}_0^s$ and $\hat{\Phi}_1^s$.

Figure 4 shows estimates of the term structure of risk for the real short-term rate, stock and bond returns. If returns were i.i.d. the annualized standard deviation would be constant across investment horizons. As expected from the sign of the VAR coefficients, the term structure of stock and bond risks are downward-slopping (after one year), suggesting mean reversion of returns. Conversely, the conditional volatility of the real short-term rate, i.e. inflation risk, is clearly mean-averting. However the mean reversion is economically weak for stocks and bonds; in fact it is not significant if we use the confidence bands as an informal test. This weak evidence of mean reversion is a consequence of our estimation strategy: pooling increases estimation precision and confirms the evidence of predictability, yet this predictability eventually appears as economically weak because the pooled coefficients *average* heterogeneous dynamics. Said otherwise, mean reverting is not a property that is robust internationally.

**[Insert Figure 4 about here]**

The widths of the confidence bands differ from one asset to another, reflecting heterogeneity in unconditional moments across countries. The gain in estimation precisions reflects in these widths and contrasts from Bec and Gollier (2009), which focus on French data only. Bec and Gollier obtain very large confidence intervals and conclude that the term structures of risk of the three asset classes intersect after 5 quarters and are therefore not significantly different from each other after that horizon. On the contrary, the term structures of risk never intersect in the pooled setup. The bottom line is that the risk of a one-year investment in stocks or bonds is, on average, the same as a ten-year investment. Stocks are riskier than bonds or bills at any horizon, while the risk of the latter more than doubles in ten years.

Figure 5 completes the picture with estimates of the term structure of correlation. The correlation between stock and cash returns is surprisingly low. It is even negative for a one-quarter horizon (-1.9%) and peaks at 21.1% for a ten-year horizon. The correlation between stocks and bonds follows a similar increasing pattern with a minimum correlation of 3.3% for one quarter and a maximum of 34.6% for ten years. It strongly contrasts from Campbell and Viceira (2005), which observes a much larger correlation between stocks and bonds, with a minimum around 20% for one quarter to a peak at 60%[6]. Finally, the correlation between bonds and cash decreases after the first year and ranges between 9% and 39%.

**[Insert Figure 5 about here]**

## 3.3 Robustness of the Term Structure Results

Based on OLS estimates of the pooled panel VAR, I have shown that the mean-reversion property of stock returns is absent, since on a ten-year horizon, the term structure of stock volatility is essentially flat. The choice of OLS is mainly based on its simplicity, in that it allows the relatively straightforward computation of confidence intervals. I examine here how robust this result is to changes in the choice of model specification and choice of data.

First I relax the intercept homogeneity hypothesis and therefore estimate (2) with the LSDV estimator, assuming that Equation (6) holds. LSDV yields biased estimates in the presence of a lagged dependent variable. This bias is typically small when the time dimension is large, but persistence may complicate this diagnostic in particular in this panel VAR setup. Moreover, since the computation of conditional moments requires to compute powers of the VAR coefficients, a small bias may also be compounded over time and become problematic. This is what happens in the bootstrap for LSDV, but luckily not in the estimation. As shown on Figure 6 the difference between conditional volatilities with OLS and LSDV is negligible.

[6] I obtain a similar pattern with my dataset when considering US data only.

Second, I relax the dynamic homogeneity assumption, i.e. assume that $\Phi_1^i = \Phi_1^\star + \Lambda^i$, where $\Lambda^i$ have zero mean but can be correlated. Since T is large, I can estimate the VAR for each country and average the results across countries. Such a mean group estimator is inefficient relative to the pooled estimator under dynamic homogeneity, but gives consistent estimates of the average dynamic effect if dynamic heterogeneity is present. By contrast, the pooled estimator would be inconsistent because the regressors are correlated with the error term (see e.g. Canova and Ciccarelli (2013); Pesaran and Smith (1995)). We see on Figure 6 that, if present, dynamic heterogeneity plays a modest role. The averaged term structure of risk is slightly mean-reverting, the ten-year volatility being approximately 3% lower than its one-quarter counterpart. Alternatively, the random coefficient model proposed by Swamy (1970) also yields consistent estimates of the average effect. I present the term structure of volatility implied by Swamy's Generalized Least Squares weighted average estimator on Figure 6. The consequence on stock volatility is almost symmetric to the mean group estimator's, pointing to some potential ambiguous effect of dynamic heterogeneity on conditional volatility. The term structure of bonds and bills is also flatter under the random coefficient estimator. Overall, the result of a flat term structure of volatility is largely unchanged.

[**Insert Figure 6 about here**]

Finally, I consider a larger dataset of 32 industrialized and a few developing countries. To the ten initial series which are available from 1971, I add 22 countries which are available on a more limited timespan. Descriptive statistics for these countries are reported on Table 4. I use the same metrology to estimate the model, except that the panel VAR is now unbalanced. The model, which is now estimated on a total of 3509 quarters, is surprisingly similar to our benchmark, as can be seen on Figure 7. The plots are almost similar, except stock volatility which is shifted upward. The addition of a larger set of countries which includes developing countries only seem to increase stock risk, without modifying the shape of volatility when investment horizon increases.

[**Insert Table 4 about here**]

[**Insert Figure 7 about here**]

In summary, the results presented in this section further suggest that national markets exhibit a very stable term structure of risk. In the succeeding section, I turn to long-horizon regressions to show that these results obtain in spite of significant short-term predictability.

# 4 Long-horizon Regressions

## 4.1 Specification and Poolability Tests

Following Ang and Bekaert (2007), I will now focus on regressions of the form

$$\tilde{r}^i_{t+k} = \alpha_i + \beta_i' z^i_t + u^i_{t+k} \tag{7}$$

for $i = 1 \dots N$ countries, subject to the restriction $\beta_i = \bar{\beta}\ \forall i$. $\tilde{r}^i_{t+k}$ are cumulated annualized k-quarters ahead stock or bond excess returns. This specification allows for freely estimated $\alpha_i$, a common feature in LSDV regression. $z^i_t$ is a $(K-1)\times 1$ matrix of instruments. For stocks, I use log dividend yield, term spread, bond excess returns, and a combination of short-term rate and log dividend yield as instruments. For bonds, I consider the term spread and the nominal short-term yield as potential predictors.

Consider the above equation. When $k > 1$, the error term $u^i_{t+k}$ follows a moving average process under the null of no predictability $\bar{\beta} = \mathbf{0}$ because of overlapping observations. The OLS estimates of $\theta = (\alpha_1 \dots \alpha_N \bar{\beta})'$ are still consistent but standard errors need to be corrected. One way to address the problem of correlated error term is to extend Hodrick (1992) standard errors to panel regression (Ang and Bekaert, 2007). The asymptotic distribution theory of GMM implies that $\sqrt{T}(\hat{\theta} - \theta) \overset{a}{\sim} N(0, (D_0' S_0^{-1} D_0)^{-1})$ where

$$D_0' = E\left(\frac{\partial h_{t+k}}{\partial \theta'}\right)$$

and $S_0 = E(h_{t+k} h'_{t+k})$. I detail in Appendix A how to compute the (thereafter) Hodrick-Ang-Bekaert estimate $\hat{S}^b_T$ of $S_0$ and the estimate $\hat{D}_T$ of $D_0$, using numerical derivatives.

I perform two tests to assess the poolability hypothesis. The first is the Hansen (1982) $\chi^2$ $J$-test of overidentifying restrictions:

$$J = T(\bar{h}'(\hat{S}^b_T)^{-1}\bar{h}) \sim \chi^2[NK - (N + K - 1)], \tag{8}$$

with

$$\bar{h} = \frac{1}{T}\sum_{t=0}^{T} h_{t+k}.$$

Ang and Bekaert (2007) use this test to assess poolability, however it does not test for poolability *per se*. A straightforward test is the Roy (1957) and Zellner (1962) test of poolability, in combination with the bootstrap procedure proposed by Bun (2004) (see Appendix B). The test allows for individual intercepts and is performed for $k = 1$.

## 4.2 Robustness of Hodrick-Ang-Bekaert Standard Errors : Simulations

The GMM adjusted standard errors are only valid asymptotically and might not be well behaved in moderate samples with a relatively large number of countries. Ang and Bekaert stress that "pooled estimations mitigate the data-mining problem plaguing US data and, under the null of no predictability, enhance efficiency because the correlations of returns across countries are not very high." If the residuals of a given year are correlated across countries (cross-sectional dependence), however, it is not clear why the above correction would not understate the true standard error. This is a quite general property that standard errors of estimates will be biased when the residuals exhibit cross-sectional dependence. The magnitude of the error is known to be increasing in the number of units (Petersen, 2008). To understand this intuition, consider the extreme case where both returns and predictors are perfectly correlated across countries. In this case, each additional country provides no additional information and will have no effect on the true standard error. However, Hodrick-Ang-Bekaert standard errors will report a gain in precision, as I show in the following simulation exercise.

I perform a Monte Carlo analysis where I vary the number of countries and the degree of cross-sectional correlation. I focus on the case of a single explanatory variable, and assume that the true data generating process is governed by Equation (7) where $k = 1$, i.e. $\tilde{r}^i_{t+1} = \alpha_i + \beta_i' z^i_t + u^i_{t+1}$. The coefficients of the process are chosen to roughly match their empirical counterpart[7].

For simulation purpose I posit that the residuals consist of a time-specific component $\gamma_t$ and an idiosyncratic component $\eta^i_t$ that is unique to each observation:

$$u^i_t = \sqrt{\tau}\gamma_t + \sqrt{1-\tau}\eta^i_t$$

I assume that the independent predictor $z$ follows an AR(1) process, which residuals also exhibit a time-specific component[8]:

$$\begin{aligned} z^i_t &= z_0 + \rho z^i_{t-1} + \epsilon^i_t \\ \epsilon^i_t &= \sqrt{\tau}\mu_t + \sqrt{1-\tau}\nu^i_t \end{aligned}$$

The components of $u$ ($\gamma$ and $\eta$) and $\epsilon$ ($\mu$ and $\nu$) are normally distributed, have zero mean and are independent of each other. Their variances are respectively $Var(\gamma_t) = Var(\eta^i_t) = \sigma^2_r$ and $Var(\mu_t) = Var(\nu^i_t) = \sigma^2_z$. This parameterization ensures that $Var(u^i_t) = \sigma^2_r$ and $Var(\epsilon^i_t) = \sigma^2_z$. The residuals volatility $\sigma^2_r$ and $\sigma^2_z$ are chosen to approximately match the properties of the

[7] More precisely, $\alpha_i = 0.1074$ and $\beta_i = 0.0284$ for all $i$.

[8] with $z_0 = -0.0975$ and $\rho = 0.9722$.

data. Finally, $\tau$ determines the level of cross-sectional correlation:

$$\begin{aligned} corr(u_t^i, u_s^j) = corr(\epsilon_t^i, \epsilon_s^j) &= 1 \text{ for } s = t \text{ and } i = j \\ &= \tau \text{ for } s = t \text{ and } i \neq j \\ &= 0 \text{ otherwise.} \end{aligned}$$

Across different simulations, I alter the level of cross-sectional correlation $\tau$ from 0% to 75% in 25% increment, 50% being roughly the average level of cross-sectional correlation of my data[9]. I also vary $N$ to illustrate the effect of increasing the number of countries.

Using the data-generating process described above, I generate 1000 artificial values of $\tilde{r}$ and $z$ for a balanced panel of $N$ countries and 169 quarters. For each simulation, I estimate the restricted model $\theta$ for $k = 1, 4, 20$ and compute Hodrick-Ang-Bekaert standard errors. I also calculate standard errors using alternative methods, namely Newey-West (where the maximum lag length is set to $k$) and time-clustered standard errors[10].

The results of the simulation are given in Table 5. Each cell is composed of four lines and three columns. I report the "true" standard error on the first line, i.e. the standard deviation of the coefficient estimate. The average standard error estimated by Hodrick-Ang-Bekaert (HAB), Newey-West (NW) and time-clustered standard errors are reported in lines 2-4. Each column is associated to a given regression horizon $k$. The first row represents the setting in our simulations in which no cross-sectional dependence is present. All standard errors produce results close to the true standard error when $k = 1$. When $k$ increases, however, only HAB standard errors are correct. This is expected: overlapping data lead to serially correlated errors, even under the null hypothesis of no return forecastability, and only HAB standard errors correct this very correlation.

**[Insert Table 5 about here]**

If HAB are robust to overlapping data, we see in the next lines that they are vulnerable to cross-sectional correlation, though not in a trivial way. While we would expect time-clustered standard errors to be the most robust in this context (Gow, Ormazabal, and Taylor, 2010; Petersen, 2008), HAB produce the most accurate standard errors for 1-period regressions. It is the interaction of long-term regression and cross-sectional correlation that generate an under-estimation of the true standard errors. For example, choosing values close to Ang and Bekaert (2007) dataset ($N = 4$ and $\tau = 0.50$), in a 20−quarter regression HAB will underestimate the true standard error by one-third ($1 - 0.0080/0.0120$). Even if the problem is worse with alternative standard errors, HAB do not seem robust enough in our configuration ($N = 10$) where the bias ranges from 7% to 53% (see Figure 8).

[9] Ang and Bekaert (2007) report pairwise correlation of excess returns in the 50% to 60% range.

[10] I use Matlab routines that estimate Newey-West and cluster-robust standard errors available form Gow, Ormazabal, and Taylor (2010) on their website http://www.people.hbs.edu/igow/GOT/.

**[Insert Figure 8 about here]**

## 4.3 Regression Results

I estimate Equation (7) by OLS and compute bootstrapped standard errors of the parameters $\theta = (\alpha_1 \dots \alpha_N \bar{\beta})'$. The bootstrap procedure is similar from the one suggested in Bun (2004) (see Appendix B). 1000 artificial datasets are generated under the null hypothesis of no predictability. The bootstrapped standard errors are the standard deviations of the 1000 realizations $\tilde{\beta}^*$ of the restricted estimator.

Equipped with these standard errors, I report $t$-stats of joint tests of the null of no predictability in Table 6. Results for stock excess returns are reported on the left panel. The evidence of predictability is scarce when considering log dividend-price ratio, alone or combined with short-term yield, in line with Rapach, Strauss, and Zhou (2013). I only find significance at the one-year horizon for the combined regression. Ang and Bekaert (2007) do not find evidence for the pooled log dividend-price ratio alone, but observe a significant relation in a combined regression up to one year, with four countries (FRA, DEU, GBR, and USA). With these four countries, I obtain a similar pattern when using GMM standard errors. However using bootstrapped standard error the one-year horizon predictability disappears and only the short-term interest rate significantly predicts the equity premium.

**[Insert Table 6 about here]**

In line with, Hjalmarsson (2010), the term spread exhibits stronger predictive ability, up to one-year horizon. I also find a positive significant relationship using excess bond returns as predictor. I accept poolability for one-quarter regressions. However, the test of the overidentifying restrictions rejects respectively at the 5% and 1% level at the one-year horizon, suggesting that pooling may not be appropriate when considering one-year regression.

The evidence of excess bond returns is much less controversial (Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005; Fama and Bliss, 1987). The pooled regression results, reported on the right panel of Table 6, confirm this evidence. The term spread is found to predict returns up to one year, alone or combined with the short-term yield. The effect of the latter is found to be positive but insignificant. Finally, I must reject the Roy-Zellner poolability tests in both cases, while the overidentifying restriction is accepted at all horizons.

Overall, the evidence of poolability is mixed. Slope homogeneity is always rejected when the log dividend-price ratio is used as predictor, contrasting from results reported in Hjalmarsson (2010). Roy-Zellner test rejection imply that assuming homogeneous, *constant*, coefficients may be too demanding. The hypothesis of constant coefficients across time is standard in the literature, although it is known to be restrictive. Dangl, Halling, and Randl (2006) and Paye and Timmermann (2006) provided formal evidence for time variation in regression coefficients.

Other papers estimated switching regime models searching for structural breaks in equity returns and forecasting variables, e.g. Pastor and Stambaugh (2001), Bec and Gollier (2010). I also give an account of parameter instability, in the context of out of sample predictions, in Section 4.4. These variations in coefficients may be explained by many factors, e.g. changes in market sentiments, regime switches in monetary policies, institutional or macroeconomic changes, disappearance of market inefficiencies, etc. We should therefore view the pooling hypothesis as stringent, but not much more than assuming constant coefficients or perfect regressors. Finally, pooled regression remain valid even if we cannot accept the hypothesis of poolability. If the slope homogeneity restrictions do not hold exactly, pooled estimates can still meaningfully capture average relationships in the data (Hjalmarsson, 2010).

## 4.4 Out of Sample Predictions

Out-of-sample evidence is now considered as an essential criterion to assess stock return predictability. For example, Welch and Goyal (2008) examined the out-of-sample performance of a large list of predictors. They found that the historical mean exhibits mostly superior performances than most predictors. Campbell and Thompson (2008) responded that short sample estimations often provided implausible coefficients or forecasts. Using simple theory-driven restrictions, they obtained out-of-sample forecasts that are slightly better than the unconditional mean, which implies that investors could have profited by using market-timing strategies. Additional robust out-of-sample evidence of return predictability has been recently provided by, e.g. Ferreira and Santa-Clara, 2011; Neely and Rapach, 2010.

A frequent argument in favor of poolability is the superior forecast delivered by panel estimates. If the fitting of individual coefficients for each country allows for more flexibility, improved out-of-sample performance can be achieved if the time-series estimates are imprecise or unstable, e.g. suffering from a limited number of observations. I evaluate out of sample predictive ability by comparing forecasts based on predictive regressions and forecasts based on historical average of country excess returns. As shown by Welch and Goyal (2008), historical average is a stringent benchmark since forecasts based on macroeconomic variables seldom achieve to outperform the historical average forecast in out-of-sample tests.

The historical average forecast corresponds to the constant expected excess return model,

$$r_{t+1}^i = \beta_0^i + u_{t+1}^i. \tag{9}$$

For each of the 10 countries, I compare this historical average forecast to a forecast generated from a competing predictive regression model, using either pooled:

$$r_{t+1}^i = \alpha_i + \bar{\beta}' z_t^i + u_{t+1}^i, \tag{10}$$

or time series regression model:

$$r_{t+1}^{i} = \alpha_i + \beta_i' z_t^i + u_{t+1}^i. \tag{11}$$

Using an expanding window after an in-sample period of 20 years, I generate time $t$ forecasts based on either Equations (9), (10) or (11) and information through $t$. For example, time series predictions are obtained as $\hat{r}_{t+1}^i = \alpha_{i,t} + \hat{\beta}_{i,t}' z_t^i$. Forming forecasts in this manner simulates the situation of an investor in real time.

I use Campbell and Thompson (2008) out-of-sample $R^2$ statistic, $R_{OS}^2$ (in percent), which measures the proportional reduction in mean-squared forecast error for the competing model relative to the historical average benchmark:

$$R_{OS}^2 = 1 - \frac{\left(\sum_{t=1}^{T} r_t^i - \hat{r}_t^i\right)^2}{\left(\sum_{t=1}^{T} r_t^i - \bar{r}_t^i\right)^2} \tag{12}$$

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

Tables 7 and 8 report the results, respectively for stocks and bonds. I use the same set of predictors as in the previous section; each panel considers the performance of a given predictive model using time series and pooled coefficients. The first two columns of each table reports the in-sample $R^2$ over the considered period, i.e. 1990-2013. The remaining columns evaluate the out-of-sample performance of the forecasts, using time series and pooled regression. '$\Delta$' columns indicate the difference between times series and pooled regression so that a positive value indicates a better forecast when using pooled coefficient.

**[Insert Tables 7 and 8 about here]**

Consider the third columns of Tables 7 and 8, which contain the results of forecasts using time series coefficients. The results of out-of-sample forecasts are overall weak: most of the prediction model fail at consistently producing positive $R_{OS}^2$ statistics, as observed by Welch and Goyal (2008). The most popular predictor of excess stock returns only manages to beat the historical average benchmark for one country (Japan) and produces large errors when combined with the short-term yield. Only the term spread achieves positive results for 5 out of 10 countries in time series predictions of equity premium. The same remark applies to excess bond return predictions, which clearly underperform the unconditional mean, surprisingly contrasting from the somehow large in-sample $R^2$ reported in the first two columns. Moreover, "large mistakes" are common with $R_{OS}^2$ below $-10\%$ in several cases, particularly for bond return regressions. For example, the term spread has been a very poor predictor of Japan returns, with a $-26.99\%$ statistic. One reason of such large mistakes is that sample estimates give conditional expected excess returns less than zero, a prediction that would recommend a short position. In the Japan example, such a negative prediction occurred on 37 quarters, i.e. more than 40% of the time.

Using pooled regressions is a way to reduce the risk of such "implausible" predictions. As reported on the fourth columns, predictions based on pooled coefficients allow an almost systematic gain in out-of-sample predictions (see column 5 in both tables). The risk of making large mistakes is also systematically reduced. Moreover the pooled coefficients are much more stable than their time series counterparts (see Figure 9). The ability to predict excess stock returns is still poor, except for the term spread, which manages to outperform its benchmark in 9 out of 10 countries, as already observed by Hjalmarsson (2010).

[**Insert Figure 9 about here**]

A more straightforward way to avoid "implausible" predictions is to impose restrictions on the estimated coefficients and on the sign of forecasts. Campbell and Thompson (2008) showed that such restrictions improve individual forecasts, allowing them to more consistently outperform the historical average when predicting the equity premium. I introduce similar restrictions, both on stock and bond predictions. I assume that rational investors would rule out a negative risk premia, so that negative predictions would be replaced by a zero forecast. When predicting equity premia, I also rule out predictions, which sign is conflicting with theory, setting the regression coefficient to zero whenever it has the "wrong" sign. More precisely, the coefficient associated to log dividend yield must always be positive. In the spirit of Ang and Bekaert (2007), I also assume that the coefficient associated to short-term interest rate must be negative. I impose these coefficients restrictions first and then the sign restriction on the forecast. The remaining columns of Tables 7 and 8 report the results under these restrictions for time series and pooled models. The improvement in $R^2_{OS}$ is almost systematic, which is not surprising since I eliminate poor forecasts by replacing them by benchmark values. Nevertheless, these restrictions are seldom sufficient to beat the unconditional mean when predicting stock returns, again with the exception of the term spread. The impact is dramatic for bonds, however. The term spread is now able to beat the historical average for all countries except Japan. Finally remark that imposing these restrictions do not exhaust the benefit of pooling. With the exception of the log dividend-price ratio (Panel A of Table 7), pooled predictions exhibit on average a lower mean-squared forecast error than time series predictions.

# 5 Conclusion

This study analyses the international predictability of stock and bond excess returns and its consequence on the term structure of risk, on a panel of ten industrialized countries. Using standard error robust to cross sectional correlation it is found that the strong predictive pattern of Ang and Bekaert (2007) is weaker than expected. Instead, only the term spread appears as a robust predictor of the equity premium in predictive regressions. I also examine the poolability hypothesis. The evidence that a common process generates returns is mixed in sample, which suggests that pooled results are best interpreted as average relations rather than point estimates. However, out-of-sample predictions based on pooled coefficients allow for an almost systematic gain in predictions, even when imposing economic restrictions on the return forecasts.

While it is generally admitted that stocks are safer on the long run, I instead obtain a rather flat term structure of risk. Therefore the "conventional wisdom" on stock investing may not robustly extend outside the US. The volatilities of stocks and bonds are almost constant on a ten-year horizon. Using bootstrapped confidence intervals of the conditional moments, I show that stocks are significantly riskier than bonds or bills at any horizon. The analysis of conditional moments also uncovers a remarkably low correlation pattern across asset classes. Long-term correlation remains low, even at a ten-year horizon. Overall, while I observe predictability with more precision than in studies focusing on a single country, the consequence of predictability is surprisingly weak.

# Appendix

## A Ang and Bekaert (2007) Approach to Cross-sectional Regressions

Ang and Bekaert (2007) show how to adapt Hodrick (1992) standard errors to pooled regressions. They also derive the Hansen (1982) $\chi^2$ $J$-test of overidentifying restrictions in this setup.

I start by estimating the system

$$\tilde{r}^i_{t+k} = \alpha_i + \beta_i' z^i_t + u^i_{t+k} \tag{A.1}$$

for $i = 1 \dots N$ countries, subject to the restriction $\beta_i = \bar{\beta}\ \forall i$. $\tilde{r}^i_{t+k}$ are cumulated k-quarter ahead excess returns. This specification allows for freely estimated $\alpha_i$, which makes the present approach very close to the more common least-squares dummy variables (LSDV) regression.

Let the dimension of $z_t$ be $(K-1)$ so that there will be a total of $K$ regressors, including the constant terms $\alpha_i$ for each of $N$ countries. In Equation (A.1), I denote the free parameters $\theta = (\alpha_1 \dots \alpha_N \bar{\beta})'$ and the unrestricted parameters stacked by each equation $\beta = (\alpha_1 \beta_1' \dots \alpha_N \beta_N')'$. I estimate Equation (A.1) subject to the restriction that $C\beta = 0$, where C is a $(N-1)(K-1) \times NK$ matrix of the form:

$$C = \begin{pmatrix} \tilde{0} & I_{K-1} & \tilde{0} & -I_{K-1} & \tilde{0} & \cdots & & & \\ \tilde{0} & O & \tilde{0} & I_{K-1} & \tilde{0} & -I_{K-1} & \cdots & & \\ \vdots & & & & & & & & \\ \tilde{0} & O & \tilde{0} & & & & & \tilde{0} & -I_{K-1} \end{pmatrix} \tag{A.2}$$

where $\tilde{0}$ is a $(K-1)\times 1$ vector of zeros and $O$ is a $(K-1)\times(K-1)$ matrix of zeros. Stacking observations across countries,

$$\begin{aligned} \tilde{r}_{t+k} &= (\tilde{r}^1_{t+k} \ldots \tilde{r}^N_{t+k})' \qquad (N \times 1) \\ x^i_t &= (1 z^{i'}_t) \qquad (K \times 1) \\ u_{t+k} &= (u^1_{t+k} \ldots u^N_{t+k})' \qquad (N \times 1) \\ X_t &= \begin{pmatrix} x^1_t & & 0 \\ & \ddots & \\ 0 & & x^N_t \end{pmatrix} \quad (NK \times N) \end{aligned}$$

the system can be written as

$$\tilde{r}_{t+k} = X'_t \beta + u_{t+k} \tag{A.3}$$

subject to $C\beta = 0$. I can rewrite the system in compact form by stacking observations across time as well. Let $R = (\tilde{r}'_{1+k} \ldots \tilde{r}'_{T+k})'$, $X = (X'_1 \ldots X'_T)'$ and $U = (u'_{1+k} \ldots u'_{T+k})'$. I can now write

$$R = X\beta + U, \tag{A.4}$$

Subject to the set of restrictions $C\beta = 0$. A consistent estimate $\tilde{\beta}$ of $\beta$ is given by

$$\tilde{\beta} = \beta^{ols} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C\beta^{ols}, \tag{A.5}$$

with $\beta^{ols} = (X'X)^{-1}X'R$. This gives us a consistent estimate $\hat{\theta}$ of the vector of restricted coefficients $\theta$.

The moment conditions of the system in Equation (A.3) are

$$E(h_{t+k}) = E(X_t u_{t+k}) = 0.$$

By standard GMM, $\hat{\theta}$ has distribution

$$\sqrt{T}(\hat{\theta} - \theta) \overset{a}{\sim} N(0, (D'_0 S_0^{-1} D_0)^{-1}),$$

with

$$D_0' = E\left(\frac{\partial h_{t+k}}{\partial \theta'}\right)$$

and

$$S_0 = E(h_{t+k}h_{t+k}').$$

The Hodrick (1992) estimate $\hat{S}_T^b$ of $S_0$ is given by

$$\hat{S}_T^b = \frac{1}{T}\sum_{t=k}^{T} wk_t wk_t', \tag{A.6}$$

where the $(NK \times 1)$ vector $wk_t$ is given by

$$wk_t = \left(\sum_{i=0}^{k-1} X_{t-i}\right) e_{t+1.}$$

Under the null hypothesis of no predictability, $u_{t+k} = e_{t+1} + \cdots + e_{t+k}$, where $e_{t+1}$ are the one-step-ahead serially uncorrelated errors, obtained by regressing 1-quarter ahead returns on a constant.

An estimate $\hat{D}_T$ of $D_0$ is given by

$$\hat{D}_T' = \frac{1}{T}\sum_{t=0}^{T} \frac{\partial h_{t+k}}{\partial \theta'},$$

where $\theta = (\alpha_1 \ldots \alpha_N \bar{\beta'})$ and

$$-\frac{\partial h_{t+k}}{\partial \theta'} = \begin{bmatrix} 1 & z_t^{1\prime} & & & & 0 & \\ & & 1 & z_t^{2\prime} & & & \\ & & & & \ddots & & \\ & & 0 & & & 1 & z_t^{N\prime} \\ z_t^1 & z_t^1 z_t^{1\prime} & z_t^2 & z_t^2 z_t^{2\prime} & \cdots & z_t^N & z_t^N z_t^{N\prime} \end{bmatrix}$$

The estimate $\hat{\theta}$ has the distribution

$$\sqrt{T}(\hat{\theta} - \theta) \overset{a}{\sim} N(0, [\hat{D_T'}(\hat{S}_T^b)^{-1}\hat{D_T}]^{-1}). \tag{A.7}$$

There are $(N + K - 1)$ free parameters in $\theta$ with $NK$ moment conditions. This gives $NK - (N+K-1)$ overidentifying restrictions. The Hansen (1982) $\chi^2$ $J$-test of overidentifying restrictions is given by

$$J = T(\bar{h}'(\hat{S}_T^b)^{-1}\bar{h}) \sim \chi^2[NK - (N + K - 1)], \tag{A.8}$$

with

$$\bar{h} = \frac{1}{T}\sum_{t=0}^{T} h_{t+k}.$$

## B Roy-Zellner Test of Poolability

I briefly present the econometrics of Roy (1957) and Zellner (1962) test of poolability. A comprehensive account together with a Monte Carlo analysis can be found in Bun (2004). We come back to the compact model described by Equation A.4: $R = X\beta + U$, with $k = 1$. A test of poolability of slope coefficients $\beta_1 = \cdots = \beta_N$ can again be expressed as $C\beta = 0$, $C$ being defined in Equation (A.2). Alternatively, a test of full poolability imposes $J = (N-1)K$ restrictions with

$$D = \begin{pmatrix} I_K & -I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & -I_K & \cdots & 0 & 0 \\ \vdots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & I_K & -I_K \end{pmatrix}$$

To account for nonspherical disturbances, the test statistic I consider is based on the feasible generalized least squares (FGLS) estimator $\hat{\beta}_{FGLS}$ of the unconstrained regression:

$$\hat{\beta}_{FGLS} = \left(X'\hat{\mathbf{\Omega}}^{-1}X\right)^{-1} X'\hat{\mathbf{\Omega}}^{-1}R \tag{B.1}$$

where $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}} \otimes I_T$ is an estimator for the unknown $\mathbf{\Omega}$, $\hat{\mathbf{\Sigma}}$ being the variance-covariance matrix of the residuals of a first step OLS regression. The generalized $F$ statistic can be written as :

$$F^g = \frac{N(T-K)}{J}\frac{q_1^g}{q_2^g} \tag{B.2}$$

where

$$q_1^g = (C\hat{\beta}_{FGLS} - r)'[C\left(X'\hat{\mathbf{\Omega}}^{-1}X\right)^{-1}C']^{-1}(C\hat{\beta}_{FGLS} - r)$$

and

$$q_2^g = (R - X\hat{\beta}_{FGLS})'\hat{\mathbf{\Omega}}^{-1}(R - X\hat{\beta}_{FGLS}).$$

My bootstrap follows the resampling scheme of Bun (2004). I first collect the unrestricted estimators, the residuals $\hat{U} = (\hat{u}_1 \ldots \hat{u}_T)$ and the restricted estimator $\tilde{\beta}$. A sample or errors $U^* = (u_1^* \ldots u_T^*)$ is drawn from the fitted residuals to construct $u^* = vec(U^{*\prime})$. I then obtain an artificial $R^* = X^*\tilde{\beta} + u^*$, estimate the model and calculate the test statistic $F^{g*}$ from the resampled data $(R^*, X^*)$.

I use $X^* = X$ when the regressors are strictly exogenous. In case of a lagged dependent variable regressor $X^*$ is generated iteratively from the first observation. I repeat these steps 999 times and construct a size-corrected test with the available 1000 realizations (including the test statistic from the original data). $H_0: \; C\beta = 0$ is rejected when the test statistic $F^g$ is larger than the $(1-\alpha)$ quantile of its bootstrapped distribution.

# C Conditional k-period Variance

This section details how to extract moments of real returns from the excess real returns in the VAR(1) model. I drop the index $i$ for convenience.

First, I derive a set of equations relating $z_{t+k}$ to its current value $z_t$ plus a weighted sum of interim shocks:

$$\mathbf{z_{t+1}} = \Phi_0 + \Phi_1 \mathbf{z_t} + \mathbf{v_{t+1}}$$

$$\begin{aligned}\mathbf{z_{t+2}} &= \Phi_0 + \Phi_1 \mathbf{z_{t+1}} + \mathbf{v_{t+2}} \\ &= \Phi_0 + \Phi_1\Phi_0 + \Phi_1\Phi_1\mathbf{z_t} + \Phi_1\mathbf{v_{t+1}} + \mathbf{v_{t+2}}\end{aligned}$$

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

Taking the sum and reordering terms yields:

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

Or more compactly :

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

I am now able to compute conditional variance:

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

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

$$\begin{aligned} V_t\left(\mathbf{z_{t+1}} + \mathbf{z_{t+2}} + ... + \mathbf{z_{t+k}}\right) &= \Sigma_v + \left(\mathbf{I} + \Phi_1\right)\Sigma_v\left(\mathbf{I} + \Phi_1\right)' \\ &\quad + \left(\mathbf{I} + \Phi_1 + \Phi_1\Phi_1\right)\Sigma_v\left(\mathbf{I} + \Phi_1 + \Phi_1\Phi_1\right)' \\ &\quad + ... \\ &\quad + \left(\mathbf{I} + \Phi_1 + ... + \Phi_1^{k-1}\right)\Sigma_v\left(\mathbf{I} + \Phi_1 + ... + \Phi_1^{k-1}\right)'. \end{aligned}$$

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

$$\frac{1}{k}V_t\begin{pmatrix} r_{0,t+k}^{(k)} \\ r_{e,t+k}^{(k)} \\ r_{b,t+k}^{(k)} \end{pmatrix} = \frac{1}{k}\mathbf{M}\,V_t\left(\mathbf{z_{t+1}} + \mathbf{z_{t+2}} + ... + \mathbf{z_{t+k}}\right)\,\mathbf{M}'. \tag{C.1}$$

$\mathbf{M}$ is a $(K-1)\times(K-1)$ selection matrix :

$$\mathbf{M} = \begin{pmatrix} 1 & 0_{1\times n} & 0_{1\times(K-n-2)} \\ \iota_{n\times 1} & \mathbf{I_{n\times n}} & 0_{(K-n-2)\times(K-n-2)} \end{pmatrix}$$

where $n$ is the number of assets beyond the short real interest rate ($n = 2$ here) and $K-1$ the number of regressors.

# References


Ang, A. and G. Bekaert (2007). "Stock Return Predictability: Is it There?" *Review of Financial Studies* 20.3, pp. 651–707.

Arellano, M. and S. Bond (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". *The Review of Economic Studies* 58.2, pp. 277–297.

Avramov, D. (2002). "Stock return predictability and model uncertainty". *Journal of Financial Economics* 64, pp. 423–458.

Barberis, N. (2000). "Investing for the long run when returns are predictable". *The Journal of Finance* 55.1, pp. 225–264.

Bec, F. and C. Gollier (2009). "Assets Returns Volatility and Investment Horizon : The French Case". *CESifo Working Paper* 2622.

— (2010). "Cyclicality and Term Structure of Value-at-Risk within a Threshold Autoregression Setup". *Toulouse School of Economics, Working Paper*.

Bun, M. J. (2004). "Testing poolability in a system of dynamic regressions with nonspherical disturbances". *Empirical Economics* 29.1, pp. 89–106.

Campbell, J. Y. and R. J. Shiller (1991). "Yield spreads and interest rate movements: A bird's eye view". *The Review of Economic* 58.3, pp. 495–514.

Campbell, J. Y. and S. B. Thompson (2008). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies* 21.4, pp. 1509–1531.

Campbell, J. Y. (1987). "Stock returns and the term structure". *Journal of financial economics* 18.2, pp. 373–399.

— (1991). "A variance decomposition for stock returns". *The Economic Journal* 101.405, pp. 157–179.

Campbell, J. Y., A. W. Lo, and C. A. MacKinlay (1997). *The econometrics of financial markets*. Princeton University Press.

Campbell, J. Y. J. and L. M. Viceira (2005). "The term structure of the risk-return tradeoff". *Financial Analysts Journal* 61.61, pp. 34–44.

Campbell, J. (1988). "The dividend-price ratio and expectations of future dividends and discount factors". *Review of financial studies* 1.3, pp. 195–228.

Campbell, J., Y. Chan, and L. Viceira (2003). "A Multivariate Model of Strategic Asset Allocation". *Journal of Financial Economics* 67, pp. 41–80.

Campbell, J. and L. Viceira (2002). "Strategic Asset Allocation".

Canova, F. and M. Ciccarelli (2013). *Panel Vector Autoregressive Models: A Survey*. 15.

Cochrane, J. and M. Piazzesi (2005). "Bond risk premia". *The American Economic Review* 95.1, pp. 138–160.

Dangl, T., M. Halling, and O. Randl (2006). "Equity Return Prediction : Are Coefficients Time Varying ?" *EFA 2006 Zurich Meetings Paper*.

Fama, E. F. and R. R. Bliss (1987). "The information in long-maturity forward rates". *The American Economic Review* 77.4, pp. 680–692.

Fama, E. F. and K. R. French (1988). "Dividend yields and expected stock returns". *Journal of Financial Economics* 22.1, pp. 3–25.

— (1989). "Business conditions and expected returns on stocks and bonds". *Journal of financial economics* 25.1, pp. 23–49.

Ferreira, M. a. and P. Santa-Clara (2011). "Forecasting stock market returns: The sum of the parts is more than the whole". *Journal of Financial Economics* 100.3, pp. 514–537.

Gow, I., G Ormazabal, and D. Taylor (2010). "Correcting for cross-sectional and time-series dependence in accounting research". *The Accounting Review* 85.2, pp. 483–512.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica* 50.4, pp. 1029–1054.

Hjalmarsson, E. (2008). "The Stambaugh bias in panel predictive regressions". *Finance Research Letters* 5.1, pp. 47–58.

— (2010). "Predicting global stock returns". *Journal of Financial and Quantitative Analysis* 45.1, pp. 49–80.

Hodrick, R. (1992). "Dividend yields and expected stock returns: Alternative procedures for inference and measurement". *Review of Financial studies* 5.3, pp. 357–386.

Hoevenaars, R. P. M. M. et al. (2013). "Strategic asset allocation for long-term investors: Parameter uncertainty and prior information". *Journal of Applied Econometrics*.

Jondeau, E. and M. Rockinger (2010). "Portfolio Allocation for European Markets with Predictability and Parameter Uncertainty". *Swiss Finance Institute Research Paper* 10.41.

Judson, R. and A. Owen (1999). "Estimating dynamic panel data models : a guide for for macroeconomists". *Economics letters* 65, pp. 9–15.

Kandel, S. and R. F. Stambaugh (1988). "Modelling expected stock returns for short and long horizons". Philadelphia.

— (1989). "Modeling expected stock returns for long and short horizons". *Rodney L. White Center for Financial Research Working Papers*. Rodney L. White Center for Financial Research Working Papers.

Keim, D. B. and R. F. Stambaugh (1986). "Predicting returns in the stock and bond markets". *Journal of financial Economics* 17.2, pp. 357–390.

Levin, A., C. Lin, and C. J. Chu (2002). "Unit root tests in panel data: Asymptotic and finite-sample properties". *Journal of econometrics* 108, pp. 1–24.

Lioui, A. and P. Poncet (2012). "Long Horizon Predictability : An Asset Allocation Perspective". *Working Paper*.

Lo, A. and A. MacKinlay (1988). "Stock market prices do not follow random walks: Evidence from a simple specification test". *Review of financial studies*.

Neely, C. and D. Rapach (2010). "Forecasting the equity risk premium: the role of technical indicators". *Federal Reserve Bank of ...*

Nickell, S. (1981). "Biases in dynamic models with fixed effects". *Econometrica* 49.6, pp. 1417–1426.

Pastor, L. and R. F. Stambaugh (2001). "The equity premium and structural breaks". *Journal of Finance* 56.4, pp. 1207–1239.

— (2012). "Are Stocks Really Less Volatile in the Long Run?" *Journal of Finance* 67.2, pp. 431–478.

Paye, B. S. and A. Timmermann (2006). "Instability of return prediction models". *Journal of Empirical Finance* 13.3, pp. 274–315.

Pesaran, H. M. and R. Smith (1995). "Estimating long-run relationships from dynamic heterogeneous panels". *Journal of Econometrics* 68.1, pp. 79–113.

Pesaran, M. and A. Timmermann (1995). "Predictability of stock returns: Robustness and economic significance". *The Journal of Finance* 50.4, pp. 1201–1228.

Petersen, M. a. (2008). "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches". *Review of Financial Studies* 22.1, pp. 435–480.

Polk, C., S. Thompson, and T. Vuolteenaho (2006). "Cross-sectional forecasts of the equity premium". *Journal of Financial Economics* 81.1, pp. 101–141.

Poterba, J. and L. Summers (1988). "Mean reversion in stock prices: Evidence and implications". *Journal of Financial Economics*.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). "International Stock Return Predictability: What Is the Role of the United States?" *The Journal of Finance* 68.4, pp. 1633–1662.

Roy, S. N. (1957). *Some aspects of multivariate analysis*. New York: Wiley.

Rozeff, M. (1984). "Dividend yields are equity risk premiums". *Journal of Portfolio management*, pp. 68–75.

Siegel, J. J. (1998). *Stocks for the long run*. McGraw Hil. New York.

Stambaugh, R. F. (1999). "Predictive Regressions". *Journal of Financial Economics* 54.3, pp. 375–421.

Swamy, P. A. (1970). "Efficient inference in a random coefficient regression model". *Econometrica: Journal of the Econometric Society*.

Welch, I. and A. Goyal (2008). "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". *Review of Financial Studies* 21.4, pp. 1455–1508.

Zellner, A (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias". *Journal of the American statistical Association* 57.298, pp. 348–368.

## Table 1: Summary Statistics, Q1-1971 - Q1-2013

| | $r_0$ | $x_e$ | $x_b$ | $y$ | $ldp$ | $spr$ |
|---|---|---|---|---|---|---|
| **Sample averages (standard deviations in parentheses).** | | | | | | |
| Australia | 2.311 | 4.539 | 1.255 | 7.784 | -3.228 | 0.538 |
| | (2.298) | (20.024) | (7.716) | (1.898) | (0.109) | (0.940) |
| Belgium | 2.244 | 6.196 | 2.014 | 6.079 | -3.087 | 0.967 |
| | (1.719) | (21.579) | (5.860) | (1.829) | (0.204) | (0.812) |
| Canada | 2.337 | 4.649 | 1.929 | 6.565 | -3.601 | 0.983 |
| | (1.572) | (18.454) | (7.799) | (1.928) | (0.179) | (0.869) |
| France | 2.036 | 5.435 | 1.987 | 6.520 | -3.531 | 0.827 |
| | (1.491) | (21.001) | (6.873) | (1.897) | (0.265) | (0.663) |
| Germany | 2.218 | 5.078 | 2.347 | 5.017 | -3.653 | 0.996 |
| | (1.218) | (20.042) | (6.812) | (1.415) | (0.186) | (0.775) |
| Italy | 1.833 | 2.047 | 1.528 | 8.550 | -3.579 | 0.618 |
| | (1.999) | (24.482) | (8.486) | (2.622) | (0.208) | (0.971) |
| Japan | 0.003 | 5.564 | 3.153 | 2.558 | -4.370 | 1.809 |
| | (2.004) | (19.925) | (7.191) | (1.150) | (0.278) | (0.392) |
| Sweden | 1.977 | 10.539 | 2.038 | 6.688 | -3.591 | 0.940 |
| | (2.145) | (23.822) | (7.342) | (1.993) | (0.230) | (0.916) |
| United Kingdom | 1.454 | 6.434 | 2.539 | 7.067 | -3.276 | 1.115 |
| | (2.643) | (20.465) | (8.586) | (1.853) | (0.140) | (0.969) |
| United States | 1.421 | 5.325 | 2.110 | 5.576 | -3.593 | 1.109 |
| | (1.562) | (16.528) | (8.981) | (1.789) | (0.222) | (0.911) |
| **Levin-Lin-Chu unit-root test** | | | | | | |
| LLC stat | -1.461 (3.6) | -30.500 (1.2) | -31.852 (1.9) | -4.069 (2.0) | -2.429 (1.2) | -4.876 (3.1) |
| $p$-value | 0.072 | 0.000 | 0.000 | 0.000 | 0.008 | 0.000 |

This table reports the summary statistics of the variables of interest for ten countries. $r_0$ is the log real return on short-term interest rates. $x_e$ and $x_b$ are the log real returns of stocks and bonds. $y$ is the nominal counterpart of $r_0$. $ldp$ is the log dividend-price ratio. $spr$ is the term spread. Levin, Lin, and Chu (2002) (LLC) panel unit root tests are a panel generalization of ADF tests and therefore are performed under the null of a unit root. The number of autoregressive lags is selected in order to minimize the Akaike information criterion and is reported in parenthesis. The test statistic is computed on demeaned series, in order to mitigate the impact of cross-sectional dependence.

**Table 2: VAR Estimation Results**

| | $r_{0,t+1}$ | $x_{e,t+1}$ | $x_{b,t+1}$ | $y_{t+1}$ | $ldp_{t+1}$ | $spr_{t+1}$ |
|---|---|---|---|---|---|---|
| Coefficient estimates | | | | | | |
| $r_{0,t}$ | 0.4810 | 0.8373 | 0.3896 | -0.0230 | -1.1326 | 0.0304 |
| | (0.0300)*** | (0.4052)** | (0.1461)*** | (0.0085)*** | (0.4677)** | (0.0334) |
| $x_{e,t}$ | 0.0023 | 0.1090 | -0.0163 | 0.0015 | -0.0792 | -0.0038 |
| | (0.0027) | (0.0458)** | (0.0159) | (0.0009)* | (0.0537) | (0.0033) |
| $x_{b,t}$ | 0.0196 | 0.2673 | 0.0802 | -0.0097 | -0.2383 | 0.0305 |
| | (0.0071)*** | (0.1183)** | (0.0429)* | (0.0023)*** | (0.1362)* | (0.0088)*** |
| $y_t$ | 0.1672 | -0.3874 | 0.1969 | 1.0136 | 0.5020 | -0.1015 |
| | (0.0386)*** | (0.5463) | (0.1927) | (0.0120)*** | (0.6325) | (0.0453)** |
| $ldp_t$ | -0.0015 | 0.0176 | -0.0008 | -0.0001 | 0.9624 | 0.0006 |
| | (0.0005)*** | (0.0072)** | (0.0028) | (0.0002) | (0.0084)*** | (0.0006) |
| $spr_t$ | -0.0322 | 0.4365 | 0.4885 | 0.0399 | -0.5806 | 0.7698 |
| | (0.0183)* | (0.2747) | (0.0969)*** | (0.0064)*** | (0.3221)* | (0.0245)*** |
| $c$ | -0.0056 | 0.0667 | -0.0084 | -0.0010 | -0.1303 | 0.0056 |
| | (0.0021)*** | (0.0291)** | (0.0113) | (0.0007) | (0.0347)*** | (0.0025)** |

Bootstrapped standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

| Annualized standard deviations and cross-correlations of residuals | | | | | | |
|---|---|---|---|---|---|---|
| | $r_0$ | $x_e$ | $x_b$ | $y$ | $ldp$ | $spr$ |
| $r_0$ | 1.5574 | -0.0956 | 0.0573 | 0.2038 | 0.0432 | -0.2612 |
| $x_e$ | | 20.3471 | 0.0335 | -0.1327 | -0.8268 | 0.1153 |
| $x_b$ | | | 7.4395 | -0.4970 | -0.0364 | -0.0089 |
| $y$ | | | | 0.5146 | 0.1192 | -0.8591 |
| $ldp$ | | | | | 24.3582 | -0.1013 |
| $spr$ | | | | | | 2.0181 |

The tables reports OLS pooled coefficient estimates of the VAR $\mathbf{z}_t^i = \Phi_0^i + \Phi_1^i \mathbf{z}_{t-1}^i + \mathbf{v}_t^i$ and the correlation matrix of residuals. Diagonal entries of the matrix are standard deviations expressed in annualized percentage points. Off-diagonal entries are correlations.

**Table 3: VAR Poolability Tests**

| Variable | Homogeneity of variances* | Homogeneity of VAR coefficients** |
|---|---|---|
| $r_0$ | 0.0000 | 0.0000 |
| $x_e$ | 0.0002 | 0.1550 |
| $x_b$ | 0.0005 | 0.0350 |
| $r_0^{nom}$ | 0.0000 | 0.0350 |
| $ldp$ | 0.0000 | 0.0000 |
| $spr$ | 0.0000 | 0.0040 |

*$p$-values, Brown-Forsythe test of homogeneity of variances.
**$p$-values, Roy-Zellner test of poolability with bootstrapped standard errors.

This table presents, for each of the VAR components, the results of tests of homogeneity of variances between countries and a test of full poolability of the VAR coefficients. The null hypotheses is respectively homogeneity of variances and dynamic homogeneity of estimates.

## Table 4: Summary Statistics of the Unbalanced Panel

| | $r_0$ | $x_e$ | $x_b$ | $y$ | $ldp$ | $spr$ |
|---|---|---|---|---|---|---|
| **Sample averages (standard deviations in parentheses).** | | | | | | |
| Austria | 1.758 | 5.877 | 2.930 | 3.971 | -3.915 | 1.153 |
| Q2-1989 - Q1-2013 | (1.487) | (31.993) | (5.640) | (1.250) | (0.209) | (0.633) |
| Chile | 0.963 | 10.388 | 3.452 | 4.159 | -3.673 | 1.921 |
| Q4-2004 - Q1-2013 | (1.807) | (18.111) | (7.391) | (0.964) | (0.081) | (0.982) |
| Czech Republic | 0.001 | 11.826 | 4.726 | 2.543 | -3.325 | 1.754 |
| Q3-2000 - Q1-2013 | (1.738) | (26.985) | (7.930) | (0.658) | (0.304) | (0.359) |
| Denmark | 2.890 | 7.204 | 3.076 | 7.480 | -3.804 | 1.177 |
| Q1-1972 - Q1-2013 | (2.005) | (20.222) | (14.160) | (2.406) | (0.276) | (1.170) |
| Estonia | -0.474 | 14.238 | 1.618 | 3.676 | -3.967 | 1.888 |
| Q2-2003 - Q4-2010 | (2.254) | (36.401) | (10.804) | (0.936) | (1.267) | (0.599) |
| Finland | 2.624 | 10.473 | 4.328 | 4.720 | -3.755 | 1.244 |
| Q1-1989 - Q1-2013 | (1.727) | (38.172) | (8.035) | (1.921) | (0.330) | (0.765) |
| Greece | -0.345 | -7.320 | 2.249 | 2.400 | -3.412 | 4.983 |
| Q1-2001 - Q1-2013 | (2.828) | (38.473) | (19.674) | (0.692) | (0.173) | (3.275) |
| Hungary | 2.776 | 3.626 | 1.047 | 8.459 | -4.064 | -1.104 |
| Q2-1999 - Q1-2013 | (1.841) | (33.475) | (11.714) | (1.050) | (0.249) | (0.905) |
| Ireland | 2.215 | 0.858 | 2.912 | 4.689 | -3.738 | 1.384 |
| Q1-1989 - Q1-2013 | (2.101) | (24.838) | (9.630) | (1.775) | (0.154) | (1.492) |
| Israel | 2.963 | 3.181 | 0.630 | 5.530 | -3.912 | 0.192 |
| Q2-1997 - Q1-2013 | (2.484) | (21.706) | (10.534) | (1.764) | (0.229) | (1.994) |
| Korea | 3.359 | 4.972 | 4.004 | 7.493 | -4.264 | 0.821 |
| Q1-1989 - Q1-2013 | (2.130) | (32.358) | (11.887) | (2.432) | (0.332) | (0.980) |
| Mexico | 2.529 | 10.401 | 16.427 | 11.490 | -4.096 | 1.582 |
| Q2-1995 - Q1-2013 | (2.839) | (23.275) | (22.854) | (4.436) | (0.153) | (0.613) |
| Netherlands | 2.136 | 6.637 | 2.695 | 4.072 | -3.391 | 1.094 |
| Q1-1986 - Q1-2013 | (1.461) | (20.500) | (5.777) | (1.177) | (0.150) | (0.571) |
| New Zealand | 4.194 | 0.832 | 3.073 | 6.644 | -2.964 | -0.086 |
| Q1-1989 - Q1-2013 | (1.399) | (16.872) | (7.869) | (1.372) | (0.102) | (0.601) |
| Norway | 3.736 | 9.013 | 1.336 | 7.574 | -3.578 | -0.389 |
| Q1-1979 - Q1-2013 | (1.789) | (29.078) | (7.917) | (2.064) | (0.198) | (0.694) |
| Poland | 2.760 | 5.260 | 4.556 | 5.380 | -3.553 | 0.535 |
| Q2-2001 - Q1-2013 | (1.990) | (27.866) | (8.062) | (1.406) | (0.246) | (0.995) |
| Portugal | 1.929 | 0.996 | 3.689 | 6.048 | -3.472 | 1.374 |
| Q1-1989 - Q1-2013 | (1.887) | (23.150) | (10.267) | (2.560) | (0.168) | (1.568) |
| Russia | -6.140 | 21.830 | 24.466 | 5.578 | -4.189 | 8.826 |
| Q2-1999 - Q1-2013 | (3.351) | (43.925) | (25.690) | (1.849) | (0.376) | (7.331) |
| Slovenia | -0.061 | 4.763 | 3.258 | 2.615 | -4.268 | 1.971 |
| Q2-2003 - Q1-2013 | (1.870) | (25.389) | (8.778) | (0.912) | (0.332) | (0.992) |
| South Africa | 3.741 | 6.075 | 2.846 | 9.539 | -3.581 | 1.152 |
| Q1-1994 - Q1-2013 | (1.999) | (19.739) | (9.641) | (1.602) | (0.117) | (0.969) |
| Spain | 2.304 | 7.375 | 2.378 | 7.884 | -3.090 | 0.790 |
| Q2-1978 - Q1-2013 | (2.270) | (23.034) | (8.519) | (2.803) | (0.307) | (1.379) |
| Switzerland | 1.210 | 6.419 | 1.823 | 3.315 | -3.836 | 0.542 |
| Q1-1974 - Q1-2013 | (1.256) | (17.194) | (5.818) | (1.357) | (0.174) | (0.855) |

This table reports extends summary statistics and timespan to the 22 countries which are used in building the unbalanced panel VAR in Section 3.3.

## Table 5: Comparing Standard Errors

| | | Number of countries | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $N=2$ | | | $N=4$ | | | $N=10$ | | |
| $Std(B_{HAB})$<br>$Avg(SE_{HAB})$<br>$Avg(SE_{NW})$<br>$Avg(SE_{Clust})$ | | | | | | | | | | |
| | | $k=1$ | $k=4$ | $k=20$ | $k=1$ | $k=4$ | $k=20$ | $k=1$ | $k=4$ | $k=20$ |
| Cross-sectional correlation | 0% | 0.0284 | 0.0267 | 0.0195 | 0.0131 | 0.0124 | 0.0108 | 0.0064 | 0.0065 | 0.0058 |
| | | 0.0282 | 0.0259 | 0.0186 | 0.0132 | 0.0126 | 0.0100 | 0.0068 | 0.0064 | 0.0051 |
| | | 0.0275 | 0.0216 | 0.0145 | 0.0122 | 0.0102 | 0.0088 | 0.0054 | 0.0046 | 0.0042 |
| | | 0.0277 | 0.0137 | 0.0058 | 0.0123 | 0.0061 | 0.0029 | 0.0054 | 0.0026 | 0.0012 |
| | 25% | 0.0129 | 0.0127 | 0.0120 | 0.0126 | 0.0119 | 0.0115 | 0.0085 | 0.0082 | 0.0086 |
| | | 0.0131 | 0.0120 | 0.0110 | 0.0126 | 0.0107 | 0.0095 | 0.0082 | 0.0063 | 0.0056 |
| | | 0.0113 | 0.0094 | 0.0074 | 0.0103 | 0.0087 | 0.0073 | 0.0061 | 0.0053 | 0.0049 |
| | | 0.0122 | 0.0059 | 0.0024 | 0.0117 | 0.0058 | 0.0024 | 0.0077 | 0.0039 | 0.0018 |
| | 50% | 0.0160 | 0.0157 | 0.0151 | 0.0131 | 0.0127 | 0.0120 | 0.0119 | 0.0116 | 0.0098 |
| | | 0.0160 | 0.0140 | 0.0123 | 0.0125 | 0.0088 | 0.0080 | 0.0111 | 0.0077 | 0.0064 |
| | | 0.0138 | 0.0113 | 0.0087 | 0.0083 | 0.0070 | 0.0059 | 0.0072 | 0.0061 | 0.0053 |
| | | 0.0158 | 0.0077 | 0.0032 | 0.0115 | 0.0057 | 0.0023 | 0.0095 | 0.0047 | 0.0021 |
| | 75% | 0.0260 | 0.0246 | 0.0226 | 0.0254 | 0.0247 | 0.0199 | 0.0116 | 0.0125 | 0.0125 |
| | | 0.0257 | 0.0201 | 0.0170 | 0.0238 | 0.0146 | 0.0117 | 0.0111 | 0.0048 | 0.0037 |
| | | 0.0187 | 0.0152 | 0.0112 | 0.0151 | 0.0122 | 0.0087 | 0.0045 | 0.0038 | 0.0031 |
| | | 0.0210 | 0.0104 | 0.0044 | 0.0242 | 0.0118 | 0.0049 | 0.0117 | 0.0057 | 0.0023 |

The table contains estimates of standard errors based on simulated panel data sets. Each data set contains $N$ countries and 169 quarters per country. In each simulation I regress Equation (7) for $k = 1, 4, 20$. The independent variable X and the residual are specified as

$$\begin{aligned} u_t^i &= \sqrt{\tau}\gamma_t + \sqrt{1-\tau}\eta_t^i \\ z_t^i &= z_0 + \rho z_{t-1}^i + \epsilon_t^i \\ \epsilon_t^i &= \sqrt{\tau}\mu_t + \sqrt{1-\tau}\nu_t^i \end{aligned}$$

The parameters of the data generating process are chosen to roughly match their empirical counterpart. $\tau$ determines the level of cross-sectional correlation and varies across the rows of the table from 0% (no time effect) to 75%. I consider $N = 2, 4, 10$ countries, the latter being the actual number of countries in my sample. Each cell is composed of four lines and three columns. Each column is associated to a given regression horizon $k$. The true standard error (i.e. the standard deviation of the coefficient estimate) varies across configurations and is reported the first line. The average standard error estimated by Hodrick-Ang-Bekaert (HAB), Newey-West (NW) and time-clustered standard errors are reported in lines 2-4.

## Table 6: Excess Return Regression

| **Stocks** | | | | | **Bonds** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $k$-qtrs | *ldp* | | $J$-test | $RZ$-test | $k$-qtrs | *spr* | | $J$-test | $RZ$-test |
| 1 | 0.0139 | | 0.233 | 0.011** | 1 | 0.3503 | | 0.428 | 0.019** |
| | (1.1440) | | | | | ( 3.8978)*** | | | |
| 4 | 0.0187 | | 0.660 | | 4 | 0.2211 | | 0.427 | |
| | (1.6233) | | | | | ( 2.9572)*** | | | |
| 20 | 0.0167 | | 0.210 | | 20 | -0.0408 | | 0.254 | |
| | (1.4265) | | | | | (-0.9300) | | | |
| $k$-qtrs | *spr* | | $J$-test | $RZ$-test | $k$-qtrs | *y* | *spr* | $J$-test | $RZ$-test |
| 1 | 0.4930 | | 0.134 | 0.572 | 1 | 0.2135 | 0.4192 | 0.689 | 0.050* |
| | (2.0652)** | | | | | (0.8218) | (3.7847)*** | | |
| 4 | 0.4354 | | 0.022** | | 4 | 0.2348 | 0.2965 | 0.690 | |
| | (2.0051)** | | | | | (0.8994) | (3.0057)*** | | |
| 20 | 0.0388 | | 0.060* | | 20 | 0.1845 | 0.0132 | 0.115 | |
| | (0.3116) | | | | | (0.6891) | (0.1776) | | |
| $k$-qtrs | *xb* | | $J$-test | $RZ$-test | | | | | |
| 1 | 0.3312 | | 0.289 | 0.531 | | | | | |
| | ( 2.6971)*** | | | | | | | | |
| 4 | 0.2885 | | 0.003*** | | | | | | |
| | ( 3.8751)*** | | | | | | | | |
| 20 | 0.0508 | | 0.748 | | | | | | |
| | ( 1.5045) | | | | | | | | |
| $k$-qtrs | *y* | *ldp* | $J$-test | $RZ$-test | | | | | |
| 1 | -0.7571 | 0.0190 | 0.000*** | 0.005*** | | | | | |
| | (-1.2396) | (1.5528) | | | | | | | |
| 4 | -0.4177 | 0.0216 | 0.164 | | | | | | |
| | (-0.6675) | (1.9139)* | | | | | | | |
| 20 | 0.5943 | 0.0114 | 0.157 | | | | | | |
| | ( 1.0160) | (1.0584) | | | | | | | |

I estimate regressions of the form $\tilde{r}^i_{t+k} = \alpha_i + \beta_i' z^i_t + u^i_{t+k}$ where $\tilde{r}^i_{t+k}$ is the cumulated and annualized $k$-quarter ahead excess return. Regressions for stocks are reported on the left panel, with instruments $z^i_t$ being log dividend yield, term spread, bond excess return, and a combination of short-term rate and log dividend yield. I report the results of bond excess return regressions on the right panel. I use term spreads or risk-free rates and term spreads as instruments. The coefficients are produced by constraining the predictive coefficients to be the same across countries. $T$-statistics in parentheses are computed using bootstrapped standard errors. The column labeled "$J$-test" reports $p$-values for a $\chi^2$ test of the overidentifying restrictions. The "$RZ$-test" column reports $p$-values for a Roy (1957) and Zellner (1962) test of slopes poolability using Bun (2004) bootstrap procedure described in the Appendix B.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

## Table 7: Out of Sample Prediction, Stocks

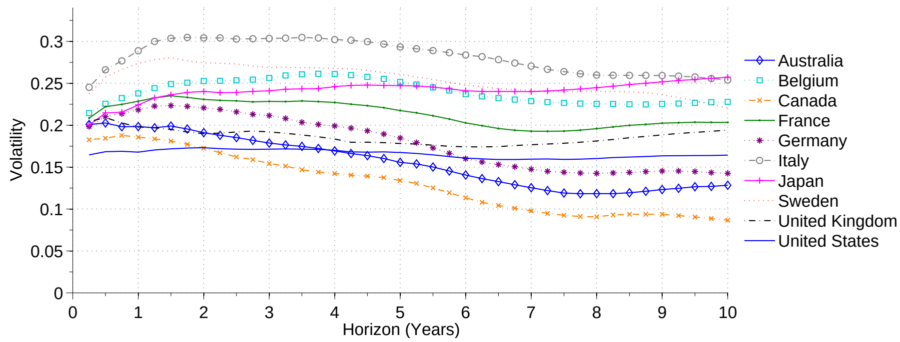
| | In-Sample | | Out-of-Sample | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Unconstrained | | | Constrained | | |
| **Panel A: log dividend price ratio** | | | | | | | | |
| | Time series | Pooled | Time series | Pooled | Δ | Time series | Pooled | Δ |
| Australia | 3.93% | 2.48% | −3.60% | 0.95% | 4.55% | 2.53% | 1.13% | −1.40% |
| Belgium | 0.87% | 0.51% | −1.64% | −1.42% | 0.22% | −1.24% | −1.21% | 0.03% |
| Canada | 3.19% | 2.31% | −6.09% | −2.13% | 3.96% | −0.85% | −0.32% | 0.53% |
| France | 0.83% | 0.66% | −0.99% | −2.09% | −1.10% | −0.38% | −0.88% | −0.49% |
| Germany | 0.90% | 0.76% | −1.34% | −1.88% | −0.54% | −0.16% | −0.60% | −0.44% |
| Italy | 1.77% | 0.24% | −2.19% | −1.52% | 0.67% | 0.50% | −0.67% | −1.18% |
| Japan | 4.19% | 3.95% | 1.38% | 1.99% | 0.61% | 1.33% | 1.95% | 0.62% |
| Sweden | 0.88% | 0.77% | −2.03% | −2.15% | −0.12% | −1.42% | −1.82% | −0.40% |
| United Kingdom | 6.29% | 1.88% | −2.83% | 0.90% | 3.73% | 1.23% | 0.90% | −0.33% |
| United States | 6.70% | 2.99% | −5.07% | −3.53% | 1.54% | −1.55% | −1.49% | 0.06% |
| Average | 2.96% | 1.66% | −2.44% | −1.09% | 1.35% | 0.00% | −0.30% | −0.30% |
| **Panel B: term spread** | | | | | | | | |
| | Time series | Pooled | Time series | Pooled | Δ | Time series | Pooled | Δ |
| Australia | 1.88% | 0.84% | −1.94% | 0.61% | 2.55% | −1.30% | 0.66% | 1.95% |
| Belgium | 6.58% | 3.95% | 3.77% | 1.62% | −2.14% | 3.80% | 1.62% | −2.18% |
| Canada | 1.94% | 1.65% | 0.59% | 0.62% | 0.03% | 0.94% | 0.71% | −0.23% |
| France | 4.01% | 3.36% | 2.38% | 1.31% | −1.07% | 3.00% | 1.34% | −1.66% |
| Germany | 6.10% | 3.98% | 3.46% | 1.40% | −2.06% | 2.62% | 1.41% | −1.21% |
| Italy | 1.00% | 0.29% | −0.65% | −0.07% | 0.59% | 0.71% | 0.86% | 0.15% |
| Japan | 3.84% | 3.75% | −5.38% | 0.74% | 6.12% | −5.36% | 0.74% | 6.10% |
| Sweden | 1.23% | 0.99% | −1.29% | 0.15% | 1.43% | −1.27% | 0.69% | 1.97% |
| United Kingdom | 1.94% | 1.81% | 0.65% | 1.04% | 0.39% | 0.99% | 1.04% | 0.05% |
| United States | 1.74% | 0.03% | −2.05% | 0.35% | 2.41% | −1.82% | 0.35% | 2.17% |
| Average | 3.03% | 2.07% | −0.05% | 0.78% | 0.83% | 0.23% | 0.94% | 0.71% |
| **Panel C: excess bond returns** | | | | | | | | |
| | Time series | Pooled | Time series | Pooled | Δ | Time series | Pooled | Δ |
| Australia | 2.07% | 1.79% | −17.17% | −9.72% | 7.45% | −11.78% | −6.39% | 5.39% |
| Belgium | 2.04% | 0.76% | 0.31% | 1.32% | 1.01% | 0.21% | 0.84% | 0.63% |
| Canada | 1.84% | 1.72% | −0.64% | −0.10% | 0.54% | 0.57% | 0.75% | 0.17% |
| France | 0.84% | 0.83% | −4.23% | −1.28% | 2.95% | −2.61% | −0.94% | 1.67% |
| Germany | 0.89% | 0.89% | −2.31% | −1.36% | 0.96% | −1.24% | −0.72% | 0.52% |
| Italy | 1.28% | 0.99% | −2.55% | −3.40% | −0.85% | −1.45% | −1.88% | −0.43% |
| Japan | 3.70% | 3.51% | −5.64% | −3.35% | 2.28% | −4.07% | −2.63% | 1.44% |
| Sweden | 1.03% | 0.89% | −1.49% | −1.47% | 0.02% | 0.21% | 0.64% | 0.43% |
| United Kingdom | 0.93% | 0.92% | −1.38% | −5.26% | −3.88% | −1.24% | −5.04% | −3.80% |
| United States | 1.96% | 1.71% | −3.02% | −5.94% | −2.92% | −1.94% | −3.24% | −1.29% |
| Average | 1.66% | 1.40% | −3.81% | −3.06% | 0.76% | −2.33% | −1.86% | 0.47% |
| **Panel D: log dividend price ratio and short term yield** | | | | | | | | |
| | Time series | Pooled | Time series | Pooled | Δ | Time series | Pooled | Δ |
| Australia | 5.19% | 2.90% | −2.81% | 2.33% | 5.14% | 1.72% | 2.33% | 0.61% |
| Belgium | 3.35% | 1.88% | −1.47% | −0.50% | 0.97% | −1.36% | −0.48% | 0.88% |
| Canada | 3.36% | 2.20% | −4.73% | −0.19% | 4.54% | −2.65% | −0.25% | 2.41% |
| France | 2.53% | 1.76% | −1.96% | −1.31% | 0.65% | −1.87% | −1.08% | 0.79% |
| Germany | 4.63% | 2.30% | −0.03% | −0.97% | −0.94% | 0.94% | −0.33% | −1.27% |
| Italy | 1.81% | −0.32% | −4.33% | −5.60% | −1.28% | 0.26% | −6.06% | −6.31% |
| Japan | 4.57% | 4.23% | −10.37% | 0.26% | 10.62% | −10.69% | 0.19% | 10.88% |
| Sweden | 0.90% | 0.51% | −4.61% | −3.35% | 1.25% | 0.00% | −3.00% | −3.00% |
| United Kingdom | 6.33% | 1.64% | −3.64% | −0.44% | 3.21% | −2.42% | −0.44% | 1.98% |
| United States | 6.98% | 1.97% | −11.99% | −2.07% | 9.93% | −1.51% | −0.91% | 0.60% |
| Average | 3.97% | 1.91% | −4.59% | −1.18% | 3.41% | −1.76% | −1.00% | 0.76% |

This table reports the Campbell and Thompson (2008) out-of-sample $R^2$ statistic, $R^2_{OS}$ (in percent), which measures the proportional reduction in mean-squared forecast error relative to the historical average benchmark. All forecasts are performed on the out-of-sample period 1990-2013. The first two columns present in-sample statistics for using times series country-specific or pooled coefficients. The "Unconstrained" columns use the benchmark predictive regression. "Constrained" impose a combination of sign restrictions on slope coefficients and forecasts. The prediction must be positive, otherwise I use zero as the forecast. When log dividend-price ratio (short-term yield) is used as predictor, the coefficient on the predictor must be of positive (negative) sign, otherwise the historical mean is used as a predictor instead. 'Δ' columns indicate the difference between times series and pooled regression so that a positive value indicates a better forecast when using pooled coefficient. "Average" is the average of the $R^2_{OS}$ statistics across the 10 countries.

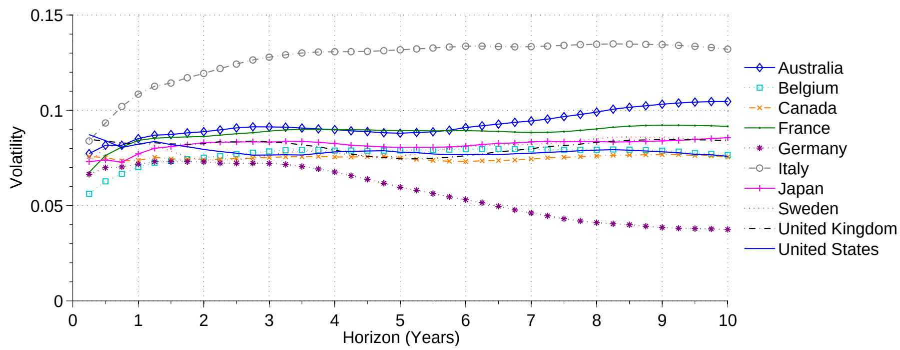
## Table 8: Out of Sample Prediction, Bonds

| | In-Sample | | Out-of-Sample | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Unconstrained | | | Constrained | | |
| **Panel A: term spread** | | | | | | | | |
| | Time series | Pooled | Time series | Pooled | Δ | Time series | Pooled | Δ |
| Australia | 5.36% | 5.35% | −0.02% | 0.94% | 0.95% | 1.08% | 2.44% | 1.36% |
| Belgium | 8.79% | 8.39% | 4.36% | 4.10% | −0.26% | 8.81% | 7.24% | −1.57% |
| Canada | 14.42% | 13.95% | 9.18% | 8.58% | −0.60% | 11.24% | 10.04% | −1.20% |
| France | 5.71% | 5.63% | −6.29% | −0.66% | 5.63% | 5.26% | 4.49% | −0.77% |
| Germany | 4.46% | 4.39% | −1.39% | −0.42% | 0.97% | 3.76% | 2.95% | −0.81% |
| Italy | 6.65% | 5.92% | 0.58% | 1.24% | 0.67% | 1.44% | 2.08% | 0.64% |
| Japan | 7.89% | 2.72% | −26.99% | −0.97% | 26.02% | −6.56% | −0.87% | 5.68% |
| Sweden | 4.98% | 3.99% | −4.14% | −5.12% | −0.98% | 0.82% | 2.70% | 1.88% |
| United Kingdom | 4.30% | 4.30% | −1.44% | −2.46% | −1.02% | 0.30% | 1.11% | 0.81% |
| United States | 6.62% | 5.59% | 4.41% | 4.31% | −0.10% | 5.81% | 5.11% | −0.70% |
| Average | 6.92% | 6.02% | −2.17% | 0.95% | 3.13% | 3.20% | 3.73% | 0.53% |
| **Panel B: term spread and short term yield** | | | | | | | | |
| | Time series | Pooled | Time series | Pooled | Δ | Time series | Pooled | Δ |
| Australia | 13.82% | 9.80% | −3.03% | −2.61% | 0.42% | 1.14% | 1.66% | 0.52% |
| Belgium | 11.92% | 11.11% | −2.27% | −1.78% | 0.50% | 6.44% | 3.76% | −2.68% |
| Canada | 16.40% | 15.98% | −11.50% | 1.88% | 13.38% | 5.10% | 5.91% | 0.81% |
| France | 8.22% | 8.13% | −13.19% | −12.92% | 0.27% | 3.27% | 0.93% | −2.34% |
| Germany | 5.72% | 5.26% | −17.26% | −3.15% | 14.11% | −1.15% | 1.25% | 2.41% |
| Italy | 9.27% | 8.82% | −2.34% | −4.09% | −1.75% | 2.09% | 3.28% | 1.18% |
| Japan | 16.12% | 11.44% | −25.86% | −17.64% | 8.22% | −5.75% | −8.45% | −2.69% |
| Sweden | 6.94% | 6.12% | −9.60% | −9.48% | 0.11% | −0.75% | 0.61% | 1.36% |
| United Kingdom | 6.52% | 6.41% | −18.27% | −15.98% | 2.29% | −2.41% | −1.29% | 1.12% |
| United States | 8.59% | 7.68% | −2.91% | 1.58% | 4.50% | 3.64% | 4.09% | 0.45% |
| Average | 10.35% | 9.08% | −10.62% | −6.42% | 4.21% | 1.16% | 1.18% | 0.01% |

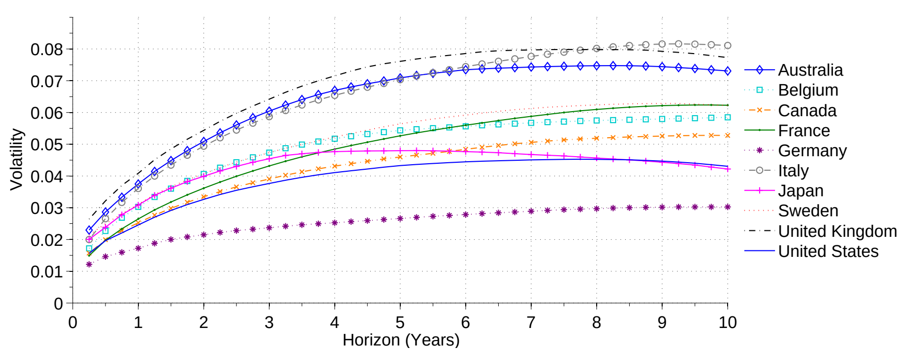
This table reports the Campbell and Thompson (2008) out-of-sample $R^2$ statistic, $R^2_{OS}$ (in percent), which measures the proportional reduction in mean-squared forecast error for the competing model relative to the historical average benchmark. All forecasts are performed on the out-of-sample period 1990-2013. The first two columns present in-sample statistics for using times series country-specific or pooled coefficients. The "Unconstrained" columns use the benchmark predictive regression. "Constrained" requires positive predictions, otherwise I use zero as the forecast. 'Δ' columns indicate the difference between times series and pooled regression so that a positive value indicates a better forecast when using pooled coefficient. "Average" is the average of the $R^2_{OS}$ statistics across the 10 countries.




**(a)** Stocks




**(b)** Bonds




**(c)** Cash

**Figure 1: Unconditional Volatilities of Real Returns**

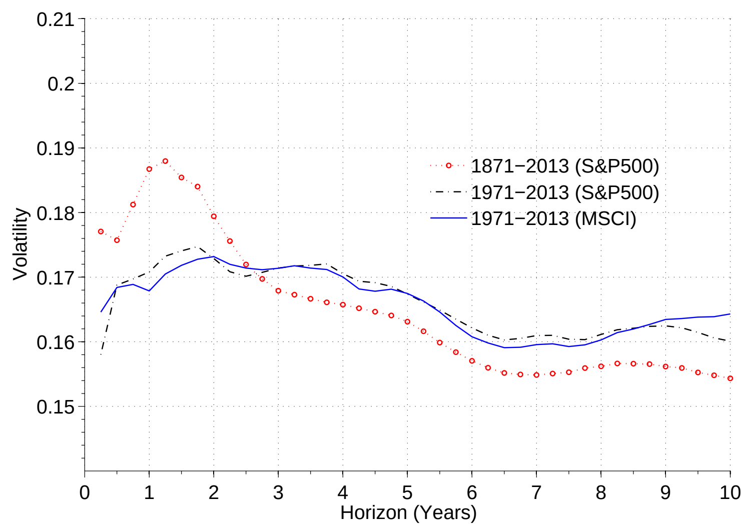
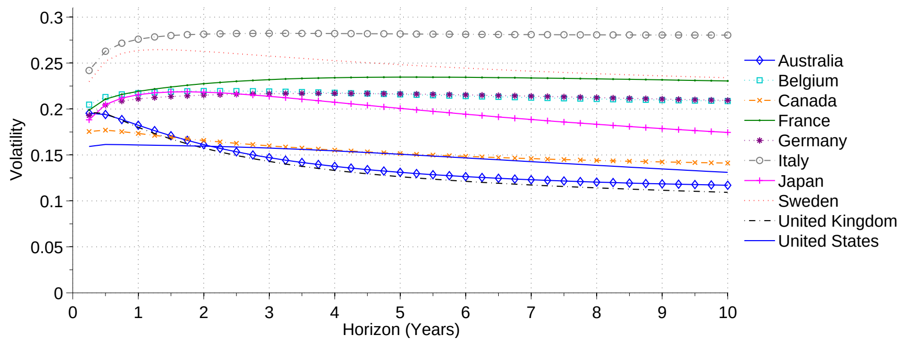
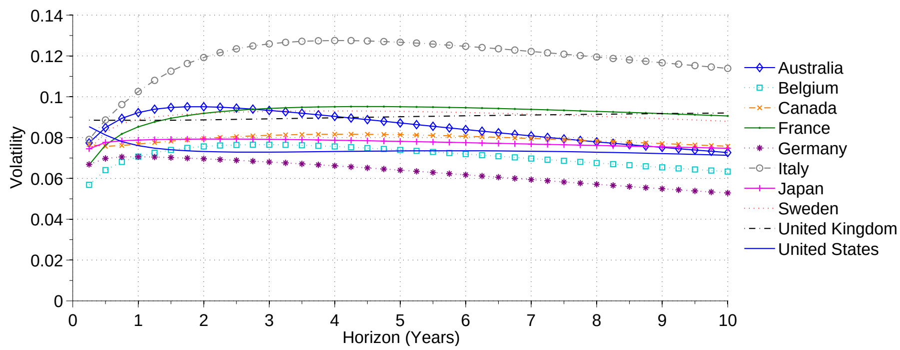


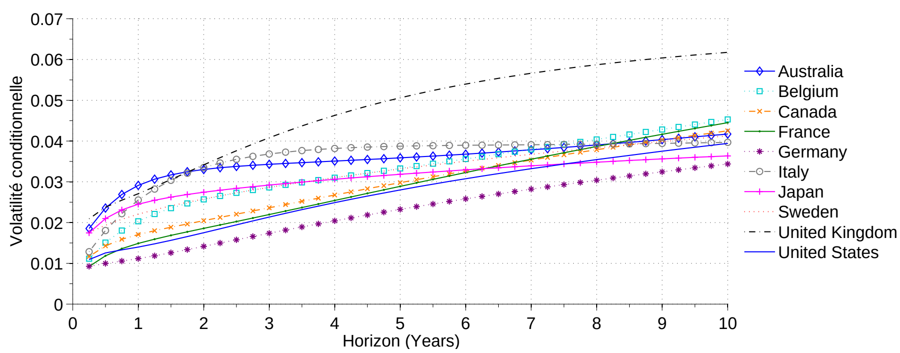

**Figure 2: Unconditional Volatilities of US Stock Returns: the Effect of Time**




**(a)** Stocks

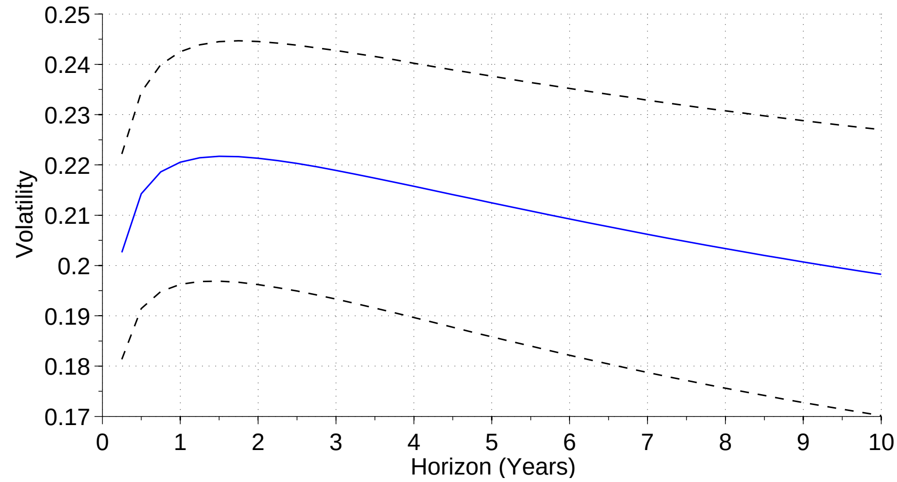



**(b)** Bonds

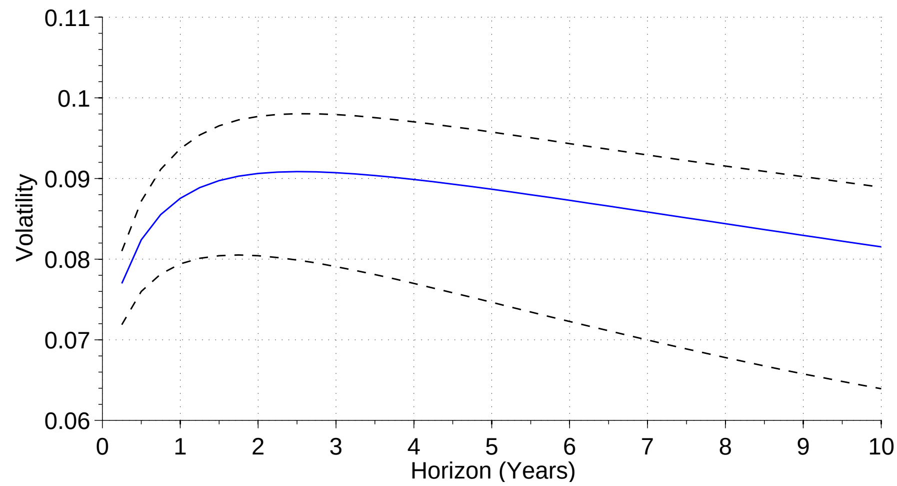



**(c)** Cash

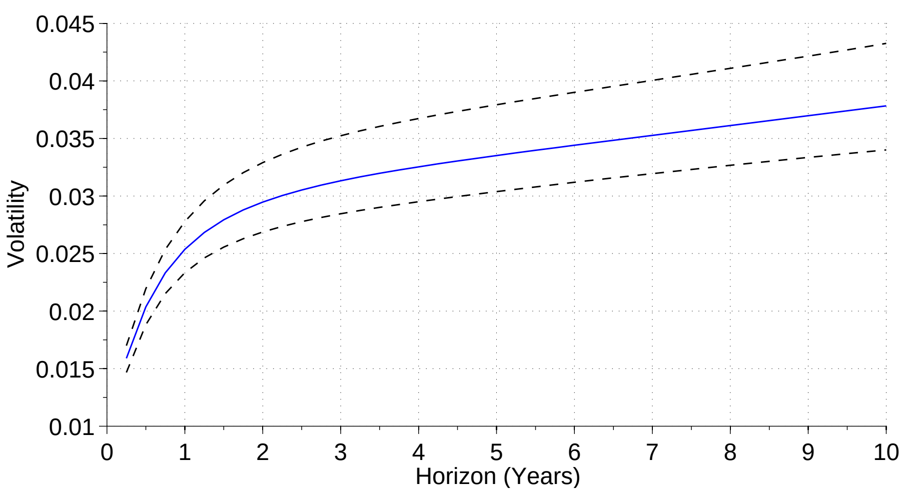
**Figure 3: Conditional Volatilities of Real Returns**




**(a)** Stocks

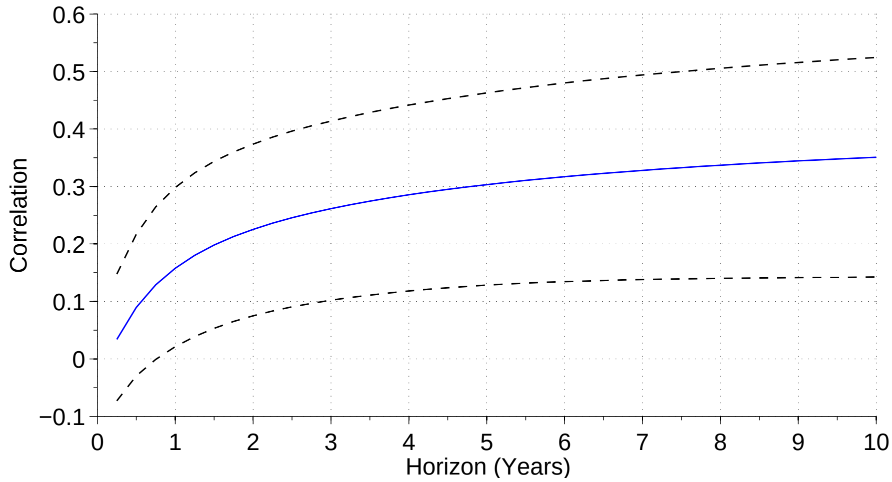



**(b)** Bonds

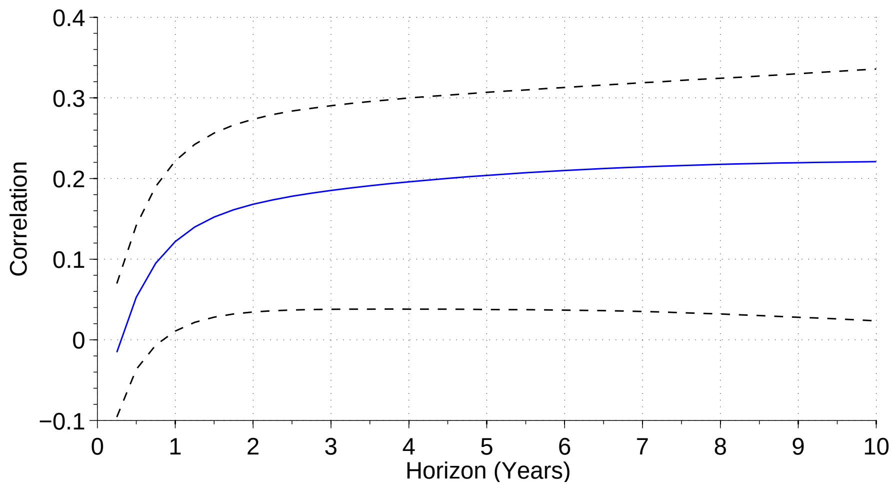



**(c)** Cash

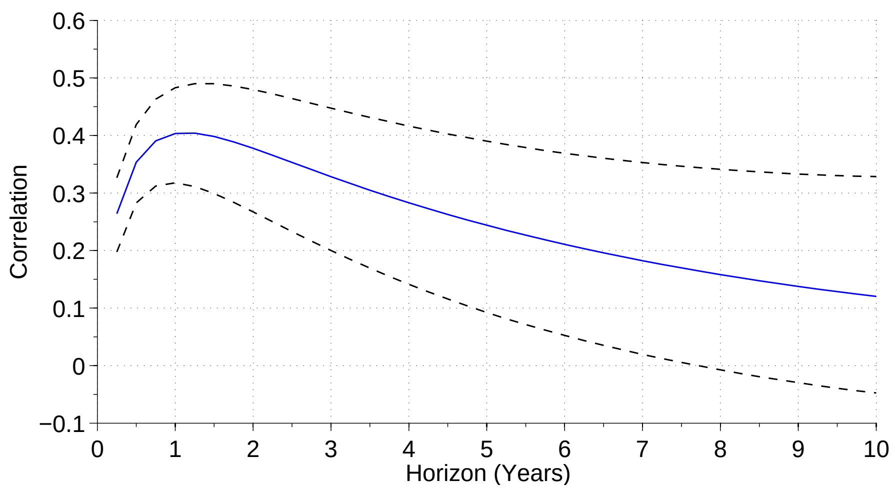
**Figure 4: Conditional Volatility of Real Returns and 95% Confidence Band (Panel)**




**(a)** (Stocks,Bonds)




**(b)** (Stocks,Cash)




**(c)** (Bonds,Cash)

**Figure 5: Correlations of Real Returns and 95% Confidence Band (Panel)**

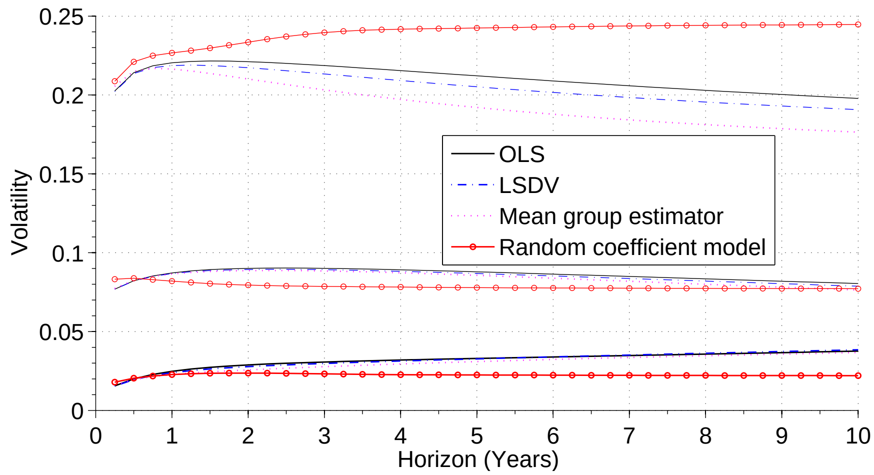



**Figure 6: Conditional Volatility of Real Returns : Robustness to Alternative Specifications**
This figures presents conditional moments under four alternative models. We use OLS to derive our main result and construct bootstrapped confidence intervals. LSDV allows for heterogeneous intercepts. The Mean group estimator and Random coefficient model are both consistent under slope heterogeneity.

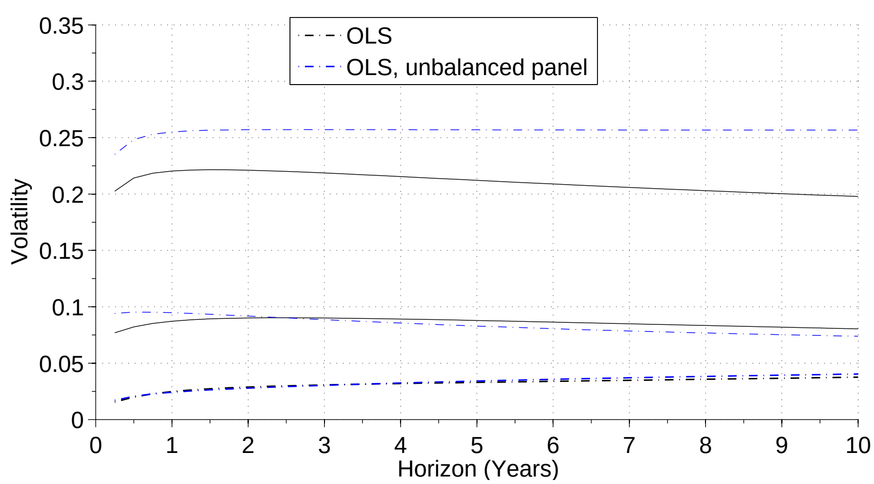



**Figure 7: Conditional Volatility of Real Returns : Robustness to Alternative Data**
This figures presents conditional moments under our base case and using a larger unbalanced panel of 32 countries.

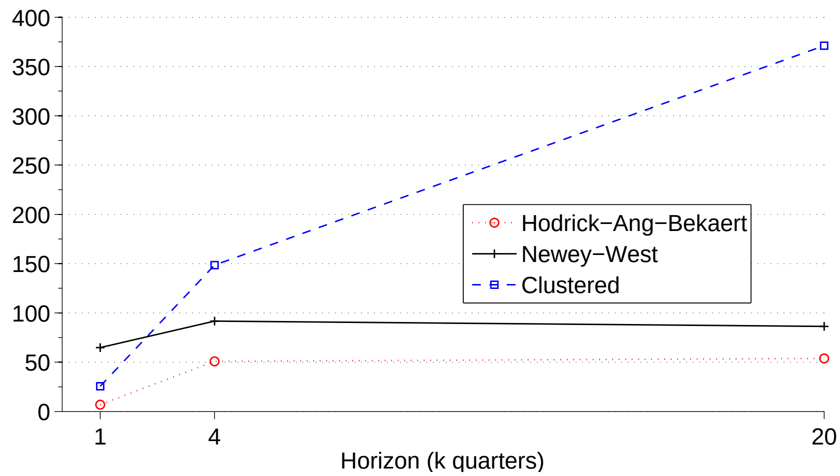
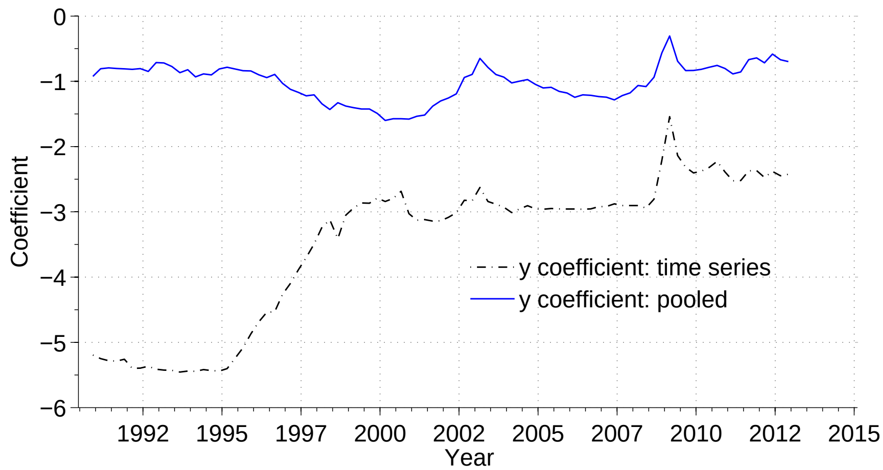
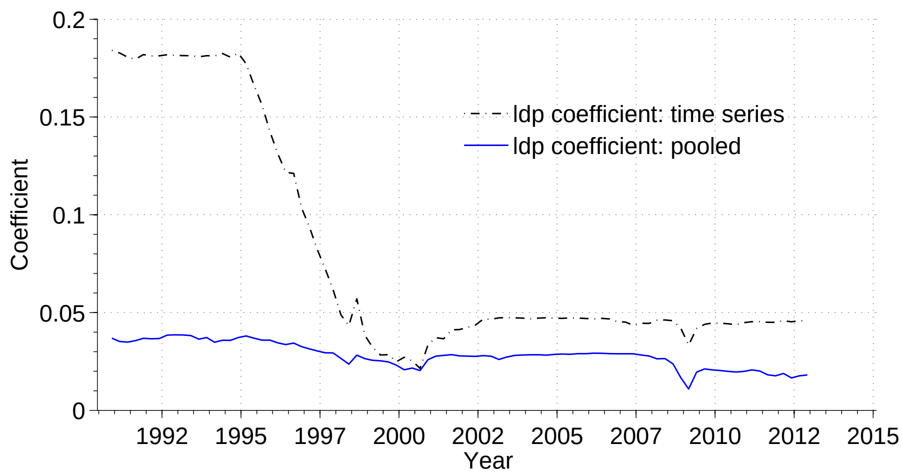



**Figure 8: Relative Performance of Standard Errors**
This figure graphs the bias as a function of regression horizon obtained from 1000 simulated data sets with $N = 10$ and cross sectional correlation $\tau = 0.5$. I compare Hodrick-Ang-Bekaert, clustered, and Newey-West standard errors. The bias is expressed in percentage term, i.e. $100 \times$ (true standard error/estimated standard error) $- 1$.




**(a)** Short term yield coefficient




**(b)** Log dividend-price ratio coefficient

**Figure 9: Coefficients Stability**
This figure shows the coefficients of a regression of 1-quarter excess stock returns on lagged short-term yield and log dividend yield. I use an expanding window after an in-sample period of 20 years. The coefficients are estimated from a pooled regression using the whole panel and from a time series regression using United States data.