Evaluating Predictability of Stock and Bond Returns with Different Density Forecasts[1]

Francisco Peñaranda
(CEMFI)

January 2002
Preliminary
Please do not quote without permission

[1] I would like to thank Manuel Arellano, Víctor López, Gabriel Pérez-Quirós, seminar participants at CEMFI, and specially Enrique Sentana, for very useful comments and suggestions. Of course, the usual caveat applies. Address for correspondence: CEMFI, Casado del Alisal 5, 28014 Madrid, Spain (tel.: +34 91 429 0551, fax: +34 91 429 1056, e-mail: francisco.pt@cemfi.es).

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.

But there is still high uncertainty about the true extent of predictability. This paper develops a careful evaluation of predictability of returns in a relevant context for portfolio management. This involves three dimensions. First, multiple assets must be included in the analysis. In this paper, monthly U.S. excess returns of bonds and stocks are jointly studied. Second, density forecasts of returns are required, not only point forecasts. Predictive distributions in a Bayesian framework are used to take into account parameter uncertainty. Third, conclusions about predictability are drawn from evaluation of density forecasts using out-of-sample checks against realizations of returns. Density forecasts are computed for normal VAR, fat tails and Markov-switching models.

Keywords: Density forecasts, Return predictability, Forecast evaluation.

JEL: G11, C53.

1 Introduction

One of the most important findings in empirical finance has been the fact that returns are not i.i.d. Predictability, which means time variation in the conditional distribution of returns, has become one of the basic ingredients of asset pricing and portfolio choice models nowadays. There are theoretical models such as the habit formation in Campbell and Cochrane (1999) or, more recently, the model with a stochastic share of labour income on consumption in Santos and Veronesi (2001) that are designed to capture time variation in risk premia.

Similarly, papers such as Campbell, Chan and Viceira (2001) incorporate predictability in the portfolio choice of an investor to reflect time-varying investment opportunities. This concept was introduced by Merton (1973), who developed the Intertemporal CAPM to relate predictability with portfolio choice and asset pricing. This model has been recovered nowadays as the reference model in portfolio management.

But there is still high uncertainty about the true extent of predictability. Campbell, Chan and Viceira (2001) 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. 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, given the weak statistical results in terms of predictive regressions. 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 density forecasts for an optimal asset allocation.

Those three papers are based on stock returns and the assumption of normality. This paper develops a careful evaluation of predictability of returns in a relevant context for portfolio management, which is defined in three dimensions.

First, multiple assets must be included in the analysis. In this paper, monthly U.S. excess returns of bonds and stocks are jointly studied in the post-war period. Since the conditional joint distribution is modelled, time-varying portfolios of bonds and stocks could be studied in this framework.

Second, density forecasts of returns are required. An investor needs density forecasts, not only first and second moments unless some restrictions are imposed on her preferences or the distribution of returns. This is clear in modern models of financial risk such as the Value-at-Risk, where interval forecasts are computed to quantify the risk of a portfolio. See for instance Berkowitz (2001). A Bayesian framework and its corresponding predictive distributions are used to take into account parameter uncertainty in the computation of density forecasts.

Third, conclusions about predictability are drawn form evaluation of density forecasts with out-of-sample checks against realizations of returns. Evaluation of density forecasts is a very active area nowadays. Diebold, Gunter and Tay (1998) advocate the use of graphical methods to study model failures and reveal useful information for the development of new models.[1] This is the approach that is going to be used is in this paper and therefore the results that are going to be shown are more qualitative than quantitative.

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. Density forecasts are computed for different models apart from the normal since this assumption can contaminate the conclusions about predictability. Two main extensions are considered: fat tails and non-linear dynamics. The former, also known as outlier models, are represented in this paper by the use of the Student-t distribution. This distribution has been already used in a Bayesian framework with stochastic volatility models applied to high-frequency data. Recent papers are Jaquier, Polson and Rossi (2001) and Chib, Nardari and Shephard (2001). The latter is going to be implemented by means of regime 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.[2] Kim and Nelson (1998) is the first implementation in a Bayesian framework to macroeconomic indicators. In finance, Pérez-Quirós and Timmermann (2000a and 2000b) and Chauvet and Potter (2001) estimate a switching model of stock returns in a classical context. The main estimation tool for these type of models in a Bayesian framework is Monte Carlo Markov chain methods jointly with data augmentation.

The focus of the literature that studies predictability in low-frequency data has been the conditional mean. Given that monthly data are going to be used in this paper, two versions of each of the previous three models will be estimated and evaluated, one with a time-varying conditional mean and another with a constant one. This paper shows that the use of predictors, where predictors are taken as the state variables that define the conditional mean of returns, produces different density forecasts as they vary much more over time than in a model with unpredictable mean. This translates into significant effects on portfolio choice as it is well known in the literature on predictability. However, the development of out-of-sample checks gives some new insights about the use of predictors. Using a model with a time-varying conditional mean in the computation of density forecasts does not improve the forecast evaluation and gives too many realizations on the left tail of the density forecast with respect to a model with unpredictable mean. Predictors are too optimistic and this can be a problem if the investor is concerned about losses. Recall that this is the main concern in Value-at-Risk models.

Some conclusions can also be drawn in terms of the different models that are implemented. Compared to the normal model, the Student-t does not improve the forecast evaluation of bonds and stocks. The Markov-switching improves the forecast evaluation of bonds since it captures better the dynamics than the normal model. However, there is no improvement for stocks with Markov-switching. It can be seen that the development of a joint model of stocks and bonds is a difficult task, but it is the necessary object for

[1] On the other hand, a parametric approach is proposed in Berkowitz (2001). See also Hong (2001) for a recent implementation of this type of analysis to high-frequency data.

[2] See Diebold (1998).

portfolio management.

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 computation of density forecasts in a Bayesian framework and their evaluation. Section 3 explains and applies the standard normal model, while sections 4 and 5 are devoted to its extensions. The first covers fat tails by means of the Student-t, while the second covers nonlinear dynamics by means of regime switching. Finally, section 6 concludes. The different Gibbs samplers, on which the estimations are based, are gathered in the appendix.

2 Density Forecasts of Returns

This section introduces the basic concepts and tools that will be used in this paper. There is a brief introduction to the literature on predictability and a description of the computation and evaluation of density forecasts of returns.

2.1 Predictability of Returns

As it has been commented in the introduction, one of the most important findings in empirical finance has been the fact that returns are not i.i.d. The conditional distributions of returns depend on past information and vary through time, that is, investment in financial assets is not like flipping coins over and over again. Merton (1973) developed the Intertemporal CAPM to relate predictability with portfolio choice and asset pricing. This model has been recovered nowadays as the reference model in portfolio management. The long run investor cares about shocks to investment opportunities, not only to wealth, when these opportunities are time-varying. This variation in interest rates and risk premia implies an optimal portfolio strategy that times the market, hedging against it with changes in the proportion of stocks and bonds.

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 (GARCH or stochastic volatility models), while studies using low-frequency data are mainly focused on first moments (predictive regressions or VARs). This paper uses monthly data and therefore starts modelling the conditional mean of returns with a VAR. However, both features are starting to be taken into account at both types of frequencies. Yong (2001) stresses the importance of modelling the conditional mean of high-frequency data which is usually forgotten. On the other hand, regime switching models as in Pérez-Quirós and Timmermann (2000a and 2000b) are implemented with monthly data, introducing time variation in the second moment of returns.

GARCH models are usually estimated in a classical framework. Anyway, there are some references in a Bayesian setting as Bawens and Lubrano (1998). On the other hand, stochastic volatility models are mainly implemented in a Bayesian framework. For instance, Uhlig (1997), Kim, Shephard, and Chib (1998) or Jaquier, Polson and Rossi (2001). The focus on second moments is very useful for high frequency data, but Hamilton and Susmel (1994) explore the issue of ARCH effects versus regime switching in variance in U.S. monthly data. They conclude that ARCH effects are not very important at the monthly frequency when regime switching is taken into account.

Anyway, the first focus in this paper will be the conditional mean of returns as it is the main focus in the literature that was commented in the introduction. Time variation in this conditional mean is taken nowadays as a stylized fact to capture with asset pricing models. A well-known approach is by means of habit formation as in Campbell and Cochrane (1999). A more recent approach is Santos and Veronesi (2001), where the cause of predictable returns is a stochastic share of labour income on consumption. Anyway, there are other explanations as simple investor irrationality and many statistical doubts remain. Those doubts will be explained later in this section.

The question of the usefulness of introducing predictors will be addressed in this paper, where predictors will mean state variables that define the conditional mean of returns. The baseline model in this paper is a VAR with returns and predictors since it is the usual way of introducing a time-varying conditional mean of returns. The most famous predictors can be classified into two categories. On the one hand, yield variables such as the dividend price ratio,[3] the price earning ratio,[4] etc. are common predictors of stock returns. A well-known example of predictor is the dividend price ratio, which has been a good predictor until the 90's. Its predictive power comes from mean reversion in stock returns. A high dividend yield must reflect low expected dividend growth or high expected returns and historically its variation has been associated with variation in expected returns.[5] An important feature of this variable is its high persistence. In fact, the usual unit root tests cannot be rejected, although it should be stationary if there are not bubbles, and expected returns and dividend growth are stationary. This may seem contradictory at a first sight. Returns are almost a random walk but its conditional expectation, represented here by the dividend yield, is very persistent. Fiorentini and Sentana (1998) show how to get this result in a multivariate process with certain restrictions. The dividend yield is one of the predictors that are used in this paper.

On the other hand, interest rates such as the term premium,[6] the short term rate,[7] the default premium,[8] etc. are used as predictors of stock returns, and of bond returns too. The first variable measures the slope in the term structure of interest rates and is used in this paper. Nowadays, new variables have been introduced into the picture: The consumption wealth ratio in Lettau and Ludvigson (2001a) and the share of labour income on consumption in Santos and Veronesi (2001). They represent some of the first successes of consumption in finance. Predictors are in general related to the business cycle and are very persistent. The consumption wealth ratio and the interest rate are good predictors for one year returns, while the dividend price ratio and the PER are

[3]See for instance Fama and French (1988 and 1989) or Campbell and Shiller (1988). The dividend yield is often computed using the sum of last year dividens to avoid seasonality.

[4]See for instance Lamont (1998).

[5]See Campbell (1991).

[6]See for instance Fama and French (1989) or Campbell and Shiller (1991).

[7]See for instance Fama and Schwert (1977) and Campbell (1987). An MA detrended series is usually computed to avoid nonstationarity.

[8]See for instance Fama and French (1988 and 1989) or Keim and Stambaugh (1986).

good for longer horizons, four or six years, in terms of R^2. These variables do not predict dividend growth, only excess or real returns.

Some recent papers about predictors are Ang and Bekaert (2001) and Cochrane and Piazzesi (2001). In the former, only short term rate is a robust predictor of excess stock returns, not only U.S. returns, and only for short horizons. They conclude that predictability is weak and only in the short run. But the short term rate is even more persistent that the dividend yield. In the latter paper, one year excess bond returns are predicted by a linear combination of forward rates.[9] They found strong predictability with R^2 of about 45% and stress that long horizons must be studied to realize it. But this predictor must be estimated.

The 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. 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, given the weak statistical results in terms of predictive regressions. They show that predictability is important in asset allocation even if the investor's prior is against predictability. Other papers in the same spirit are Stambaugh (1999) and Barberis (2000). Stambaugh studies the sensitivity of asset allocation to horizon and predictor effects and gives new insights about the impact of estimation risk. For instance, there can be an optimal decrease in the share of stocks after an increase in the dividend yield, although it signals higher expected returns, because it can induce negative skewness in the predictive distribution. In addition, Barberis studies portfolios of stocks and the safe asset and shows that there are still horizon effects after incorporating estimation risk. However, this risk lowers the increase in the optimal share of stocks with horizon and the sensitivity to the predictor.

Campbell, Chan and Viceira (2001) show the effects of adding predictors in the portfolio choice between stocks and bonds but in a classical framework. The hedging demands are driven by the correlation between shocks to the state variables that define the conditional distribution of returns and shocks to the returns themselves. In the case of stocks, this correlation is negative (stocks are mean-reverting) and therefore they have a positive hedging demand, while in the case of bonds it is positive (bonds are mean-averting), which implies a negative hedging demand.[10] This is the usual result, a long-run investor increase its position in stocks and decreases its position on bonds with respect to a short-run investor. They conclude that the state variable with the strongest effect on the portfolio choice is the dividend yield.

Now, the statistical doubts that still remain are going to be commented. In terms of classical predictive regressions and VARs, many doubts remain and there is still high uncertainty about the true extent of predictability. The predictors are clearly not exogenous and usually very persistent. At the same time, long-horizon returns are often used. Lack of exogeneity implies a nonstandard finite sample distribution of the parameters, while persistence and long-horizon imply a nonstandard asymptotic

[9] While in Fama and Bliss (1987), each bond was regressed on the corresponding forward.

[10] This is reinforced by the fact that the residual covariance between stocks and bonds is positive.

distribution.

The predictor is predetermined, not exogenous, which implies a nonstandard finite sample distribution under normality of returns, which has been derived by Stambaugh (1999). Some of its finite sample features are an upward bias, although consistency still holds, positive skewness and kurtosis in the case of stocks. In this respect, Campbell et al. (2001) comment that this relation between the biases in the predictive regressions and the signs of coefficient and the residual covariance implies different biases for stocks and bonds. In a regression of stock returns on the dividend yield, the slope coefficient is positive and the residual correlation is negative, which implies an upward bias. In a regression of bond returns on the term premium, the slope coefficient is positive and the residual correlation is positive, which implies a downward bias.

On the other hand, Lewellen (2001) arrives to the opposite conclusion for stocks, the predictability in stocks is stronger than predictive regressions show. By means of simulations, he shows a strong negative correlation between the slope in the predictive regression and the persistence of the predictor. Therefore, he proposes the use of the conditional distribution of the former given the latter instead of the unconditional one, which is almost normal and has a lower variance. This avoids taking into account levels of persistence that are not in the sample when making the bias correction. This adjustment is lower than Stambaugh's one, but it needs the assumption of knowing some parameters. Even in the most conservative setting against predictability when choosing the unknown parameters, he finds evidence of predictability.

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 are stationary. In this case, a big sample does not help. It is the well-known problem of spurious regressions applied to predictive regressions. In this context, the right-hand side variable is very persistent and the left-hand side variable is stationary but, if the expected return is very persistent and is an important component of the return, then a significant coefficient may show up although both persistent variables, the true conditional mean and the predictor, are independent.

Another problem is the use of long-horizon returns. The empirical results about predictability are often shown for long-horizon regressions of stock returns on the dividend yield, that is, accumulating monthly returns up to one year or more. A small short-run predictability jointly with the predictor's persistence imply that the R^2 increases with horizon, getting a maximum at four or six years. But there is a problem of small sample in this context since there are few nonoverlapping observations, as it happened in the univariate analysis, which translates also in nonstandard asymptotics. There are few available decades with increases and decreases in the dividend yield. Both finite sample problems, lack of exogeneity and long horizon returns, are explained in Nelson and Kim (1993). VAR methods can be used to avoid long run returns, and then long run predictions can be estimated by means of short run returns. Hodrick (1992) advocates the use of the VAR approach.

Turous and Yan (1999) underscore that uncertainty about integration plus the use of long horizon regressions imply the need of nonstandard asymptotics that depend on unknown nuisance parameters, which makes difficult to interpret the results about predictability. They focus on the effects of the uncertainty about the integration order of the predictor and the horizon, the problem of a few nonoverlapping periods. The former is formalized using a local-to-unity approach in a similar spirit to local alternatives in asymptotics and the latter is formalized as a ratio of the horizon and the sample size that does not tend to zero. After controlling for both effects, they conclude that there is no predictability.

Finally, there seems to be problems of parameter instability. Goyal and Welch (1999) face in-sample and out-of-sample success of predictive regressions and notice that the latter worsens. They associate this fact to parameter instability and fit a model with coefficients that are functions of time. This let them see a decrease in the predictive parameter and in its predictive power in the final 90's. A related work on out-of-sample success is Bossaerts and Hillion (1999). They explore statistical model selection criteria to select the best out-of-sample behaviour and avoid over-fitting. They use classical measures as adjusted R^2, the Akaike's criterion and the Schwarz's criterion, jointly with new ones and one of their own. The selected models have good in-sample behaviour, which would indicate predictability, but it worsens out-of-sample, which contradicts predictability. They conclude that there is a problem of nonstationarity in the model so that time-varying parameters should be used[11].

To sum up, the need of nonstandard distributions depending on unknown parameters translates into high uncertainty about the interpretation of results from predictive regressions. This paper tries to shed some light about predictability using a different framework and different tools.

2.2 Data on Returns and Predictors

This paper uses monthly nominal data from January 1954 to December 1994, which gives 492 observations.[12] 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 1970, which gives 300 observations for this out-of-sample check. 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, returns R^S_{t+1} are based on the value weighted NYSE stock index (CRSP tapes). They are converted into continuously compounded rates in monthly percentage, $r^S_{t+1} = 100 \ln\left(1 + R^S_{t+1}\right)$. The other ingredient is a short 1-month T-bill rate from the Fama-Bliss risk-free rates file (CRSP tapes), with yield $Y_t^{(1)}$. The notation $Y_t^{(n)}$ means the yield of a discount bond that pays off in n months. Again, the series is converted into

[11] Another possibility is nonlinear models.

[12] Stock market data and the safe asset are the same as the data in Pérez-Quirós and Timmerman (1996). Bond market data are the same as in Cochrane and Piazzesi (2001).

continuously compounded rates in monthly percentage, $r_{t+1}^F = y_t^{(1)} = 100 \ln \left(1 + Y_t^{(1)}\right)$. Finally, the excess stock return is given by $er_{t+1}^S = r_{t+1}^S - r_{t+1}^F$.

It can be observed the huge fall in October 1987 in figure 2.1. In table 2.1, there are some descriptive statistics of the stock and bond excess returns. The mean is 0.48%, while the standard deviation is 4.17%, which gives an annualized Sharpe ratio of about 0.4. The skewness is -0.55 and the kurtosis 5.90, that is, asymmetry to the left and fat tails. The Jarque-Bera test clearly rejects the hypothesis of normality. To study temporal dependence, the autocorrelogram of powers of the demeaned series are shown in figure 2.2. The level does not show autocorrelation, only the fifth lag is marginally significant, neither the third and forth powers, while it can be found on the second power, from the first to the fifth lag and from the ninth to the eleventh in terms of the Ljung-Box test.[13] However, some of the correlation that is found in even powers of the stocks is due to two opposite extreme values in the series with 9 periods between them. One is the 12.25% on January 1987 and the other is the -24.75% on October 1987. Anyway, the correlation in the square will be much clearer for bonds.

On the other hand, excess bond returns are computed from returns of the 5 year discount bond from the Fama-Bliss discount bond yields (CRSP tapes). Using the previous notation of discount bonds, $y_t^{(n)} = 100 \ln \left(1 + Y_t^{(n)}\right)$ and the corresponding bond return is $r_{t+1}^B = 60 y_t^{(60)} - 59 y_{t+1}^{(59)}$, where $y_{t+1}^{(59)}$ is approximated by $y_{t+1}^{(60)}$. This is the only approximation needed here since the bonds do not have coupons. Finally, the excess bond return is $er_{t+1}^B = r_{t+1}^B - r_{t+1}^F$.

The series, figure 2.3, shows a huge increase in the volatility at the beginning of the 80's due to the commented change in the monetary policy. The mean is 0.05%, while the standard deviation is 1.72%, which gives an annualized Sharpe ratio of about 0.1, four times lower than the stocks. The skewness is 0.25 and the kurtosis 7.00, asymmetry of the opposite sign to stocks and even fatter tails. The Jarque-Bera test clearly rejects the hypothesis of normality. As it can be seen in figure 2.4, there is some autocorrelation in the level of the series from the eleventh lag, while the first lag is only marginally significant. There is a clear dependence in the second power at any lag in terms of the Ljung-Box test.[14] The third power does not show autocorrelation for the first lag, but it can be found in the rest of lags. Specifically, it is significantly negative for the second and sixth lags. The forth power shows also correlation from the second lag. The second, forth and sixth lags are significantly positive. There is a peak in these correlograms at the 19^{th} lag due to the two maxima of the series in April 1980, 8.84%, and November 1981, 9.08%. Since the sign is common, they appear in the correlograms of all the powers. The strongest evidence of time dependence is in the square of the bonds and this will become a key feature in the evaluation of density forecasts in the following sections.

Two predictors are used, the dividend yield and the term premium. The dividend yield is derived from the value weighted NYSE returns with and without dividends

[13] The absolute value does not show autocorrelation at the first lag, but it is even stronger for the rest of lags.

[14] Also in the absolute value

(CRSP tapes), $dy_{t+1} = \ln(1 + DY_{t+1})$. It can be seen in figure 2.5 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 (Fama-Bliss discount bond yields) and the yield of the 1-month T-bill (obtained from the Fama-Bliss risk-free rates file), $tp_{t+1} = 12\left(y_{t+1}^{(60)} - y_{t+1}^{(1)}\right)$, where it is annualiezed to get values of similar order to dy_{t+1}. This series, see figure 2.6, is also persistent but not as much as the divided yield.

These four series will be stacked together in a VAR as the following vector,

$$y_{t+1} = \begin{pmatrix} r_{t+1} \\ x_{t+1} \end{pmatrix} = \begin{pmatrix} er_{t+1}^{S} \\ er_{t+1}^{B} \\ dy_{t+1} \\ tp_{t+1} \end{pmatrix}.$$

This is similar to the VAR in Campbell, Chan and Viceira (2001).[15] 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.

2.3 Computation of Density Forecasts

As it has been commented in the introduction, a Bayesian framework is going to be used. See Geweke (1998) for a description of the following methods. In this paper, density forecasts of future returns given past returns and predictors are the object of interest. This joint distribution of stocks and bonds will be denoted as $p\left(r_{T+1} \mid y^{T}\right)$. We condition onto the past as we are predicting, not on unknown parameters and the interest is not on the sampling variability in the observed past. This is the way investors construct their beliefs about future returns since they do not know their distribution[16] and this question is naturally addressed in a Bayesian setting with the concept of predictive distribution.

A Bayesian model is constructed with a prior and a likelihood. Both define a posterior as[17]

$$p\left(\theta \mid y^{T}\right) \propto p(\theta)\, p\left(y^{T} \mid \theta\right)$$

where $y^{T} = \{y_t\}_{t=1}^{T}$ represents the data, returns and predictors in this paper, and θ is the vector of parameters. This is only a definition of the joint distribution of the observables, y^{T}, and unobservables, θ.

It is not very important that the likelihood is the "true" one or not, surely it will be false. The posterior enables to compute our main goal, the predictive distribution, which deals only with observables

$$p\left(y_{T+1} \mid y^{T}\right) = \int_{\Theta} p\left(y_{T+1} \mid y^{T}, \theta\right) p\left(\theta \mid y^{T}\right) d\theta$$

[15] Other papers using VAR's are Kandel and Stambaugh (1996), Campbell (1991), Hodrick (1992) or Barberis (2000).

[16] Note the different context with respect to other questions such as the estimation of a coefficient of risk aversion. In that context, the agent knows the parameter, but the econometrician not.

[17] The notation $p(\cdot)$ will represent density or mass depending on the associated random variable.

where the parameters θ are integrated out and the first two entries of y_{T+1} are the stock and bond excess returns, the object of interest. This is how the density forecasts are computed in this paper.

This is the way to take into account estimation risk by means of the posterior, that is, we state our uncertainty about a forecast by including parameter uncertainty and taking the observed data as given. The important point is if it gives an accurate forecast or not. For instance, Uhlig (1994) stresses that pretesting for unit roots and proceeding as if one is sure about its conclusion can be misleading in calculating the uncertainty with regard to density forecasts.

It is important to develop a sensitivity analysis with respect to the prior, to give an idea of the mapping from priors to posteriors for a given likelihood. A wide-spread prior is Jeffreys prior, which is invariant to reparametrizations. It is usually taken as uninformative because it often gives results similar to the classical ones. Therefore, it provides a good benchmark. Its kernel is given by the square root of the determinant of the information matrix so that it does not integrate to one. Anyway, proper priors are going to be used in this paper. It is better to use proper priors with regime switching models and therefore they will be used for the rest of the models for continuity. But those priors will not be too informative.

When complex models are implemented, there are not closed solutions in terms of posterior distributions and simulation-based methods have to be used, specifically Markov Chain Monte Carlo (MCMC) methods.[18] They have been a key part of the development of Bayesian statistics in the last years. The Gibbs sampler is very popular among the MCMC methods and will be use here for its simplicity. See for instance Gelfand and Smith (1990) or Smith and Roberts (1993). Many times, we do not know how to sample from the posterior, but we can define a block structure with the parameters, for instance two blocks, $\theta = \left(\theta_1', \theta_2'\right)'$, such that we can draw from $p\left(\theta_1 \mid y^T, \theta_2\right)$ and $p\left(\theta_2 \mid y^T, \theta_1\right)$. This structure defines a Markov chain with the desired invariant measure $p\left(\theta_1, \theta_2 \mid y^T\right)$, and therefore it will converge to that invariant measure, our goal, if the chain is ergodic. That is, if we sample many times from the chain we know that in the limit we are sampling form the desired distribution. The conditions of convergence should be used or at least we should check the convergence of the simulations.[19]

Data augmentation is a useful tool when we do not know how to sample from the posterior of θ, but we know how to do it from the posterior of $\left(\theta, \lambda^T\right)$, adding nuisance parameters λ^T to the sampler. See Tanner and Wong (1987). For instance, we can implement it in terms of a Gibbs sampler if we know how to draw from $p\left(\theta \mid y^T, \lambda^T\right)$ and $p\left(\lambda^T \mid y^T, \theta\right)$. Then, if the chain is ergodic, we will be sampling from $p\left(\theta, \lambda^T \mid y^T\right)$ in the limit. Finally, looking only to the realizations of θ, we have the desired distribution from that marginal. Shephard (1994b) shows how to go from non-Gaussian models to a Gaussian state-space context, where their are lots of well-known results to work with, by means of data augmentation. He calls this general setting as "Partial Non-Gaussian State Space". He gives some examples such as outlier and Markov-switching models.

[18] See Chib and Greenberg (1996).

[19] See Tierney (1994).

These type of models will be studied in this paper.

The practical implementation of these methods is the following. The different Gibbs samplers, which are developed in the appendix, give draws $\{\theta_j\}_{j=B+1}^{I}$ from the posterior $p\left(\theta \mid y^T\right)$ in the limit, where we are discarding the first $\{\theta_j\}_{j=1}^{B}$ draws. Then, drawing from $p\left(y_{T+1} \mid y^T, \theta_j\right)$ will give a sample $\{y_{T+1,j}\}_{j=B+1}^{I}$ which constitutes an approximation to the predictive $p\left(y_{T+1} \mid y^T\right)$. The first two entries are the excess returns and approximate $p\left(r_{T+1} \mid y^T\right)$.

2.4 Evaluation of 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. On the other hand, we can make a relative comparison between pairs of models with the usual posterior odds analysis in a Bayesian framework. In this paper, the former approach is taken.

Conclusions about predictability are drawn from evaluation of density forecasts with out-of-sample checks against realizations of returns. The goal of this paper is not to find a model of density forecasts but to study the quality of adequacy of the current models. Nevertheless, this analysis reveals useful information for developing new models. The concern here is about accurate predictions not model misspecification, that is, this paper is not a search for the "true" likelihood of the data. If our concern is in-sample success we can end up overfitting the data, and we will likely lose predictive power.

Evaluation of density forecasts is a very active area nowadays. Anyway, it is based on an old result by Rosenblatt (1952) about the Probability Integral Transform (PIT) of the density forecasts evaluated in the realizations of the returns. The sample period $t = 1, ..., T_1$ can be split in two parts. The first part, $t = 1, ..., T_0$, can be used for a first estimation and the second one, $t = T_0 + 1, ..., T_1$, will be added observation by observation to the previous data set to estimate recursively the model and construct the corresponding density forecast of returns.

After the estimation of a Bayesian model, we can compute the corresponding sequence of one-step ahead predictive distributions $\left\{p\left(r_{t+1} \mid y^t\right)\right\}_{T_0}^{T_1-1}$. In addition, the realizations of returns for that period are $\{r_{t+1}\}_{T}^{T^*-1}$. The joint distribution of stocks and bonds is computed since it is the necessary object for portfolio management. However, the current evaluation methods are designed for univariate density forecasts.

If the c.d.f. of each return predictive distribution is evaluated in its corresponding realization, that is,

$$d_{i,t+1} = \int_{-\infty}^{r_{i,t+1}} p\left(u_i \mid y^t\right) du_i$$

where the subindex i is used to separate stocks and bonds , we have two new series $\{d_{i,t+1}\}_{T_0}^{T_1-1}$. Each one of these series is called the PIT. The following property is very useful to asses the accuracy of the density forecasts from a model. If these density forecasts are accurate, that is, if they are equal to the true conditional densities of future returns, then[20]

$$\{d_{i,t+1}\}_{T_0}^{T_1-1} \overset{iid}{\sim} U(0,1).$$

This is the commented result about the PIT. It is important to stress again that this a result for univariate distributions. That is, the joint distribution of bonds and stocks is computed because it is the relevant object for portfolio management, but the evaluation of the forecasts is done in terms of the univariate distributions since there are not available results for the joint distribution. For instance, the probability integral transform from a bivariate distribution is not uniform. Therefore, in this paper the marginal density forecasts of each asset are the objects of this test.

There are two features to test, uniform distribution and independence, and this is not easy. The approach can be nonparametric by means of the Kolmogorov-Smirnoff test, a goodness-of-fit test of some distribution under the assumption of i.i.d. observations, but it is not very useful for evaluation of density forecasts. For instance, it has low power against a bad description of the tails of the distribution. In addition, it is interesting to realize where the most important failures are and not to finish the analysis with a simple rejection of the model Therefore, it is not used in evaluation of density forecasts. Diebold, Gunter and Tay (1998) advocate the use of graphical methods to study model failures. On the other hand, a parametric approach is proposed in Berkowitz (2001). See also Hong (2001) for a recent implementation of this type of analysis.

Berkowitz (2001) emphasizes the evaluation of density forecasts give that this is the main interest in risk models instead of only point forecasts. He proposes a new parametric approach based on LR tests with good power properties. He builds on the previous work of Crnkovic and Drachman (1996) and Diebold et al. (1998) which also use the probability integral transform. The former is based on the Kuiper statistic. The problem with nonparametric tests such as the Kuiper statistic or the Kolmogorv-Smirnoff is that they have low power in samples that have much less than 1000 observations. On the other hand, the latter is a graphical approach more than a formal test with the goal of studying the model failures.

Berkowitz (2001) transforms the PIT by means of the c.d.f. of the standard normal

$$z_{i,t+1} = \Phi^{-1}(d_{i,t+1})$$

to get a similar result

$$\{z_{i,t+1}\}_{T_0}^{T_1-1} \overset{iid}{\sim} N(0,1)$$

and then advocates the use of likelihood ratio tests of certain hypothesis of model failure such as an AR(1). He also points out that this approach has power only against nonnormality in mean and variance. The problem with this approach is that for extreme cases, such as the crash in the stock market in October 1987, too many simulations should be run to avoid a 0 or a 1 in the PIT.

Diebold, Gunther, and Tay (1998) study the accuracy of density forecasts. They work in the classical framework but give results that can be applied in a Bayesian setting. The "primitive" in Rosenblatt's result is the density forecast independently of

[20] Assuming a nonsingular Jacobian and continuous partial derivatives for the density forecasts.

the way it was computed.[21] That is, it does not matter if it was computed in a classical or Bayesian framework. They advocate the use of a graphical analysis, because the main interest is not a final decision about rejection, but the reason of rejection. This is the approach that is taken in this paper.

On the one hand, the uniform distribution of the PIT can be checked by means of a histogram, jointly with confidence intervals constructed for each bin under the null of an i.i.d uniform PIT. This may give an idea of the performance of the forecasts in terms of marginal distribution of the returns. For instance, an histogram with peaks at zero and one would indicate that the model fails to capture fat tails. However, histograms can be misleading as they depend on the number of bins that are used.[22] Because of that, Q-Q plots should also be shown. They are constructed by pairs of quantiles from the real distribution and the theoretical distribution, the uniform in this application, and therefore a perfect fit would correspond to the 45^oline. On the other hand, independence of the PIT can be checked with correlograms of different moments, $\left(d_{i,t+1}-\bar{d}_i\right)$, $\left(d_{i,t+1}-\bar{d}_i\right)^2$, $\left(d_{i,t+1}-\bar{d}_i\right)^3$ and $\left(d_{i,t+1}-\bar{d}_i\right)^4$, jointly with the corresponding Barlett confidence intervals. This may give an idea of the performance of the forecasts in terms of time dependence or dynamics of the returns.[23]

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 the corresponding return since they are not linked in general. Hong (2001) points out that if the innovation distribution belongs to the local-scale family[24] and the conditional location and scale are well specified then the PIT will be i.i.d. but not uniform if the innovation distribution is misspecified. However, the failure in the i.i.d. or uniform features of the PIT can be due to a bad specification of the innovation distribution or/and the dynamics in other cases. Anyway, failures in the histogram and correlograms of the PIT may suggest directions of model improvement as have been commented that can be checked once implemented.

Now, the practical implementation of this methodology is going to be explained. The main interest is the one-period ahead density forecast that is computed recursively in the sample period selected for the out-of-sample check. Its computation in terms of simulated draws was explained in the previous subsection. Then, by means of each realized $r_{i,T+1}$ of stocks or bonds, the value of the corresponding PIT $d_{i,T+1}$ can be computed with the number of simulations that fall below the realization. Recall that the marginalized performance for each asset is evaluated. This computation is repeated for $T = T_0, ..., T_1 - 1$ which is the sample period left for an out-of-sample check. This generates the sample $\{d_{i,t}\}_{t=T_0+1}^{T_1}$ for each return which is used to evaluate the different models. In this paper, the chosen period for the out-of-sample check of the density fore-

[21] For instance, Diebold et al. "freeze" the estimated parameters and use the same values in all the forecasts.

[22] In fact, as a extreme case, one single bin would give a perfect uniform.

[23] There are many other predictions that can be tested such as longer horizon forecasts. See Diebold et al. (1998).

[24] Examples of this family are the normal, the uniform, the double exponential, the logistic and the Cauchy distributions.

casts starts with the forecast of January 1970, that is, T_0 is December 1969. Therefore, we have 300 observations of the PIT while 192 data observations are let for the first estimation in the recursive procedure.

Note that the estimation being used in each density forecast is not freezed as the first period estimation. There is real time learning in the beliefs about the parameters of each model. In the following sections, the estimation and evaluation of each model is going to be commented. The estimation is shown for the whole sample period to interpret the model and compare it to the usual results in the literature. On the other hand, the evaluation is done by means of the commented recursive estimation.

Another paper about evaluation of density forecasts but applied to high-frequency data is Hong (2001). He uses daily data of the S&P 500 and concludes that is equally important to model the conditional mean, the conditional variance and the distribution of the innovations. The mean is usually forgotten in the computations of density forecasts because it seems weaker that the conditional variance. He studies models from a normal GARCH(1,1) to a Student-t MA(1)-GARCH(1,1)[25] and does not find any model that satisfies uniformity and independence in the PIT. He also points out that introducing MA components for this task improves the dynamics but worsens the marginal distribution.

3 Standard Model: Normal VAR

This section will focus on the usual Bayesian setting, which is generally based on a VAR with a diffuse prior and a Normal likelihood. Extensions will be developed and implemented in the next sections.

The usual way of defining the likelihood of the data is by means of conditional mean, a conditional variance and a distribution of the innovations. 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. Homoskedasticy is also often assumed and therefore the model shows time-varying investment opportunities in terms of conditional means only. This is mainly due to the fact that the predictors that are used for the conditional mean do not predict risk.

3.1 Model

As it has been commented above, a Bayesian model is composed of a prior and a likelihood. This section is devoted to the definition of the simplest model that is going to be implemented. The different models that have been used in the literature are summarized.

Recalling that y_t is a 4×1 vector composed by the two excess returns and the two predictors, the likelihood is given by a Normal VAR[26],

[25] Plus others such as RiskMetrics and EWMA.

[26] The introduction of different retruns is also studied in Tamayo (2001).

$$
\begin{aligned}
y_t &= a + By_{t-1} + u_t \\
u_t \mid y^{t-1} &\sim N(0, \Sigma)
\end{aligned}
$$

where a is a 4×1 vector of intercepts, B is a 4×4 matrix of slope coefficients and u_t represents the shocks, which are independent. The VAR captures the relation between the expected returns and the lagged predictors. The distributional assumption of innovations is normality and independence. Cross-sectional residual correlation is allowed to take into account the feed-back among the variables, but homoskedasticity and independence over time are assumed.

We can write also

$$
y_t = \Pi' z_t + u_t = (I_N \otimes z_t)\, \pi + u_t
$$

where

$$
\begin{aligned}
z_t &= \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix} \\
\Pi &= \begin{pmatrix} \pi_1 & \dots & \pi_N \end{pmatrix} = \begin{pmatrix} a' \\ B' \end{pmatrix}.
\end{aligned}
$$

On the other hand, we can parametrize Σ in terms of Σ_{rr}, Φ, and Γ,

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} I & 0 \\ \Gamma & I \end{pmatrix} \begin{pmatrix} \Sigma_{rr} & 0 \\ 0 & \Phi \end{pmatrix} \begin{pmatrix} I & \Gamma' \\ 0 & I \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{rr} & \Sigma_{rr}\Gamma' \\ \Gamma\Sigma_{rr} & \Phi + \Gamma\Sigma_{rr}\Gamma' \end{pmatrix}
\end{aligned}
$$

which is only the triangular factorization, or Cholesky decomposition, of Σ.

We can also parametrize Σ_{rr} in terms of σ_{11}, ϕ, and γ,

$$
\begin{aligned}
\Sigma_{rr} &= \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \phi \end{pmatrix} \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11} & \gamma\sigma_{11} \\ \gamma\sigma_{11} & \phi + \gamma^2\sigma_{11} \end{pmatrix}
\end{aligned}
$$

which is again the triangular factorization, or Cholesky decomposition, of Σ_{rr}. This parametrization of Σ is used for continuity with the Student-t and regime switching models. Therefore the parameters to estimate in this model are

$$
\theta = \{(\sigma_{11}, \phi, \Phi), (\gamma, \Gamma), \pi\}
$$

We can write the sample in matrix notation as

$$
Y = Z\Pi + U
$$

where Y is $T \times N$ and Z is $T \times (N+1)$, or vectorized

$$\begin{aligned} y &= \tilde{Z}\pi + u \\ u &\sim N\left(0, \Sigma \otimes I_T\right) \\ y = vec\left(Y\right), &\quad \tilde{Z} = I_N \otimes Z \\ \pi = vec\left(\Pi\right), &\quad u = vec\left(U\right) \end{aligned}$$

where I_T is the identity matrix of order T.

The likelihood for a given y_0 is given by

$$\begin{aligned} p\left(y^T \mid \theta, y_0\right) &= p\left(y_1 \mid \theta, y_0\right) p\left(y_2 \mid \theta, y_1\right) ... p\left(y_T \mid \theta, y_{T-1}\right) \\ &\propto \left|\Sigma^{-1}\right|^{\frac{T}{2}} \exp\left\{-\frac{1}{2}\left(y - \tilde{Z}\pi\right)' \left(\Sigma^{-1} \otimes I_T\right)\left(y - \tilde{Z}\pi\right)\right\}, \end{aligned}$$

that is,

$$y \mid \theta, y_0 \sim N\left(\tilde{Z}\pi, \Sigma \otimes I_T\right)$$

which is the same as a SUR model although the assumptions behind each model are different.

The focus of the literature that studies predictability in low-frequency data has been the conditional mean. Given that monthly data are going to be used in this paper, two versions of this model will be estimated and evaluated, one with a time-varying conditional mean and another with a constant one. In homoskedastic models like this one, the second option also implies unpredictable returns in general. The case of unpredictable mean is defined as a constant conditional mean of returns,

$$\begin{aligned} r_t &= a_r + u_{r,t} \\ x_t &= a_x + B_x y_{t-1} + u_{x,t}, \end{aligned}$$

that is,

$$a = \begin{pmatrix} a_r \\ a_x \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ B_{xr} & B_{xx} \end{pmatrix}.$$

The VAR can be expressed as a SURE system to incorporate the case of unpredictable mean with

$$\tilde{Z} = \begin{pmatrix} I_2 \otimes \ell_T & 0 \\ 0 & I_2 \otimes Z \end{pmatrix}$$

where ℓ_T is a vector of ones of dimension $T \times 1$. The vector π is 20×1 with predictors and 12×1 without predictors.

To close the model, the priors[27] are

$$y_0 \sim N\left(\underline{\mu}, \underline{\Sigma}\right)$$
$$\sigma_{11} \sim IG\left(\underline{\alpha}_1, \underline{\beta}_1\right), \quad \phi \sim IG\left(\underline{\alpha}_2, \underline{\beta}_2\right), \quad \Phi \sim IW\left(\underline{\nu}, \underline{S}\right)$$
$$\gamma \sim N\left(\underline{\gamma}_r, \underline{\Gamma}_r\right), \quad vec\left(\Gamma'\right) \sim N\left(\underline{\gamma}_x, \underline{\Gamma}_x\right)$$
$$\pi \sim TN_{\mathcal{I}(\|B\|<1)}\left(\underline{\pi}, \underline{C}\right)$$

where $IG(\cdot)$ denotes the inverse gamma distribution and $IW(\cdot)$ denotes the inverse Wishart distribution. On the other hand, $TN_{\mathcal{I}(\cdot)}(\cdot)$ means a truncated normal where $\mathcal{I}(\cdot)$ defines the support. The notation $\|B\| < 1$ means that the eigenvalues of B are inside the unit circle, that is, the VAR is restricted to be stable. This prior reflects our beliefs about the stationarity of returns and predictors. However, the initial condition y_0 is not taken form the stationary distribution to simplify the computations.[28] These priors are informative for continuity with the regime switching models where it is better to use informative priors.

The estimation uses data augmentation with y_0, that is, it is treated as an additional parameter to be sampled,

$$p\left(\theta, y_0 \mid y^T\right) \propto p(\theta)\, p(y_0 \mid \theta)\, p\left(y^T \mid \theta, y_0\right)$$
$$p(y_0 \mid \theta) = p(y_0)$$
$$p\left(y^T \mid \theta, y_0\right) = \prod_{t=1}^{T} p\left(y_t \mid \theta, y_{t-1}\right).$$

This configuration is used in the corresponding Gibbs sampler in the appendix to approximate the posterior $p\left(\theta, y_0 \mid y^T\right)$. Finally, the one-period ahead predictive distribution $p\left(y_{T+1} \mid y^T\right)$ is sampled as

$$y_{T+1} \mid y^T, y_0, \theta \sim N\left(a + By_T, \Sigma\right)$$

for each (y_0, θ) sampled from the posterior. The first elements of this vector are the returns.

Now, I am going to explain the different models that have been already implemented. The priors are usually diffuse, often imposing stationarity in the VAR. There are also informative priors, for instance against predictability, defining a normal prior of $vec(B)$ centred at zero. Kandel and Stambaugh (1996) use a VAR(1) for return and predictive variables, where the first equation is the prediction equation. They use two priors, a

[27] The parametrization is such that if $x \sim G(\alpha, \beta)$ then $E(x) = \alpha/\beta$, and if $X \sim W(\nu, S)$ then $E(X) = \nu S$. For instance, if X is scalar then $W(\nu, S) = G(\nu/2, 1/2S) = \nu\chi^2(\nu)$.

[28] The Gibbs sampler would not be enough and a Metropolis-Hastings sampler would be required within. See for instance Stambaugh (1999).

diffuse one and another centred at $B = 0$. The latter is Normal-Wishart, that is, Wishart for the inverse of Σ and Normal for π given Σ. But this prior imposes some restrictions on the variance of π. This constitutes a natural conjugate prior,[29] and therefore it can be interpreted as a diffuse prior jointly with previous data with a sample R^2 equal to zero in the predictive equation. In both cases, they truncate the matrix of coefficients B to the stationary region, that is, where eigenvalues lie inside the unit circle.[30] The posteriors are Normal-Wishart and the predictive distributions of r_{T+1} are Student-t for both uninformative and informative priors.

Stambaugh (1999) combines different priors and likelihoods in a related model with a return and a predictor where only lags of the predictor are on the left side of the VAR. The three priors differ on the coefficient of the AR(1) of the predictor, truncating or not to the stationary region, and the way of computation, although are always uninformative The last prior is an approximation to Jeffreys prior for a stochastic initial condition.

The likelihoods that Stambaugh uses differ on the treatment of the initial condition as fixed or stochastic although in both cases are Normal. We can use the conditional likelihood or the exact one, where the initial condition is drawn from the stationary distribution of the predictor. Finally, he comments the application of this study to a VAR with returns and predictors.

He studies several combinations. The diffuse prior jointly with a fixed initial condition gives the standard Normal-Wishart posterior, while the marginal of π is a Student-t with $T-3$ degrees of freedom. The truncated prior to the stationary region jointly with a fixed initial condition is like the previous case but with truncation. Therefore, it is the same but using accept-reject sampling. In this case, this means to store only the outcomes that fall into the stationary region. This prior jointly with a stochastic initial condition needs the use of the Metropolis-Hastings sampler[31]. The final combination is the diffuse prior that takes into account the initial condition jointly with a stochastic initial condition. Again, the Metropolis-Hastings algorithm has to be used.

Barberis (2000) also uses a diffuse prior.[32] He studies two models, one without predictability and another with predictability, the latter as the previous VAR. In the case of predictability, the posteriors are like Stambaugh's ones for the first configuration.

A different approach is given in Demos and Sentana (2000). They use results from Fiorentini and Sentana (1998) about the joint process of returns and their conditional expectations. They study a bivariate system with both variables, treating the expected returns as a latent variable. This is the main point of the paper, the use of return data only. It is implemented with a Gibbs sampling for parameters and expected returns using a Bayesian Kalman filter. They work with informative priors, which are Normal for coefficients and residual covariance and Gammas for variances. This is a natural

[29] The prior, the likelihood and the posterior belong to the same familiy of distributions.

[30] Uhlig (1994) discusses the reasonable priors to use in a time series context, specifically, the Bayesian treatment of unit roots. He shows other priors but concludes that the Normal-Wishart prior centred around a random walk is reasonable. However, he recommends also a sensitivity analysis with respect to the prior treatment of explosive roots. He shows that the tails of the predictive distribution can be sensitive to the prior treatment of explosive roots.

[31] It is an MCMC version of the accept-reject sampling.

[32] However, he comments about the use of a prior that imposes a positive expected market premium.

conjugate prior.

To summarize, only a Normal likelihood has been considered, and the priors are usually diffuse and restricted to stationarity.

3.2 Estimation and Out-of-sample Check

In this section and in the following ones, the empirical results of each model are shown in two dimensions. First, the estimation is shown for the whole sample period to interpret the model and compare it to the usual results in the literature. 15000 iterations were run in the computation of each posterior distribution and the first 5000 were burnt for convergence reasons. Second, the evaluation of the density forecasts is shown from the commented recursive estimation starting with the forecast of January 1970. The evaluation is done separately for stocks and bonds.

3.2.1 Estimation of the Model

Tables 3.1.1 and 3.1.2 show the estimation of the model with predictable mean for the whole sample, while tables 3.2.1 and 3.2.2 show the estimation of the model with unpredictable mean . They describe the posterior of each parameter, giving the mean, the standard deviation, the median, and the last two columns are the 95% bands. The priors are proper but not too informative. The prior of Σ is given in table 3.1.0 to check that it is not very tight. The definition of $vech\,(\Sigma)$ in the tables is

$$vech\,(\Sigma) = (\sigma_{r_1,r_1}, (\sigma_{r_2,r_1}, \sigma_{r_2,r_2}), (\sigma_{x_1,r_1}, \sigma_{x_1,r_2}, \sigma_{x_1,x_1}), (\sigma_{x_2,r_1}, \sigma_{x_2,r_2}, \sigma_{x_2,x_1}, \sigma_{x_2,x_2}))'.$$

The Markov chain is run for 15000 simulations, burning the first 5000 to avoid the effects of the initial conditions.[33]

Table 3.1.1 describes the posterior of the initial condition of the VAR and the residual variance. The residual covariance between stocks and returns is positive. There is a strongly negative residual covariance between the stocks and the dividend yield as it is found in the literature of predictability. This is a signal of mean-reversion in stocks. Another feature that is found in Campbell, Chan and Viceira (2001) 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 sign of the covariance is not clearly defined, with a low negative median. Maybe this is due to the fact that zero coupon bonds are used in this paper so that it is not necessary an approximation to compute the bonds returns as in the case of the coupon bonds in Campbell et al. (2001).

Table 3.1.2 describes the posterior of the VAR intercepts and slopes. For the stocks, π_1, there is not a clear effect of its own lag, but the lagged bond return has a positive effect. Both predictors have a positive coefficient as it is found in the literature. In the equation for bonds, π_2, there is a negative effect of the lagged stock return and a

[33] The ratio of numerical efficiency in the estimation of the posterior mean is near one for all the parameters in both tables. That is, the chain is close to i.i.d. sampling in terms of numerical standard error.

positive effect of its own lag. There is an uncertain effect of the dividend yield, while the positive effect of the term premium is clear. The sign of the predictors in both equations is the same as in the literature of predictability. Another well-known fact is that the predictors are almost single persistent AR(1)'s. The dividend yield is very persistent, while the term premium is also persistent but not so much.

To sum up the signs that are clear for stocks and bonds, stocks are positively related with the lagged dividend yield and term premium, while the corresponding residual covariances are negative. On the other hand, bonds are positevely related with the term premium while their residual covariance with the dividend yield is negative.

Table 3.2.1 starts the analysis of the model with unpredictable mean. The residual variances of stocks and bonds increase only a bit with respect to table 3.1.1, that is, their posterior distributions shift only a bit to the right. This is expected as the R^2 using predictors is very low in classical estimations. The sign of the covariance between the excess stock return and the term premium is not clear now.

In table 3.2.2., the median of the intercepts is positive for stocks and for bonds, although the intercept is not clearly positive in the latter case. They are similar to their sample means, see table 2.1. The persistence of the dividend yield and the term premium increases a bit.

3.2.2 Evaluation of Density Forecasts

The density forecasts of stocks, shown in figure 3.1 as the median and the 50 and 95% bands, depend clearly on the assumption about the conditional mean of returns. As it can be expected, the quantile bands of the forecasts are very stable for the model with unpredictable mean, while the one with predictable mean shows more time variation. Nevertheless, the former has two widenings, the first during the first half of the 70's and the second during the second half of the 80's. In both cases, there is clear outlier in October 1987. Given the big difference between both density forecasts, it is obvious that there are very different implications in terms of portfolio choice as it is found in the literature. An important feature of these forecasts is that the bands of the model with predictable mean are above the corresponding ones of the model with unpredictable mean in general and only a few times they are below, around 1974 and 1980.

In terms of evaluation of these density forecasts by means of the PIT, the histograms do not seem very bad. Comparing the model with predictable mean, figure 3.3, and the model with unpredictable mean, figure 3.5, the left tail of the PIT seems to be lower in the second case. Anyway, the histograms can be misleading and because of that the Q-Q plots are also shown. They confirm the previous idea since the first model is above the 45^oline, while the second is very close to it.

On the other hand, the dynamic behaviour of the PIT is shown in terms of the correlograms of powers of the demeaned PIT. Figure 3.4 shows the results for the model with predictable mean and figure 3.6 for the model with unpredictable mean. The correlogram of the level shows only a marginally significant value at the first lag with predictable mean. The square shows correlation foir both models from the third lag in terms of the Ljung-Box test.[34] The third power gives the same result as the levels, with only a marginally significant first lag with predictable mean. Finally, the forth power shows correlation for both models from the third lag as in the second power.

The density forecasts of the bonds, figure 3.2, can be commented in the same way as stocks. However, now it is clear that the bands show a widening that starts at the beginnig of the 80's and holds for the rest of the period, being very sharp in the case of the 95% bands. There are clear outliers in the first 80's due to the sharp change in the U.S. monetary policy. As in the case of stocks, an important feature of these forecasts is that the bands of the model with predictable mean are above the corresponding ones of the model with unpredictable mean in general and only a few times they are below, around 1974 and 1980.

The histograms of the PIT of bonds are worse with respect to stocks. There are two clear tails in both models, figures 3.7 and 3.9. This translates into Q-Q plots where both models start above the 45^oline and finish bellow it. The second model cuts before the line and then falls a bit deeper. The common result for stocks and bonds is that the left tail of the PIT decreases with an unpredictable mean.

In term of serial dependence of the demeaned PIT, figures 3.8 and 3.10, there is not a clear serial correlation in levels and the third power. Only the model with unpredictable mean shows a marginal correlation in the levels of the PIT at the first lag. Now there is a clearer picture of correlation in the second and forth powers. There is correlation in the second power of the demeaned PIT from the second lag in terms of the Ljung-Box test.[35] The correlation becomes stronger in the forth power. and it is also found from the second lag.

To sum up, there are some clear failures of the normal VAR in terms of both the marginal distribution and dynamics of the PITs. The next sections will try to diminish them. On the other hand, the model with predictable mean gives too many PIT realizations on the left tail.

4 Fat Tails: Student-t

This section and the next are devoted to the development and implementation of extensions to the standard normal VAR model explained in section 3. I will enrich the previous VAR, specially in terms of the likelihood of the data to abandon the normal specification. We should be concerned about the use of the normal distribution since it is known that the returns are not normal due to their fat tails and negative skewness. Data augmentation in the MCMC is a key tool because it brings us back to the normal environment as it will be shown. The particular extensions that will be studied are fat tails and regime switching. The latter will be studied in the following section.

Outliers models are usually specified by means of a Student-t distribution or a Gaussian mixture. This section will deal with the former approach. The issue of fat tails

[34]This correlation starts at the same lag with the absolute value of the demeaned PIT instead of squares.

[35]In the absoulte value of the PIT, the correlation starts to be signifcative form the 7^{th} lag.

has been already treated in the stochastic volatility literature. Specifically, Jaquier, Polson and Rossi (2001) introduce fat tails with errors distributed as a Student-t.[36] They add fat tails as

$$r_t = \sqrt{h_t}\sqrt{\lambda_t} z_t$$
$$z_t \sim N(0,1), \quad \lambda_t \mid \nu \sim \frac{\nu}{\chi^2_\nu}$$

where h_t is the conditional variance, its logarithm is modelled as an AR(1), and ν is the degrees of freedom of the Student-t. They propose an algorithm that uses data augmentation with λ^T. If we condition on λ_t then we can work with $r_t/\sqrt{\lambda_t}$, which is normally distributed. They do not take ν as given and estimate it with a uniform prior on $(5, 30)$. It is also estimated in this paper.

4.1 Model

This new model is only a different definition of the distribution of the innovations in the previous VAR. To define the Bayesian model, we can introduce Student-t errors with the following specification,[37]

$$y_t = a + By_{t-1} + u_t$$
$$u_t = \begin{pmatrix} u_{r,t} \\ u_{x,t} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\nu-2}{\xi_t}}\tilde{u}_{r,t} \\ \tilde{u}_{x,t} \end{pmatrix}$$
$$\xi_t \mid y^{t-1} \sim \chi^2(\nu)$$
$$\tilde{u}_t \mid y^{t-1} \sim N(0, \Sigma)$$

where ξ_t and $\tilde{u}_t$ are independent at all leads and lags[38] The Student-t is associated only to the return innovations in this simple model. This means that, conditioning on ξ_t,

$$V(u_t) = \Sigma^* = \begin{pmatrix} \left(\frac{\nu-2}{\xi_t}\right)\Sigma_{rr} & \sqrt{\frac{\nu-2}{\xi_t}}\Sigma_{rx} \\ \sqrt{\frac{\nu-2}{\xi_t}}\Sigma_{xr} & \Sigma_{xx} \end{pmatrix}$$
$$= \begin{pmatrix} \sqrt{\frac{\nu-2}{\xi_t}}I & 0 \\ \Gamma & I \end{pmatrix} \begin{pmatrix} \Sigma_{rr} & 0 \\ 0 & \Phi \end{pmatrix} \begin{pmatrix} \sqrt{\frac{\nu-2}{\xi_t}}I & \Gamma' \\ 0 & I \end{pmatrix}$$

and that

$$u_{r,t} = \sqrt{\frac{\nu-2}{\nu}} t(0, \Sigma_{rr}, \nu)$$

[36] Recall that only moments of order lower than the number of degrees of freedom are defined for the Student-t and that it converges to a normal as it goes to infinity. For instance, the case of one degree of freeedom is the Cauchy distribution where no moments are defined.

[37] This is similar to an extension explained in Kim, Shephard and Chib (1998).

[38] The parametrization is such that if $x \sim \chi^2(\nu)$ then $E(x) = \nu$. In fact, $\chi^2(\nu) = G(\nu/2, 1/2)$.

with $E(u_{r,t}) = 0$, if $\nu > 1$, and $V(u_{r,t}) = \Sigma_{rr}$, if $\nu > 2$.. This is why the Student-t is scaled, as the residual variance is still Σ_{rr} and can be directly compared with the normal model. See for instance Lange, Little and Taylor (1989) for these type of results.

The parameters to estimate are

$$\theta = \{\nu, (\sigma_{11}, \phi, \Phi), (\gamma, \Gamma), \pi\}.$$

There is only one additional parameter, the degrees of freedom ν.

As in the normal VAR model, two versions of this model will be estimated and evaluated, one with a time-varying conditional mean and another with a constant one. In homoskedastic models like this one, the second option also implies unpredictable returns in general. The case of unpredictable mean is defined with the same restrictions on B as in the previous normal model,

$$B = \begin{pmatrix} 0 & 0 \\ B_{xr} & B_{xx} \end{pmatrix}.$$

The priors are defined for the degrees of freedom

$$p(\nu = m) = p_m,$$

for $m = 4, 6, 8, 10, 15, 20, 30, 60$, and similarly to the normal model,

$$y_0 \sim N\left(\underline{\mu}, \underline{\Sigma}\right)$$

$$\sigma_{11} \sim IG\left(\underline{\alpha}_1, \underline{\beta}_1\right), \quad \phi \sim IG\left(\underline{\alpha}_2, \underline{\beta}_2\right), \quad \Phi \sim IW\left(\underline{\nu}, \underline{S}\right)$$

$$\gamma \sim N\left(\underline{\gamma}_r, \underline{\Gamma}_r\right), \quad vec\left(\Gamma'\right) \sim N\left(\underline{\gamma}_x, \underline{\Gamma}_x\right)$$

$$\pi \sim TN_{\mathcal{I}(\|B\|<1)}\left(\underline{\pi}, \underline{C}\right)$$

For ease of computation, this particular prior on ν has been chosen. Bawens and Lubrano (1998) stress the problem with a flat prior on $(0, \infty)$, since it would give a nonintegrable posterior. Because of that, they propose a prior of order $O\left(\nu^{-2}\right)$. Other alternatives are a flat prior on a bounded support, an exponential prior, etc.

This model is estimated with the Gibbs sampler as it is explained in the appendix. The key is to condition on ξ^T,

$$\begin{pmatrix} \sqrt{\frac{\xi_t}{\nu-2}} u_{r,t} \\ u_{x,t} - \Gamma u_{r,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{rr} & 0 \\ 0 & \Phi \end{pmatrix}\right).$$

It is natural to apply data augmentation by means of $\left(\xi^T, y_0\right)$, and sample from the posterior

$$p\left(\theta, y_0, \xi^T \mid y^T\right) \propto p(\theta)\, p(y_0 \mid \theta)\, p\left(\xi^T \mid \theta, y_0\right) p\left(y^T \mid \theta, y_0, \xi^T\right)$$
$$p(y_0 \mid \theta) = p(y_0)$$
$$p\left(\xi^T \mid \theta, y_0\right) = p\left(\xi^T \mid \theta\right)$$
$$p\left(y^T \mid \theta, y_0, \xi^T\right) = \prod_{t=1}^{T} p(y_t \mid \theta, y_{t-1}, \xi_t)$$

because $p\left(y^T \mid \theta, y_0, \xi^T\right)$ is Normal again. Specifically,

$$y_t \mid \theta, y_{t-1}, \xi_t \sim N\left((I_N \otimes z_t)\, \pi, \Sigma^*\right).$$

The next step is to compute the one-period ahead predictive distribution. Given that

$$\xi_{T+1} \mid \left(y^T, y_0, \xi^T, \theta\right) \sim \chi^2(\nu)$$
$$y_{T+1} \mid \left(y^T, y_0, \xi^T, \theta, \xi_{T+1}\right) \sim N\left(a + By_T, \Sigma^*\right),$$

, where Σ^* is defined as before but now with the corresponding ξ_{T+1}, a draw from this distribution can be computed for each draw from the posterior of $\left(y_0, \xi^T, \theta\right)$. These draws define the desired predictive distribution. The first entries of y_{T+1} are the corresponding returns.

4.2 Estimation and Out-of-sample Check

Again, the empirical results of each model are shown in two dimensions. First, the estimation is shown for the whole sample period to interpret the model and compare it to the usual results in the literature. 15000 iterations were run in the computation of each posterior distribution and the first 5000 were burnt for convergence reasons. Second, the evaluation of the density forecasts is shown from the commented recursive estimation starting with the forecast of January 1970. The evaluation is done separately for stocks and bonds.

4.2.1 Estimation of the Model

The priors are the same as in the normal model. The only additional prior is the degrees of freedom and a uniform prior over a finite set of values is defined. They range from 4 to 60. Tables 4.1.1 and 4.1.2 show the estimation of the model with predictable returns for the whole sample, while tables 4.2.1 and 4.2.2 show the estimation of the model with unpredictable returns. The Markov chain is again run for 15000 simulations, burning the first 5000 to avoid the effects of the initial conditions.[39]

[39] The ratio of numerical efficiency in the estimation of the posterior mean is near one for all the parameters in both tables. That is, the chain is close to i.i.d. sampling in terms of numerical standard error.

Table 4.1.1 shows a posterior of the degrees of freedom that is concentrated on the lowest and highest values, that is, concentrated on evidence of nonnormality and normality. This can be due to the joint analysis of stocks and bonds or a better description of returns with a mixture of a normal and a Student-t.[40] In table, 4.1.2, the 95% bands of the residual variance of stocks are a bit wider with respect to 3.1.1, while the median of the residual variance of bonds decreases a bit. The covariance of stocks and bonds shifts a bit to the right. In table 4.1.3, it can be seen that the coefficients of stocks and bonds do not change much with respect to the normal model. The equations of the predictors are almost equal as the Student-t is modelled only over the returns.

The model with unpredictable mean gives a similar posterior of the degrees of freedom. Again, comparing table 4.2.2 and 3.2.1 now, the 95% bands of the residual variance of stocks are a bit wider, while the posterior of the residual variance of bonds shifts a bit to the left. Finally, comparing tables 4.2.3 and 3.2.2, the coefficient of stocks shifts to the right, while the bond one shifts to the left with a twice lower mean and median. The equations of the predictors are almost equal to the previous model.

4.2.2 Evaluation of Density Forecasts

Again, in terms of the density forecasts of stocks, figure 4.1, the quantile bands of the forecasts are very stable for the model with unpredictable mean, while with predictable mean there is more time variation as it can be expected.

There is not a big difference between the forecasts from the normal and the Student-t models with predictable mean. When the conditional mean is constant, the Student-t model shows tighter 50% bands and wider 95% bands. This behaviour can be expected as the Student-t increases kurtosis. In terms of the comparison of the forecasts of the models with time-varying and constant conditional mean, the comments of the normal model still hold.

The histograms of the PIT for stocks, figures 4.3 and 4.4, do not seem to change much with respect to the normal model. The corresponding Q-Q plots reflect this similarity. Again, the model with unpredictable mean is closer to the diagonal. The comments about the correlograms of powers of the demeaned PIT for stocks, figures 4.4 and 4.6, are the same as for the normal model, finding correlation in the second and forth powers of the PIT.

The density forecasts of bonds are similar to the normal model with a predictable mean. In the version with unpredictable mean, the 50 and 95% bands are tighter and the median is lower. In terms of the comparison of the forecasts of the models with time-varying and constant conditional mean, the comments of the normal model still hold.

Compared to the normal model, both tails in the histogram of the PIT for bonds seem to increase, figures 4.7 and 4.9. But the left tail is still fatter with predictable mean. This is reflected in Q-Q plots, figures 4.8 and 4.10, that start higher and then fall lower with respect to the diagonal.

[40] A Markov-switching between a normal and a Student-t is advocated in Pérez-Quirós and Timmermann (2000b).

In terms of correlations of powers of the demeaned PIT, figures 4.8 and 4.10, there is only a small change in the square. Now the correlation starts a bit later in terms of the Ljung-Box test, from the sixth lag with predictable mean and the third lag with unpredictable mean.[41]

To sum up, the Student-t does not help much in terms of the marginal distribution and dynamics of the PITs. Maybe different chi-squares are needed for stocks and bonds. Anyway, this distribution is often used in models with time-varying volatility such as GARCH and stochastic volatility, not on its own. Note also that the Student-t does not allow asymmetry and there is a clear negative asymmetry in stocks and positive in returns. But this model was only applied to enrich the distribution of the innovations and does not add anything in terms of dynamics. This is addressed in the following section. Anyway, the model with predictable mean still gives too many PIT realizations on the left tail.

5 Nonlinear Dynamics: Markov-switching

Now, time-varying risk is introduced into the picture by means of nonlinear dynamics. The investment opportunities will vary in terms of conditional means and variances, not only as conditional means as in the previous sections. Time-varying volatility can be modelled as in GARCH and stochastic volatility models or as regime switching models. The second ones are going to be studied.

Hamilton and Susmel (1994) explore the issue of ARCH effects versus regime switching in variance in U.S. monthly data. It is related to Perron's idea that unit roots may reflect mean shifts. They remark that ARCH models show high persistence and poor predictive performance and that a SWARCH, switching plus ARCH effects, does better. The persistence comes mainly from the regimes and the high volatility regime is related with recessions.

This type of models adds 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 or/and the variance matrix. The Markov chain is used to represent persistence in economic regimes and differentiates this model from a simple gaussian mixture, which is used to describe fat tails, not nonlinear dynamics. The unobservability of the state is the approach to take if we want to be able to make real time forecasts. Hamilton (1989) develops the reference model with an unobserved state variable that follows a Markov chain, while Kim (1994) translates that idea to general state-space models and improves some technical issues. Kim and Nelson (1998) show its implementation in a Bayesian framework to macroeconomic indicators, while Pérez-Quirós and Timmermann (2000a and 2000b) estimate it for stock returns in a classical context. Current extensions of the basic setting are time-varying transition probabilities for instance, where these two papers are also examples.

This type of models lets us separate the asymmetric behaviour of returns in different

[41] The absolute value of the demeaned PIT shows correlation from the 8^{th} lag with predictable mean, while it arrives later with unpredictable mean.

states as it is emphasized in conditional asset pricing models and modern nonstructural forecasting procedures in macroeconomics (see Diebold (1998)). In asset pricing, it is a fact that people seem more risk averse in recessions than in expansions in terms of a countercyclical Sharpe ratio. Lettau and Ludvigson (2001b) use the consumption wealth ratio as the state variable to measure different states and implement conditional asset pricing models with that variable. In macroeconomic applications, for instance, an economy can be modelled with high or low growth depending on the outcome from the Markov process. This is a basic part of the modern treatment of business cycles. Kim and Nelson (1998) is the basic reference in a Bayesian framework. They stress that, in general, maximum likelihood requires approximations in this context. Therefore, they advocate the use of the Bayesian approach because it is approximation-free thanks to the Gibbs-sampler. This is really pseudo-exact inference, exact up to simulation error, because the estimation relies on MCMC methods.

5.1 Model

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, 0 or 1. The model can be decomposed as

$$
\begin{aligned}
y_t &= a_{s_t} + B_{s_t} y_{t-1} + u_t \\
a_{s_t} &= \begin{pmatrix} a_{r,s_t} \\ a_x \end{pmatrix}, \quad B_{s_t} = \begin{pmatrix} B_{rr,s_t} & B_{rx,s_t} \\ B_{xr} & B_{xx} \end{pmatrix} \\
u_t \mid y^{t-1}, s^t &\sim N\left(0, \Sigma_{s_t}\right) \\
\Sigma_{s_t} &= \begin{pmatrix} \Sigma_{rr,s_t} & \Sigma_{rr,s_t}\Gamma' \\ \Gamma\Sigma_{rr,s_t} & \Phi + \Gamma\Sigma_{rr,s_t}\Gamma' \end{pmatrix} \\
\Sigma_{rr,s_t} &= \begin{pmatrix} \sigma_{11,s_t} & \gamma_{s_t}\sigma_{11,s_t} \\ \gamma_{s_t}\sigma_{11,s_t} & \phi_{s_t} + \gamma_{s_t}^2\sigma_{11,s_t} \end{pmatrix}
\end{aligned}
$$

where the distribution of u_t is conditional on the regimes. In this model, the switching is mostly concentrated to returns to avoid increasing the number of parameters unnecessarily.

Therefore, this model focus on regime switching of the returns, in their relation with the predictors and in their variance. The unobservable regime follows a binary Markov chain,

$$s_t = j \sim p\left(s_t = j \mid s_{t-1} = i\right) = p_{ij}, \quad i, j = 0, 1$$

which captures the idea of persistence in Economics instead of a Gaussian mixture only. The transition probabilities can be expressed in a compact way as

$$P = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}.$$

The identification of regimes will be defined by the variance of the first term of the VAR, calling regime one the one with a higher value for that variance,

$$\sigma_{11,t} = \sigma_{11}^0 \left(1 + hs_t\right)$$
$$\bar{h} = 1 + h, \quad \bar{h} > 1$$

and that is why the residual variance is parametrized this way from the first model. On the other hand the regime switching in π will be defined as

$$\pi_{r,s_t} = \pi_r^0 + \delta s_t$$

where δ is a vector with the same dimension as π_{r,s_t} and then

$$r_t = a_{r,s_t} + B_{r,s_t} y_{t-1} + u_{r,t} = \left(I_2 \otimes z_{r,t}^s\right) \pi_r + u_{r,t}$$
$$x_t = a_x + B_x y_{t-1} + u_{x,t} = \left(I_2 \otimes z_{x,t}\right) \pi_x + u_{x,t}$$

where

$$\pi_r = \begin{pmatrix} \pi_1^0 \\ \delta_1 \\ \pi_2^0 \\ \delta_2 \end{pmatrix}, \quad z_{r,t}^s = z_{r,t} \otimes \begin{pmatrix} 1 \\ s_t \end{pmatrix}$$

which let us define the system of equations as

$$\begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} I_2 \otimes Z_r^s & 0 \\ 0 & I_2 \otimes Z_x \end{pmatrix} \begin{pmatrix} \pi_r \\ \pi_x \end{pmatrix} + \begin{pmatrix} u_r \\ u_x \end{pmatrix}$$

and we are back to the SURE context again. Although there is a difference, the residual variance changes over time. The way to estimate this model is shown in the appendix.

The parameters to estimate are now

$$\theta = \left\{ \left(p_{00}, p_{11}\right), \left(\sigma_{11}^0, \phi^0, \phi^1, \Phi\right), \bar{h}, \left(\gamma^0, \gamma^1, \Gamma\right), \pi \right\}.$$

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 unpredictable mean is defined as

$$a_{s_t} = \begin{pmatrix} a_r \\ a_x \end{pmatrix}, \quad B_{s_t} = \begin{pmatrix} 0 & 0 \\ B_{xr} & B_{xx} \end{pmatrix},$$

where there is not switching in the intercept, which implies

$$\pi_r = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$$

and the system is again like the original VAR with unpredictable mean

$$\begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} I_2 \otimes \ell_T & 0 \\ 0 & I_2 \otimes Z \end{pmatrix} \begin{pmatrix} \pi_r \\ \pi_x \end{pmatrix} + \begin{pmatrix} u_r \\ u_x \end{pmatrix},$$

but with a time-varying conditional variance of the innovations.

The priors have the following structure,[42]

$$s_0 \sim Bernoulli\,(p_0)$$
$$p_{00} \sim B\left(\underline{n}_{00}, \underline{n}_{01}\right), \quad p_{11} \sim B\left(\underline{n}_{11}, \underline{n}_{10}\right)$$
$$y_0 \sim N\left(\underline{\mu}, \underline{\Sigma}\right)$$
$$\sigma_{11}^0 \sim IG\left(\underline{\alpha}_{10}, \underline{\beta}_{10}\right), \quad \phi^0 \sim IG\left(\underline{\alpha}_{20}, \underline{\beta}_{20}\right), \quad \phi^1 \sim IG\left(\underline{\alpha}_{21}, \underline{\beta}_{21}\right), \quad \Phi \sim IW\left(\underline{\nu}, \underline{S}\right)$$
$$\bar{h} \sim TIG_{\mathcal{I}(\bar{h}>1)}\left(\underline{\alpha}_{11}, \underline{\beta}_{11}\right)$$
$$\gamma^0 \sim N\left(\underline{\gamma}_{r0}, \underline{\Gamma}_{r0}\right), \quad \gamma^1 \sim N\left(\underline{\gamma}_{r1}, \underline{\Gamma}_{r1}\right), \quad vec\,(\Gamma') \sim N\left(\underline{\gamma}_x, \underline{\Gamma}_x\right)$$
$$\pi \sim TN_{\mathcal{I}(\|B^0\|<1, \|B^1\|<1)}\left(\underline{\pi}, \underline{C}\right)$$

where the distribution of $\bar{h}$ is denoted by $TIG_{\mathcal{I}(\bar{h}>1)}(\cdot)$, this means a truncated inverse gamma with support[43] $\bar{h} > 1$. Now the truncation of the prior of π to the stationarity region is made in terms of both implied B^0 and B^1, which are equal in the case of unpredictable mean. Note that the initial condition of the regimes s_0, like y_0, is not drawn form the stationary distribution. 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 estimation is based on Kim and Nelson (1998). Data augmentation with regimes $\left(s_0, s^T, y_0\right)$ is the key element,

$$p\left(\theta, y_0, s_0, s^T \mid y^T\right) \propto p(\theta)\, p(s_0 \mid \theta)\, p\left(s^T \mid \theta, s_0\right) p\left(y_0 \mid \theta, s_0, s^T\right) p\left(y^T \mid \theta, s_0, s^T, y_0\right)$$
$$p(s_0 \mid \theta) = p(s_0 \mid p_0)$$
$$p\left(y_0 \mid \theta, s_0, s^T\right) = p(y_0 \mid \theta)$$
$$p\left(y^T \mid \theta, s_0, s^T, y_0\right) = \prod_{t=1}^{T} p(y_t \mid \theta, y_{t-1}, s_t)\,.$$

The vector y_0 is defined as independent of s_0. The main point in this methodology is to construct a filter that serves as an algorithm to infer probabilities about the underlying state. It is similar to the Kalman filter, but it is nonlinear and is defined for a discrete variable. As well, as a by-product, it enables us to evaluate the log-likelihood and compute the maximum likelihood estimator in a classical framework. Kim (1994) generalizes this methodology to state-space models[44]. He allows switching in the measurement and transition equations and develops the corresponding filtering and

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

[43] The notation $\mathcal{I}\left(\bar{h} > 1\right)$ refers to the indicator function that gives 1 if $\bar{h} > 1$ and 0 otherwise.

[44] So that we can handle MA terms for instance.

smoothing algorithms. In a general context, this filters are only an approximation since there are two unobservable variables, the state variable in the transition equation that defines the dynamics of the system and the variable that rules the Markov-switching. This is why normality is lost for the state variable and the Kalman filter that he uses no longer gives the conditional expectation.

Finally, the one-period ahead predictive distribution is computed from

$$p\left(s_{T+1} \mid y^T, s_0, s^T, \theta\right) = p\left(s_{T+1} \mid s_T, \theta\right)$$
$$y_{T+1} \mid \left(y^T, s_0, s^T, \theta, s_{T+1}\right) \sim N\left(a_{s_{T+1}} + B_{s_{T+1}} y_T, \Sigma_{s_{T+1}}\right)$$

for each $\left(y_0, s_0, s^T, \theta\right)$ from the Gibbs sampler.

A different implementation of regime switching is developed in Chauvet and Potter (2001) by means of a factor model as a nonlinear proxy to market risk premia. They use regime switching to study the joint behaviour of mean and variance of monthly excess returns of stocks, introducing the switching in the factor mean and variance. They identify two regimes, a bear market regime with negative mean and high volatility, and a bull market regime with positive mean and low volatility, which has a longer duration.

5.2 Estimation and Out-of-sample Check

Now, there is not the possibility of unpredictable returns since at least the conditional variance will be time-varying. But a model with a predictable mean will be faced with a model with an unpredictable mean as in the previous sections. Again, the empirical results of each model are shown in two dimensions. First, the estimation is shown for the whole sample period to interpret the model and compare it to the usual results in the literature. 15000 iterations were run in the computation of each posterior distribution and the first 5000 were burnt for convergence reasons. Second, the evaluation of the density forecasts is shown from the commented recursive estimation starting with the forecast of January 1970. The evaluation is done separately for stocks and bonds.

5.2.1 Estimation of the Model

The chosen priors are not too tight. The prior of the residual variance matrices, table 5.1.0, is such that the median of the stock variance is twice higher in the high-volatility regime, while it is a bit less in the case of bonds. A model with identification of the regimes in terms of bonds has also been estimated and the results are very similar. The Markov chain is again run for 15000 simulations, burning the first 5000 to avoid the effects of the initial conditions.[45]

The probability of the high-volatility regime is shown in figure 5.1 for the model with predictable mean and in figure 5.2 for the model with unpredictable mean, jointly with the NBER recession index. The regimes are similar and very persistent. The high-volatility regime is roughly related to recessions until the end of the 70's, but the figures

[45] The ratios of numerical efficiency in the estimation of the posterior means are near one in general. There are some exceptions close to two.

show evidence of that regime since the end of the 70's. It is mainly driven by the bond volatility.

Tables 5.1.1. and 5.1.2 show the results for the model with predictable mean. As it can be inferred from the previous graph, both transition probabilities are very high. The switch is clear in the residual variances of stocks and specially bonds. The numbers in the table increase roughly twice from one regime to the other in the case of stocks, but they increase ten times in the case of bonds. Their residual covariance goes from an undefined sign to a positive sign. The numbers corresponding to the residual covariance of the dividend yield with stocks turn to be twice more negative in the high-volatility regime, while with bonds go from an undefined sign to a negative one. The residual covariance of the them premium and the bonds increases its bands and its mean and median become ten times more negative in the high-volatility regime.

The intercepts and slopes are shown in table 5.1.2. There is not a clear evidence of switching in any parameter. In the case of stocks, the results for the normal model are closer to the high-volatility regime, while in the other regime the signs of the lagged bond return and the term premium are not clear. In the case of bonds, the results for the normal model are again closer to the high-volatility regime, while in the other regime there are not clear signs. The equations of predictors do not change much with respect to the normal case as it should be expected from the model, where there is not switching in those coefficients.

Tables 5.2.1 and 5.2.2. show the results for the model without unpredictable mean. Looking at the transition probabilities, both regimes are again very persistent. The comments about the residual variance matrix are similar to the model with predictable mean. The intercept of stocks is positive and similar to the normal model while the sign for bonds is still undefined, but it shifts to the left now and both the mean and the median are negative. The coefficients of the predictors do not change much with respect to the normal case as it should be expected from the model.

5.2.2 Evaluation of Density Forecasts

The density forecasts of stocks are very similar to the normal model in the version with predictable mean, figure 5.3. If the conditional mean is constant then the bands are wider and the median is higher in general, depending on the widenings and contractions of the bands. Recall that now there is a high volatility regime. The comparison between the two versions of regime switching, predictable and unpredictable mean, is defined by the same comments as in the normal model. Nevertheless, the unpredictable mean model shows more time variation than in the normal or the Student-t cases since the volatility changes over time. There is a series of widenings and contractions of the bands until a final widening from the beginnig of the 80's, being sharper for the 95% bands.

The histograms of the PIT for stocks are similar to the normal model, figures 5.5 and 5.7. Although now there is one bin on the left side that is clearly too high. Anyway, we should trust more on Q-Q plots and they confirm the similarity with the normal model. The correlograms of powers of the PIT are very similar the normal model too, figures 5.6 and 5.8. Therefore, switching does not help much with respect to the normal model in the case of stocks. But this is different with bonds.

There is a clear change in the density forecasts of bonds, figure 5.4. The predictable mean version has wider 95% bands than the corresponding normal model, although the 50% bands are similar. With a constant conditional mean, the median is lower and the 95 and 50% bands are wider, the latter a bit less, but there are lots of cuts with the bands of the normal model. Comparing the two versions of regime switching, the movements in the bands are similar, although they are not so sharp for the 50% bands with unpredictable mean. The tendencey of the predictable model to give higher bands still holds from the 80's.

The histograms of the PIT for bonds seem to improve with respect to the normal model, as both tails decrease, figures 5.9 and 5.11. Again, the left tail of the model with predictable mean is higher than the model with unpredictable mean. But histograms can be misleading and the Q-Q plots show that there is not a big difference with the normal model in general terms. Anyway, the difference on the left tail between both regime switching models is confirmed by these plots.

In terms of correlograms of powers of the demeaned PIT, figures 5.10 and 5.12, the level and the third power show no correlation, similar to the normal model. But now there is a clear improvement in terms of the second and forth powers. There is not correlation in the square[46] of both switching models at any lag in terms of the Ljung-Box test. About the forth power, there is not correlation in the model with predictable mean while it only appears at long lags in the model with unpredictable mean. Therefore, in terms of PIT dynamics, the switching model represents an improvement with respect to the normal model. 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 and forth moments of the corresponding PIT.

To sum up, Markov-switching has improved only the performance in terms of the PIT dynamics in second and forth moments. But there is no improvement with stocks. Maybe the identification of a common regime is bad in terms of mixing outliers of both returns or the transition probabilities should be modelled to track the business cycle.

6 Conclusions

This paper shows that the use of predictors, where predictors are taken as the state variables that define the conditional mean of returns, produces different density forecasts as they vary much more over time than in a model with unpredictable mean. This translates into significant effects on portfolio choice as it is well known in the literature on predictability. However, the development of out-of-sample checks gives some new insights about predictors. Using a model with a time-varying conditional mean in the computation of density forecasts does not improve the forecast evaluation and gives too many realizations on the left tail of the density forecast with respect to a model with

[46]Neither in the absolute value of the demeaned PIT. Only the first lag is marginally significant in the model with unpredictable mean.

unpredictable mean. Predictors are too optimistic and, although in terms of density evaluation this is as bad as too many realizations on the right tail, this can be a problem if the investor is concerned about losses. Recall that losses are the main concern in Value-at-Risk models. But this must be stated carefully since the data that has been used in the paper ends in 1994. The boom in the stock market during the second half of the 90's, which has decreased the predictive power of the dividend yield, may change this conclusion. Anyway, this situation has been corrected in the last years.

Some conclusions can also be drawn in terms of the different models that are implemented. Compared to the normal model, the Student-t does not improve the forecast evaluation of bonds and stocks. The Markov-switching improves the forecast evaluation of bonds, since it captures better the dynamics than the normal model. However, there is no improvement for stocks with Markov-switching. This could be expected since the correlation in the second moment is clearer for the bond series.

Specifically, the evaluation results for stocks have shown that the three models and both versions of conditional means give a similar dynamic behaviour of the Probability Integral Transform (PIT), with a clear failure in its second and forth moments. The Q-Q plots of the PIT are similar among the models but different between each version of the conditional mean. A predictable mean gives a Q-Q plot above the 45^oline, which indicates too many realizations on the left side of the density forecast, while an unpredictable mean gives one very close to the diagonal.

On the other hand, the evaluation results for bonds reflect a bad performance of all the models at the tails, with too many realizations on them. Anyway, a constant conditional mean decreases the left tail of the PIT as in the case of stocks. In terms of the dynamic behaviour of the PIT, there is no difference in using a time-varying or constant conditional mean. However, the Markov-switching model kills the correlation that is found in the second and forth moments with the normal and Student-t models. It was noted in the paper that conclusions about dynamics in returns cannot be directly drawn form the dynamics of the PIT, but this is an example where the correlation in the second moment of the PIT has been solved by means of a richer specification of the volatility of returns. This may suggest that stochastic volatility could be introduced to solve the correlation in the square of the PIT of stocks since Markov-switching did not help on this dimension.

Future avenues of research are described in the following paragraphs. First, the data should be expanded to the second half of the 90's and other countries[47]. Second, in terms of the Bayesian framework, Bayes factors can be computed as they represent the standard way of Bayesian comparison of models. This would give quantitative results instead of the qualitative ones that have been explained.

Finally, in terms of the models that have been implemented, the fat tails model with Student-t may be used with different degrees of freedom for each return and/or combined with nonlinear dynamics.[48] The models with nonlinear dynamics, say time-varying

[47] In this direction, differnt types of stocks could also be studied, say growth and value stocks. In addition, housing and labor income could be introduced.

[48] The Student-t is usually found as the distribution of the innovation in GARCH or stochastic volatility models.

volatility, can be expanded in two ways. The regime switching model in this paper could be enriched with time-varying transition probabilities to track the business cycle instead of outliers. For instance, Pérez-Quirós and Timmermann (2000a and 2000b) introduce this variation as dependence of the transition probabilities on the composite leading indicator, as a variable that measures the state of the economy. Generally, a model with constant transition probabilities track low and high volatility states as in this paper. After an outlier is observed, the model usually assigns that observation to the high volatility regime and this may decrease its predictive power. Another approach may be the introduction of stochastic volatility given that the time dependence in the PIT of the bonds has disappeared by means of regime switching but this has not happened with stocks. Recent papers on stochastic volatility are Aguilar and West (2000) and Chib, Nardari and Shephard (2001).[49]

It can be seen that the development of a joint model of stocks and bonds is a difficult task, but it is the necessary object for portfolio management, and this paper has tried to shed some light on the directions to follow.

[49] They use a factor structure to deal with multivariate models.

References

Aguilar, O., and M. West, 2000, Bayesian Dynamic Factor Models and Portfolio Allocation, *Journal of Business and Economic Statistics* 18, 338-357.

Ang, A., and G. Bekaert, 2001, Stock Return Predictability: Is it There?, mimeo, Columbia University.

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

Bawens, L., and M. Lubrano, 1998, Bayesian Inference on GARCH Models Using the Gibbs Sampler, *Econometrics Journal* 1, 23-46.

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, ?-?.

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

Campbell, J.Y., 1991, A Variance Decomposition of Stock Returns, *The Economic Journal* 101, 157-179.

Campbell, J.Y., 1999, Asset Prices, Consumption, and the Business Cycle, in Taylor , J.B. and M. Woodford, *Handbook of Macroeconomics* vol. 1, Elsevier Science Publishers.

Campbell, J.Y., Y.L. Chan, and L.M. Viceira, 2001, A Multivariate Model of Strategic Asset Allocation, mimeo, Harvard University.

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* 4.

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

Chib, S., F. Nardari, and N. Shephard, 2001a, Markov Chain Monte Carlo Methods for Stochastic Volatility Models, mimeo, Nuffield College, Oxford.

Chib, S., F. Nardari, and N. Shephard, 2001b, Analysis of High Dimensional Multivariate Stochastic Volatility Models, mimeo, Nuffield College, Oxford.

Cochrane, J.H., and M. Piazzesi, 2001, Bond Risk Premia, mimeo, GSB, University of Chicago.

Demos, A., and E. Sentana, 2000, How does the Future Change our Past Views of the Present?, mimeo, CEMFI.

Diebold, F.X., 1998, The Past, Present, and Future of Macroeconomic Forecasting, *Journal of Economic Perspectives* 12, 175-192.

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.

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.

Ferson, W.E., S. Sarkissian, and T. Simin, 1999, Spirious Regressions in Financial Economics?, mimeo.

Fiorentini, G., and E. Sentana, 1998, Conditional Means of Time Series Processes and Time Series Processes for Conditional Means, *International Economic Review* 39, 1101-1118.

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.

Geweke, J., 1998, Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication, Staff Report 249, Ferderal Reserve Bank of Minneapolis.

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

Goyal, A., and I. Welch, 1999, Predicting the Equity Premium, UCLA, WP.

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.

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

Hong, Y., 2001, Evaluation of Out-of-Sample Probability Density Forecasts with Applications to Stock Prices, mimeo, Cornell University.

Jacquier, E. N.G. Polson, and P. E. Rossi, 2001, An Extended Stochastic Volatility Model with Fat-tails and Leverage Effect: A Bayesian Analysis, mimeo.

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.

Kothari, S.P., and J. Shanken, 1997, Book-to-market, Dividend Yield, and Expected Market Returns: A Time-series Analysis, *Journal of Financial Economics* 44, 169-203.

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

Lange, K.L., R.J.A. Little, and J.M.G. Taylor, 1989, Robust Statistical Modeling Using the t Distribution, *Journal of the American Statistical Association* 84, 881-896.

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

Lettau, M., and S. Ludvigson, 2001b, Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying, *Journal of Political Economy* 109, 1238-1287.

Lewellen, J., 2001, Predicting Returns with Financial Ratios, mimeo, MIT Sloan School.

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

Nelson, C.R., and M.J. Kim, 1993, Predictable Stock Returns: The Role of Small Sample Bias, *Journal of Finance* 48, 641-661.

Pérez-Quirós, G., and A. Timmermann, 1996, On Business Cycle Variation in the Mean, Volatility and Conditional Distribution of Stock Returns, mimeo, University of California.

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

Pérez-Quirós, G., and A. Timmermann, 2000b, Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities, mimeo.

Pesaran, M.H., and A. Timmermann, 1995, Predictability of Stock Returns: Robustness and Economic Significance, *Journal of Finance* 50, 1201-1228.

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

Santos, T., and P. Veronesi, 2001, Labor Income and Predictable Stock Returns, NBER WP 8309.

Shephard, N., 1994a, Local Scale Models, *Journal of Econometrics* 60, 181-202.

Shephard, N., 1994b, 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.

Tamayo, A., 2001, Stock Return Predictability, Conditional Asset Pricing Models and Portfolio Selection, mimeo, Simon Graduate School.

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.

Tierney, L., 1994, Markov Chains for Exploring Posterior Distributions, *Annals of Statistics* 22, 1701-1762.

Torous, W., and S. Yan, 1999, Predictive Regressions Revisited, UCLA, WP.

Uhlig, H., 1997, Bayesian Vector Autoregressions with Stochastic Volatility, *Econometrica* 65, 59-73.

Xia, Y., 2001, Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation, *Journal of Finance* 56, 205-246.

Appendix

A Gibbs Samplers

The Gibbs samplers that have been used to approximate the posterior distributions for the different models are developed in the following pages. The conditional posteriors of the blocks of parameters on which the Gibbs sampler iterates are described.

A.1 Normal VAR

- Sampling y_0:

Its conditional posterior follows from

$$p\left(y_0 \mid y^T, \theta\right) \propto p\left(y_0\right) p\left(y_1 \mid \theta, y_0\right),$$

which gives

$$\begin{aligned} y_0 \mid \left(y^T, \theta\right) &\sim N\left(\bar{\mu}, \bar{\Sigma}\right) \\ \bar{\mu} &= \bar{\Sigma}\left[\underline{\Sigma}^{-1}\underline{\mu} + B'\Sigma^{-1}\left(y_1 - a\right)\right] \\ \bar{\Sigma} &= \left[\underline{\Sigma}^{-1} + B'\Sigma^{-1}B\right]^{-1}. \end{aligned}$$

- Sampling $(\sigma_{11}, \phi, \Phi)$:

We can define a block $\theta_S = (\sigma_{11}, \phi, \Phi)$ and given that we can write

$$u_t = y_t - \left(I_N \otimes z_t\right)\pi = \begin{pmatrix} u_{r,t} \\ u_{x,t} \end{pmatrix} = \begin{pmatrix} u_{r_1,t} \\ u_{r_2,t} \\ u_{x,t} \end{pmatrix}$$

$$\begin{pmatrix} u_{r_1,t} \\ u_{r_2,t} - \gamma u_{r_1,t} \\ u_{x,t} - \Gamma u_{r,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \Phi \end{pmatrix}\right),$$

conditioning on the rest of parameters and the data, each element of the block can be sampled independently of the others. The variance of the first return is sampled as

$$\begin{aligned} \sigma_{11} \mid \left(y^T, y_0, \theta_{\neq S}\right) &\sim IG\left(\bar{\alpha}_1, \bar{\beta}_1\right) \\ \bar{\alpha}_1 = \underline{\alpha}_1 + \frac{T}{2}, &\quad \bar{\beta}_1 = \underline{\beta}_1 + \frac{u_{r_1}' u_{r_1}}{2}, \end{aligned}$$

the residual variance of the second returns as

$$
\begin{gathered}
\phi \mid \left(y^T, y_0, \theta_{\neq S}\right) \sim IG\left(\bar{\alpha}_2, \bar{\beta}_2\right) \\
\bar{\alpha}_2 = \underline{\alpha}_2 + \frac{T}{2}, \quad \bar{\beta}_2 = \underline{\beta}_2 + \frac{\tilde{u}_{r_2}' \tilde{u}_{r_2}}{2} \\
\tilde{u}_{r_2} = u_{r_2} - \gamma u_{r_1},
\end{gathered}
$$

and finally the residual variance of the predictors as

$$
\begin{gathered}
\Phi \mid \left(y^T, y_0, \theta_{\neq S}\right) \sim IW\left(\bar{\nu}, \bar{S}\right) \\
\bar{\nu} = \underline{\nu} + T, \quad \bar{S} = \left[\underline{S}^{-1} + \tilde{U}_x' \tilde{U}_x\right]^{-1} \\
\tilde{U}_x = U_x - U_r \Gamma'.
\end{gathered}
$$

- Sampling (γ, Γ):

  Defining $\theta_G = (\gamma, \Gamma)$, we can use the same idea of independence and focus on

$$
\begin{pmatrix} u_{r_2,t} - \gamma u_{r_1,t} \\ u_{x,t} - \Gamma u_{r,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \phi & 0 \\ 0 & \Phi \end{pmatrix}\right),
$$

conditioning on the rest of parameters and the data. The coefficient of the projection of the second return is drawn from

$$
\begin{gathered}
\gamma \mid \left(y^T, y_0, \theta_{\neq G}\right) \sim N\left(\bar{\gamma}_r, \bar{\Gamma}_r\right) \\
\bar{\gamma}_r = \bar{\Gamma}_r \left[\underline{\Gamma}^{-1} \underline{\gamma} + \phi^{-1} u_{r_1}' u_{r_2}\right] \\
\bar{\Gamma}_r = \left[\underline{\Gamma_r}^{-1} + \phi^{-1} u_{r_1}' u_{r_1}\right]^{-1},
\end{gathered}
$$

while the matrix of the projection of the predictors is sampled from

$$
\begin{gathered}
vec\left(\Gamma'\right) \mid \left(y^T, y_0, \theta_{\neq G}\right) \sim N\left(\bar{\gamma}_x, \bar{\Gamma}_x\right) \\
\bar{\gamma}_x = \bar{\Gamma}_x \left[\underline{\Gamma_x}^{-1} \underline{\gamma_x} + vec\left[U_r' U_x \Phi^{-1}\right]\right] \\
\bar{\Gamma}_x = \left[\underline{\Gamma_x}^{-1} + \left(\Phi^{-1} \otimes U_r' U_r\right)\right]^{-1}.
\end{gathered}
$$

- Sampling π:

Since the likelihood implies the following kernel for π

$$\exp\left\{-\frac{1}{2}\left(y-\tilde{Z}\pi\right)' V^{-1}\left(y-\tilde{Z}\pi\right)\right\} \propto \exp\left\{-\frac{1}{2}\left(\pi-\hat{\pi}\right)'\tilde{Z}'V^{-1}\tilde{Z}\left(\pi-\hat{\pi}\right)\right\}$$
$$\hat{\pi}=\left(\tilde{Z}'V^{-1}\tilde{Z}\right)^{-1}\left(\tilde{Z}'V^{-1}y\right), \quad V^{-1}=\Sigma^{-1}\otimes I_T$$

we can sample it as

$$\pi \mid \left(y^T, y_0, \theta_{\neq\pi}\right) \sim TN_{\mathcal{I}(\|B\|<1)}\left(\bar{\pi}, \bar{C}\right)$$
$$\bar{\pi}=\bar{C}\left[\underline{C}^{-1}\underline{\pi}+\tilde{Z}'V^{-1}y\right]$$
$$\bar{C}=\left[\underline{C}^{-1}+\tilde{Z}'V^{-1}\tilde{Z}\right]^{-1}$$

Of course, this is only a compact way of expressing the terms, because the dimension of V^{-1} would make it infeasible. In the case of predictability, $Z_r = Z_x$, the computations are easier as in a SURE where the right hand variables are the same in all equations.

A.2 Fat Tails: Student-t

- Sampling $\left(\nu, \xi^T\right)$:

The block $\left(\nu, \xi^T\right)$ can be decomposed as

$$p\left(\nu, \xi^T \mid y^T, y_0, \theta_{\neq\nu}\right)=p\left(\nu \mid y^T, y_0, \theta_{\neq\nu}\right) p\left(\xi^T \mid y^T, y_0, \theta\right)$$

because it turns out that it is better to integrate out ξ^T when sampling ν. See Chib, Nardari and Shephard (2001b).

Since

$$\begin{pmatrix} \sqrt{\frac{\xi_t}{\nu-2}}u_{r,t} \\ u_{xt}-\Gamma u_{r,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{rr} & 0 \\ 0 & \Phi \end{pmatrix}\right),$$

conditioning on the rest of parameters and the data, only $u_{r,t}$ gives information about ν

$$u_{r,t}=\sqrt{\frac{\nu-2}{\nu}}t\left(0, \Sigma_{rr}, \nu\right)$$

and it can be drawn from

$$p\left(\nu=m \mid y^T, y_0, \theta_{\neq\nu}\right) \propto \pi_m p\left(y^T \mid y_0, \theta_{\neq\nu}, \nu=m\right)$$
$$\propto p_m\left[\frac{\Gamma\left(\frac{m+2}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}\sqrt{\frac{1}{m\left(m-2\right)}}\right]^T\left[\prod_{t=1}^{T}\left(1+\frac{1}{m-2}u_{r,t}'\Sigma^{-1}u_{r,t}\right)\right]^{-\frac{m+2}{2}}$$

where $\Gamma(\cdot)$ is the gamma function. After that, we can see that the conditional distribution of the series of chi-squares is given by

$$
\begin{gathered}
p\left(\xi^T \mid y^T, y_0, \theta\right) \propto p\left(\xi^T \mid \theta\right) p\left(y^T \mid \theta, \xi^T\right) \\
\propto \prod_{t=1}^{T}\left[\xi_t^{(\nu/2)-1} \exp\left(-\frac{1}{2}\xi_t\right)\right]\left[\xi_t^{-1} \exp\left(-\frac{1}{2}\xi_t q_t\right)\right] \\
q_t = \frac{1}{\nu-2} u_{r,t}' \Sigma^{-1} u_{r,t}
\end{gathered}
$$

and each one can be sampled as

$$
\xi_t \mid \left(y^T, \theta\right) \sim G\left(\frac{1}{2}(\nu+2), \frac{1}{2}(1+q_t)\right)
$$

independently of the rest.

- Sampling y_0:

Recalling the definition of Σ^* in the paper, the conditional distribution of interest is proportional to

$$
\exp\left\{-\frac{1}{2}\left(y_0 - \underline{\mu}\right)' \underline{\Sigma}^{-1}\left(y_0 - \underline{\mu}\right)\right\} \exp\left\{-\frac{1}{2}\left(y_1 - a - By_0\right)'\left(\Sigma^*\right)^{-1}\left(y_1 - a - By_0\right)\right\}
$$

and therefore we can sample from

$$
\begin{gathered}
y_0 \mid \left(y^T, \xi^T, \theta\right) \sim N\left(\bar{\mu}, \bar{\Sigma}\right) \\
\bar{\mu} = \bar{\Sigma}\left[\underline{\Sigma}^{-1}\underline{\mu} + B'\left(\Sigma^*\right)^{-1}\left(y_1 - a\right)\right] \\
\bar{\Sigma} = \left[\underline{\Sigma}^{-1} + B'\left(\Sigma^*\right)^{-1} B\right]^{-1}.
\end{gathered}
$$

- Sampling $(\sigma_{11}, \phi, \Phi)$:

Again, we can sample the block $\theta_S = (\sigma_{11}, \phi, \Phi)$ as three independent pieces. Defining

$$
u_{r_1,t}^* = \sqrt{\frac{\xi_t}{\nu-2}} u_{r_1,t}, \quad u_{r_2,t}^* = \sqrt{\frac{\xi_t}{\nu-2}} u_{r_2,t}
$$

we can return to a similar structure to the normal model,

$$
\begin{pmatrix} u_{r_1,t}^* \\ u_{r_2,t}^* - \gamma u_{r_1,t}^* \\ u_{x,t} - \Gamma u_{r,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \Phi \end{pmatrix}\right),
$$

conditioning on the rest of parameters and the data. Therefore, we can sample

$$\sigma_{11} \mid \left(y^T, y_0, \xi^T, \theta_{\neq S}\right) \sim IG\left(\bar{\alpha}_1, \bar{\beta}_1\right)$$
$$\bar{\alpha}_1 = \underline{\alpha}_1 + \frac{T}{2}, \quad \bar{\beta}_1 = \underline{\beta}_1 + \frac{\left(u^*_{r_1}\right)' u^*_{r_1}}{2},$$

and

$$\phi \mid \left(y^T, y_0, \xi^T, \theta_{\neq S}\right) \sim IG\left(\bar{\alpha}_2, \bar{\beta}_2\right)$$
$$\bar{\alpha}_2 = \underline{\alpha}_2 + \frac{T}{2}, \quad \bar{\beta}_2 = \underline{\beta}_2 + \frac{\left(\tilde{u}^*_{r_2}\right)' \tilde{u}^*_{r_2}}{2}$$
$$\tilde{u}^*_{r_2} = u^*_{r_2} - \gamma u^*_{r_1},$$

while the last variance does not need any adjustment in the residuals,

$$\Phi \mid \left(y^T, y_0, \xi^T, \theta_{\neq S}\right) \sim IW\left(\bar{\nu}, \bar{S}\right)$$
$$\bar{\nu} = \underline{\nu} + T, \quad \bar{S} = \left[\underline{S}^{-1} + \tilde{U}_x' \tilde{U}_x\right]^{-1}$$
$$\tilde{U}_x = U_x - U_r \Gamma'.$$

- Sampling (γ, Γ):

  Defining $\theta_G = (\gamma, \Gamma)$, we can use the previous adjusted residuals and write

$$\begin{pmatrix} u^*_{r_2,t} - \gamma u^*_{r_1,t} \\ u_{x,t} - \Gamma u_{r,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \phi & 0 \\ 0 & \Phi \end{pmatrix}\right),$$

conditioning on the rest of parameters and the data. Therefore, similar to the previous model,

$$\gamma \mid \left(y^T, y_0, \xi^T, \theta_{\neq G}\right) \sim N\left(\bar{\gamma}_r, \bar{\Gamma}_r\right)$$
$$\bar{\gamma}_r = \bar{\Gamma}_r \left[\underline{\Gamma}^{-1} \underline{\gamma} + \phi^{-1}\left(u^*_{r_1}\right)' u^*_{r_2}\right]$$
$$\bar{\Gamma}_r = \left[\underline{\Gamma_r}^{-1} + \phi^{-1}\left(u^*_{r_1}\right)' u^*_{r_1}\right]^{-1},$$

and

$$vec\left(\Gamma'\right) \mid \left(y^T, y_0, \xi^T, \theta_{\neq G}\right) \sim N\left(\bar{\gamma}_x, \bar{\Gamma}_x\right)$$
$$\bar{\gamma}_x = \bar{\Gamma}_x \left[\underline{\Gamma_x}^{-1} \underline{\gamma_x} + vec\left[U_r' U_x \Phi^{-1}\right]\right]$$
$$\bar{\Gamma}_x = \left[\underline{\Gamma_x}^{-1} + \left(\Phi^{-1} \otimes U_r' U_r\right)\right]^{-1}.$$

- Sampling π:

The original system

$$r = (I_2 \otimes Z_r)\,\pi_r + u_r$$
$$x = (I_2 \otimes Z_x)\,\pi_x + u_x$$

can be transformed as it has been shown in the previous draws into

$$r^* = (I_2 \otimes Z_r^*)\,\pi_r + u_r^*$$
$$x = (I_2 \otimes Z_x)\,\pi_x + u_x$$
$$\left(\begin{array}{c} u_{r,t}^* \\ u_{x,t} \end{array} \right) \sim N(0, \Sigma)$$

to return to a normal model. Note that only the equations of the returns need to be adjusted

$$r_t^* = \sqrt{\frac{\xi_t}{\nu - 2}} r_t, \quad z_{r,t}^* = \sqrt{\frac{\xi_t}{\nu - 2}} z_{r,t}$$

This is a SURE with different variables even in the case of predictability in the conditional mean. Then, defining

$$y^* = \left(\begin{array}{c} r^* \\ x \end{array} \right), \quad \tilde{Z}^* = \left(\begin{array}{cc} I_2 \otimes Z_r^* & 0 \\ 0 & I_2 \otimes Z_x \end{array} \right)$$

we can finally draw from

$$\pi \mid \left(y^T, y_0, \xi^T, \theta_{\neq \pi}\right) \sim TN_{\mathcal{I}(\|B\|<1)}\left(\bar{\pi}, \bar{C}\right)$$
$$\bar{\pi} = \bar{C}\left[\underline{C}^{-1}\underline{\pi} + \left(\tilde{Z}^*\right)' V^{-1} y^*\right]$$
$$\bar{C} = \left[\underline{C}^{-1} + \left(\tilde{Z}^*\right)' V^{-1} \tilde{Z}^*\right]^{-1}.$$

A.3 Nonlinear Dynamics: Markov-switching

- Sampling s^T:

The sample regimes are drawn following Carter and Khon (1994) who advocate the use of a multimove sampler instead of a singlemove one since it is more efficient and the convergence is faster. The required distribution is

$$p\left(s^T \mid y^T, y_0, s_0\right) \propto p\left(s_T \mid y^T, y_0, s_0\right) \prod_{t=1}^{T-1} p\left(s_t \mid y^T, y_0, s_0, \{s_j\}_{j=t+1}^T\right)$$
$$= p\left(s_T \mid y^T, y_0, s_0\right) \prod_{t=1}^{T-1} p\left(s_t \mid y^t, y_0, s_0, s_{t+1}\right)$$
$$p\left(s_t = j \mid y^t, y_0, s_0, s_{t+1}\right) \propto p\left(s_t = j \mid y^t, y_0, s_0\right) p\left(s_{t+1} \mid s_t = j\right)$$

where, for ease of notation, θ has been suppressed from all the conditioning sets.

To implement that sampler, the filter developed by Hamilton (1989) must be run before. The inference about the unobserved regime can be computed for given parameter values by iterations , $t = 1, ..., T$, on the system

$$\begin{gathered} p\left(s_t \mid y^{t-1}\right) = p\left(s_t \mid s_{t-1}\right) p\left(s_{t-1} \mid y^{t-1}\right) \\ p\left(s_t \mid y^t\right) \propto p\left(s_t \mid y^{t-1}\right) p\left(y_t \mid s_t, y^{t-1}\right) \end{gathered}$$

where we are conditioning on (y_0, s_0) and θ. Finally, it is easy to sample from s_T to s_1 given the previous computed probabilities.

- Sampling s_0:

To sample the regime in $t = 0$ is easy since we condition on the rest of regimes

$$\begin{gathered} p\left(s_0 = j \mid y^T, y_0, s^T, \theta\right) = p\left(s_0 = j \mid s_1, \theta\right) \\ \propto p\left(s_0 = j \mid \theta\right) p\left(s_1 \mid \theta, s_0 = j\right) \propto p_j p\left(s_1 \mid \theta, s_0 = j\right). \end{gathered}$$

- Sampling (p_{00}, p_{11}):

Defining the block $\theta_P = (p_{00}, p_{11})$, the sampler of probabilities is based on

$$p\left(\theta_P \mid y^T, y_0, s_0, s^T, \theta_{\neq P}\right) = p\left(\theta_P \mid s_0, s^T\right) \propto p\left(\theta_P\right) p\left(s^T \mid \theta_P, s_0\right).$$

For instance, in the case of p_{00}, the kernel is

$$p_{00}^{\underline{n}_{00}-1} \left(1 - p_{00}\right)^{\underline{n}_{01}-1} p_{00}^{n_{00}} \left(1 - p_{00}\right)^{1-n_{01}}$$

where n_{ij} is the number of transitions from $s_t = i$ to $s_{t+1} = j$.

Therefore,.we can draw it from

$$\begin{gathered} p_{00} \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq P}\right) \sim B\left(\bar{n}_{00}, \bar{n}_{01}\right) \\ \bar{n}_{00} = \underline{n}_{00} + n_{00}, \quad \bar{n}_{01} = \underline{n}_{01} + n_{01} \end{gathered}$$

By a similar argument,

$$\begin{gathered} p_{11} \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq P}\right) \sim B\left(\bar{n}_{11}, \bar{n}_{10}\right) \\ \bar{n}_{11} = \underline{n}_{11} + n_{11}, \quad \bar{n}_{10} = \underline{n}_{10} + n_{10}. \end{gathered}$$

- Sampling y_0:

Noting that

$$p\left(y_0 \mid y^T, s_0, s^T, \theta\right) = p\left(y_0 \mid y_1, s_1, \theta\right) \propto p\left(y_0 \mid \theta\right) p\left(y_1 \mid \theta, s_1, y_0\right)$$

then for $s_1 = j$,

$$\begin{aligned} y_0 \mid \left(y^T, s_0, s^T, \theta\right) &\sim N\left(\bar{\mu}, \bar{\Sigma}\right) \\ \bar{\mu} &= \bar{\Sigma}\left[\underline{\Sigma}^{-1}\underline{\mu} + B_j'\Sigma_j^{-1}\left(y_1 - a_j\right)\right] \\ \bar{\Sigma} &= \left[\underline{\Sigma}^{-1} + B_j'\Sigma_j^{-1}B_j\right]^{-1}. \end{aligned}$$

- Sampling $\left(\sigma_{11}^0, \phi^0, \phi^1, \Phi\right)$:

For the variance, we can get independence again in the sampling of the block $\theta_S = \left(\sigma_{11,0}, \phi^0, \phi^1, \Phi\right)$, now between σ_{11}^0, $\left(\phi^0, \phi^1\right)$, and Φ. The distribution of u_{1t} given s^T can be decomposed as

$$\begin{pmatrix} u_{r_1,t} \\ u_{r_2,t} - \gamma_{s_t} u_{r_1,t} \\ u_{x,t} - \Gamma u_{r,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11,s_t} & 0 & 0 \\ 0 & \phi_{s_t} & 0 \\ 0 & 0 & \Phi \end{pmatrix}\right),$$

conditioning on the rest of parameters and the data. Given that

$$\begin{aligned} u_{r_1,t} &\sim N\left(0, \sigma_{11,s_t}\right) \\ \sigma_{11,s_t} &= \sigma_{11}^0\left(1 + h_1 s_t\right) \\ \sigma_{11}^1 &= \sigma_{11}^0\bar{h}, \quad \bar{h} > 1 \end{aligned}$$

it is easy to see that

$$u_{r_1,t}^* = \frac{u_{r_1,t}}{\sqrt{1 + h_1 s_t}} \sim N\left(0, \sigma_{11}^0\right).$$

Therefore,

$$\begin{aligned} \sigma_{11}^0 \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq S}\right) &\sim IG\left(\bar{\alpha}_{10}, \bar{\beta}_{10}\right) \\ \bar{\alpha}_{10} = \underline{\alpha}_{10} + \frac{T}{2}, \quad \bar{\beta}_{10} &= \underline{\beta}_{10} + \frac{\left(u_{r_1}^*\right)' u_{r_1}^*}{2} \end{aligned}$$

The residual variance of the second return needs the division of the sample by means of the regime variable $s_t = 0$ and 1. We can split the sample in $\left(Y^0, Z^0, U^0\right)$ and $\left(Y^1, Z^1, U^1\right)$, where $U^j = Y^j - Z^j\Pi^j$. 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 posterior.

$$\phi^j \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq S}\right) \sim IG\left(\bar{\alpha}_{2j}, \bar{\beta}_{2j}\right)$$
$$\bar{\alpha}_{2j} = \underline{\alpha}_{2j} + \frac{T_j}{2}, \quad \bar{\beta}_{2j} = \underline{\beta}_{2j} + \frac{\left(\tilde{u}^j_{r_2}\right)' \tilde{u}^j_{r_2}}{2}$$
$$\tilde{u}^j_{r_2} = u^j_{r_2} - \gamma^j u^j_{r_1}.$$

On the other hand, the draw of the residual variance of the predictors uses the whole sample

$$\Phi \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq S}\right) \sim IW\left(\bar{\nu}, \bar{S}\right)$$
$$\bar{\nu} = \underline{\nu} + T, \quad \bar{S} = \left[\underline{S}^{-1} + \tilde{U}_x' \tilde{U}_x\right]^{-1}$$
$$\tilde{U}_x = U_x - U_r \Gamma'.$$

- Sampling $\bar{h}$:

By similar arguments as in the sampling of σ^0_{11}, conditioning on the rest of prameters,

$$u^{**}_{r_1,t} = \frac{u_{r_1,t}}{\sqrt{\sigma^0_{11}}} \sim N\left(0, 1 + h_1 s_t\right),$$

with $1 + h_1 s_t$ equal to 1 or $\bar{h}$. Then, we can draw from

$$\bar{h} \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq \bar{h}}\right) \sim TIG_{\mathcal{I}(\bar{h}>1)}\left(\bar{\alpha}_{11}, \bar{\beta}_{11}\right)$$
$$\bar{\alpha}_{11} = \underline{\alpha}_{11} + \frac{T_1}{2}, \quad \bar{\beta}_{11} = \underline{\beta}_{11} + \frac{\left(u^{**}_{r_1,1}\right)' u^{**}_{r_1,1}}{2},$$

where T_1 is the number of observations such that $s_t = 1$ and $u^{**}_{r_1 1,1}$ is the corresponding part of $u^{**}_{r_1}$. If $s_t = 0$ for every t in a particular iteration of the Gibbs sampler then we would sample directly from the prior.

- Sampling $\left(\gamma^0, \gamma^1, \Gamma\right)$:

Defining $\theta_G = \left(\gamma^0, \gamma^1, \Gamma\right)$, the coefficient of the projection of the second return relies also on the division of the sample as seen for $\left(\phi^0, \phi^1\right)$ and is drawn from

$$\gamma^j \mid \left(y^T, y_0, \theta_{\neq G}\right) \sim N\left(\bar{\gamma}_{rj}, \bar{\Gamma}_{rj}\right)$$
$$\bar{\gamma}_{rj} = \bar{\Gamma}_{rj}\left[\underline{\Gamma}^{-1}\underline{\gamma} + \left(\phi^j\right)^{-1}\left(u^j_{r_1}\right)' u^j_{r_2}\right]$$
$$\bar{\Gamma}_{rj} = \left[\underline{\Gamma}_{rj}{}^{-1} + \left(\phi^j\right)^{-1}\left(u^j_{r_1}\right)' u^j_{r_1}\right]^{-1},$$

while the matrix of the projection of the predictors is drawn from the whole sample

$$\begin{aligned} vec\left(\Gamma'\right) \mid \left(y^T, y_0, \theta_{\neq G}\right) &\sim N\left(\bar{\gamma}_x, \bar{\Gamma}_x\right) \\ \bar{\gamma}_x &= \bar{\Gamma}_x \left[\underline{\Gamma}_x{}^{-1} \underline{\gamma}_x + vec\left[U_r' U_x \Phi^{-1}\right]\right] \\ \bar{\Gamma}_x &= \left[\underline{\Gamma}_x{}^{-1} + \left(\Phi^{-1} \otimes U_r' U_r\right)\right]^{-1}. \end{aligned}$$

Anyway, note that U_r must be computed with the switching given by s^T.

- Sampling π:

Since the likelihood can be decomposed for π as

$$\exp\left\{-\frac{1}{2}\left(y^0 - \tilde{Z}^0\pi\right)'\left(V^0\right)^{-1}\left(y^0 - \tilde{Z}^0\pi\right)\right\}\exp\left\{-\frac{1}{2}\left(y^1 - \tilde{Z}^1\pi\right)'\left(V^1\right)^{-1}\left(y^1 - \tilde{Z}^1\pi\right)\right\},$$

we have something like two different SURE systems and the sampling is

$$\begin{aligned} \pi \mid \left(y^T, y_0, s_0, s^T, \theta_{\neq\pi}\right) &\sim TN_{\mathcal{I}(\|B^0\|<1, \|B^1\|<1)}\left(\bar{\pi}, \bar{C}\right) \\ \bar{\pi} &= \bar{C}\left[\underline{C}^{-1}\underline{\pi} + \left(\tilde{Z}^0\right)'\left(V^0\right)^{-1} y^0 + \left(\tilde{Z}^1\right)'\left(V^1\right)^{-1} y^1\right] \\ \bar{C} &= \left[\underline{C}^{-1} + \left(\tilde{Z}^0\right)'\left(V^0\right)^{-1}\tilde{Z}^0 + \left(\tilde{Z}^1\right)'\left(V^1\right)^{-1}\tilde{Z}^1\right]^{-1}. \end{aligned}$$

B Tables

Table 2.1: Descriptive statistics of excess returns in monthly percentage

	Stocks	*Bonds*
Mean	0.480	0.047
Median	0.721	−0.053
Maximum	15.007	9.081
Minimum	−24.751	−7.182
S.D.	4.169	1.721
Skewness	−0.549	0.248
Kurtosis	5.900	6.998
Jarque − Bera	197.111	332.808
	(0.000)	(0.000)

The priors used for tables 3.1.1 and 3.1.2 are

$$y_0 \sim N(0_4, I_4)$$
$$\sigma_{11} \sim IG(3, 30), \quad \phi \sim IG(3, 3), \quad \Phi \sim IW(5, I_2)$$
$$\gamma \sim N(0, 0.25), \quad vec(\Gamma') \sim N(0_4, 0.25I_4)$$
$$\pi \sim TN_{\mathcal{I}(\|B\|<1)}(0_{20}, I_{20})$$

which translates into the following prior of Σ,

Table 3.1.0: Prior of Σ

	Mean	*SD*	*Median*	95%	*Bands*
$vech(\Sigma)$	15.115	15.553	11.187	4.190	47.798
	0.067	11.369	0.025	−18.998	18.172
	5.286	9.077	2.919	0.648	23.189
	0.057	12.298	0.007	−20.042	21.427
	−0.034	7.190	−0.010	−12.085	11.061
	5.577	11.305	2.481	0.278	28.343
	0.247	11.264	−0.013	−19.645	21.560
	0.011	7.010	−0.023	−11.719	12.075
	−0.021	6.524	−0.010	−11.291	11.631
	5.437	9.963	2.405	0.290	28.326

where $vech(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 3.1.1: Posterior - Normal model and predictable mean

	Mean	*SD*	*Median*	95%	*Bands*
y_0	0.212	0.992	0.216	−1.692	2.161
	0.168	0.984	0.164	−1.756	2.070
	5.613	0.111	5.612	5.396	5.832
	0.811	0.599	0.817	−0.373	1.977
$vech\,(\Sigma)$	16.474	1.049	16.431	14.559	18.615
	1.294	0.319	1.288	0.688	1.930
	2.853	0.182	2.844	2.518	3.231
	−0.638	0.045	−0.637	−0.730	−0.556
	−0.049	0.015	−0.049	−0.078	−0.021
	0.035	0.002	0.035	0.031	0.040
	−0.224	0.116	−0.222	−0.456	−0.001
	−0.072	0.049	−0.072	−0.168	0.022
	0.006	0.005	0.006	−0.005	0.016
	0.393	0.025	0.392	0.347	0.445

where $vech\,(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 3.1.2: Posterior - Normal model and predictable mean

	Mean	*SD*	*Median*	95%	*Bands*
π_1	−1.591	0.680	−1.595	−2.940	−0.245
	0.003	0.045	0.003	−0.085	0.090
	0.324	0.109	0.323	0.114	0.539
	0.393	0.170	0.394	0.061	0.730
	0.494	0.153	0.493	0.195	0.791
π_2	−0.277	0.363	−0.279	−0.968	0.444
	−0.059	0.019	−0.059	−0.096	−0.021
	0.107	0.045	0.107	0.016	0.196
	0.031	0.089	0.031	−0.143	0.202
	0.196	0.065	0.196	0.069	0.322
π_3	0.111	0.035	0.111	0.043	0.179
	0.039	0.002	0.039	0.035	0.043
	−0.010	0.005	−0.010	−0.020	0.000
	0.972	0.009	0.972	0.955	0.989
	−0.024	0.007	−0.024	−0.038	−0.010
π_4	0.198	0.146	0.200	−0.087	0.483
	−0.014	0.007	−0.014	−0.028	−0.001
	0.093	0.017	0.093	0.060	0.126
	0.000	0.035	0.000	−0.069	0.069
	0.839	0.024	0.839	0.793	0.887

The priors for tables 3.2.1 and 3.2.2 are

$$y_0 \sim N(0_4, I_4)$$
$$\sigma_{11} \sim IG(3, 30), \quad \phi \sim IG(3, 3), \quad \Phi \sim IW(5, I_2)$$
$$\gamma \sim N(0, 0.25), \quad vec(\Gamma') \sim N(0_4, 0.25 I_4)$$
$$\pi \sim TN_{\mathcal{I}(\|B\|<1)}(0_{12}, I_{12})$$

which translate into the same prior of Σ as in table 3.1.0.

Table 3.2.1: Posterior - Normal model and unpredictable mean

	Mean	*SD*	*Median*	95%	*Bands*
y_0	0.210	1.000	0.210	-1.736	2.215
	0.105	0.995	0.114	-1.878	2.011
	5.613	0.111	5.614	5.396	5.824
	0.765	0.601	0.772	-0.434	1.922
$vech(\Sigma)$	17.366	1.101	17.328	15.343	19.616
	1.536	0.334	1.532	0.883	2.203
	2.960	0.187	2.954	2.610	3.348
	-0.672	0.047	-0.671	-0.771	-0.586
	-0.058	0.015	-0.058	-0.089	-0.028
	0.037	0.002	0.036	0.032	0.041
	-0.240	0.124	-0.239	-0.485	0.002
	-0.076	0.050	-0.076	-0.176	0.022
	0.006	0.006	0.006	-0.005	0.018
	0.394	0.025	0.393	0.347	0.446

where $vech(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 3.2.2: Posterior - Normal model and unpredictable mean

	Mean	*SD*	*Median*	95%	*Bands*
π_1	0.462	0.183	0.462	0.100	0.823
π_2	0.045	0.076	0.045	-0.104	0.195
π_3	0.031	0.024	0.031	-0.014	0.079
	0.039	0.001	0.039	0.037	0.042
	0.002	0.003	0.002	-0.003	0.008
	0.987	0.006	0.988	0.976	0.998
	-0.005	0.004	-0.005	-0.013	0.003
π_4	0.169	0.146	0.168	-0.117	0.452
	-0.015	0.007	-0.015	-0.030	-0.002
	0.099	0.017	0.099	0.065	0.133
	0.005	0.035	0.005	-0.064	0.073
	0.849	0.025	0.848	0.800	0.897

The priors used in tables 4.1.1, 4.1.2 and 4.1.3 are

$$y_0 \sim N\left(0_4, I_4\right)$$
$$\sigma_{11} \sim IG\left(3, 30\right), \quad \phi \sim IG\left(3, 3\right), \quad \Phi \sim IW\left(5, I_2\right)$$
$$\gamma \sim N\left(0, 0.25\right), \quad vec\left(\Gamma'\right) \sim N\left(0_4, 0.25 I_4\right)$$
$$\pi \sim TN_{\mathcal{I}(\|B\|<1)}\left(0_{20}, I_{20}\right)$$

which translate into the same prior of Σ as in table 3.1.0, plus a prior on ν that is shown in table 4.1.1.

Table 4.1.1: Degrees of freedom - Student-t model and predictable mean

ν	*Prior*	*Posterior*
4	0.125	0.232
6	0.125	0.231
8	0.125	0.164
10	0.125	0.000
15	0.125	0.001
20	0.125	0.030
30	0.125	0.223
60	0.125	0.118

Table 4.1.2: Posterior - Student-t model and predictable mean

	Mean	*SD*	*Median*	95%	*Bands*
y_0	0.199	1.005	0.193	−1.768	2.171
	0.220	0.987	0.226	−1.720	2.153
	5.638	0.120	5.636	5.407	5.878
	0.841	0.597	0.842	−0.334	2.006
$vech\,(\Sigma)$	16.732	1.726	16.454	13.965	20.744
	1.339	0.340	1.332	0.697	2.038
	2.610	0.220	2.600	2.208	3.074
	−0.648	0.070	−0.638	−0.809	−0.535
	−0.051	0.015	−0.051	−0.082	−0.023
	0.036	0.003	0.035	0.031	0.043
	−0.230	0.118	−0.227	−0.473	−0.005
	−0.067	0.044	−0.067	−0.156	0.017
	0.006	0.005	0.006	−0.005	0.017
	0.393	0.025	0.392	0.347	0.446

where $vech\,(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 4.1.3: Posterior - Student-t model and predictable mean

	Mean	*SD*	*Median*	95%	*Bands*
π_1	−1.602	0.645	−1.600	−2.849	−0.324
	−0.009	0.045	−0.009	−0.098	0.082
	0.304	0.108	0.303	0.096	0.519
	0.413	0.162	0.414	0.092	0.720
	0.473	0.146	0.471	0.189	0.759
π_2	−0.272	0.329	−0.269	−0.921	0.372
	−0.046	0.019	−0.046	−0.083	−0.011
	0.112	0.045	0.112	0.024	0.200
	0.018	0.082	0.017	−0.142	0.180
	0.211	0.062	0.211	0.091	0.330
π_3	0.118	0.035	0.119	0.049	0.188
	0.040	0.002	0.040	0.036	0.044
	−0.010	0.005	−0.010	−0.020	−0.001
	0.970	0.009	0.970	0.953	0.987
	−0.023	0.007	−0.023	−0.037	−0.009
π_4	0.200	0.145	0.203	−0.086	0.481
	−0.014	0.007	−0.014	−0.028	−0.001
	0.093	0.017	0.093	0.060	0.126
	0.000	0.035	−0.001	−0.067	0.069
	0.839	0.024	0.839	0.790	0.886

The priors used in tables 4.2.1, 4.2.2, and 4.2.3 are

$$y_0 \sim N(0_4, I_4)$$
$$\sigma_{11} \sim IG(3, 30), \quad \phi \sim IG(3, 3), \quad \Phi \sim IW(5, I_2)$$
$$\gamma \sim N(0, 0.25), \quad vec(\Gamma') \sim N(0_4, 0.25 I_4)$$
$$\pi \sim TN_{\mathcal{I}(\|B\|<1)}(0_{12}, I_{12})$$

which translate into the same prior of Σ as in table 3.1.0, plus a prior on ν that is shown in table 4.2.1.

Table 4.2.1: Degrees of freedom - Student-t model and unpredictable mean

ν	*Prior*	*Posterior*
4	0.125	0.249
6	0.125	0.259
8	0.125	0.079
10	0.125	0.000
15	0.125	0.002
20	0.125	0.043
30	0.125	0.255
60	0.125	0.114

Table 4.2.2: Posterior - Student-t model and unpredictable mean

	Mean	*SD*	*Median*	95%	*Bands*
y_0	0.192	1.009	0.196	-1.813	2.137
	0.100	0.985	0.088	-1.818	2.018
	5.640	0.122	5.638	5.403	5.883
	0.766	0.597	0.769	-0.422	1.930
$vech\,(\Sigma)$	17.519	1.761	17.260	14.751	21.539
	1.569	0.358	1.556	0.905	2.302
	2.712	0.220	2.700	2.321	3.184
	-0.675	0.071	-0.666	-0.838	-0.561
	-0.059	0.016	-0.059	-0.091	-0.030
	0.037	0.003	0.036	0.031	0.044
	-0.242	0.125	-0.239	-0.497	-0.004
	-0.072	0.046	-0.071	-0.163	0.017
	0.006	0.006	0.006	-0.005	0.018
	0.394	0.025	0.392	0.348	0.447

where $vech\,(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 4.2.3: Posterior - Student-t model and unpredictable mean

	Mean	*SD*	*Median*	95%	*Bands*
π_1	0.497	0.170	0.498	0.157	0.830
π_2	0.021	0.070	0.022	-0.113	0.160
π_3	0.038	0.026	0.037	-0.012	0.089
	0.040	0.001	0.040	0.037	0.042
	0.000	0.003	0.000	-0.006	0.006
	0.986	0.006	0.987	0.974	0.998
	-0.006	0.004	-0.006	-0.014	0.003
π_4	0.170	0.145	0.172	-0.111	0.456
	-0.015	0.007	-0.015	-0.029	-0.002
	0.098	0.017	0.098	0.065	0.131
	0.005	0.035	0.004	-0.065	0.074
	0.849	0.024	0.848	0.800	0.897

The priors used for tables 5.1.1 and 5.1.2 are

$$
\begin{gathered}
s_0 \sim Bernoulli\,(0.5) \\
p_{00} \sim B\,(1,1)\,, \quad p_{11} \sim B\,(1,1) \\
y_0 \sim N\,(0_4, I_4) \\
\sigma_{11}^0 \sim IG\,(3,20)\,, \quad \phi^0 \sim IG\,(3,3)\,, \quad \phi^1 \sim IG\,(3,3)\,, \quad \Phi \sim IW\,(5, I_2) \\
\bar{h} \sim TIG_{\mathcal{I}(\bar{h}>1)}\,(3,4) \\
\gamma^0 \sim N\,(0, 0.25)\,, \quad \gamma^0 \sim N\,(0, 0.25)\,, \quad vec\left(\Gamma'\right) \sim N\,(0_4, 0.25 I_4) \\
\pi \sim TN_{\mathcal{I}(\|B^0\|<1, \|B^1\|<1)}\,(0_{30}, I_{30})
\end{gathered}
$$

which translate into the following prior of Σ^0 and Σ^1

Table 5.1.0: Prior of Σ^0 and Σ^1

	Mean	*SD*	*Median*	95%	*Bands*
$vech\left(\Sigma^0\right)$	10.052	14.096	7.452	2.733	32.156
	−0.108	6.441	−0.037	−12.498	12.335
	3.861	4.761	2.383	0.614	15.843
	−0.025	7.401	−0.004	−13.995	13.701
	0.026	4.395	0.016	−8.071	8.404
	3.981	6.784	1.973	0.261	19.778
	−0.044	7.399	−0.062	−14.553	14.278
	−0.002	4.431	0.004	−7.940	8.061
	−0.007	4.247	−0.012	−7.975	7.844
	4.041	6.554	2.002	0.264	21.349
$vech\left(\Sigma^1\right)$	23.673	34.725	14.535	4.034	98.650
	0.047	20.204	0.019	−31.841	32.506
	7.441	17.440	3.289	0.665	37.803
	−0.140	20.855	−0.056	−37.105	35.365
	0.127	12.829	0.011	−18.235	17.768
	8.247	19.013	3.009	0.291	47.908
	0.175	22.043	−0.056	−36.131	37.315
	−0.213	13.409	0.013	−19.578	18.344
	−0.036	12.587	0.007	−18.793	18.431
	8.427	19.811	3.090	0.301	49.264

where $vech\,(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 5.1.1: Posterior - Markov-switching model and predictable mean

	Mean	*SD*	*Median*	95%	*Bands*
s_0	0.218	0.413	0.000	0.000	1.000
p_{00}	0.927	0.026	0.929	0.868	0.969
p_{11}	0.946	0.025	0.951	0.885	0.982
y_0	0.180	0.994	0.172	−1.778	2.122
	0.196	0.998	0.203	−1.776	2.142
	5.611	0.111	5.612	5.397	5.830
	0.850	0.596	0.848	−0.324	2.029
$vech\left(\Sigma^0\right)$	10.980	1.340	10.885	8.579	13.845
	−0.026	0.208	−0.033	−0.415	0.394
	0.480	0.093	0.467	0.348	0.687
	−0.426	0.053	−0.423	−0.542	−0.332
	0.001	0.008	0.002	−0.015	0.016
	0.027	0.002	0.027	0.023	0.032
	−0.131	0.080	−0.129	−0.291	0.024
	−0.009	0.009	−0.009	−0.028	0.008
	0.002	0.004	0.002	−0.006	0.010
	0.390	0.025	0.389	0.344	0.441
$vech\left(\Sigma^1\right)$	20.725	2.075	20.551	17.219	25.152
	2.300	0.628	2.267	1.150	3.597
	4.562	0.507	4.513	3.761	5.637
	−0.803	0.084	−0.796	−0.980	−0.660
	−0.087	0.027	−0.086	−0.143	−0.036
	0.042	0.004	0.041	0.035	0.049
	−0.294	0.150	−0.292	−0.602	−0.006
	−0.119	0.078	−0.117	−0.280	0.028
	0.008	0.006	0.008	−0.004	0.021
	0.396	0.026	0.395	0.348	0.448

where $vech\,(\Sigma) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 5.1.2: Posterior - Markov-switching model and predictable mean

	Mean	*SD*	*Median*	95%	*Bands*
π_1^0	−1.564	0.682	−1.573	−2.876	−0.218
	−0.042	0.059	−0.042	−0.161	0.073
	0.235	0.237	0.228	−0.211	0.718
	0.456	0.182	0.456	0.086	0.808
	0.416	0.244	0.415	−0.053	0.902
π_2^0	−0.181	0.265	−0.186	−0.678	0.354
	−0.014	0.016	−0.014	−0.045	0.017
	0.056	0.090	0.053	−0.114	0.243
	−0.007	0.074	−0.005	−0.157	0.131
	0.134	0.100	0.131	−0.051	0.337
π_3	0.114	0.035	0.113	0.046	0.183
	0.039	0.002	0.039	0.034	0.043
	−0.010	0.005	−0.010	−0.021	0.000
	0.970	0.009	0.970	0.953	0.986
	−0.023	0.007	−0.023	−0.037	−0.008
π_4	0.188	0.145	0.190	−0.097	0.472
	−0.015	0.007	−0.015	−0.028	−0.001
	0.093	0.017	0.093	0.059	0.126
	0.002	0.035	0.002	−0.066	0.071
	0.840	0.024	0.840	0.792	0.888
π_1^1	−1.675	0.850	−1.678	−3.332	−0.004
	0.051	0.053	0.051	−0.053	0.157
	0.330	0.122	0.331	0.090	0.569
	0.446	0.206	0.444	0.040	0.844
	0.469	0.174	0.470	0.130	0.804
π_2^1	−0.166	0.560	−0.166	−1.242	0.943
	−0.070	0.029	−0.069	−0.126	−0.014
	0.118	0.062	0.118	−0.003	0.240
	0.031	0.132	0.030	−0.229	0.290
	0.191	0.092	0.190	0.010	0.371

The prior for tables 5.2.1 and 5.2.2 are

$$
\begin{gathered}
s_0 \sim Bernoulli\,(0.5) \\
p_{00} \sim B\,(1,1)\,, \quad p_{11} \sim B\,(1,1) \\
y_0 \sim N\,(0_4, I_4) \\
\sigma_{11}^0 \sim IG\,(3,20)\,, \quad \phi^0 \sim IG\,(3,3)\,, \quad \phi^1 \sim IG\,(3,3)\,, \quad , \Phi \sim IW\,(5, I_2) \\
\bar{h} \sim TIG_{\mathcal{I}(\bar{h}>1)}\,(3,4) \\
\gamma^0 \sim N\,(0,0.25)\,, \quad \gamma^0 \sim N\,(0,0.25)\,, \quad vec\left(\Gamma'\right) \sim N\,(0_4, 0.25 I_4) \\
\pi \sim TN_{\mathcal{I}(\|B\|<1)}\,(0_{12}, I_{12})
\end{gathered}
$$

which translate into the same prior of Σ^0 and Σ^1 as in table 5.1.0.

Table 5.2.1: Posterior - Markov-switching model and unpredictable mean

	Mean	*SD*	*Median*	95%	*Bands*
s_0	0.141	0.348	0.000	0.000	1.000
p_{00}	0.937	0.023	0.939	0.884	0.974
p_{11}	0.959	0.020	0.962	0.911	0.986
y_0	0.212	0.996	0.214	-1.733	2.154
	0.120	0.991	0.111	-1.786	2.080
	5.611	0.110	5.611	5.394	5.830
	0.766	0.596	0.771	-0.407	1.935
$vech\left(\Sigma^0\right)$	11.489	1.315	11.398	9.107	14.277
	-0.041	0.193	-0.042	-0.418	0.340
	0.450	0.068	0.442	0.343	0.606
	-0.445	0.053	-0.442	-0.557	-0.351
	0.002	0.008	0.002	-0.013	0.016
	0.028	0.002	0.028	0.024	0.033
	-0.139	0.083	-0.137	-0.307	0.018
	-0.008	0.008	-0.008	-0.025	0.008
	0.002	0.004	0.002	-0.006	0.011
	0.390	0.025	0.389	0.345	0.443
$vech\left(\Sigma^1\right)$	21.331	1.928	21.158	18.017	25.450
	2.555	0.609	2.533	1.412	3.805
	4.612	0.439	4.571	3.852	5.587
	-0.826	0.079	-0.820	-0.994	-0.691
	-0.098	0.027	-0.097	-0.153	-0.048
	0.043	0.004	0.042	0.036	0.050
	-0.308	0.152	-0.308	-0.611	-0.019
	-0.119	0.078	-0.119	-0.277	0.031
	0.009	0.007	0.009	-0.004	0.022
	0.396	0.026	0.395	0.350	0.450

where $vech\left(\Sigma\right) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, ..., \sigma_{N1}, \sigma_{N2}, ..., \sigma_{NN})'$.

Table 5.2.2: Posterior - Markov-switching model and unpredictable mean

	Mean	*SD*	*Median*	95%	*Bands*
π_1	0.490	0.179	0.492	0.138	0.844
π_2	-0.049	0.050	-0.050	-0.148	0.052
π_3	0.030	0.024	0.029	-0.016	0.079
	0.039	0.001	0.039	0.037	0.042
	0.002	0.003	0.002	-0.003	0.008
	0.988	0.006	0.988	0.976	0.998
	-0.005	0.004	-0.005	-0.013	0.003
π_4	0.167	0.148	0.167	-0.124	0.452
	-0.015	0.007	-0.015	-0.029	-0.002
	0.099	0.017	0.099	0.066	0.132
	0.006	0.036	0.006	-0.064	0.076
	0.849	0.024	0.849	0.800	0.896

C Figures

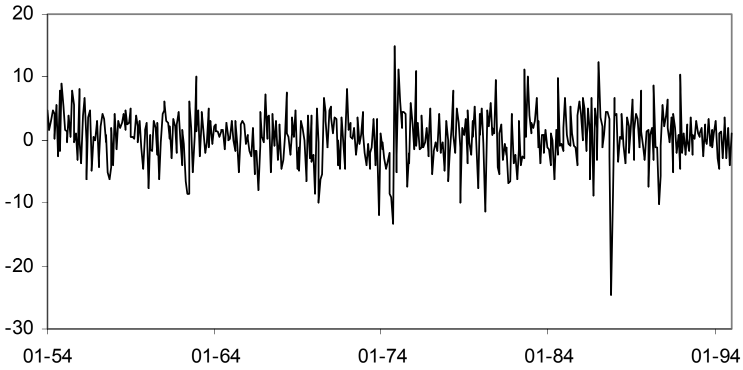

Figure 2.1: Excess stock returns in monthly percentage.

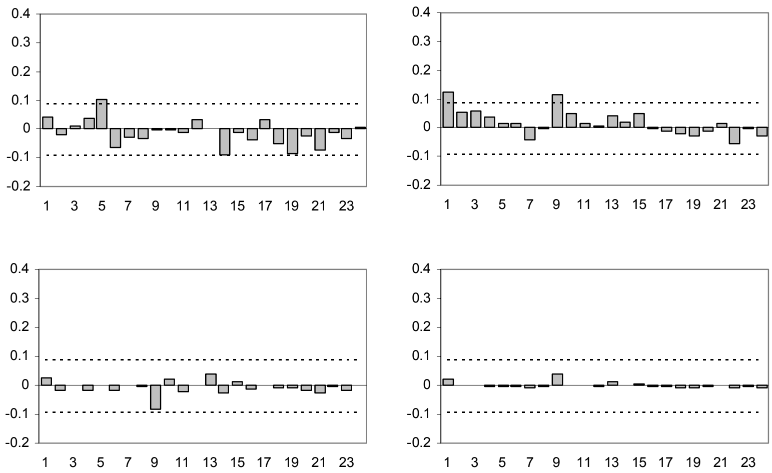

Figure 2.2: Correlograms of powers of the demeaned excess stock returns. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

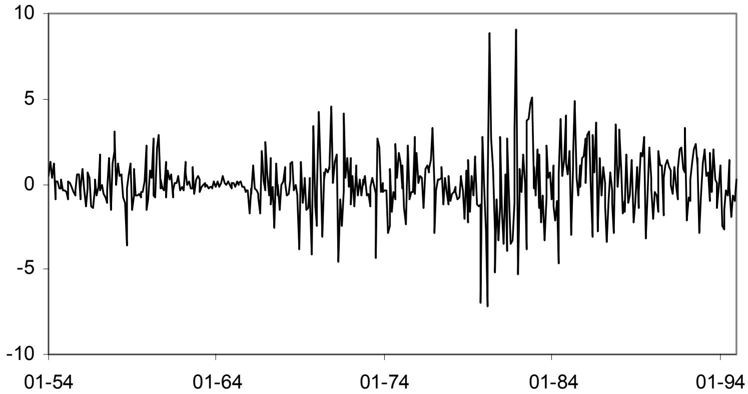

Figure 2.3: Excess bond returns in monthly percentage.

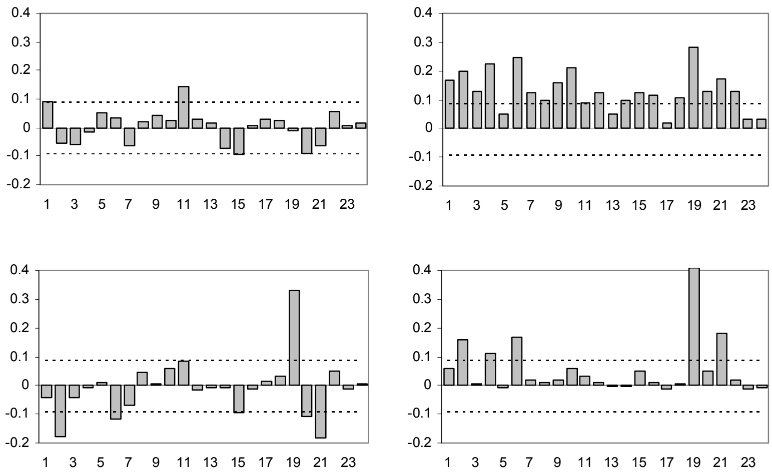

Figure 2.4: Correlograms of powers of the demeaned excess bond returns. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

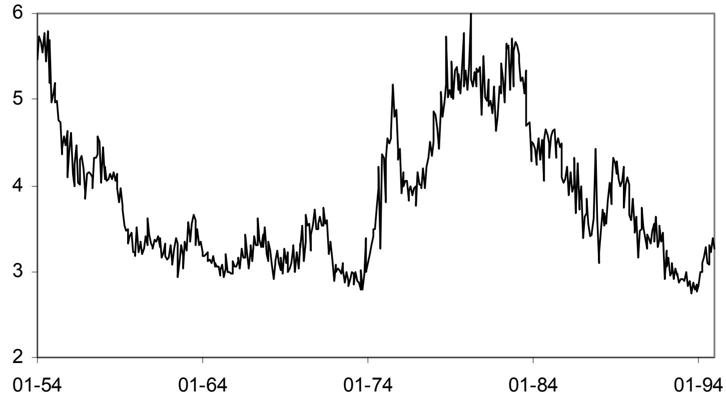

Figure 2.5: Dividend yield in percentage.

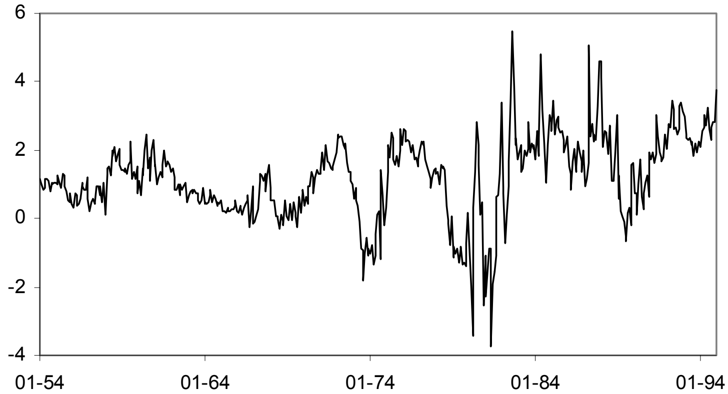

Figure 2.6: Term premium in annualized percentage.

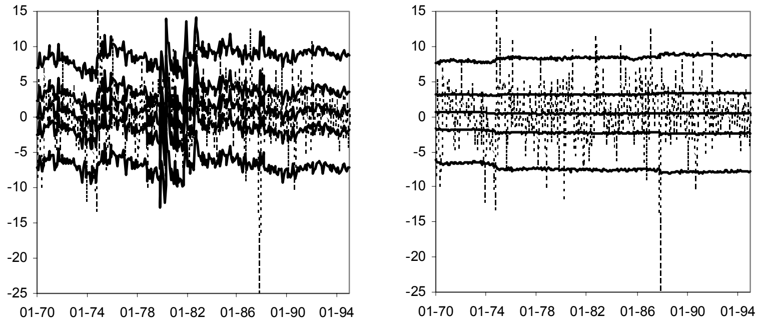

Figure 3.1: Density forecasts of excess stock returns from the normal model. The solid lines are the 2.5, 25, 50, 75, and 97.5% quantiles of the forecasts, while the dashed line is the corresponding realization. The figure on the left corresponds to the model with predictable mean, while the figure on the right corresponds to the model with unpredictable mean.

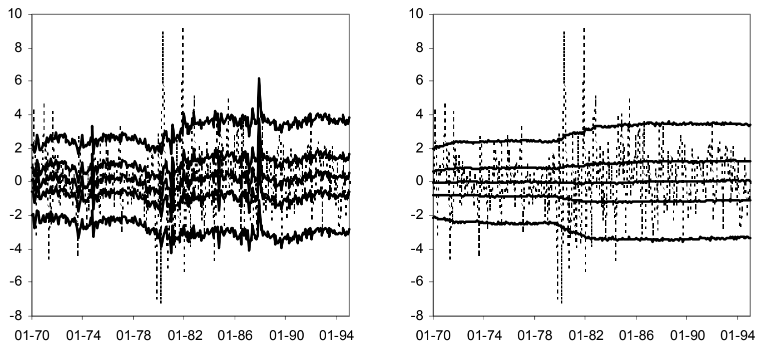

Figure 3.2: Density forecasts of excess bond returns from the normal model. The solid lines are the 2.5, 25, 50, 75, and 97.5% quantiles of the forecasts, while the dashed line is the corresponding realization. The figure on the left corresponds to the model with predictable mean, while the figure on the right corresponds to the model with unpredictable mean

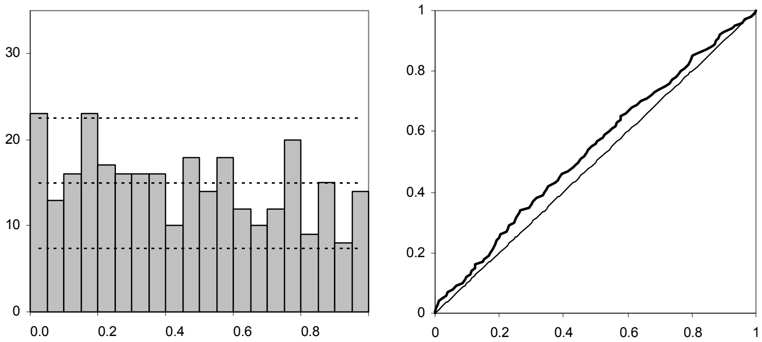

Figure 3.3: Histogram and Q-Q plot of the PIT for excess stock returns from the normal model with predictable mean.

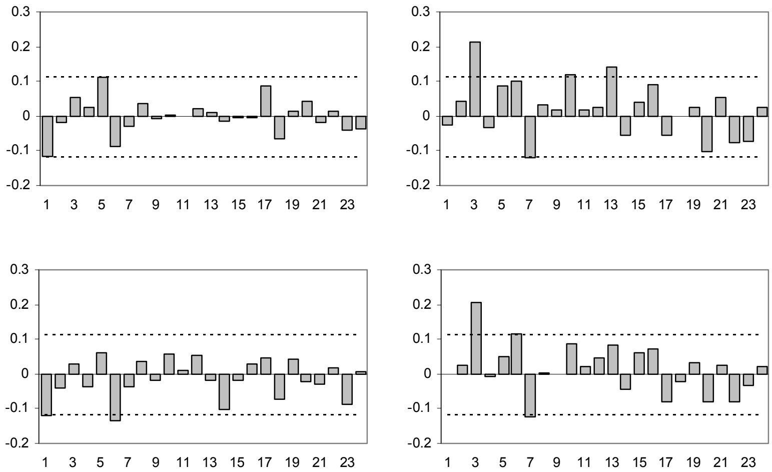

Figure 3.4: Correlograms of powers of the demeaned PIT for excess stock returns from the normal model with predictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

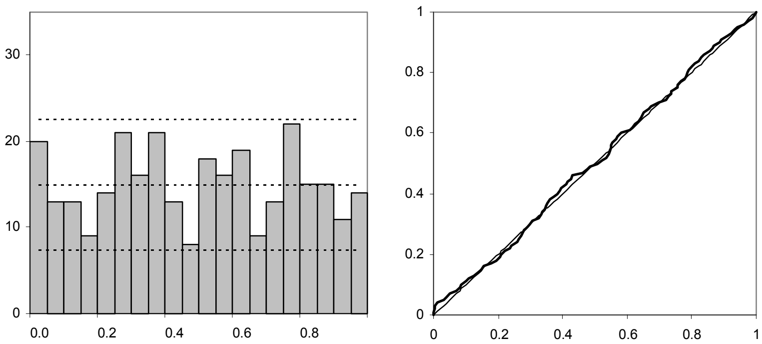

Figure 3.5: Histogram and Q-Q plot of the PIT for excess stock returns from the normal model with unpredictable mean.

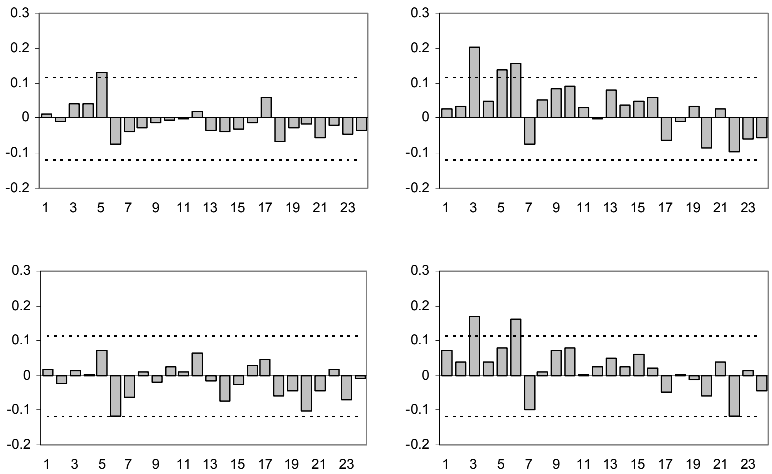

Figure 3.6: Correlograms of powers of the demeaned PIT for excess stock returns from the normal model with unpredictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

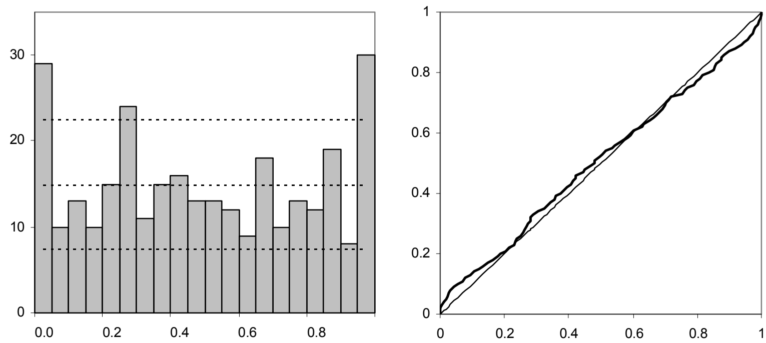

Figure 3.7: Histogram and Q-Q plot of the PIT for excess bond returns from the normal model with predictable mean.

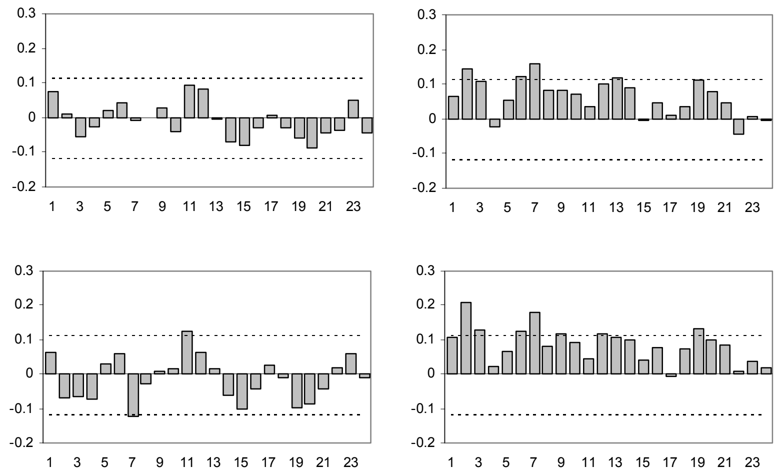

Figure 3.8: Correlograms of powers of the demeaned PIT for excess bond returns from the normal model with predictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

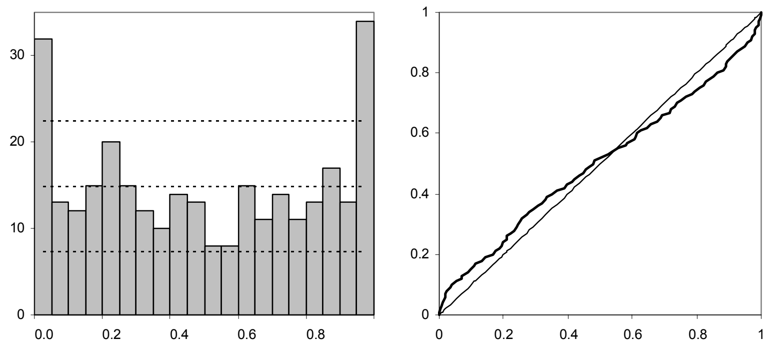

Figure 3.9: Histogram and Q-Q plot of the PIT for excess bond returns from the normal model with unpredictable mean.

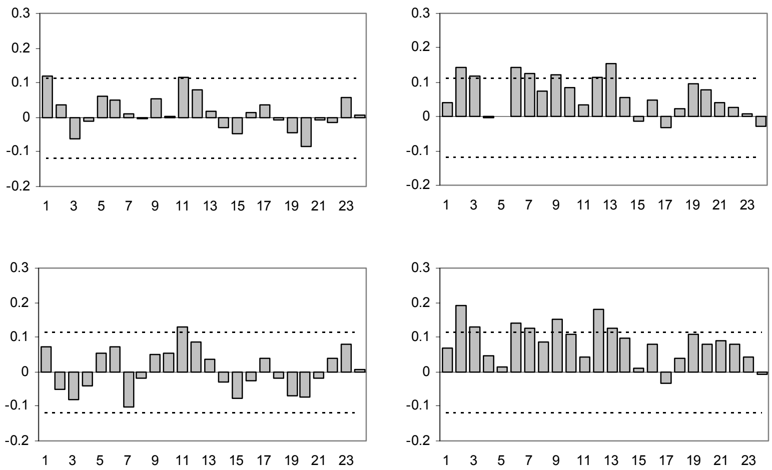

Figure 3.10: Correlograms of powers of the demeaned PIT for excess bond returns form the normal model with unpredictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

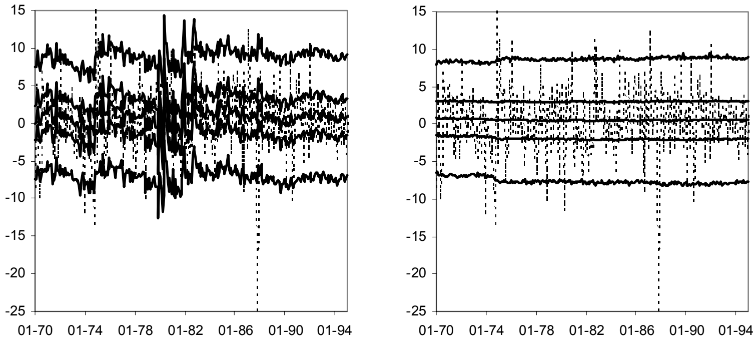

Figure 4.1: Density forecasts of excess stock returns from the Student-t model. The solid lines are the 2.5, 25, 50, 75, and 97.5% quantiles of the forecasts, while the dashed line is the corresponding realization. The figure on the left corresponds to the model with predictable mean, while the figure on the right corresponds to the model with unpredictable mean.

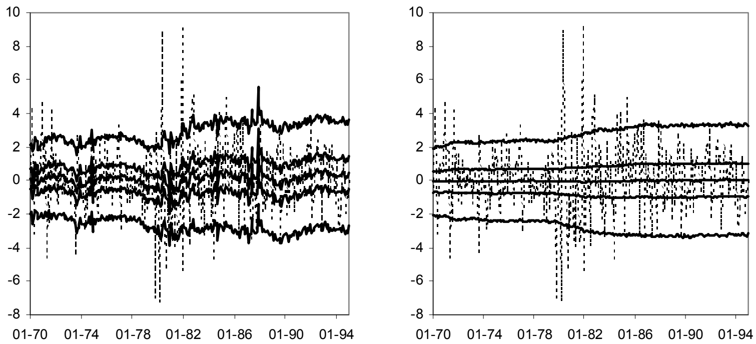

Figure 4.2: Density forecasts of excess bond returns from the Student-t model. The solid lines are the 2.5, 25, 50, 75, and 97.5% quantiles of the forecasts, while the dashed line is the corresponding realization. The figure on the left corresponds to the model with predictable mean, while the figure on the right corresponds to the model with unpredictable mean.

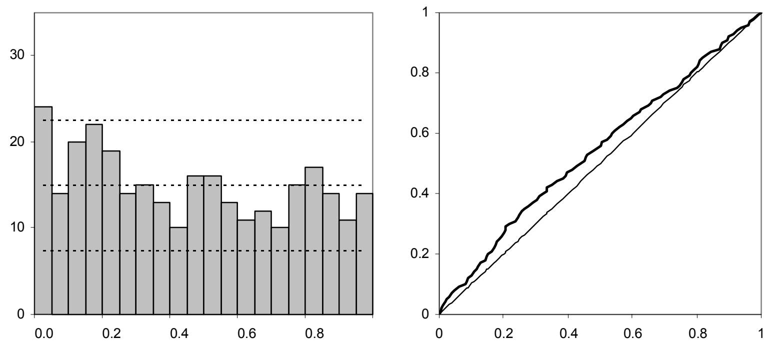

Figure 4.3: Histogram and Q-Q plot of the PIT for excess stock returns from the Student-t model with predictable mean.

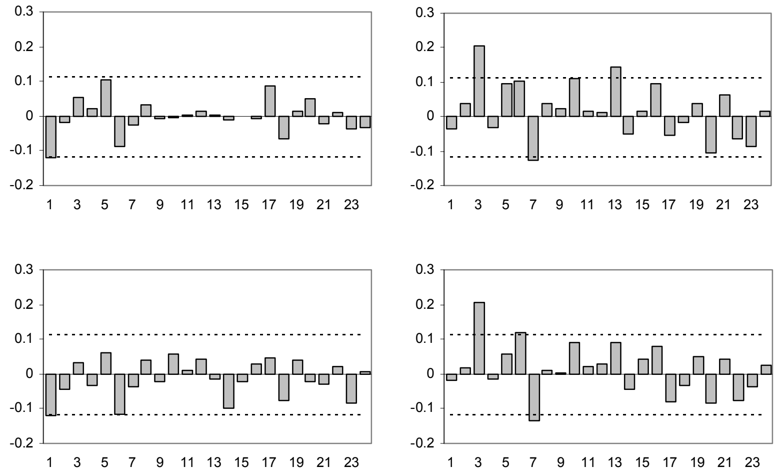

Figure 4.4: Correlograms of powers of the demeaned PIT for excess stock returns from the Student-t model with predictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

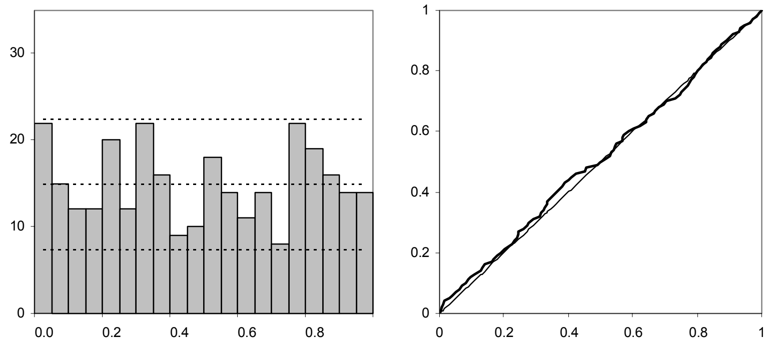

Figure 4.5: Histogram and Q-Q plot of the PIT for excess stock returns from the Student-t model with unpredictable mean.

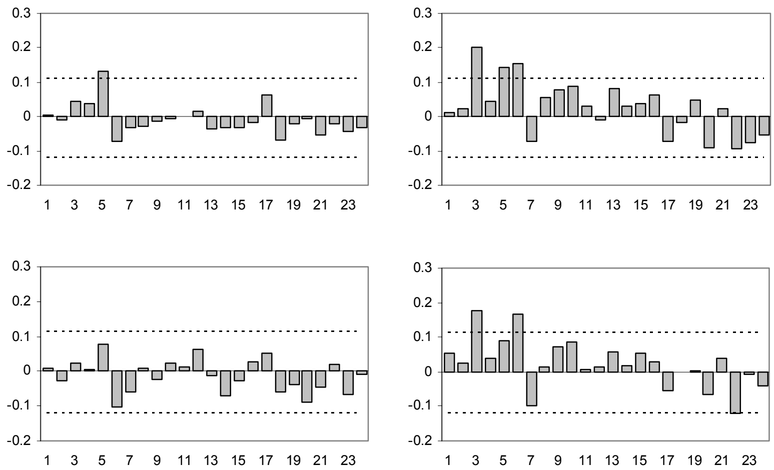

Figure 4.6: Correlograms of powers of the demeaned PIT for excess stock returns form the Student-t model with unpredictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

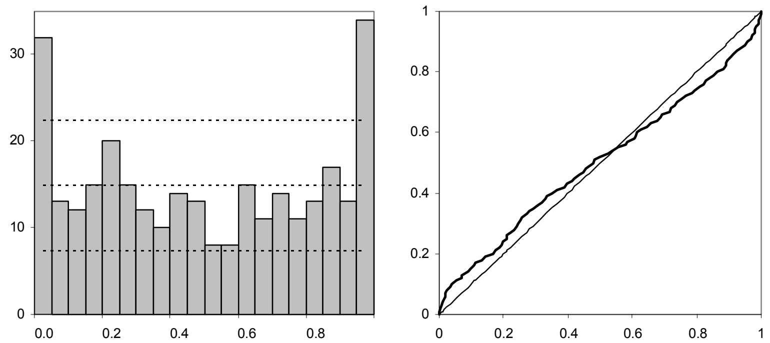

Figure 4.7: Histogram and Q-Q plot of the PIT for excess bond returns from the Student-t model with predictable mean.

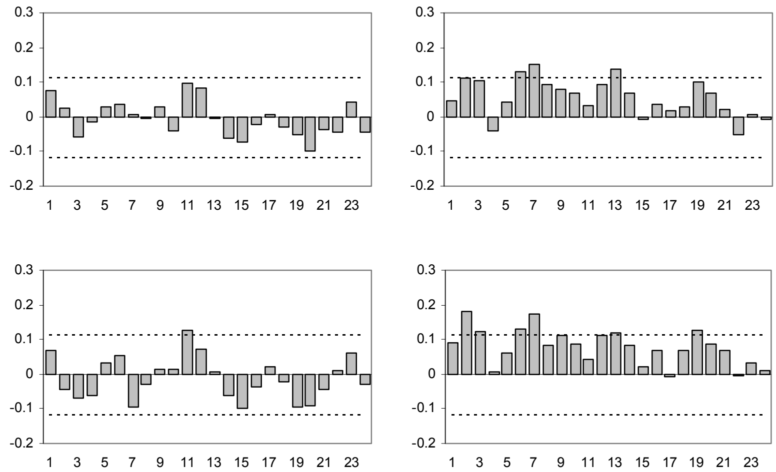

Figure 4.8: Correlograms of powers of the demeaned PIT for excess bond returns from the Student-t model with predictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

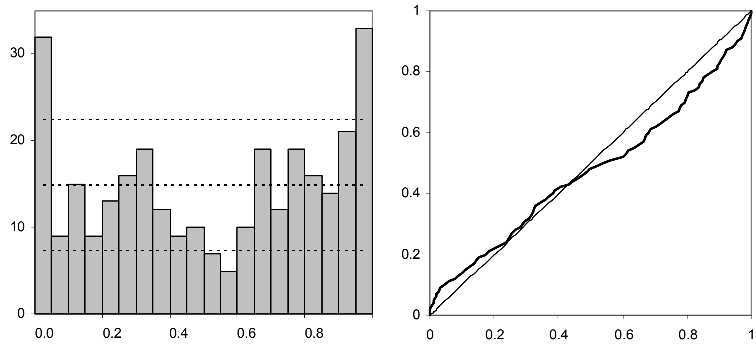

Figure 4.9: Histogram and Q-Q plot of the PIT for excess bond returns from the Student-t model with unpredictable mean.

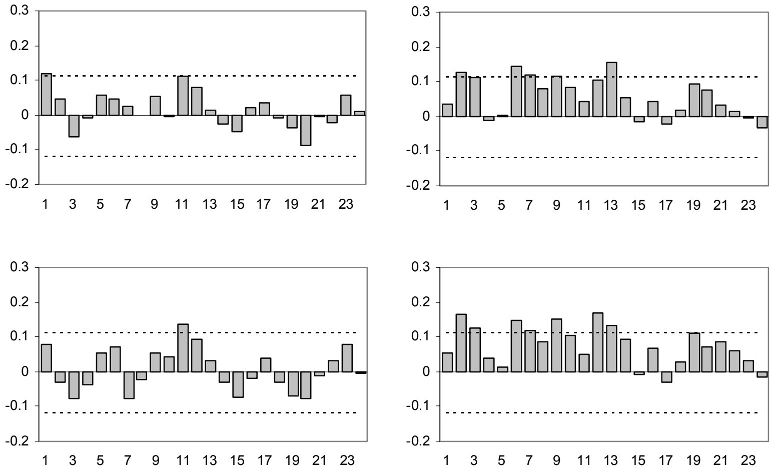

Figure 4.10: Correlograms of powers of the demeaned PIT for excess bond returns from the Student-t model with unpredictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

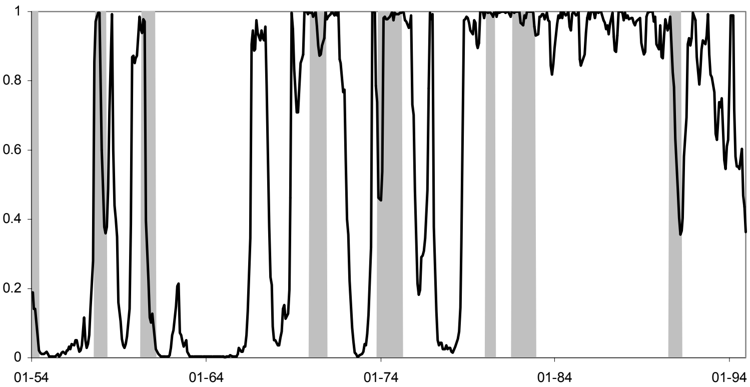

Figure 5.1: Probability of high-volatility regime from the model with predictable mean. The shaded regions are defined by the NBER recession indicator.

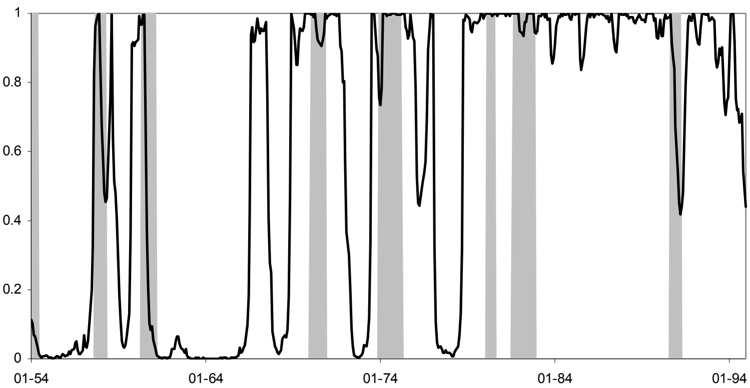

Figure 5.2: Probability of high-volatility regime from the model with unpredictable mean. The shaded regions are defined by the NBER recession indicator.

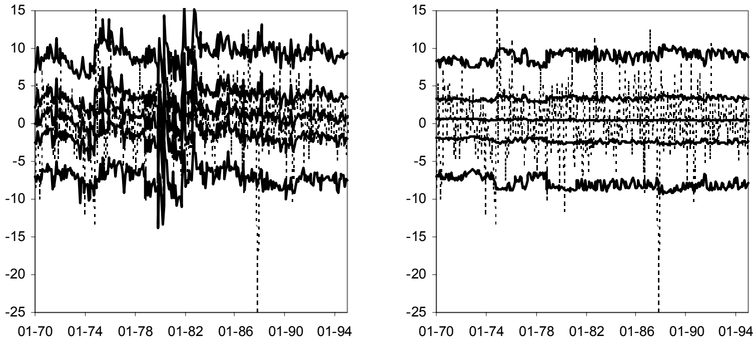

Figure 5.3: Density forecasts of excess stock returns from the Markov-switching model. The solid lines are the 2.5, 25, 50, 75, and 97.5% quantiles of the forecasts, while the dashed line is the corresponding realization. The figure on the left corresponds to the model with predictable mean, while the figure on the right corresponds to the model with unpredictable mean.

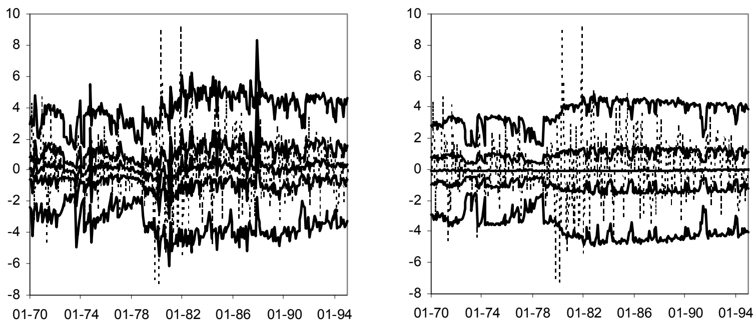

Figure 5.4: Density forecasts of excess bond returns from the Markov-switching model. The solid lines are the 2.5, 25, 50, 75, and 97.5% quantiles of the forecasts, while the dashed line is the corresponding realization. The figure on the left corresponds to the model with predictable mean, while the figure on the right corresponds to the model with unpredictable mean.

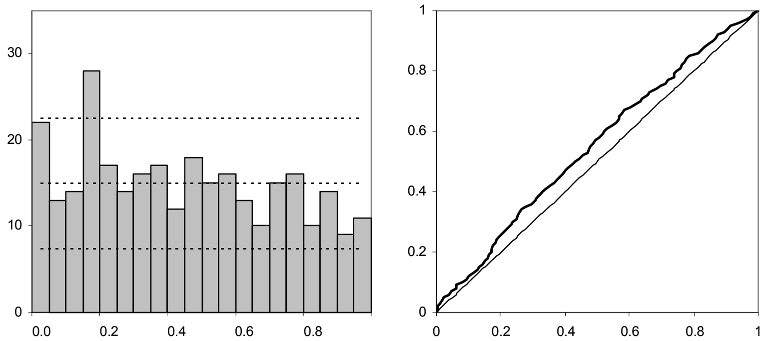

Figure 5.5: Histogram and Q-Q plot of the PIT for excess stock returns from the Markov-switching model with predictable mean.

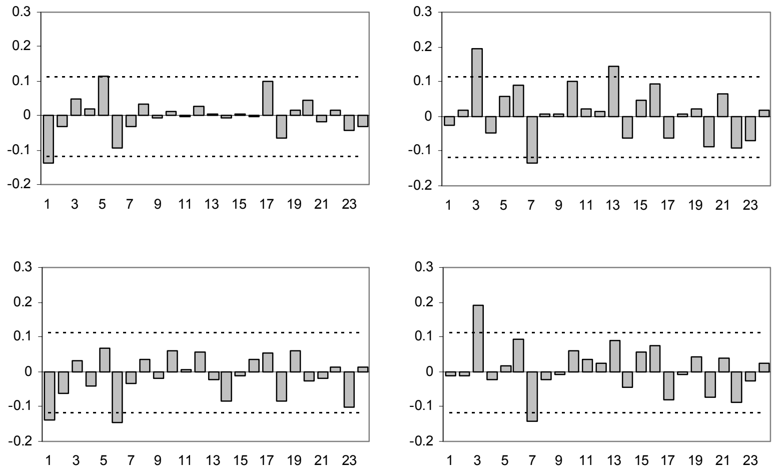

Figure 5.6: Correlograms of powers of the demeaned PIT for excess stock returns from the Markov-switching model with predictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

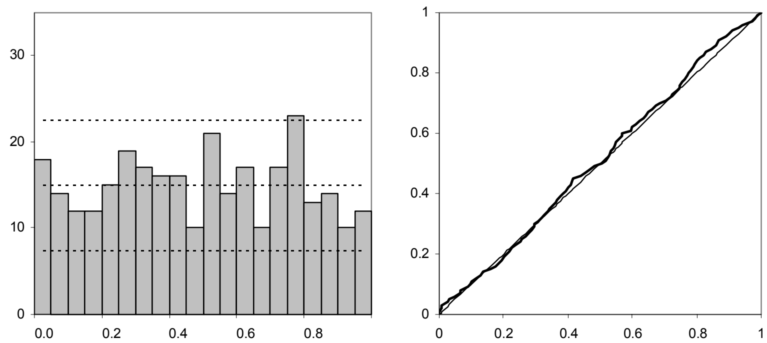

Figure 5.7: Histogram and Q-Q plot of the PIT for excess stock returns from the Markov-switching model with unpredictable mean.

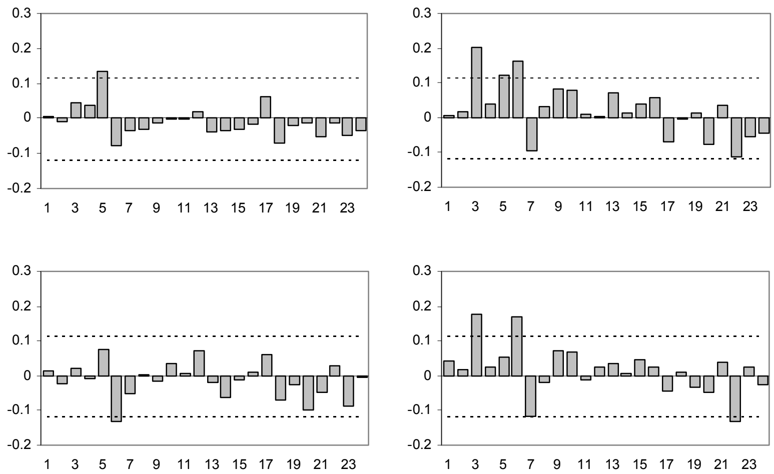

Figure 5.8: Correlograms of powers of the demeaned PIT for excess stock returns from the Markov-switching model with unpredictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

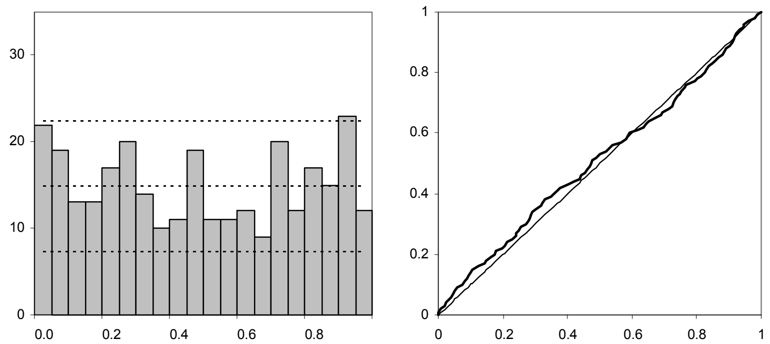

Figure 5.9: Histogram and Q-Q plot of the PIT for excess bond returns from the Markov-switching model with predictable mean.

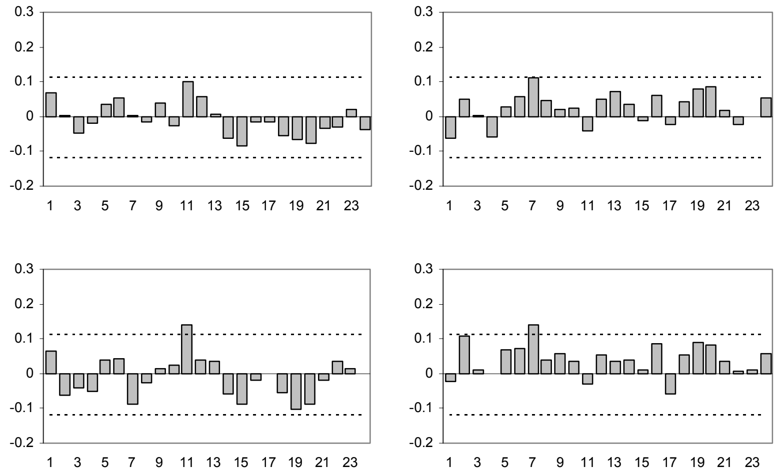

Figure 5.10: Correlograms of powers of the demeaned PIT for excess bond returns from the Markov-switching model with predictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.

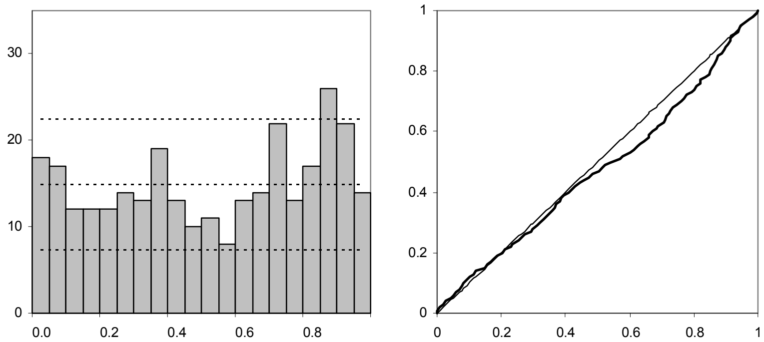

Figure 5.11: Histogram and Q-Q plot of the PIT for excess bond returns from the Markov-switching model with unpredictable mean.

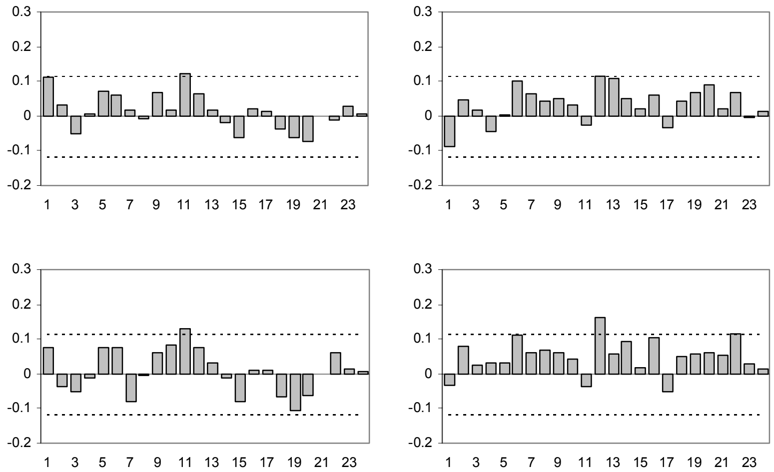

Figure 5.12: Correlograms of powers of the demeaned PIT for excess bond returns from the Markov-switching model with unpredictable mean. The top left figure corrresponds to the level, the top right figure corrresponds to the square, the bottom left figure corrresponds to the third power, and the bottom right figure corrresponds to the forth power.