Evaluation of Joint Density Forecasts of Stock and Bond Returns: Predictability and Parameter Uncertainty[1]

Francisco Peñaranda[2]

This version: June 2003

[1]I thank Manuel Arellano, Víctor López, Gabriel Pérez-Quirós, Luis Viceira, and specially Enrique Sentana, for very useful comments and suggestions. I also thank seminar participants at Alicante, Carlos III, CEMFI, ESEM (Venice 2002), FEDEA, LSE, and Pompeu Fabra. I am responsible for any remaining errors.

[2]FMG/LSE. E-mail: F.Penaranda@lse.ac.uk.

Abstract

One of the most important findings in empirical finance has been the fact that returns are not i.i.d. Predictability, or time variation in the conditional distribution of returns, is one of the basic ingredients of asset pricing and portfolio choice models nowadays. Under the current renewed interest in its implications for portfolio management, there is a growing literature on the computation of optimal portfolios for a given utility function and a given estimation of a particular model of returns. But there is an obvious problem with that approach. If the estimated model is not an accurate description of the distribution of returns the conclusions may be misleading. The approach in this paper is to study the properties of a particular model by means of the evaluation of its joint density forecasts of stock and bond returns. The focus is not on model testing, the procedure is based instead on out-of-sample checks of real-time density forecasting rules against realizations of returns. That is the relevant context for portfolio management. In addition, two important and controversial questions are addressed for each model: time-variation in risk premia and parameter uncertainty.

Keywords: Density forecasts, Forecast evaluation, Return predictability.

JEL: G11, C53.

1 Introduction

One of the most important findings in empirical finance has been the fact that returns are not independent neither identically distributed over time. This fact means that there is time variation in the conditional distribution of returns and it is called predictability. Its economic meaning is that the investor faces time-varying investment opportunities. This concept was introduced by Merton (1973), who developed the Intertemporal CAPM to describe the implications of predictability for portfolio choice and asset pricing. In terms of portfolio management, predictability implies additional terms in the optimal portfolio of an investor with respect to the usual ones of risk-return trade-off. These terms are called hedging demands as they are driven by the correlation between shocks to the returns and shocks to the state variables. This portfolio choice has implications in terms of asset pricing. The market portfolio is no longer the only priced risk factor and a multifactor model should be used for pricing.

Predictability has become one of the basic ingredients of asset pricing and portfolio choice models nowadays.[1] There is a renewed interest in its portfolio management implications. This is shown by a growing literature on the computation of optimal portfolios for a given utility function and the estimation of a particular model of returns. The latter is required to evaluate the expected utility of the agent for any portfolio choice. An investor needs density forecasts, not only first and second moments, unless some restrictions are imposed on her preferences or the distribution of returns. In addition, a joint density forecast of several assets is required to be useful for portfolio management.

Campbell et al. (2003) is an example of this literature using classic inference. On the other hand, Kandel and Stambaugh (1996), Stambaugh (1999) and Barberis (2000) study portfolio selection based on Bayesian inference. Those papers are based on a normal vector autoregression (VAR) of returns and predictors. Although new models are being introduced, for instance the Markov-switching VAR as in Ang and Bekaert (2001), and its portfolio implications studied.

But there is an obvious problem with that approach since conclusions may be misof joint density forecasts of stock and bond returns computed from different models. Multiple assets must be included in the analysis and therefore monthly U.S. excess returns of bonds and stocks are jointly studied in the post-war period.

The evaluation is going to be implemented in a relevant context for portfolio management. The focus is not on model testing, the procedure is based instead on out-of-sample checks of real-time density forecasting rules against realizations of returns.[2] In addition, my evaluation is not going to use comparisons of models in terms of their performance for a given loss function. It is based on the probability integral transform (PIT), which is defined as the cumulative distribution function from the density forecast evaluated at the final realization. The PIT should be uniform and independent over time if the forecasts are accurate. This methodology is advocated in Diebold, Gunther and Tay (1998).

The baseline model in this paper is a Bayesian VAR that represents the dynamic behaviour of returns and predictors, where predictors are specifically state variables that define the conditional mean of returns. In addition, non-linear dynamics models are also evaluated. It is implemented by means of Markov-switching models to separate the asymmetric behaviour of returns in different states as it is emphasized in conditional asset pricing models and macroeconomic forecasting nowadays. In finance, Pérez-Quirós and Timmermann (2000, 2001) and Chauvet and Potter (2001) are some examples of this type of models.

Some conclusions can be drawn in terms of the different models that are implemented. Compared to the normal model, the Markov-switching does not improve the evaluation results. It only improves the dynamics of the second moment of the PIT of bonds. Therefore, the standard normal VAR model does not seem to be a bad description of the data. This could explain why the portfolio choices in Brandt (1999), which do not rely in a statistical model of returns, are not far away from the choices previously computed for the normal VAR model in the literature.

In addition, two important and controversial questions are addressed for each model, as it is pointed out in the title of the paper.[3] On the one hand, time-variation in risk premia is studied. Monthly data are used in this paper and the focus of the literature that studies predictability in low-frequency data has been the conditional mean. There are two estimated versions of each model, with constant and time-varying conditional mean of returns. On the other hand, parameter uncertainty is also studied. There are two forecasting rules from each estimated model, the Bayesian predictive distribution versus a plug-in rule. The last one is easier to compute than the fully Bayesian approach but may suffer from lack of a proper accounting of the uncertainty.

In terms of time-variation in risk premia, it is clear that models with or without time variation in the conditional mean of returns imply very different density forecasts. They vary much more over time in a model with a predictable mean, and this translates into significant effects in portfolio choice as it is well known in the literature on predictability.

[2] Evaluation of density forecasts is a very active area nowadays. The evaluation of density forecasts is not only useful for portfolio choice, but also in another contexts. This is clear in modern models of financial risk such as the Value-at-Risk, where quantiles are computed to quantify the risk of a portfolio.

[3] Both questions are also studied in Barberis (2000), but in terms of their portfolio choice implications.

The results in terms of the PIT are not so clear. With respect to a constant conditional mean, a time-varying one worsens the forecast evaluation for stocks but improves for bonds. Finally, about parameter uncertainty, the fully Bayesian approach and the plug-in forecasting rule give similar density forecasts and similar properties of the PIT.

The rest of the paper is organised as follows. In section 2 there is a brief summary of the literature on predictability and a description of the data. Section 3 explains the computation of density forecasts in a Bayesian framework and their evaluation. Section 4 defines and applies the standard normal model, while section 5 is devoted to nonlinear dynamics by means of Markov-switching. Finally, section 7 concludes. The different Gibbs samplers used in the estimations are gathered in the appendix, jointly with tables and figures.

2 Data: Excess Returns and Predictors

This section introduces the data that will be used in this paper. But it will start with a brief introduction to the literature on predictability. The usual way to characterize the opportunity set of the investor is to model the conditional mean and variance of excess returns. Studies using high-frequency data are mainly focused on second moments since variables such as the consumption wealth ratio in Lettau and Ludvigson (2001) have been introduced in the literature.

There are mainly two approaches to study predictability in the conditional mean of returns. In the classical approach, predictive regressions or VARs are studied in terms of significance tests. There are many statistical problems in these classic tests of predictability and therefore it is difficult to interpret their results. The predictors are clearly not exogenous and usually very persistent. At the same time, long-horizon returns are often used.[7]

Lack of exogeneity implies a nonstandard finite sample distribution of the parameters under normality of returns. It has been derived by Stambaugh (1999).[8] On the other hand, persistence and long-horizon imply a nonstandard asymptotic distribution. About persistence of predictors, Ferson et al. (1999) point out that since predictors are very persistent, there is a possibility of spurious regression although returns themselves are stationary. Another problem is the use of long-horizon returns since there are few nonoverlapping observations. Hodrick (1992) advocates the use of the VAR approach. VAR methods can be used to avoid long run returns, and then long run predictions can be estimated by means of short run returns.

The need of nonstandard distributions, which also depend on unknown parameters, translates into high uncertainty about the interpretation of results from classic inference. Kandel and Stambaugh (1996) propose the portfolio choice based on Bayesian inference as a more interesting metric for predictability than the usual statistical ones of significance. They show that predictability is important in asset allocation even if the investor's prior is against predictability. This second approach will be commented in Section 4.

After this extremely brief summary of the literature on predictability, the description of the data starts. The data are similar to the ones used in Campbell et al. (2003) and are taken from CRSP. This paper uses US monthly nominal data from January 1954 to December 2001, which gives 576 observations. The post-war sample period is used to conform with the period after the Fed-Treasury accord and the presidential election in 1952 after which the Fed stopped pegging interest rates. This changed fundamentally the process of the nominal interest rates. The chosen period for the out-of-sample check of the density forecasts starts with the forecast of January 1975, which gives 324 observations for that check. They represent enough observations, 252, without letting too few ones for the first estimation. During this period, there was a sharp change in the monetary policy in the U.S. Specifically, Volcker arrived at the Fed in October 1979 and followed a tighter monetary policy by means of monetary aggregate controls which increased the volatility of interest rates. Therefore, it is an interesting period for forecast evaluation.

Excess returns are computed for stocks and bonds. In the computation of excess stock returns (ESR), returns are based on the value weighted stock index of the NYSE, NASDAQ and AMEX markets including dividends. They are converted into continu-

[7] In addition, there seem to be problems of parameter instability. This is the interpretation that Bossaerts and Hillion (1999) give to their results in terms of out-of-sample success of different models.

[8] Goetzmann and Jorion (1993) also study the issue of bias in the predictive regressions.

ously compounded rates in monthly percentage. The other ingredient is the 1-month T-bill return from the US Treasury and Inflation Series (CTI). Again, the series is converted into continuously compounded rates in monthly percentage. Finally, the excess stock return e_t^S is given by the difference between both returns.

A huge fall can be observed in October 1987 in Panel A of Figure 1 and also a less important fall in August 1998 due to the Russian financial crisis. Apart form that, there are not sharp changes as in the series of bonds. In Table 1, Panel A, there are some descriptive statistics of the stock and bond excess returns for the whole data period and for the two periods given by the out-of-sample check. The period from January 1954 to December 1974 corresponds to the first estimation period while the period form January 1975 to December 2001 corresponds to the out-of-sample check. First, the statistics for the whole period are explained. The mean is 0.49%, while the standard deviation is 4.35%, which gives an annualized Sharpe ratio of about 0.4. The coefficient of skewness is -0.77 and the one of kurtosis 2.93, that is, there is asymmetry to the left and fat tails. The Jarque-Bera test clearly rejects the hypothesis of normality with a p-value of 0.00. To approximate the temporal dependence in the series, the autocorrelograms of powers of the demeaned series are shown in Panels B and C of Figure 1. The level does not show autocorrelation, neither the third and forth powers, while it can be found only marginally in the second power for the first lag (a p-value of 0.04) even in terms of the Ljung-Box test, see Panel B in Table 1. Anyway, the correlation in the square will be much clearer for bonds.

The period that is left for the out-of-sample check, January 1975 to December 2001, shows a much higher mean and kurtosis than the previous period, while the standard deviation is only slightly higher. The negative skewness also increases. It is also noteworthy that there is not autocorrelation in the first and second moments during the last period, while the correlation of the second moment is clear in the first period.

Excess bond returns (EBR) are computed from returns of the 5 year bond from the US Treasury and Inflation Series (CTI), which are converted into continuously compounded rates in monthly percentage. Finally, the excess bond return e_t^B is given by the difference between that return and the continuously compounded 1-month T-bill return.

The time series, Panel A in Figure 2, shows a huge increase in the volatility at the beginning of the 80's due to the commented Fed experiment. There is also an increase during the first half of the 70's due to the OPEC oil crises, but is much less sharp. Anyway, the range of the series is narrower than for stocks. The mean is 0.10%, while the standard deviation is 1.49%, see Panel A in Table 1, which gives an annualized Sharpe ratio of about 0.2, half the ratio of stocks. The skewness is 0.07 and the kurtosis 4.12, there is asymmetry of the opposite sign to stocks and even fatter tails. The Jarque-Bera test clearly rejects the hypothesis of normality with a p-value of 0.000. As it can be seen in Figure 2, Panels B and C, there is autocorrelation in the level of the series at the first lag, the p-value is 0.00. There is a clear dependence in the second power with zero p-values at any lag in terms of the Ljung-Box test, see Panel B in Table 1. The third and forth powers also show correlation, although not at the first lag.

In the out-of-sample period, the mean is much higher while the skewness is no longer positive. The standard deviation and the kurtosis also increase, specially the former.

As opposed to the case of stocks, bonds show a higher autocorrelation in the first and second moments during the second period, specially in the first moment.

Two predictors are used, the dividend yield and the term premium. The dividend yield is derived from the previous value weighted index of stocks without dividends and the corresponding return with dividends, summing up previous year dividends. The dividend yield is often computed using the sum of last year dividends to avoid seasonality. Finally, the logarithm is taken. It can be seen in Panel A of Figure 3 that it is a highly persistent variable. The term premium is defined as the difference between the continuously compounded monthly yield of a 5 year zero coupon bond, taken form the Fama-Bliss discount bond yields, and the yield of the 1-month risk free asset, obtained from the risk-free rates file. This series, see Panel B in Figure 3, is also persistent but not as much as the divided yield.

These four series, two excess returns and two predictors, will be stacked together in a vector,

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{e}_t \\ \mathbf{z}_t \end{pmatrix}, \quad t = 1, ..., T,$$

where $\mathbf{e}_t$ and $\mathbf{z}_t$ are 2×1 vectors containing the excess returns and the predictors. This vector will be the basis of a VAR. This is similar to the VAR in Campbell et al. (2003). They use two additional variables. As a return, they add the real short term rate, and as a predictor, they add the nominal short term rate. Other papers that use a VAR in this context are Kandel and Stambaugh (1996), Hodrick (1992) and Barberis (2000).

3 Computation and Evaluation of Joint Density Forecasts

This section introduces the basic tools that will be used in this paper, the computation and evaluation of density forecasts of returns.[9] In this paper, density forecasts of future returns given present and past returns and predictors are the object of interest. This conditional joint distribution of stocks and bonds at a given point in time will be denoted as $p\left(\mathbf{e}_{T+1} \mid \mathbf{y}^T\right)$, where $\mathbf{y}^T = \{\mathbf{y}_t\}_{t=1}^T$ represents the data up to T. The notation $p(\cdot)$ will represent density or mass depending on the associated random variable. In addition, $P(\cdot)$ will be the corresponding cumulative distribution function (CDF) The notation $E_{p(\cdot)}[\cdot]$ means the expectation with respect to $p(\cdot)$.

3.1 Computation of Joint Density Forecasts

The main goal is the computation of a density forecast $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^T\right)$ at a given point in time T, where the first two entries of $\mathbf{y}_{T+1}$ are the object of interest, from a statistical model that defines $p\left(\mathbf{y}_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)$, where $\boldsymbol{\theta}$ is the vector of parameters and s_{T+1} is a latent variable, and the dynamics of that unobservable variable. In this paper, two ways of linking both objects are going to be implemented. Both forecasting rules start from a Bayesian estimation of the model, but only one is fully Bayesian when computing

[9] Tay and Wallis (2000) is a survey about computation and evaluation of density forecasts, but in a classic context. It does not say much about the Bayesian approach that will be used in this paper.

the density forecast. We want to forecast future observables given only present and past observables, not given unknown parameters, and this question is naturally addressed in a Bayesian setting with the concept of predictive distributions.

First of all, the Bayesian estimation of a model is going to be explained. Many papers use a Bayesian approach to compute portfolio choice, for instance Kandel and Stambaugh (1996), Stambaugh (1999) and Barberis (2000). A Bayesian model is constructed with a prior and a likelihood, which define a posterior of the parameters as

$$p\left(\boldsymbol{\theta} \mid \mathbf{y}^T\right) \propto p\left(\boldsymbol{\theta}\right) p\left(\mathbf{y}^T \mid \boldsymbol{\theta}\right).$$

The estimation of all the models is going to be conditional on a initial observation $\mathbf{y}_0$ which is omitted from the expressions for simplicity. The support of the posterior will be denoted by $\boldsymbol{\Theta}$. Proper priors are going to be used in this paper, but those priors will not be too informative, since they are necessary for regime-switching models. In addition, they will be related to the corresponding maximum likelihood estimator (MLE). The priors of some parameters will be centred at the MLE. This type of priors is known as sample priors and has been used for instance by Kandel and Stambaugh (1996) and Avramov (2002).

I rely on simulation-based methods to get a simulated sample $\{\boldsymbol{\theta}_i\}$ that approximates the posterior distribution. The basis of the estimation of the different models will be the Gibbs sampler jointly with data augmentation. The Gibbs sampler is very popular draws. The convergence of all the chains in this paper was checked and did not show problems. Specifically, in the case of the normal VAR, 101000 iterations were run in the computation of the posterior distribution and the first 1000 were burnt for convergence reasons. These numbers also correspond to the estimation of other two regime-switching models commented in the conclusions of the paper. Given the cost of its computations, 51000 iterations were run and the first 1000 were burnt for the Markov-switching model.

Since computing the density forecast for a grid of points would be very expensive in terms of time, I will focus on some descriptive statistics of the density to get an idea of its shape. Specifically, I will compute the mean, standard deviation, and the coefficients of skewness and kurtosis. Therefore, I have to compute the first four central moments, which can be expressed in terms of the noncentral ones. If the particular random variable is normally distributed then the noncentral moments have well known formulas. The density forecasts are not normal, but we can still use those formulas due to the data augmentation approach in the Bayesian estimation. The noncentral moments of the density forecast can be easily expressed as

$$E_{p^{\mathrm{I}}\left(\mathbf{y}_{T+1}|\mathbf{y}^{T}\right)}\left[\left(e_{T+1}^{S}\right)^{k}\right] = \tag{1}$$
$$E_{p(\boldsymbol{\theta},s_T|\mathbf{y}^T)}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^{T}, \boldsymbol{\theta}, s_{T}\right) E_{p\left(\mathbf{y}_{T+1}|\mathbf{y}^{T},\boldsymbol{\theta},s_{T+1}\right)}\left[\left(e_{T+1}^{S}\right)^{k}\right]\right],$$

for the case of stocks and for a particular power k, which can be approximated with the MCMC output. The Gibbs output gives draws from $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right)$ and we only have to average the corresponding computations. In every model $p\left(\mathbf{y}_{T+1}| \mathbf{y}^{T}, \boldsymbol{\theta}, s_{T+1}\right)$ is normal and therefore the corresponding moment is easy to compute, while $p\left(s_{T+1} \mid \mathbf{y}^{T}, \boldsymbol{\theta}, s_{T}\right)$ is discrete and therefore easy to integrate. The counterparts for the plug-in rule do not integrate with respect to $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right)$ and only evaluate at the corresponding posterior means

$$E_{p^{\mathrm{P}}\left(\mathbf{y}_{T+1}|\mathbf{y}^{T}\right)}\left[\left(e_{T+1}^{S}\right)^{k}\right] = \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^{T}, \bar{\boldsymbol{\theta}}, \bar{s}_{T}\right) E_{p\left(\mathbf{y}_{T+1}|\mathbf{y}^{T},\bar{\boldsymbol{\theta}},s_{T+1}\right)}\left[\left(e_{T+1}^{S}\right)^{k}\right]. \tag{2}$$

3.2 Evaluation of Joint Density Forecasts

Discrimination among different models can be done in two dimensions. On the one hand, each model can be applied to the data and its "coherence" can be analysed by means of out-of-sample checks. In terms of density forecasts, procedures for that analysis can be found in Diebold, Gunther and Tay (1998), Berkowitz (2001) and Wallis (2001). On the other hand, we can make a relative comparison between pairs of models given a loss function.[11] In this paper, the former approach is taken. The lack of a loss function implies that the conclusions of the evaluation are not conditional in a particular function. The cost is that we cannot get the same type of clear-cut conclusions. However, if our density forecasts are equal to the true conditional distribution of excess returns then the corresponding portfolios will be optimal for any loss function. Therefore, it is interesting to study if our density forecast is close to the true distribution. In addition, the quoted literature on Bayesian portfolio choice has relied on a loss function and this paper contributes in a different direction.

Diebold, Gunther and Tay (1998) advocate the use of graphical methods to study model failures and reveal useful information for the development of new models. The

[11] Diebold and Mariano (1995) show asymptotic and finite sample tests for that comparison.

main interest is not a final decision about rejection, but the reason of rejection. This is the approach that is taken in this paper, jointly with formal tests such as the ones in Wallis (2001). The evaluation is based on the probability integral transform (PIT), which is defined as the CDF from the density forecast evaluated at the final realization. Let us think for a moment in a particular scalar random variable x with conditional density $p\left(x_{t+1} \mid x^t\right)$, then the PIT is

$$d_{t+1} = \int_{-\infty}^{x_{t+1}} p\left(u \mid x^t\right) du = P\left(x_{t+1} \mid x^t\right),$$

and this transformation is finally uniform and independent over time

$$\{d_{t+1}\} \overset{iid}{\sim} U\left(0,1\right),$$

assuming a nonsingular Jacobian and continuous partial derivatives for this transformation. This an old result by Rosenblatt (1952) and it has a clear application in checking the accuracy of a particular density forecast as Diebold, Gunther and Tay (1998) show.[12] If our density forecast is an accurate description of the true conditional density then its corresponding PIT should be uniformly distributed and independent over time. That is, we should not see a tendency to have more or less likely realizations of returns, and we should not see any pattern along time of this likeliness. Both things look very intuitive. The PIT is interesting by itself, since it shows if the final realization was more or less likely from the point of view of our forecast. We can summarize the performance of the density forecast in only one number. But here the evaluation of interest is a joint distribution and there is not a similar result for the PIT of the joint distribution. Hothne

(r)TJ4i(a)1169(t)947(e)-312.3(t29.4(n)-314.5(m4-021(u)4.5l)76.8(t29.4i)76.8v)6111((r)TJ4i(a)1169(t)947(e)-312.3dea
t ##Þ t71.8(ui04.8(r)-0.1nr)-369
.1523.2(.)13016,Te# be

have 324 observations of the previous objects while 252 data observations are left for the first estimation in the recursive procedure. The estimation used in each density forecast is not frozen at the first period estimation and a rolling approach is used instead. The posterior of the parameters is only updated each three months instead of each month since there should be not much difference in the corresponding output while the saving in computing time is high.

The next stage is to compare both series to evaluate the forecasts. In this paper, there are two excess returns to predict and therefore four PIT's to compute. Recall $\mathbf{y}_{T+1} = \left(\mathbf{e}'_{T+1}, \mathbf{z}'_{T+1}\right)'$ and $\mathbf{e}_{T+1} = \left(e^S_{T+1}, e^B_{T+1}\right)'$ so that we can compute

$$\begin{aligned} d^{f,S}_{T+1} &= P^f\left(e^S_{T+1} \mid \mathbf{y}^T\right), \quad f = I, P, \\ d^{f,B|S}_{T+1} &= P^f\left(e^B_{T+1} \mid \mathbf{y}^T, e^S_{T+1}\right), \end{aligned}$$

and similarly $d^{f,B}_{T+1}$ and $d^{f,S|B}_{T+1}$.

Two ways of computing density forecasts are faced in this paper. The fully Bayesian approach was called the integrated rule and its corresponding PIT's can be computed as

$$d^{I,S}_{T+1} = E_{p(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T)}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_T\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)\right], \tag{3}$$

$$d^{I,B|S}_{T+1} = E_{p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T, e^S_{T+1}\right)}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_T\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}, s_{T+1}\right)\right],$$

where $P\left(e^S_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}, s_{T+1}\right)$ and $P\left(e^S_{T+1} \mid \mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}, s_{T+1}\right)$ are normal CDF's and therefore easy to compute. The previous expectations are approximated by the corresponding sample mean from the Gibbs output, for instance

$$d^{I,S}_{T+1} \simeq \frac{1}{I}\sum_{i=1}^{I}\sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}_i, s_{T,i}\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, \boldsymbol{\theta}_i, s_{T+1}\right).$$

The conditional PIT is computed under the approximation $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T, e^S_{T+1}\right) \simeq p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right)$ to use directly the previous Gibbs output.

On the other hand, the PIT's of the plug-in rule are easily computed as

$$d^{P,S}_{T+1} = \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, \bar{s}_T\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, s_{T+1}\right), \tag{4}$$

$$d^{P,B|S}_{T+1} = \sum_{s_{T+1}} p\left(s_{T+1} \mid \mathbf{y}^T, \bar{\boldsymbol{\theta}}, \bar{s}_T\right) P\left(e^S_{T+1} \mid \mathbf{y}^T, e^S_{T+1}, \bar{\boldsymbol{\theta}}, s_{T+1}\right).$$

The PIT's properties imply that if the density forecasts are accurate then there are 6 series that should be independent and identically uniform for each forecasting rule,

$$\begin{gathered} \left\{d^{f,S}_{t+1}\right\}, \left\{d^{f,B}_{t+1}\right\}, \left\{d^{f,S|B}_{t+1}\right\}, \left\{d^{f,B|S}_{t+1}\right\}, \\ \left\{d^{f,S}_{t+1}, d^{f,B|S}_{t+1}\right\}, \left\{d^{f,B}_{t+1}, d^{f,S|B}_{t+1}\right\} \overset{iid}{\sim} U(0,1). \end{gathered}$$

Therefore, if the evaluation is based on the PIT then we have to compute only 4 numbers, the 4 PIT's, for each density forecast. The ideal object to study would be probability regions for joint events concerning stocks and bonds. This is theoretically possible since I can compute the density in each point, but that would be very expensive to compute. In the results of this paper, it turns out that the performance of $d_{t+1}^{f,S|B}$, $d_{t+1}^{f,B|S}$ and $\left(d_{t+1}^{f,B}, d_{t+1}^{f,S|B}\right)$ is similar to $d_{t+1}^{f,S}$, $d_{t+1}^{f,B}$ and $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$, respectively. Therefore, only tables and figures concerning $d_{t+1}^{f,S}$, $d_{t+1}^{f,B}$, and $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$ are shown.

There are two features to test for each series, uniformity and independence. I will use graphs and tests for the different density forecasts. I will show the CDF and correlograms of the PIT's. The uniform distribution of the PIT can be checked by means of a histogram. However, histograms can be misleading as they depend on the number of bins that are used. Because of that, I am going to show CDF's. They will be shown jointly with the 5% critical value of the Kolmogorov-Smirnov test. But that may not be very useful for evaluation of density forecasts. For instance, this test has low power against a bad description of the tails of the distribution since it focuses mainly on the median. Because of that, other tests such as Kupier and Anderson-Darling are computed, which have more power against bad performance in the tails. They will be called empirical distribution function (EDF) tests in the rest of the paper. The critical values of the three tests are taken from Stephens (1974). However, all of them have the independence assumption as a maintained hypothesis. So we should look at them after studying the independence of the series. The independence of the PIT can be checked with correlograms of different moments, for instance $\left(d_{t+1}^{f,S} - \overline{d}^{f,S}\right)$ and $\left(d_{t+1}^{f,S} - \overline{d}^{f,S}\right)^2$, where $\overline{d}^{f,S}$ is the corresponding time series mean, jointly with the corresponding Barlett confidence intervals. The Ljung-Box statistic was also computed to have a joint tests of the different lags.

Finally, I compute the tests proposed by Wallis (2001).[14] His tests are designed in principle for the density forecasts themselves not for the PIT, but that would require the storage of a huge number of computations and therefore I apply them to the PIT. We have to choose a particular number of bins to divide the support of the PIT in the implementation of the tests. In this paper, results will be shown for 3 and 4 bins. He develops three likelihood ratio tests:[15] unconditional coverage, independence and conditional coverage. The null of the first test is i.i.d. uniformity against i.i.d. sampling from any other distribution, while in the second test we face any i.i.d. distribution against a first order Markov chain with unknown distribution. The final test collapses the previous ones in a joint test and its null hypothesis is i.i.d. uniformity against a first order Markov chain with unknown distribution.

Care must be taken in going from conclusions about the marginal and dynamic behaviour of the PIT to conclusions about the marginal and dynamic behaviour of

[14] Another approach is proposed in Berkowitz (2001). That approach was also used during the development of this paper, but it gives very low p-values in general for all the models, which does not help much to draw conclusions. This happens also in Clements and Smith (2000) and they interpret it as higher sensitivity to outliers.

[15] In the same spirit as the interval forecast tests in Christoffersen (1998).

the corresponding return since they are not linked in general. Anyway, failures in the uniformity and independence of the PIT may suggest directions of model improvement that can be checked once implemented.

4 Normal VAR

This section will focus on the standard Bayesian setting, which is generally based on a VAR with a diffuse prior and a Normal likelihood. Usually, the priors are diffuse and impose stationarity in the VAR. However, there are also informative priors. For instance against predictability as in Kandel and Stambaugh (1996). The VAR is a simple way of introducing time-varying risk premia and because of that it is very used in the literature on predictability. Homoskedasticity is also often assumed and therefore the model only shows time-varying investment opportunities in terms of conditional means. This paper studies if this particular model is a good description of stock and bond returns.

A recent example of this model is Campbell et al. (2003), although their application is portfolio choice not forecast evaluation, and the estimation is based on classic inference. They show the effects of adding predictors, taken as state variables that define the conditional distribution of returns, in the portfolio choice between stocks and bonds. They assume that the investor knows the parameters of the model and point out that their values have important effects on the portfolio choice. Some papers have studied asset allocation in a Bayesian setting to take into account parameter uncertainty. Given the weak statistical results in terms of predictive regressions, Kandel and Stambaugh (1996) propose portfolio selection based on Bayesian inference as a more interesting metric for predictability than the usual statistical ones of significance. They show that predictability is important in asset allocation even if the investor's prior is against predictability. In a similar spirit, Stambaugh (1999) and Barberis (2000) show that it is important to take into account jointly predictability and estimation risk when computing an optimal asset allocation.

4.1 Model

Recall that $\mathbf{y}_t$ is a 4×1 vector composed by the two excess returns and the two predictors. The likelihood of this model is given by a Normal VAR,

$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}\mathbf{y}_{+}^{\mathrm{y}}$$

to be used in this paper, two versions of this model will be estimated and evaluated to study the predictability in the first moment. One with a time-varying conditional mean and another with a constant one. In homoskedastic models like this one, the second option implies unpredictable returns. The case of unpredictable mean is defined as a constant conditional mean of returns, which imposes the following restrictions on the general model

$$\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{ze} & \mathbf{B}_{zz} \end{pmatrix}, \tag{7}$$

where $\mathbf{B}$ has been partitioned in terms of $\mathbf{e}_t$ and $\mathbf{z}_t$.

A compact notation for the VAR intercepts and slopes will be used. Specifically,

$$\mathbf{\Pi} = \begin{pmatrix} \mathbf{a}' \\ \mathbf{B}' \end{pmatrix}, \quad \boldsymbol{\pi} = vec\left(\mathbf{\Pi}\right),$$

and the vector $\boldsymbol{\pi}$ is 20×1 for the unrestricted version of the model and 12×1 for the restricted version. The parameters to estimate in this model are

$$\boldsymbol{\theta} = \{\boldsymbol{\pi}, \mathbf{\Sigma}\},$$

while the likelihood is

$$p\left(\mathbf{y}^T \middle| \boldsymbol{\theta}\right) = \prod_{t=1}^{T} p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}\right),$$

where each $p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}\right)$ is normal. Recall that we are coditioning on an initial $\mathbf{y}_0$. It is not written in all the expressions for simplicity.

To close the model, the priors are

$$\boldsymbol{\pi} \sim N\left(\underline{\boldsymbol{\pi}}, \underline{\mathbf{C}}\right),$$
$$\mathbf{\Sigma} \sim IW\left(\underline{\nu}, \underline{\mathbf{S}}\right),$$

where $IW(\cdot)$ denotes the inverse Wishart distribution.[16] I do not restrict the eigenvalues of $\mathbf{B}$ to the interior of the unit circle, i.e. the VAR is not restricted to be stable. We should restrict the VAR if we believe that our model is the true data generating process, since we expect stationary returns and predictors. But if the true data generating process is say a Markov-switching model then it may be the case that the estimation of this simple VAR shows nonstationarity although the process is really stationary. These priors are informative for continuity with the regime switching models where it is better to use informative priors. The corresponding Gibbs sampler can be found in the appendix.

[16] The parametrization is such that if $\mathsf{X} \sim IW(\nu, \mathsf{S})$, where X is a $d \times d$ positive definite matrix, then $E(\mathsf{X}) = \mathsf{S}^{-1}/(\nu - d - 1)$. This is equivalent to $\mathsf{X}^{-1} \sim W(\nu, \mathsf{S})$ with $E\left(\mathsf{X}^{-1}\right) = \nu\mathsf{S}$.

The specific values of the hyperparameters are the following.[17] I choose $\underline{\boldsymbol{\pi}} = \mathbf{0}$ and $\underline{\mathbf{C}} = 100\mathbf{I}$, where $\mathbf{I}$ is the corresponding identity matrix. Therefore, the prior on $\boldsymbol{\pi}$ is not very informative. On the other hand, the prior on $\boldsymbol{\Sigma}$ of each version of the model will be computed from its MLE for the first estimation period, up to December 1974. Specifically,

$$\underline{\nu} = \underline{T}, \quad \underline{\mathbf{S}} = \left(\underline{T}\hat{\boldsymbol{\Sigma}}\right)^{-1},$$

where $\underline{T} = 25 \simeq T/10$. This prior represents a small sample centred at the MLE.

In terms of forecasting, there are not latent variables is this model and we can directly rely on normal distributions. Since

$$\mathbf{y}_{T+1} | \left(\mathbf{y}^T, \boldsymbol{\theta}\right) \sim N\left(\mathbf{a} + \mathbf{B}\mathbf{y}_T, \boldsymbol{\Sigma}\right),$$

we have that

$$e^S_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}\right) \sim N\left(a^S + \left(\mathbf{b}^S\right)' \mathbf{y}_T, \sigma^{SS}\right),$$

where a^S is the corresponding entry in $\mathbf{a}$, $\left(\mathbf{b}^S\right)'$ is the corresponding row in $\mathbf{B}$, and σ^{SS} is the corresponding entry in $\boldsymbol{\Sigma}$.

In addition,

$$e^B_{T+1} \mid \left(\mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}\right) \sim N\left(a^B + \left(\mathbf{b}^B\right)' \mathbf{y}_T + \frac{\sigma^{SB}}{\sigma^{SS}} u^S_{T+1}, \sigma^{BB} - \frac{\left(\sigma^{SB}\right)^2}{\sigma^{SS}}\right),$$

where the notation is similar to the previous one. The distributions of $e^B_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}\right)$ and $e^S_{T+1} \mid \left(\mathbf{y}^T, e^B_{T+1}, \boldsymbol{\theta}\right)$ are defined in the same way. The required moments can be computed following (1) and (2), while the PIT's are described in (3) and (4).

Finally, I use a slight approximation in the previous computations since I do not reestimate the model in each period of time. I do it each three periods, that is, if I use $p\left(\boldsymbol{\theta} \mid \mathbf{y}^T\right)$ at period T then I use the approximations $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T+1}\right) \simeq p\left(\boldsymbol{\theta} \mid \mathbf{y}^T\right)$ and $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T+2}\right) \simeq p\left(\boldsymbol{\theta} \mid \mathbf{y}^T\right)$ in the next two following periods. Afterwards, I estimate $p\left(\boldsymbol{\theta} \mid \mathbf{y}^{T+3}\right)$ and the previous procedure is repeated again. This does not represent a big change in the final computations while it really decreases the required time for the recursive out-of-sample check.

4.2 Estimation and Out-of-sample Evaluation

In this section, the empirical results of the model are shown in two dimensions. First, the estimation is shown for the first and last estimation periods to interpret the model and compare it to the usual results in the literature. Second, the evaluation of the density forecasts is shown for the commented recursive estimation starting with the forecast of January 1975.

[17] The priors for each model are available upon request from the author.

4.2.1 Estimation of the Model

Table 2 shows part of the estimation of the two versions of the model. Specifically, the tables and the corresponding comments are focused on the posterior mean of some parameters. The estimation is shown for the first and last estimation periods, defined by data up to December 1974 and November 2001. The following comments also explain results that are not shown in the tables. The comments on the significance of the parameters are based on the comparison of the posterior mean and standard deviation, although the latter is not displayed in the tables. The posterior distributions of the conditional mean parameters are very close to a normal distribution in general.

In the case of stocks, both predictors contribute with a positive coefficient to the conditional mean, as it is found in the literature. But the dividend yield cannot be regarded as significant in the last estimation. It is well known that the dividend yield has lost part of its predictive power during the second half of the 90's. The lagged bond return, which is not shown in the table, has a significant positive coefficient in both estimations. The constant mean version shows a positive mean that is only significant in the last estimation. If we turn to bonds, the dividend yield coefficient changes sign. But the last value cannot be regarded as significant and therefore it has lost predictive power for bonds too. On the other hand, the term premium has a positive coefficient in both estimations, as it is usual in the literature of predictability, being significant the last one. Both lagged excess returns, which are not shown in the table, have a significant coefficient in the last estimation, negative for the stock and positive for the bond. The intercept in the constant mean version is not significant in both estimations. In fact, it is negative in some periods. A well-known feature is that the predictors are almost single persistent AR(1)'s. This is also found here. During the last years, the estimated value of the autoregressive coefficient of the dividend yield has increased and got closer to 1.

In terms of the residual variances of excess returns, there is an increase from one period to the other in the covariance, which is positive, and both variances. The covariance suffers a huge increase in the first estimation years and then becomes more stable. The bond variance increases sharply in the beginning and then falls smoothly. The correlation, computed at the posterior mean values of covariances and variances, has increased from 0.04 to 0.15 in the predictable mean version. The residual covariance between the ESR and the dividend yield is always negative and decreasing. This sign is found in the literature of predictability and it is a signal of mean-reversion in stocks. Another feature that is found in Campbell et al. (2003) is a positive residual covariance of bonds and the term premium, which is a signal of mean-aversion in bonds. That is not clear in this estimation, where the residual covariance between the EBR and the term premium changes sign form negative to positive and finally becomes negative again. The residual variances in the constant mean version do not show a huge increase with respect to the predictable version. This is expected as the classical R^2 using predictors is low in monthly predictive regressions.

4.2.2 Evaluation of Density Forecasts

The density forecasts from each version of this model are shown in Figures 4 and 5. In the case of stocks and the predictable mean version, Panels A and B in Figure 4, bellow the 45 degrees line, which may be due to a higher mean. Panel A in Table 3 shows that the means of the PIT's are around 0.52, while the predictable mean versions had a mean of around 0.51. The main difference is that the asymmetry doubles its negative value in the constant mean version. Now we see that the right tail is too thick and the left tail too thin in the constant mean PIT's. It was commented that the time series means of the density forecasts where lower for the case of constant mean and we wondered if that was overpessimism. The PIT gives now the answer.

In terms of correlograms, see Figure 8 again, the level correlation in ESR increases, but the second power correlation decreases. This is confirmed by the numbers in Panel B of Table 3. The level correlation of EBR increases at the first lag. The EDF tests worsen, Panel C in Table 3, with the Anderson-Darling rejected at the 2.5% in the case of the joint PIT. Wallis improves for ESR, Panel D, but worsens for EBR because of the independence test. The joint PIT has lower p-values than before. However, there are no rejections at the 2.5%.

The difference between the time series of the PIT's from both versions is shown in Figure 9. The predictable mean version gives lower PIT's in general until the beginning of the 90's, see Panel A and B, 1993 for ESR and 1995 for EBR. This is due to a higher mean in its density forecasts. Then the tendency changes and it usually gives higher PIT's because it predicted negative returns. Apart from the mean, the rest of features of the forecasts are similar. The correlation between the PIT's ranges from 0.94 to 0.97.

The previous results are based on the integrated forecasting rules but the plug-in rule was also evaluated. The difference between the time series of the PIT's from both forecasting rules can be seen in Figure 10 for the predictable version. That difference is low and not systematic, with higher volatility during unstable periods, mainly the beginning of the 80's and the last years. The difference is small in the case of the constant mean version, as it can be expected since the conditional version drives many of the results. All the evaluation tests give similar results when applied to both forecasting rules. The correlation between the corresponding PIT's is 1. The correlation between the 4 moments of the density forecasts that are computed in this paper are also 1 or very close. The parameter uncertainty is not able to create by itself a significant amount of asymmetry or kurtosis.

To sum up, the performance of the PIT's from the normal VAR is good in general. There is only a failure in the dependence of the second power of the EBR. It is not clear if a constant or time-varying conditional mean gives better forecasting performance. The constant mean version tends to give overpessimistic forecasts. On the other hand, using the integrated or the plug-in forecasting rules does not matter in terms of this type of evaluation.

5 Markov-Switching VAR

This section is devoted to the evaluation of an extension to the standard normal VAR model explained in section 4. Time-varying risk is introduced by means of nonlinear dynamics. The investment opportunities will vary in terms of conditional means and variances, not only conditional means as in the previous section. Time-varying volatility can be modelled as in GARCH and stochastic volatility models or as regime switching models. The latter are the ones that will be studied in this paper.[18] Hamilton and Susmel (1994) explore the issue of ARCH versus regime switching effects in variance for U.S. monthly data. They conclude that ARCH effects are not very important at the monthly frequency after taking into account regime switching.

This type of models separates the asymmetric behaviour of returns in different states as it is emphasized in conditional asset pricing models and modern forecasting procedures in macroeconomics. For instance, in terms of asset pricing, a countercyclical Sharpe ratio is interpreted as higher risk aversion in recessions than in expansions. Ang and Bekaert (2002) point out the advantage of regime-switching models to capture the asymmetry that is found in stock correlations during bull and bear markets.[19]

On the other hand, the portfolio choice implications of these models have just begun to be analysed. Ang and Bekaert (2001) have applied this model to portfolio choice, although in a very simple framework. They compute the mean-variance optimal portfolio in each regime and switch the final portfolio between both depending on a threshold in the probability of regimes.

This type of models introduces an unobserved discrete variable that follows a Markov chain and which outcome defines, in the context of a VAR, the values of the vector of constants, the autoregressive matrices and/or the residual variance matrix. The Markov chain is used to represent persistence in economic regimes and differentiates this model from a simple gaussian mixture. In addition, the unobservability of the state represents the approach to take if we want to be able to make real-time forecasts. Hamilton (1989) develops the reference model[20] and Kim and Nelson (1998) show its implementation in a Bayesian framework to macroeconomic indicators. In finance, Pérez-Quirós and Timmermann (2000, 2001) and Chauvet and Potter (2001) are some examples.

5.1 Model

This new model is only a different distribution of the innovations with respect to the previous VAR. The switching is concentrated to the residual variance matrix. The model is defined by two pieces, the distribution of the VAR in each regime and the evolution of the regime, called s_t, as a Markov chain. Only two regimes are allowed, so that the regime variable s_t can take only two values, say 1 or 2. The model starts with the same VAR as (5), i.e. the parameters of the conditional mean are kept constant in both regimes, but changes the distribution of the perturbation. Instead of (6), now we have

$$\mathbf{u}_t \mid \left(\mathbf{y}^{t-1}, s_t = j, \boldsymbol{\theta}\right) \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_j\right), \quad j = 1, 2 \tag{8}$$

[18] There is also a growing literature on the implementation of this model to explain the behaviour of the short-term interest rate, see for instance Gray (1996) or Ang and Bekaert (2002). There seems to be correlation between the regimes of stocks and interests rates, which are also related to the business cycle.

[19] See for instance Ang and Chen (2002).

[20] Kim (1994) translates that idea to general state-space models and improves some technical issues.

where the distribution of $\mathbf{u}_t$ is expressed conditional on the regime. The vector $\boldsymbol{\theta}$ represents all the parameters of the model as it will be shown later. This model does not allow asymmetry in the returns. Anyway, a model with switches in the conditional mean was estimated in a previous version of this paper and there was not a clear evidence of switching in any parameter. This a Markov-switching VAR with 4 variables, but it is not usual to see these types of VAR's with more than 2 or 3 variables. For instance, Pérez-Quirós and Timmermann (2000, 2001) apply the model to 1 or 2 series of stock returns, but without introducing the dynamics of the predictors. On the other hand, Ang and Bekaert (2001) apply it to 3 returns, short-term rate, bonds and stocks, but their only predictor is then one of the returns, specifically the short term rate. The joint modelling of returns and predictors will be crucial in the identification of regimes.

Again, two versions of this model will be estimated and evaluated, one with a time-varying conditional mean and another with a constant one. Even in the last case, returns are predictable in this model since the conditional second moment is still time-varying. The case of constant mean is defined with the same restrictions on $\mathbf{B}$ as in the normal model (7).

The unobservable regime follows a binary Markov chain,

$$\begin{aligned} p\left(s_t \mid s^{t-1}, \mathbf{y}^{t-1}, \boldsymbol{\theta}\right) &= p\left(s_t \mid s_{t-1}, \boldsymbol{\theta}\right), \\ p\left(s_t = j \mid s_{t-1} = i, \boldsymbol{\theta}\right) &= p_{ij}, \quad i, j = 1, 2, \end{aligned} \tag{9}$$

which captures the idea of persistence in economic variables. The parameters to estimate are now

$$\boldsymbol{\theta} = \left\{\left(p_{11}, p_{22}\right), \boldsymbol{\pi}, \left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right)\right\},$$

and the likelihood is

$$p\left(\mathbf{y}^T \middle| \boldsymbol{\theta}\right) = \prod_{t=1}^{T} \sum_{s_t} p\left(s_t \mid y^{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}, s_t\right),$$

where $p\left(s_t \mid \mathbf{y}^{t-1}, \boldsymbol{\theta}\right)$ is given by the filter in Hamilton (1989). The component $p\left(\mathbf{y}_t \middle| \mathbf{y}^{t-1}, \boldsymbol{\theta}, s_t\right)$ is normal.

The priors have the following structure,

$$\begin{gathered} p_{11} \sim B\left(\underline{n}_{11}, \underline{n}_{12}\right), \quad p_{22} \sim B\left(\underline{n}_{22}, \underline{n}_{21}\right), \\ \boldsymbol{\pi} \sim N\left(\underline{\boldsymbol{\pi}}, \underline{\mathbf{C}}\right), \\ \boldsymbol{\Sigma}_1 \sim IW\left(\underline{\nu}_1, \underline{\mathbf{S}}_1\right), \quad \boldsymbol{\Sigma}_2 \middle| \boldsymbol{\Sigma}_1 \sim TIW_{\mathcal{I}(\sigma_{2,11} > \sigma_{1,11})}\left(\underline{\nu}_2, \underline{\mathbf{S}}_2\right) \end{gathered}$$

While the prior on $\left(p_{11}, p_{22}\right)$ is constructed as independent, the prior on $\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right)$ is defined jointly to identify the regimes during the MCMC. The notation $B\left(\cdot, \cdot\right)$ means a beta distribution.[21] The prior of $\mathbf{\Sigma}_2$ given $\mathbf{\Sigma}_1$ is truncated to region where the first entry of $\mathbf{\Sigma}_2$ is higher that the first entry of $\mathbf{\Sigma}_1$. The notation used for the truncated inverse Wishart is $TIW_{\mathcal{I}(\cdot)}(\cdot,\cdot)$, where $\mathcal{I}(\cdot)$ defines the truncation region. Informative priors are needed in these type of models because it could be the case that some regimes are not visited in some iterations. The corresponding Gibbs sampler of this model is shown in the appendix. The estimation is based on Kim and Nelson (1998), where data augmentation with regimes s^T is the key element to return to the normal world.

The specific values of the hyperparameters are the following. I choose $\bar{n}_{ij} = 1$, for all i and j, and $\underline{\boldsymbol{\pi}} = \mathbf{0}$ and $\underline{\mathbf{C}} = 100\mathbf{I}$. On the other hand, the prior on $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ will be computed from its MLE of each version of the model for the first estimation period, up to December 1974,

$$\underline{\nu}_j = \underline{T}, \quad \underline{\mathbf{S}}_j = \left(\underline{T}\hat{\mathbf{\Sigma}}_j\right)^{-1}, \quad j = 1, 2,$$

where $\underline{T} = 25$ again.

In terms of forecasting, there are latent variables in this model and we cannot directly rely on normal distributions. However, we are back to the normal world conditioning on the latent variables. Since

$$\mathbf{y}_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}, s_{T+1} = j\right) \sim N\left(\mathbf{a} + \mathbf{B}\mathbf{y}_T, \mathbf{\Sigma}_j\right),$$

we have that

$$e^S_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}, s_{T+1} = j\right) \sim N\left(a^S + \left(\mathbf{b}^S\right)' \mathbf{y}_T, \sigma_j^{SS}\right),$$

where a^S is the corresponding entry in $\mathbf{a}$, $\left(\mathbf{b}^S\right)'$ is the corresponding row in $\mathbf{B}$, and σ_j^{SS} is the corresponding entry in $\mathbf{\Sigma}_j$. In addition,

$$e^B_{T+1} \mid \left(\mathbf{y}^T, e^S_{T+1}, \boldsymbol{\theta}, s_{T+1} = j\right) \sim N\left(a^B + \left(\mathbf{b}^B\right)' \mathbf{y}_T + \frac{\sigma_j^{SB}}{\sigma_j^{SS}} u^S_{T+1}, \sigma_j^{BB} - \frac{\left(\sigma_j^{SB}\right)^2}{\sigma_j^{SS}}\right).$$

The distributions of $e^B_{T+1} \mid \left(\mathbf{y}^T, \boldsymbol{\theta}, s_{T+1} = j\right)$ and $e^S_{T+1} \mid \left(\mathbf{y}^T, e^B_{T+1}, \boldsymbol{\theta}, s_{T+1} = j\right)$ are defined in a similar way. The required moments can be computed following (1) and (2), while the PIT's are described in (3) and (4).

Finally, I use an approximation in the previous computations since I do not reestimate the model in each period of time. If I use $p\left(\boldsymbol{\theta}, s^T \mid \mathbf{y}^T\right)$ at period T, its exact distribution, then I use approximations at $T+1$ and $T+2$ to take advantage of the Gibbs from T and decrease the number of required computations. I use the exact expression

[21] The parametrization is such that if $x \sim B(\alpha, \beta)$ then $E(x) = \alpha/(\alpha + \beta)$.

(1) at T, but at period $T+1$ I use

$$E_{p^{\text{I}}\left(\mathbf{y}_{T+2}|\mathbf{y}^{T+1}\right)}\left[\left(e_{T+2}^{S}\right)^{k}\right] =$$

$$E_{p(\boldsymbol{\theta},s_T|\mathbf{y}^{T+1})}\left[\sum_{s_{T+1}} p\left(s_{T+1} \mid \boldsymbol{\theta}, s_T\right) \sum_{s_{T+2}} p\left(s_{T+2} \mid \boldsymbol{\theta}, s_{T+1}\right) E_{p\left(\mathbf{y}_{T+2}|\mathbf{y}^{T+1},\boldsymbol{\theta},s_{T+2}\right)}\left[\left(e_{T+2}^{S}\right)^{k}\right]\right]$$

and the only approximation is the use of $p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^T\right) \simeq p\left(\boldsymbol{\theta}, s_T \mid \mathbf{y}^{T+1}\right)$ to take advantage of the Gibbs from the previous period. The sums over regimes can be directly computed by means of powers of the matrix of transition probabilities. Something similar is applied at period $T+2$. Afterwards, I estimate $p\left(\boldsymbol{\theta}, s^{T+3} \mid \mathbf{y}^{T+3}\right)$ and the previous procedure is repeated again. Finally, in the case of the plug-in rule, I evaluate the previous expressions at the corresponding posterior means.

5.2 Estimation and Out-of-sample Evaluation

Again, the empirical results of the model are shown in two dimensions, estimation of the parameters and evaluation of real-time density forecasts.

5.2.1 Estimation of the Model

Table 4 shows the posterior mean of some parameters of the two versions of the model. Again, the estimation is shown for the first and last estimation periods, December 1974 and November 2001, and there will be comments that refer to some computations that are not shown in the tables.

The values of the transition probabilities imply persistent regimes. Both regimes are similarly persistent in the first estimation, but the high-volatility regime persistence decreases in the last estimation. The persistence of the low-volatility regime decreases at the beginning but afterwards increases so that ends at a similar level, while the other regime persistence is decreasing along the whole period. The posterior means for the predictable mean version would imply a steady state probability for the low-volatility regime that goes from 0.51 to 0.77 during the data period.

The values of the conditional mean parameters are similar to the normal VAR model. Maybe it is strange to see now a negative intercept for bonds in the first estimation of the constant mean version, but it is not really significant.

In terms of the residual variance, recall that the only identifying assumption was that the second regime had a higher first entry. The rest of entries are left unrestricted. Anyway, the model identifies a high-volatility regime for both assets. In the first estimation, stocks approximately double its variance form one regime to the other. Bond variance is approximately eight times higher, but its values are much lower than stock's ones. The covariance is not really significant in both regimes. In the second estimation, the switching is about three times for stocks and six times for bonds. All the values are higher than the first estimation, specially the high-volatility regime. The stock variance has increased along the sample period in both regimes. The bond variance in the first regime grows smoothly in general but has a jump at the beginning of the 90's. The corresponding variance of the second regime jumps at the beginning of the 80's and the 90's.

About covariances, only the low-volatility regime one can be regarded as significantly positive. In fact, the corresponding correlations in each regime, computed at the posterior mean of the parameters, are 0.19 versus 0.12 for the predictable mean version. The covariance in the first regime starts being negative but becomes positive and grows very soon, while in the second regime starts growing and then becomes stable. About the residual covariances between the returns and the predictors, the covariance between ESR and dividend yield is negative in both regimes and both periods. In the final estimation, it turns to be about three times more negative in the high-volatility regime. On the other hand, the covariance between EBR and term premium is negative in the first regime, but the sign is not clear in the second. The high-volatility regime covariance is significantly positive in the constant mean version.

Finally, the probability of the high-volatility regime from the last estimation can be seen in Panel A of Figure 11, jointly with the NBER recession index. The model identifies regimes that are closely related to the business cycle, because the high-volatility regime is usually associated with recessions. This is a noteworthy result because I am not using macroeconomic variables such as GDP or inflation and I am not using time-varying transition probabilities. Pérez-Quirós and Timmermann (2000, 2001) introduce this variation as dependence of the transition probabilities on the composite leading indicator. That is their way to track the business cycle instead of outliers, but it is not necessary in this paper. This may be due to the fact of modelling a joint VAR of excess returns and predictors, not only a VAR of excess returns.

5.2.2 Evaluation of Density Forecasts

The density forecasts from each version of this model are shown in Figures 12 and 13. The case of stocks and the predictable mean version is displayed in Panels A and B of Figure 12. The standard deviation shows more action now, while the mean shows a bit lower variation. The coefficient of asymmetry is still close to zero all the time, but the coefficient of kurtosis varies through time and increase during the 90's. The EBR density forecast changes more with respect to the normal VAR, see Panels C and D of Figure 12. It shows more variation in the standard deviation compared to the mean, while the mean shows smaller peaks. The coefficient of asymmetry is again close to zero all the time, but kurtosis now changes very sharply and is higher than ESR's kurtosis. The constant mean version of the model, Figure 13, does not need additional comments.

In terms of numbers, the time series means of the mean, standard deviation, and the coefficients of asymmetry and kurtosis are 0.71, 4.20, -0.00, and 0.36, respectively, for the ESR's density forecasts with predictable mean. The corresponding EBR's values are 0.10, 1.48, 0.00 and 1.97. In the case of the constant mean version, the values are 0.48, 4.32, -0.00, and 0.47 for ESR, and 0.01, 1.50, -0.00, and 2.01 for EBR. The main difference between both versions of forecasts is that the time series mean is clearly lower for both assets. Therefore, the previous result found for the normal VAR model is robust. In both versions, the main difference with the previous normal VAR forecasts is the increase in the kurtosis, specially in the EBR. The correlation between the means from the predictable mean version of both models is around 0.99 for ESR, and around 0.85 for EBR. These values are similar for other two regime-switching models that were also evaluated. They are commented in the conclusions.

Now I turn to the study of the PIT's. The time series of the PIT's are not shown since they do not show many differences at first sight with respect to the normal VAR. The correlation between the corresponding PIT's from each model is 1 for the ESR and around 0.97-0.98 for the EBR. Again, these values are similar for other two regime-switching models that were also evaluated. If we compute the time series of the differences of the PIT's from the predictable mean versions of both models then we can see a similar pattern to Figure 9, the difference between the predictable and constant mean versions of the normal model. It exhibits a negative value before the 90's in general and a positive value afterwards. In fact, the difference between the PIT's can also be related with the difference in the mean of the density forecasts. Anyway, the constant mean version shows a lower difference, there is a higher range for EBR again, and does not have a clear sign.

Figure 14 shows some PIT's from the predictable mean version. The CDF's do not show a significant departure form uniformity, although the EBR in Panel B looks worse that the normal VAR. It has a fatter right tail, which translates also into the joint PIT in Panel C. The noteworthy feature is the lack of correlation in the second moment of the EBR, although there is still some in the joint PIT. It has been stressed that there is not a direct relation between the marginal and dynamic properties of the PITs and the returns, but in this case a richer specification of the conditional variance of returns has killed the correlation in the second moment of the corresponding PIT.

The corresponding numbers and tests can be found in Table 5. Panel A shows some descriptive statistics that are close to its uniform counterparts, while Panel B shows the commented lack of second moment correlation in the EBR. The EDF tests of EBR, Panel C, worsen with respect to the normal VAR, specifically the Kuiper test of the joint PIT is rejected at the 2.5%. The Wallis test, Panel D, shows a p-value of only 0.001 for ESR with 4 bins due to the independence test. The EBR's p-values decrease, mainly due to the uniformity test. The unconditional coverage p-value for EBR with 4 bins is only 0.012.

Figure 15 displays some PIT's from the constant mean version. In terms of CDF's there seems to be a worsening with respect to the predictable version. The correlograms do not require additional comments. Table 5 shows the corresponding tests too. First, looking at Panel A, the PIT's mean from the constant mean version is a bit higher and its asymmetry is more negative, as it was the case in the normal VAR model. Therefore, this feature of the PIT's is robust to the model. The EDF tests worsen, Panel C, with the Anderson-Darling rejected at the 1% in the case of the EBR and the joint PIT. Wallis improves for ESR, Panel D, but worsens for EBR leading to a clear rejection with 4 bins. The joint PIT has lower p-values than before, with rejections of conditional coverage at the 1%.

In terms of comparing the time series of the PIT's from the two versions, the results are similar to the normal VAR, which was plotted in Figure 9. In terms of comparing the corresponding PIT's of the two forecasting rules, there is neither a big difference. Recall that these time series were shown in figure 10 for the normal model. One of the new features in the Markov-switching model is kurtosis. The kurtosis of both forecasting rules also have a correlation of 1.

To sum up, the main difference with the density forecasts from the normal VAR is the increase in the kurtosis, specially for bonds. The Markov-switching model has killed the dependence of the second power of the EBR's PIT, but has worsen other dimensions with respect to the normal VAR such as the uniformity of the EBR's PIT. In addition, it has been found that the relation between the PIT's from the two mean versions and the relation between the PIT's from the two forecasting rules are robust to the model.

6 Conclusions

This paper has evaluated different joint density forecasts of stocks and bonds. Some conclusions can be drawn in terms of the different models that have been implemented. The normal model gives uniform probability integral transforms (PIT's) and only fails in the dynamics of the second moment of the bond's PIT.

The Markov-switching increases the kurtosis of the density forecasts, specially for the bond. This model does not improve the evaluation results in general. It only improves the dynamics of the second moment of the PIT of bonds, while it worsens the uniformity of that PIT. Anyway, in a previous version of this paper that rely on data up to 1994, the Markov-switching model really improved the evaluation for bonds and it seemed a good choice to capture changes in monetary policy. It may be the case that the addition of new years have decreased the importance of the dramatic changes in bonds behaviour during the beginning of the 80's. A noteworthy feature of the estimation of this model is that the regime variable tracks the business cycle, not outliers of excess returns, without the introduction of macroeconomic variables.

In addition to the commented models, two alternative regime-switching models were evaluated. They may be called an independent-switching and a predictor-switching model. The former is an independent mixture of two regimes and therefore it may be seen as a restricted version of the Markov-switching model. It only introduces fat tails with respect to the normal model. The predictor-switching model defines directly the weights of the mixture as a function of the predictors by means of the normal CDF. That avoids the need of filtering an unobservable regime. Both models are much easier to estimate than the Markov-switching one so it is interesting to study their performance.

Their corresponding results are not shown to shorten the extension of the paper,[22] apart from their probabilities of the high-volatility regime in Figure 11. The independent-switching model identifies regimes that are not closely related to the business cycle. The behaviour of the probabilities in the predictor-switching model lies in between the other two regime-switching models. That is, they are not as smooth as the Markov-switching ones but closer to the business cycle than the independent-switching model ones. The

[22] Of course, they are available upon request from the author.

big difference with the previous models is that it labels the end of the sample as a low volatility regime although it is a recession for the NBER index. This is due to the strong effect of the dividend yield in the probability. These additional two models create more kurtosis in general than the Markov-switching. However, the kurtosis decreases during the 90's because of the dividend yield in the predictor-switching model. Both models fail to give a uniform PIT in the case of bonds, with too many realizations on both tails of the density forecast. The predictor-switching model is able to improve the dynamics of the second moment of the PIT of bonds.

Therefore, the standard normal VAR model does not seem to be a bad description of the data. This could explain why the portfolio choices in Brandt (1999), which do not rely in a statistical model of returns, are not far away from the choices previously computed in the literature for the normal VAR model. For instance, in the case of models with time-varying risk premia, the correlation of the PIT's is 1 for the stocks and approximately 0.98 for bonds. This is mainly driven by the correlation in the conditional means, approximately 0.99 for stocks, while lower for bonds, 0.85.

In terms of time variation in risk premia, it is clear that the versions of the models with or without that time variation imply very different density forecasts. They vary much more over time in a model with predictable mean, and this translates into significant effects in portfolio choice as it is well known in the literature on predictability. The results in terms of the PIT are not so clear. With respect to a constant conditional mean, a time-varying one worsens the forecast evaluation for stocks but improves for bonds. The PIT's from the constant mean version show that the density forecasts of this version are overpessimistic. Those PIT's have a thicker right tail and a thinner left tail than a uniform distribution. This is a robust result that can be found in all the studied models.

Finally, about parameter uncertainty, the fully Bayesian approach and the plug-in forecasting rule give similar density forecasts and similar properties of the PIT. This uncertainty does not create much asymmetry or kurtosis in the forecasts and the correlation between the corresponding descriptive statistics of the forecasts and PIT's are close to 1. This is also a robust result across the different models. The literature on portfolio choice has pointed out the importance of this uncertainty, but it may be due to the choice of a particular utility function or the use of long-horizon returns instead of the monthly forecasts of this paper.

This paper has relied on two returns and two predictors an clear avenues of future research would be the study of other returns, predictors and/or countries. In this direction, different types of stocks could also be studied, for instance growth and value stocks. In terms of models, it would be interesting to see the performance of stochastic volatility or GARCH models against the Markov-switching estimated here. It would also be interesting the introduction of asymmetry in the models along the lines of Fernández and Steel (1998) and Hansen (1994). Finally, the results of this paper are relevant for an investor that rebalances her portfolio each month. But it may be the case that the investor's rebalancing frequency is lower. Therefore, it would be interesting to study the performance of the models in terms of long-horizon returns in addition to the monthly forecasts presented here. However, the overlapping forecasts make the evaluation more difficult.

References

Ang, A., and G. Bekaert (2001): How Do Regimes Affect Asset Allocation?, mimeo, Columbia University.

Ang, A., and G. Bekaert (2002): Short Rate Nonlinearities and Regime Switches, Journal of Economic Dynamics and Control 26, 1243-1274.

Ang, A., and J. Chen (2002): Asymmetric Correlations of Equity Portfolios, Journal of Financial Economics 63, 443-494.

Avramov, D. (2002): Stock Return Predictability and Model Uncertainty, Journal of Financial Economics 64, 423-458.

Bai, J. (2002): Testing Parametric Conditional Distributions of Dynamic Models, forthcoming in Review of Economics and Statistics.

Barberis, N. (2000): Investing in the Long Run when Returns are Predictable, Journal of Finance 55, 225-264.

Berkowitz, J. (2001): Testing Density Forecasts, with Applications to Risk Management, Journal of Business and Economic Statistics 19, 465-474.

Bossaerts, P., and P. Hillion (1999): Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, Review of Financial Studies 12, 405-428.

Campbell, J.Y. (1987): Stock Returns and the Term Structure, Journal of Financial Economics 18, 373-399.

Campbell, J.Y., Y.L. Chan, and L.M. Viceira (2003): A Multivariate Model of Strategic Asset Allocation, Journal of Financial Economics 67, 41-80.

Campbell, J.Y. and J.H. Cochrane (1999): By Force of Habit: A Consumption-Based Explaination of Aggregate Stock Market Behavior, Journal of Political Economy 107, 205-251.

Campbell, J.Y., and R.J. Shiller (1988): The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, Review of Financial Studies 1, 195-228.

Campbell, J.Y., and R.J. Shiller (1991): Yield Spreads and Interest Rate Movements: A Bird's Eye View, Review of Economic Studies 58, 495-514.

Carter, C.K., and R. Kohn (1994): On Gibbs Sampling for State Space Models, Biometrika 81, 541-553.

Chauvet, M., and S. Potter (2001): Nonlinear Risk, Macroeconomic Dynamics 5, 621-646.

Chib, S., and E. Greenberg (1996): Markov Chain Monte Carlo Simulation Methods in Econometrics, Econometric Theory 12, 409-431.

Christoffersen, P. (1998): Evaluating Interval Forecasts, International Economic Review 39, 841-862.

Clements, M.P., and J. Smith (2000): Evaluating the Forecast Densities of Linear and Non-linear Models: Applications to Output Growth and Unemployment, Journal of Forecasting 19, 255-276.

Cremers, M. (2002): Stock Return Predictability: A Bayesian Model Selection Perspective, forthcoming Review of Financial Studies 15, 1223-1249.

Dawid, A.P. (1984): Statistical Theory: The Prequential Approach, Journal of the Royal Statistical Society (Series A) 147, 278-292.

Diebold, F.X., T.A. Gunther, and A. S. Tay (1998): Evaluating Density Forecasts with Applications to Financial Risk Management, International Economic Review 39, 863-883.

Diebold, F.X., J. Hahn, and A.S. Tay (1999): Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns of Foreign Exchange, Review of Economics and Statistics 81, 661-673.

Fama, E.F., and R.R. Bliss (1987): The Information in Long-Maturity Forward Rates, American Economic Review 77, 680-692.

Fama, E.F., and K.R. French (1988): Dividend Yields and Expected Stock Returns, Journal of Financial Economics 22, 3-27.

Fama, E.F., and K.R. French (1989): Business Conditions and Expected Returns on Stocks and Bonds, Journal of Financial Economics 25, 23-49.

Fama, E.F., and G.W. Schwert (1977): Asset Returns and Inflation, Journal of Financial Economics 5, 115-146.

Fernández, C., and M.F.J. Steel (1998): On Bayesian Modelling of Fat Tails and Skewness, Journal of the American Statistical Association 93, 359-371.

Ferson, W.E., S. Sarkissian, and T. Simin (2003): Spurious Regressions in Financial Economics?, forthcoming in the Journal of Finance.

Gelfand, A.E., and A.F.M. Smith (1990): Sampling-Based Approaches to Calculating Marginal Densities, Journal of the American Statistical Association 85, 398-409.

Goetzmann, W.N., and P. Jorion (1993): Testing the Predictive Power of Dividend Yields, Journal of Finance 48, 663-679.

Gray, S.F. (1996): Modeling the Conditional Distribution of Interest Rates as a Regime-switching Process, Journal of Financial Economics 42, 27-62.

Hamilton, J.D. (1989): A New Approach to the Economic Analysis of Nonstationay Time Series and the Business Cycle, Econometrica 57, 357-384.

Hamilton, J.D., and R. Susmel (1994): Autoregressive Conditional Heteroskedasticity and Changes in Regime, Journal of Econometrics 64, 307-333.

Hansen, B. (1994): Autoregressive Conditional Density Estimation, International Economic Review 35, 705-730.

Hodrick, R.J. (1992): Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, Review of Financial Studies 5, 357-386.

Kandel, S., and R.F. Stambaugh (1996): On the Predictability of Stock Returns: An Asset-Allocation Perspective, Journal of Finance 51, 385-424.

Keim, D.B., and R.F. Stambaugh (1986): Predicting Returns in the Stock and Bond Markets, Journal of Financial Economics 17, 357-390.

Kim, C.-J. (1994): Dynamic Linear Models with Markov-switching, Journal of Econometrics 60, 1-22.

Kim, C.-J., and C.R. Nelson (1998): Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching, Review of Economics and Statisitics 80, 188-201.

Kim, S., N. Shephard, and S. Chib (1998): Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, Review of Economic Studies 65, 361-393.

Lamont, O. (1998): Earnings and Expected Returns, Journal of Finance 53, 1563-1587.

Lettau, M., and S. Ludvigson (2001a): Consumption, Aggregate Wealth and Expected Stock Retruns, Journal of Finance 56, 815-849.

Merton, R.C. (1973): An Intertemporal Capital Asset Pricing Model, Econometrica 41, 867-887.

Pérez-Quirós, G., and A. Timmermann (2000): Firm Size and Cyclical Variations in Stock Returns, Journal of Finance 55, 1229-1263.

Pérez-Quirós, G., and A. Timmermann (2001): Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities, Journal of Econometrics 103, 259-306.

Rosenblatt, M. (1952): Remarks on Multivariate Transformations, Annals of Mathematical Statistics 23, 470-472.

Shephard, N. (1994): Partial Non-Gaussian State Space, Biometrika 81, 115-131.

Smith, A.F.M., and G.O. Roberts (1993): Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, Journal of the Royal Statistical Society (Series B) 55, 3-23.

Stambaugh, R.F. (1999): Predictive Regressions, Journal of Financial Economics 54, 375-421.

Stephens, M.A. (1974): EDF Statistics for Goodness of Fit and Some Comparisons, Journal of the American Statistical Association 69, 730-737.

Tanner, M.A., and W.H. Wong (1987): The Calculation of Posterior Distributions by Data Augmentation, Journal of the American Statistical Association 82, 528-550.

Tay, A.S., and K.F. Wallis (2000): Density Forecasting: A Survey, Journal of Forecasting 19, 235-254.

Wallis, K.F. (2001): Chi-squared Tests of Interval and Density Forecasts, and the Bank of England's Fan Charts, mimeo, University of Warwick.

Appendix

A Gibbs Samplers

The Gibbs samplers that have been used to approximate the posterior distributions are explained in the following pages. There is a description of the conditional posteriors of the blocks of parameters on which the Gibbs sampler iterates are described.

A.1 Normal VAR

Defining

$$\mathbf{x}_t = \begin{pmatrix} 1 \\ \mathbf{y}_{t-1} \end{pmatrix},$$

we can write the sample data at a particular point in time T in matrix notation as

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{U},$$

where $\mathbf{Y}$ is $T \times 4$ and $\mathbf{X}$ is $T \times 5$. The following notation will be convenient,

$$\mathbf{y} = vec\left(\mathbf{Y}\right), \quad \widetilde{\mathbf{X}} = \mathbf{I}_4 \otimes \mathbf{X},$$

where $\mathbf{I}_4$ is the identity matrix of order 4. In the case of constant mean, the matrix $\widetilde{\mathbf{X}}$ takes the form

$$\widetilde{\mathbf{X}} = \begin{pmatrix} \mathbf{I}_2 \otimes \ell_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \otimes \mathbf{X} \end{pmatrix}$$

where ℓ_T is a $T \times 1$ vector of ones.

- Sampling $\boldsymbol{\pi}$:

Since the likelihood implies the following kernel for $\boldsymbol{\pi}$

$$p\left(\mathbf{y}^T \mid \boldsymbol{\theta}\right) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \widetilde{\mathbf{X}}\boldsymbol{\pi}\right)' \mathbf{V}^{-1}\left(\mathbf{y} - \widetilde{\mathbf{X}}\boldsymbol{\pi}\right)\right\},$$
$$\mathbf{V}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T,$$

we can sample it as

$$\boldsymbol{\pi} \mid \left(\mathbf{y}^T, \boldsymbol{\Sigma}\right) \sim N\left(\bar{\boldsymbol{\pi}}, \bar{\mathbf{C}}\right),$$
$$\bar{\boldsymbol{\pi}} = \bar{\mathbf{C}}\left[\underline{\mathbf{C}}^{-1}\underline{\boldsymbol{\pi}} + \left(\widetilde{\mathbf{X}}\right)' \mathbf{V}^{-1}\mathbf{y}\right],$$
$$\bar{\mathbf{C}} = \left[\underline{\mathbf{C}}^{-1} + \left(\widetilde{\mathbf{X}}\right)' \mathbf{V}^{-1}\widetilde{\mathbf{X}}\right]^{-1}.$$

Of course, this is only a compact way of expressing the terms, because the dimension of $\mathbf{V}^{-1}$

- Sampling $\mathbf{\Sigma}$:

The corresponding kernel from the likelihood can be expressed as

$$p\left(\mathbf{y}^T \mid \boldsymbol{\theta}\right) \propto \left|\mathbf{\Sigma}^{-1}\right|^{\frac{T}{2}} \exp\left\{-\frac{1}{2} tr\left(\mathbf{\Sigma}^{-1}\mathbf{U}'\mathbf{U}\right)\right\},$$

which implies the conditional posterior

$$\mathbf{\Sigma} \mid \left(\mathbf{y}^T, \boldsymbol{\pi}\right) \sim IW\left(\bar{\nu}, \bar{\mathbf{S}}\right),$$
$$\bar{\nu} = \underline{\nu} + T, \quad \bar{\mathbf{S}} = \left[\underline{\mathbf{S}}^{-1} + \mathbf{U}'\mathbf{U}\right]^{-1},$$
$$\mathbf{U} = \mathbf{Y} - \mathbf{X}\mathbf{\Pi}.$$

A.2 Markov-Switching VAR

- Sampling s^T:

The sample regimes are drawn following Carter and Khon (1994) and Shephard (1994) who advocate the use of a multimove sampler instead of a singlemove one. The required distribution is

$$p\left(s^T \mid \mathbf{y}^T\right) \propto p\left(s_T \mid \mathbf{y}^T\right) \prod_{t=1}^{T-1} p\left(s_t \mid \mathbf{y}^t, s_{t+1}\right),$$
$$p\left(s_t = j \mid \mathbf{y}^t, s_{t+1}\right) \propto p\left(s_t = j \mid \mathbf{y}^t\right) p\left(s_{t+1} \mid s_t = j\right)$$

where, for ease of notation, $\boldsymbol{\theta}$ has been suppressed from all the conditioning sets. To implement that sampler, the filter developed by Hamilton (1989) must be run before to compute $p\left(s_t \mid \mathbf{y}^t\right)$.

- Sampling (p_{11}, p_{22}):

The sampler of probabilities is based on

$$p\left(p_{11}, p_{22} \mid \mathbf{y}^T, s^T, \boldsymbol{\pi}, (\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)\right) = p\left(p_{11}, p_{22} \mid s^T\right) \propto p\left(p_{11}, p_{22}\right) p\left(s^T \mid p_{11}, p_{22}\right).$$

Therefore, we can draw p_{11} from

$$p_{11} \mid \left(\mathbf{y}^T, s^T, \boldsymbol{\pi}, (\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)\right) \sim B\left(\bar{n}_{11}, \bar{n}_{12}\right),$$
$$\bar{n}_{11} = \underline{n}_{11} + n_{11}, \quad \bar{n}_{12} = \underline{n}_{12} + n_{12},$$

where n_{ij} is the number of transitions from $s_t = i$ to $s_{t+1} = j$. We are ignoring the dependence of $p\left(s_1 \mid p_{11}, p_{22}\right)$ on the transition probabilities and rely only on $p\left(s_t \mid s_{t-1}\right)$. This is just a small approximation that simplifies much the computations.

By a similar argument, and given the independence of the priors,

$$p_{22} \mid \left(\mathbf{y}^T, s^T, \boldsymbol{\pi}, (\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)\right) \sim B\left(\bar{n}_{22}, \bar{n}_{21}\right)$$
$$\bar{n}_{22} = \underline{n}_{22} + n_{22}, \quad \bar{n}_{21} = \underline{n}_{21} + n_{21}.$$

- Sampling $\boldsymbol{\pi}$:

By means of the regimes, the likelihood can be decomposed for π as

$$\exp\left\{-\frac{1}{2}\left(\mathbf{y}_1-\widetilde{\mathbf{X}}_1\boldsymbol{\pi}\right)'\mathbf{V}_1^{-1}\left(\mathbf{y}_1-\widetilde{\mathbf{X}}_1\boldsymbol{\pi}\right)\right\}\exp\left\{-\frac{1}{2}\left(\mathbf{y}_2-\widetilde{\mathbf{X}}_2\boldsymbol{\pi}\right)'\mathbf{V}_2^{-1}\left(\mathbf{y}_2-\widetilde{\mathbf{X}}_2\boldsymbol{\pi}\right)\right\},$$

which implies that the conditional posterior is

$$\begin{aligned}
\boldsymbol{\pi} \mid \left(\mathbf{y}^T, s^T, (p_{11}, p_{22}), (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)\right) &\sim N\left(\bar{\boldsymbol{\pi}}, \bar{\mathbf{C}}\right), \\
\bar{\boldsymbol{\pi}} &= \bar{\mathbf{C}}\left[\underline{\mathbf{C}}^{-1}\underline{\boldsymbol{\pi}} + \left(\widetilde{\mathbf{X}}_1\right)'\mathbf{V}_1^{-1}\mathbf{y}_1 + \left(\widetilde{\mathbf{X}}_2\right)'\mathbf{V}_2^{-1}\mathbf{y}_2\right], \\
\bar{\mathbf{C}} &= \left[\underline{\mathbf{C}}^{-1} + \left(\widetilde{\mathbf{X}}_1\right)'\mathbf{V}_1^{-1}\widetilde{\mathbf{X}}_1 + \left(\widetilde{\mathbf{X}}_2\right)'\mathbf{V}_2^{-1}\widetilde{\mathbf{X}}_2\right]^{-1}.
\end{aligned}$$

- Sampling $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$:

The distribution of $\mathbf{u}_t$ given s^T is

$$\mathbf{u}_t \mid s_t = j \overset{iid}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_j\right), \quad j = 1, 2,$$

conditioning on the rest of parameters and the data.

This sampling needs the division of the data by means of the regime variable $s_t = 1$ and 2. We can define $\mathbf{U}_j = \mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\Pi}$, $j = 1, 2$, for the data that are assigned to regime j in one particular iteration of the Gibbs sampler. This splits the information from the likelihood for the parameters in each regime. If all the realizations of the regimes are equal in a particular iteration, the parameters for the other regime are sampled directly from the prior.

The residual variance in the first regime is sampled as

$$\begin{aligned}
\boldsymbol{\Sigma}_1 \mid \left(\mathbf{y}^T, s^T, (p_{11}, p_{22}), \boldsymbol{\pi}\right) &\sim IW\left(\bar{\nu}_1, \bar{\mathbf{S}}_1\right), \\
\bar{\nu}_1 = \underline{\nu}_1 + T_1, \quad \bar{\mathbf{S}}_1 &= \left[\underline{\mathbf{S}}_1^{-1} + \mathbf{U}_1'\mathbf{U}_1\right]^{-1},
\end{aligned}$$

and the one from the second regime follows a truncated distribution given the previous draw

$$\begin{aligned}
\boldsymbol{\Sigma}_2 \mid \left(\mathbf{y}^T, s^T, (p_{11}, p_{22}), \boldsymbol{\pi}, \boldsymbol{\Sigma}_1\right) &\sim TIW_{\mathcal{I}(\sigma_{2,11} > \sigma_{1,11})}\left(\bar{\nu}_2, \bar{\mathbf{S}}_2\right), \\
\bar{\nu}_2 = \underline{\nu}_2 + T_2, \quad \bar{\mathbf{S}}_2 &= \left[\underline{\mathbf{S}}_2^{-1} + \mathbf{U}_2'\mathbf{U}_2\right]^{-1}.
\end{aligned}$$

The truncated inverse Wishart is sampled using an accept-reject with the inverse Wishart.

B Tables

Table 1. Descriptive Statistics of Excess Stock and Bond Returns.

Descriptive statistics are shown for three periods. The first estimation period, 252 observations, the out-of-sample period, 324 observations, and the whole period. For each series, its distribution is described by its mean, standard deviation, and its coefficients of skewness and kurtosis. Finally, its dynamic properties are described by the p-values of then Ljung-Box test of the level and the second power of the demeaned series. That test is shown for 1, 3, and 6 lags.

		1954/01-1974/12		1975/01-2001/12		1954/01-2001/12	
		ESR	EBR	ESR	EBR	ESR	EBR
Panel A. Descriptive Statistics.							
Mean		0.347	0.011	0.608	0.169	0.494	0.100
Stan. Deviation		4.053	1.176	4.572	1.692	4.351	1.489
Coef. of Skewness		-0.423	0.076	-0.971	-0.001	-0.772	0.068
Coef. of Kurtosis		0.865	3.013	3.873	3.515	2.928	4.118
Panel B. Ljung-Box Test. P-values of 1, 3, and 6 Lags.							
Level	1	0.043	0.753	0.403	0.004	0.079	0.003
	3	0.151	0.953	0.331	0.008	0.252	0.008
	6	0.022	0.410	0.392	0.015	0.310	0.041
Second Power	1	0.000	0.139	0.557	0.015	0.041	0.000
	3	0.000	0.031	0.946	0.000	0.065	0.000
	6	0.000	0.001	0.998	0.000	0.188	0.000

Table 2. Estimation of Normal VAR. Posterior Mean of some Parameters. The normal VAR model is defined by (5) and (6). The posterior means of some parameters are shown for estimations at December 1974 and November 2001. For each date, results are shown for the predictable and constant mean versions of the model. In the case of the former version, the first parameters are the slopes of the excess stock return (ESR) and the excess bond return (EBR) with respect to the dividend yield (DY) and the term premium (TP). In the latter version, the intercepts are shown. Finally, the residual variances and covariances of ESR and EBR are shown.

		1974/12		2001/11	
		Predictable	Constant	Predictable	Constant
$\mathbf{\Pi}$	ESR,DY	2.620	0.333	0.827	0.485
	ESR,TP	1.052		0.478	
	EBR,DY	0.897	0.012	-0.008	0.102
	EBR,TP	0.196		0.192	
$\mathbf{\Sigma}$	ESR,ESR	15.169	16.716	18.333	19.104
	ESR,EBR	0.166	0.324	0.944	1.130
	EBR,EBR	1.382	1.410	2.105	2.240

Table 3. PIT's of Normal VAR.

The normal VAR model is defined by (5) and (6). The PIT's $d_{t+1}^{f,S}$, $d_{t+1}^{f,B}$, and $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$ are studied for the predictable and constant mean versions of the model. Panel A shows four descriptive statistics of each PIT. The values of these statistics are 0.50, 0.29, 0.00, and -1.20 for a uniform distribution. Panel B studies the autocorrelation of each PIT by means of the Ljung-Box test. Panel C shows some empirical distribution function (EDF) tests, where the statistic is shown with $*$ if it is significant at the 5%, $**$ if it is significant at the 2.5%, and $***$ if it is significant at the 1%. Finally, Panel D shows the Wallis tests of unconditional coverage (UC), independence (IN), and conditional coverage (CC).

		Predictable Mean			Constant Mean		
		ESR	EBR	Joint	ESR	EBR	Joint
Panel A. Descriptive Statistics.							
Mean		0.506	0.508	0.505	0.524	0.523	0.522
Stan. Deviation		0.299	0.286	0.291	0.285	0.286	0.285
Coef. of Skewness		-0.074	-0.051	-0.047	-0.181	-0.116	-0.137
Coef. of Kurtosis		-1.308	-1.110	-1.202	-1.155	-1.090	-1.121
Panel B. Ljung-Box Test. P-values of 1, 3, and 6 Lags.							
Level	1	0.375	0.030	0.293	0.834	0.006	0.012
	3	0.221	0.186	0.289	0.865	0.049	0.077
	6	0.046	0.228	0.082	0.668	0.102	0.128
Second Power	1	0.890	0.092	0.135	0.251	0.087	0.025
	3	0.172	0.000	0.000	0.081	0.000	0.000
	6	0.023	0.000	0.000	0.001	0.000	0.000
Panel C. EDF Tests. Statistic and Significance.							
Kolmogorov-Smir.		0.875	0.761	0.643	1.273	0.995	1.384*
Kuiper		1.520*	1.163	1.120	1.700*	1.288	1.787*
Anderson-Darling		1.039	0.996	1.158	1.929*	1.933*	3.274**
Panel D. Wallis Tests. P-values of UC, IN, and CC.							
3 Bins	UC	0.046	0.796	0.292	0.294	0.213	0.156
	IN	0.034	0.154	0.025	0.099	0.029	0.137
	CC	0.011	0.309	0.034	0.115	0.031	0.098
4 Bins	UC	0.587	0.972	0.826	0.116	0.381	0.056
	IN	0.018	0.536	0.447	0.230	0.035	0.355
	CC	0.039	0.768	0.634	0.128	0.049	0.131

Table 4. Estimation of Markov-Switching VAR. Posterior Mean of Some Parameters.

The Markov-switching VAR model is defined by (5), (8), and (9). The posterior means of some parameters are shown for estimations at December 1974 and November 2001. For each date, results are shown for the predictable and constant mean versions of the model. First, the transition probabilities are shown. In the case of the predictable mean version, the first parameters are the slopes of the excess stock return (ESR) and the excess bond return (EBR) with respect to the dividend yield (DY) and the term premium (TP). In the second version, the intercepts are shown. Finally, the residual variances and covariances of ESR and EBR are shown for each regime.

		1974/12		2001/11	
		Predictable	Constant	Predictable	Constant
p_{11}		0.932	0.939	0.951	0.943
p_{22}		0.928	0.937	0.838	0.838
$\mathbf{\Pi}$	ESR,DY	2.791	0.339	0.677	0.600
	ESR,TP	0.965		0.383	
	EBR,DY	0.370	-0.050	-0.052	0.083
	EBR,TP	0.155		0.190	
$\mathbf{\Sigma}_1$	ESR,ESR	11.367	12.000	13.470	12.934
	ESR,EBR	-0.321	-0.315	0.726	0.719
	EBR,EBR	0.352	0.345	1.036	1.030
$\mathbf{\Sigma}_2$	ESR,ESR	20.251	22.655	36.058	38.133
	ESR,EBR	0.762	1.063	1.750	2.364
	EBR,EBR	2.665	2.690	5.936	5.921

Table 5. PIT's of Markov-Switching VAR.

The Markov-switching VAR model is defined by (5), (8), and (9). The PIT's $d_{t+1}^{f,S}$, $d_{t+1}^{f,B}$, and $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$ are studied for the predictable and constant mean versions of the model. Panel A shows four descriptive statistics of each PIT. The values of these statistics are 0.50, 0.29, 0.00, and -1.20 for a uniform distribution. Panel B studies the autocorrelation of each PIT by means of the Ljung-Box test. Panel C shows some empirical distribution function (EDF) tests, where the statistic is shown with $*$ if it is significant at the 5%, $**$ if it is significant at the 2.5%, and $***$ if it is significant at the 1%. Finally, Panel D shows the Wallis tests of unconditional coverage (UC), independence (IN), and conditional coverage (CC).

		Predictable Mean			Constant Mean		
		ESR	EBR	Joint	ESR	EBR	Joint
Panel A. Descriptive Statistics.							
Mean		0.504	0.518	0.509	0.522	0.536	0.528
Stan. Deviation		0.299	0.304	0.301	0.288	0.305	0.296
Coef. of Skewness		-0.068	-0.089	-0.071	-0.182	-0.148	-0.158
Coef. of Kurtosis		-1.341	-1.343	-1.330	-1.205	-1.329	-1.267
Panel B. Ljung-Box Test. P-values of 1, 3, and 6 Lags.							
Level	1	0.169	0.003	0.120	0.988	0.001	0.006
	3	0.258	0.028	0.302	0.875	0.013	0.050
	6	0.090	0.097	0.245	0.719	0.040	0.103
Second Power	1	0.950	0.752	0.584	0.426	0.613	0.183
	3	0.212	0.305	0.063	0.168	0.630	0.124
	6	0.020	0.397	0.009	0.004	0.629	0.007
Panel C. EDF Tests. Statistic and Significance.							
Kolmogorov-Smir.		0.965	1.259	1.282	1.242	1.601*	1.782**
Kuiper		1.632*	1.817*	2.013**	1.717*	1.798*	2.009**
Anderson-Darling		1.023	1.828*	1.957*	1.676*	3.982***	4.107***
Panel D. Wallis Tests. P-values of UC, IN, and CC.							
3 Bins	UC	0.137	0.068	0.120	0.133	0.006	0.002
	IN	0.033	0.238	0.054	0.547	0.253	0.265
	CC	0.025	0.092	0.035	0.311	0.015	0.007
4 Bins	UC	0.118	0.011	0.012	0.079	0.006	0.006
	IN	0.012	0.372	0.280	0.062	0.003	0.043
	CC	0.008	0.051	0.039	0.028	0.000	0.003

C Figures

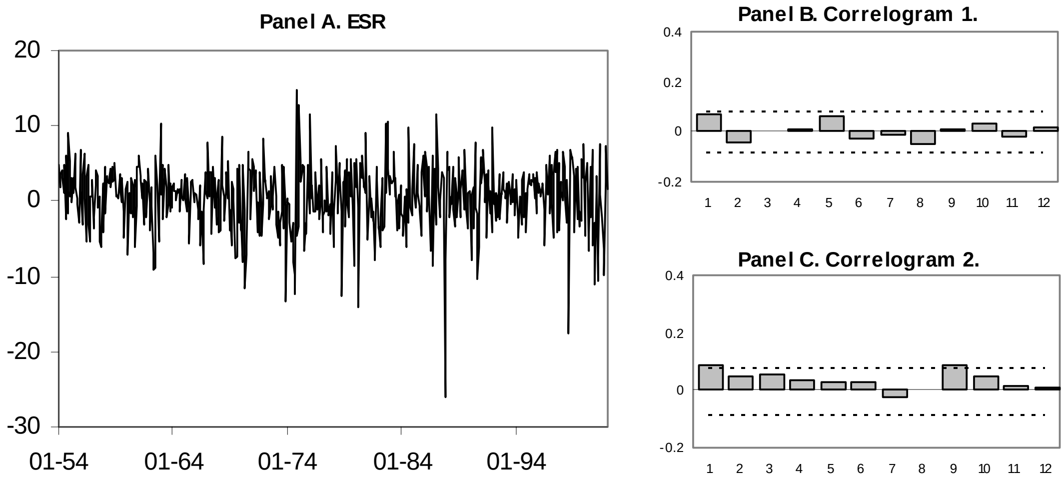


Figure 1. Time Series and Correlograms of Excess Stock Return. Panel A shows the time series of ESR in monthly percentage. Panel B shows the autocorrelation of the level of the series from the 1^{st} to the 12^{th} lag jointly with 5% critical values (dotted lines). Panel C shows the autocorrelation of the second power of the demeaned series.

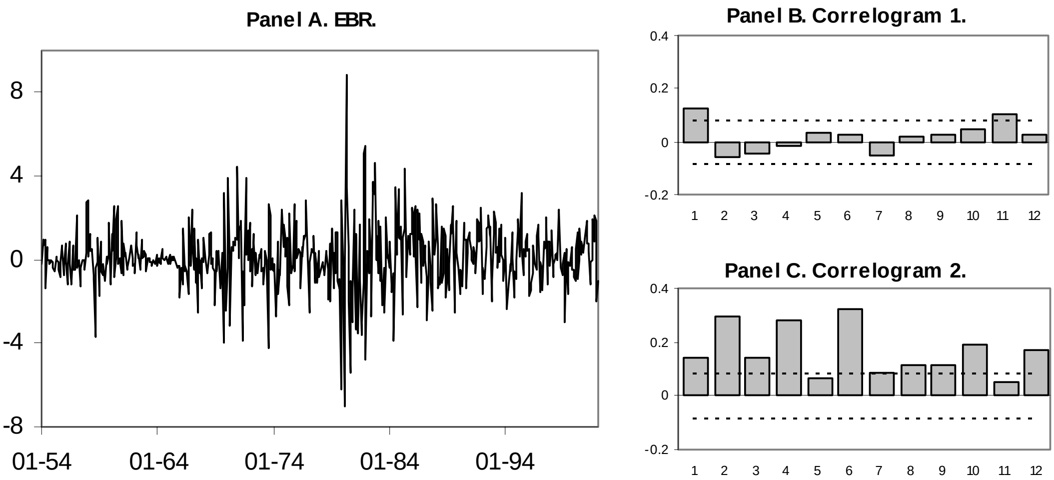


Figure 2. Time Series and Correlograms of Excess Bond Return. Panel A shows the time series of EBR in monthly percentage. Panel B shows the autocorrelation of the level of the series from the 1^{st} to the 12^{th} lag jointly with 5% critical values (dotted lines). Panel C shows the autocorrelation of the second power of the demeaned series.

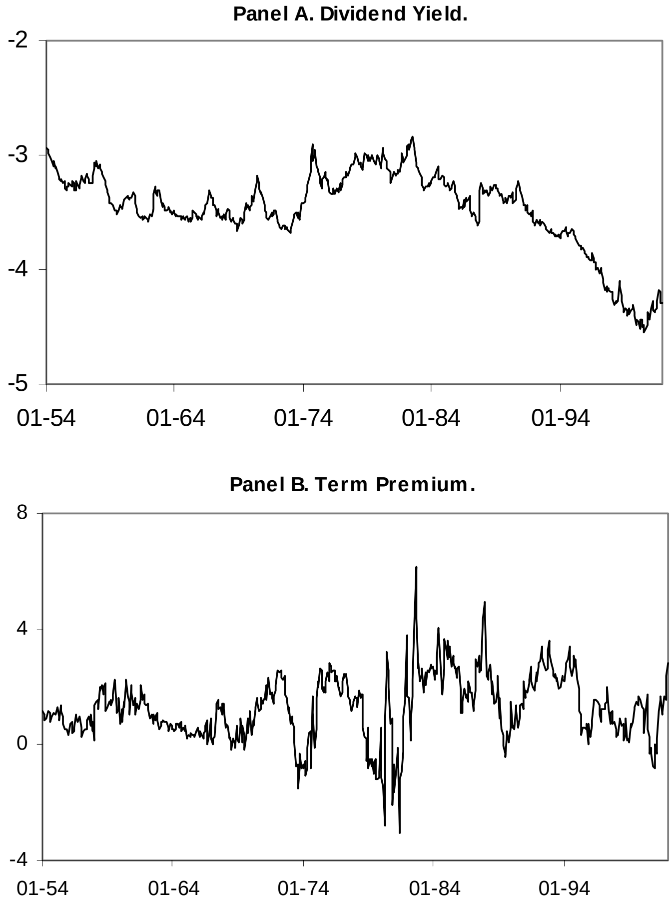


Figure 3. Time Series of Predictors. Panel A shows the time series of the dividend yield. Panel B shows the time series of the term premium.

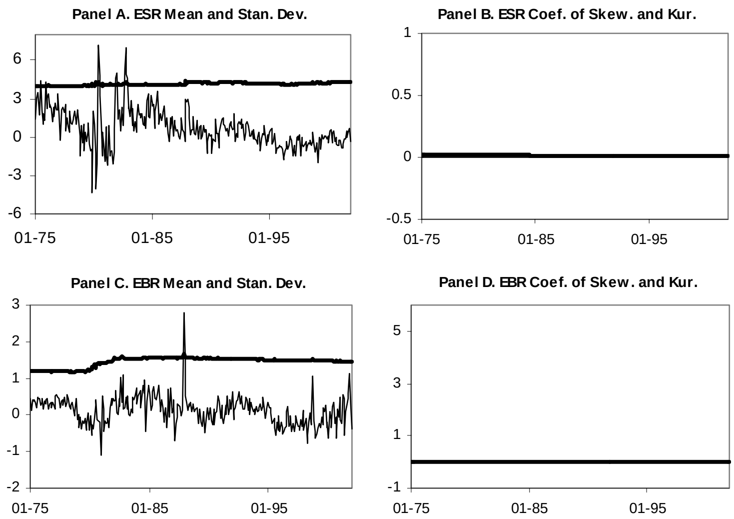


Figure 4. Density Forecasts from Normal VAR with Predictable Mean. Panel A shows the mean and standard deviation (thick line) of ESR. Panel B shows the coefficients of asymmetry and kurtosis (thick line) of ESR. Panel C shows the mean and standard deviation (thick line) of EBR. Panel D shows the coefficients of asymmetry and kurtosis (thick line) of EBR.

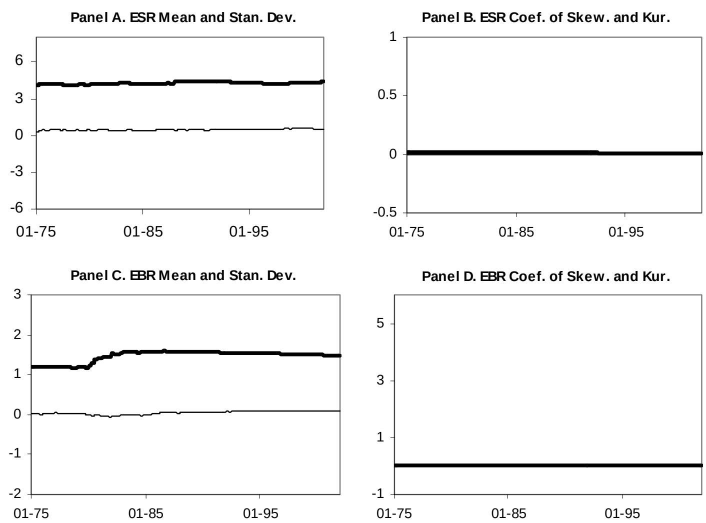


Figure 5. Density Forecasts from Normal VAR with Constant Mean. Panel A shows the mean and standard deviation (thick line) of ESR. Panel B shows the coefficients of asymmetry and kurtosis (thick line) of ESR. Panel C shows the mean and standard deviation (thick line) of EBR. Panel D shows the coefficients of asymmetry and kurtosis (thick line) of EBR.

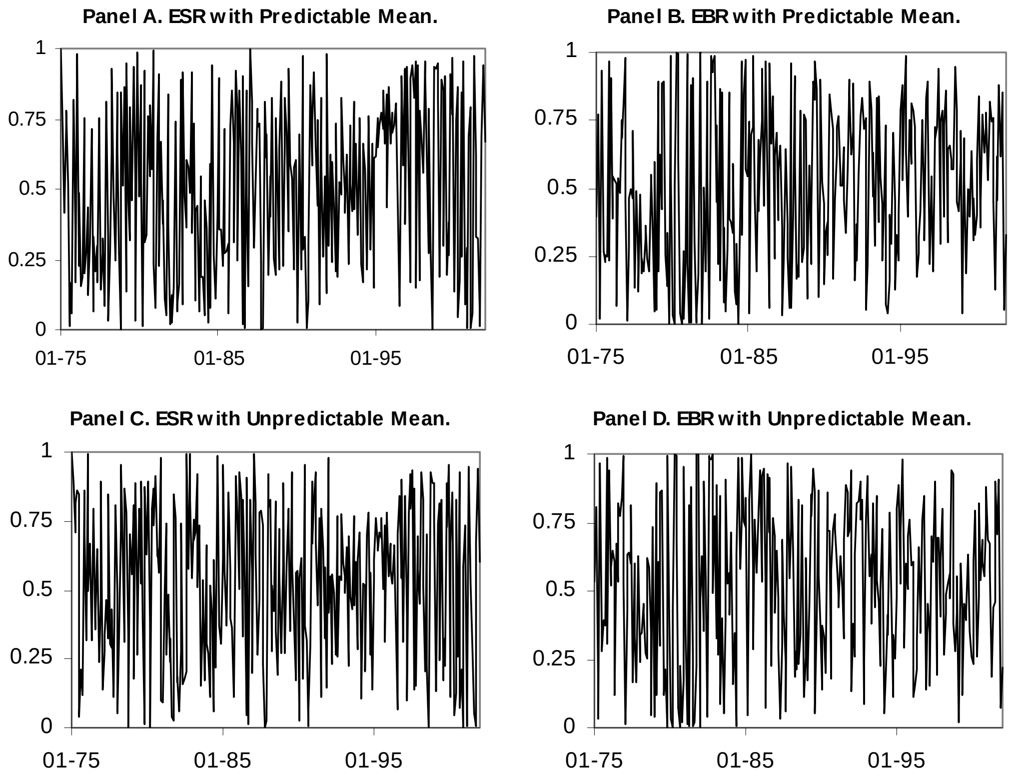


Figure 6. PIT's from Normal VAR. Panel A shows the time series of ESR's PIT with predictable mean. Panel B shows the time series of EBR's PIT with predictable mean. Panel C shows the time series of ESR's PIT with unpredictable mean. Panel D shows the time series of EBR's PIT with constant mean.

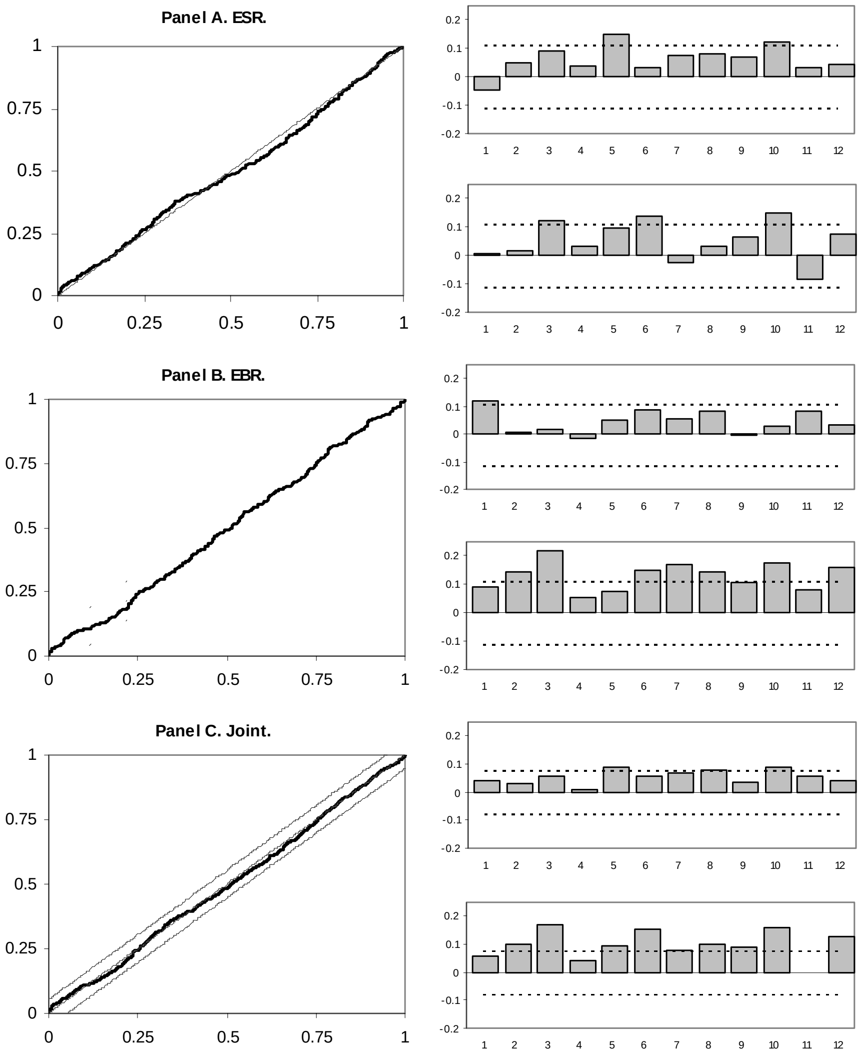
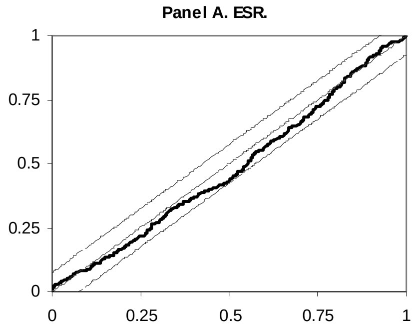
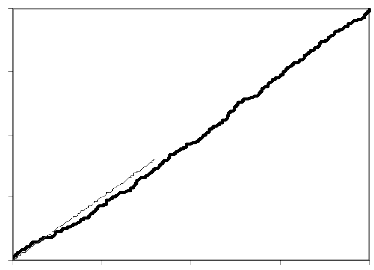


Figure 7. CDF and Correlograms of PIT's from Normal VAR with Predictable Mean. The left figure in each panel is the CDF of a particular PIT (thick line) jointly with Kolmogorov-Smirnov 5% critical values. The two figures on the right show the autocorrelation of the level (top figure) and the second power of the demeaned PIT from the 1^{st} to the 12^{th} lag jointly with 5% critical values (dotted lines). Panel A shows those figures for the ESR's PIT $d_{t+1}^{f,S}$, Panel B for the EBR's PIT $d_{t+1}^{f,B}$, and Panel C for the joint PIT $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$.


Panel A. ESR.
1
0.75
0.5
0.25
0
0
0.25
0.5
0.75
1



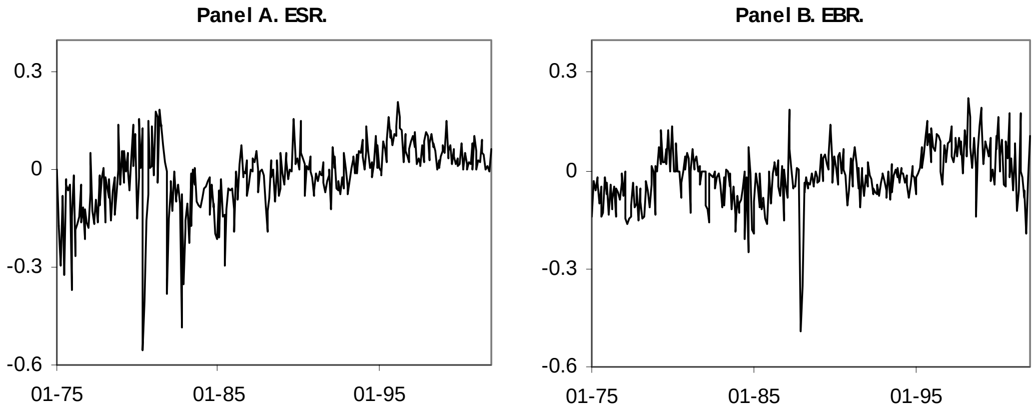


Figure 9. Difference between PIT's of Predictable and Constant Mean versions of Normal VAR. The PIT's are computed from the integrated forecasting rules. Panel A shows the difference for the ESR's PIT. Panel B shows the difference for the EBR's PIT.

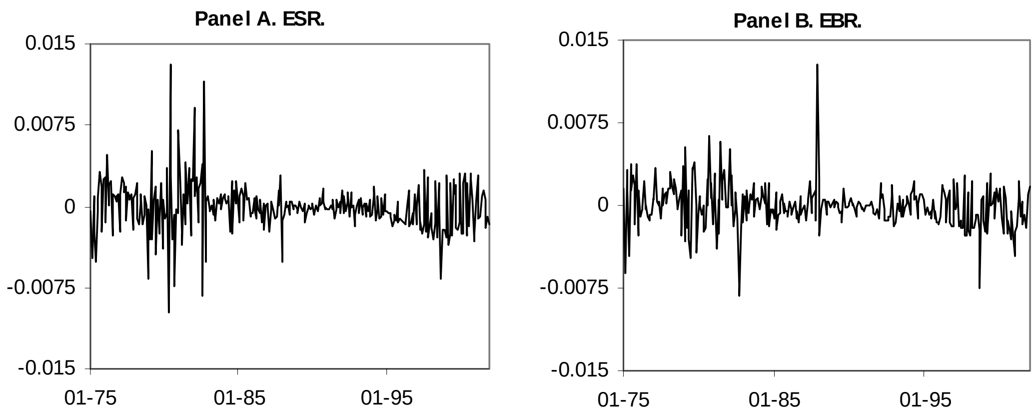


Figure 10. Difference between PIT's of Integrated and Plug-in Forecasting Rules of Normal VAR. The PIT's are computed from the predictable mean version. Panel A shows the difference for the ESR's PIT. Panel B shows the difference for the EBR's PIT.

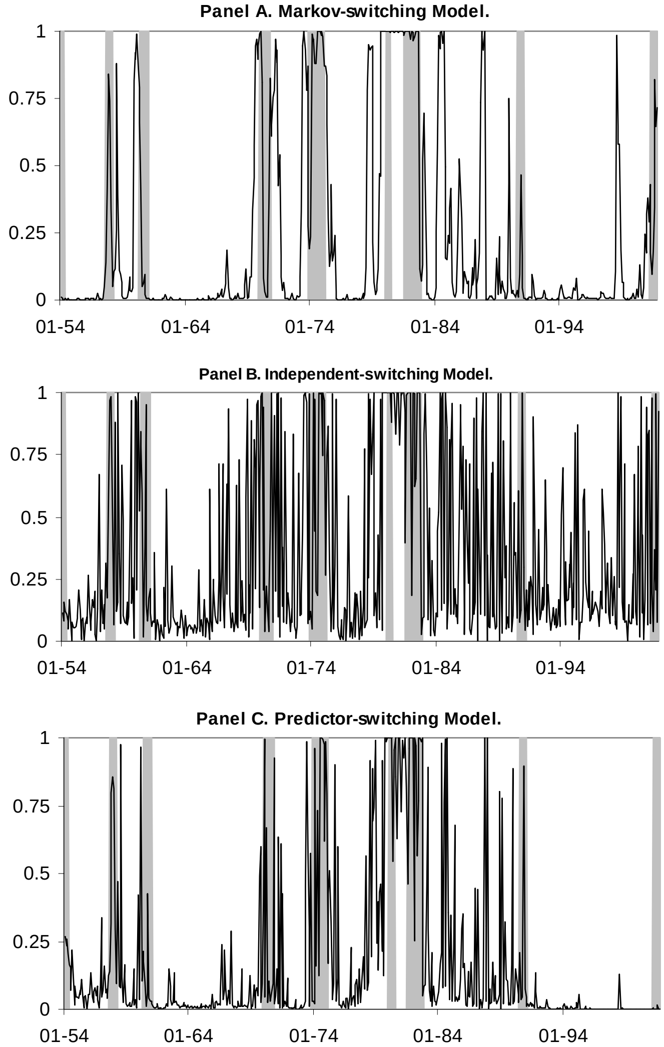


Figure 11. Probability of High-Volatility Regime and NBER Recession Index. The shaded regions are labelled as recessions by the NBER. The probabilities are taken from the last estimation, November 2001, of the predictable mean version of each model. Panel A shows the probability from the Markov-switching VAR, Panel B from the Independent-switching VAR, and Panel C from the Predictor-switching VAR.

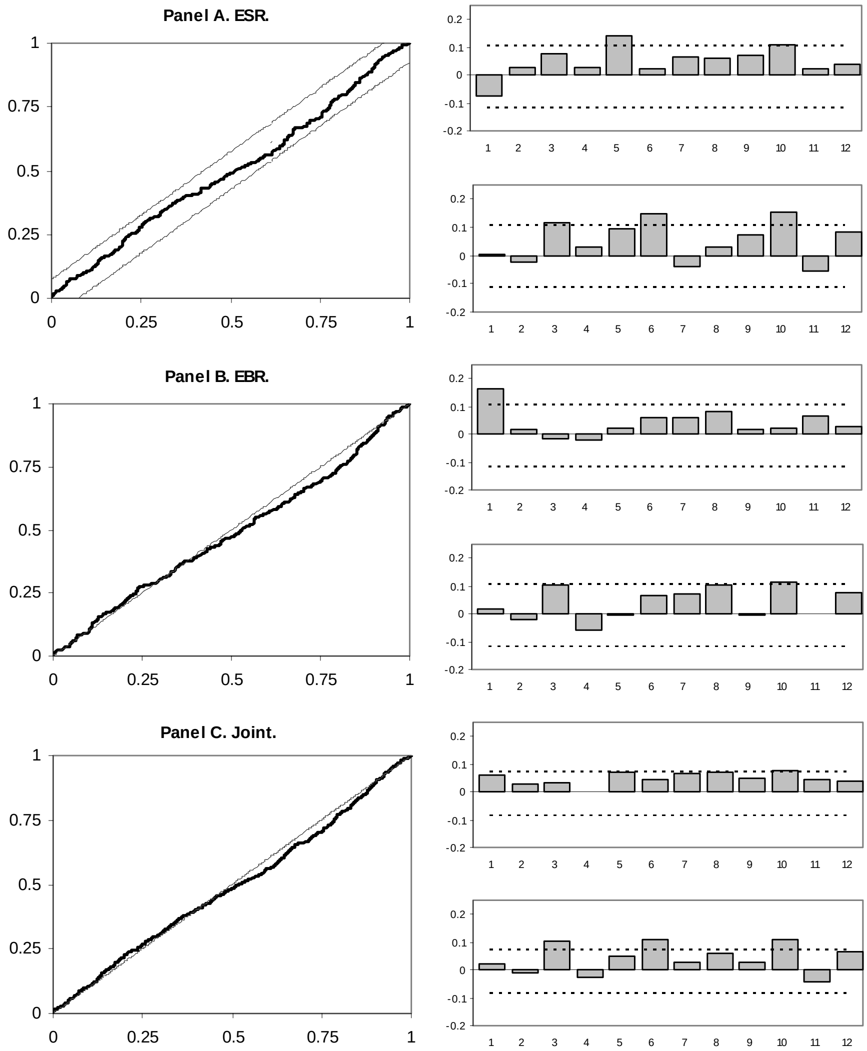


Figure 14. CDF and Correlograms of PIT's from Markov-Switching VAR with Predictable Mean. The left figure in each panel is the CDF of a particular PIT (thick line) jointly with Kolmogorov-Smirnov 5% critical values. The two figures on the right show the autocorrelation of the level (top figure) and the second power of the demeaned PIT from the 1^{st} to the 12^{th} lag jointly with 5% critical values (dotted lines). Panel A shows those figures for the ESR's PIT $d^{f,S}_{t+1}$, Panel B for the EBR's PIT $d^{f,B}_{t+1}$, and Panel C for the joint PIT $\left(d^{f,S}_{t+1}, d^{f,B|S}_{t+1}\right)$.

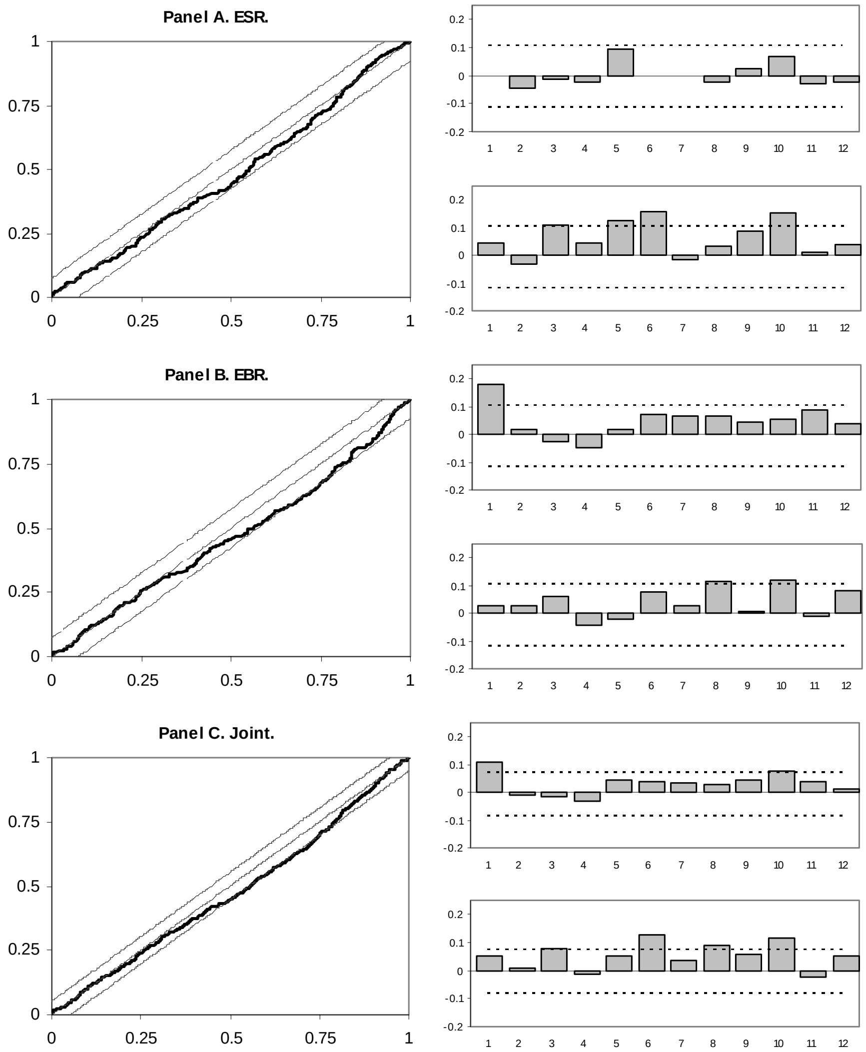


Figure 15. CDF and Correlograms of PIT's from Markov-Switching VAR with Constant Mean. The left figure in each panel is the CDF of a particular PIT (thick line) jointly with Kolmogorov-Smirnov 5% critical values. The two figures on the right show the autocorrelation of the level (top figure) and the second power of the demeaned PIT from the 1^{st} to the 12^{th} lag jointly with 5% critical values (dotted lines). Panel A shows those figures for the ESR's PIT $d_{t+1}^{f,S}$, Panel B for the EBR's PIT $d_{t+1}^{f,B}$, and Panel C for the joint PIT $\left(d_{t+1}^{f,S}, d_{t+1}^{f,B|S}\right)$.