BAYESIAN STATISTICAL APPROACH TO MARKET RISK MODELING

Sergei Tkachenko

Trabajo de investigación 026/013

Master en Banca y Finanzas Cuantitativas

Tutores: Dr. Antonio Alberto Rodríguez Sousa

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

Universidad de Castilla-La Mancha

www.finanzascuantitativas.com

Master's Thesis in Banking and Quantitative Finance

Bayesian statistical approach to market risk modeling

Author:

Sergei Tkachenko

Supervisor:

Dr. Antonio Alberto Rodríguez Sousa

VNIVERSITAT DE VALÈNCIA

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

UCLM

UNIVERSIDAD DE CASTILLA-LA MANCHA

February 2026

Abstract

With an objective of developing Bayesian models to measure market risk, a normal, mixture of normals, a factor-driven Bayesian t-GARCH(1,1) and a three-level hierarchical normal model are studied. The last two models are a distinct contribution of this work; they are applied to market data of 1003 stocks from 19 markets. Value-at-Risk and Excepted Shortfall are estimated for daily, weekly and monthly logarithmic returns. Results for an equivalent frequentist model and formal backtesting procedures are provided. The Bayesian version renders a high precision of market risk estimation and outperforms the frequentist approach, especially in small training samples. The efficient market hypothesis is tested on an individual, market and global levels concluding that it mostly holds.

Keywords: Bayesian statistical methods, market risk estimation, VaR and ES backtesting, Bayesian GARCH, hierarchical model.

Contents

Introduction

Motivation

Quantitative finance has long been grounded in probability theory, as financial models typically involve uncertainty about future outcomes or unknown parameters. This probabilistic foundation was firmly established by Nobel Prize–winning contributions such as Markowitz's mean–variance portfolio model (Markowitz, 1952), Sharpe's Capital Asset Pricing Model (Sharpe, 1964), and the Black–Scholes–Merton model for European option pricing (Black and Scholes, 1973). While these models are theoretically self-contained, their practical implementation requires knowledge of parameters—such as expected returns and variances, beta coefficients, or volatility—that are unknown in reality. Even if these parameters were known, an additional challenge would remain: assessing how well a model represents observable financial data, a question that depends on statistical analysis.

Statistical inference is commonly framed within two schools of thought: frequentist and Bayesian. The frequentist approach treats probabilities as long-run frequencies and views data as realizations from a fixed probabilistic model with unknown but constant parameters. In this sense, it is essentially pre-experimental: the model is specified in advance and data are used solely to estimate its parameters, functioning as a deterministic mechanism embedded in randomness (Fornacon-Wood et al., 2022). By contrast, the Bayesian approach is inherently post-experimental and based on sequential learning. Parameters are treated as random variables, and beliefs about them are updated as new data become available. Although both approaches require a probabilistic model for the data, the Bayesian interpretation differs philosophically, a distinction explored extensively in the literature (Hartmann and Sprenger, 2013).

Modeling parameters as random variables leads to the central Bayesian concepts of prior and posterior distributions. The posterior combines prior beliefs with observed data, yielding comprehensive information about parameters and enabling posterior predictive distributions that are essential for model checking and prediction. Despite the randomness of parameters, Bayesian point estimators can be derived by minimizing the posterior expected loss (Lehmann and Casella, 1998). From this perspective, Bayesian inference provides a coherent framework for updating beliefs via Bayes' rule, optimal in terms of mean squared error (Savage, 1954), accommodating informative, weakly informative, or non-informative priors, the latter often regarded as an objective starting point (Gelman et al., 2013).

The alleged subjectivity of Bayesian statistics is largely a misconception. Both Bayesian and frequentist methods require the specification of a probabilistic model, a choice that inevitably reflects researcher judgment. Consequently, Bayesian analysis is no more subjective than its frequentist counterpart. Moreover, results from one framework can often be evaluated within the other. From a frequentist perspective, appropriately chosen priors can yield Bayesian estimators with lower mean squared error and variance than maximum-likelihood estimators, leading to greater stability (Hoff, 2009). Under general conditions, Bayesian posteriors and point estimators are also asymptotically equivalent to frequentist results, as established by the Bernstein–von Mises theorem (van der Vaart, 2012).

Beyond philosophical considerations, Bayesian methods enable classes of models that are difficult or impossible to handle within a frequentist framework, notably hierarchical and latent

variable models. These models allow parameters to be estimated jointly, capturing interactions that improve estimation accuracy. From a frequentist viewpoint, such joint estimation benefits from Stein's phenomenon, which guarantees lower mean squared error than separate estimation (Stein, 1956). Finally, Bayesian methods may exhibit superior predictive performance (Duan et al., 2013), although the extent of this advantage is problem-dependent (Karami et al., 2025).

Objectives and resources

The first objective of this work is theoretical: to develop Bayesian one-dimensional versions of a normal model, a mixture of normals model and a Student-t GARCH(1,1) model with a conditional mean model with a latent variable selection in order to model stock returns distribution and apply them to observed market data. A treatment of a generalization of a one-dimensional normal model into a Bayesian hierarchical three-level normal model is also the aim of this work.

The second objective is empirical: to carry out an estimation of Value-at-Risk (VaR) and Expected Shortfall (ES), perform formal backtesting procedures and compare the results obtained by the means of the Bayesian statistical methods with their frequentist counterparts from observations of daily, weekly and monthly logarithmic returns under different time windows. Testing of a version of the efficient market hypothesis on an individual, a market and a global level is a transversal objective.

Globally, the thesis aims at contributing to the *Decent Work and Economic Growth* among the 17 Sustainable Development Goals by 2030 proposed by the United Nations.

The theoretical background of this work can be found in References. The empirical applications have been conducted with the use of R Studio and Python software. The data has been obtained from Refinitiv Eikon data provider using an academic license.

Literature review

One of the earliest systematic studies of the empirical properties of asset returns was conducted by Rama Cont (2000). He identified key stylized facts, including heavy tails, gain–loss asymmetry, aggregational Gaussianity, intermittency, and volatility clustering. Although not framed within either a frequentist or Bayesian paradigm, this work put the light on essential properties that any adequate statistical model of asset returns should capture.

A short time before Alan Agresti and Brent Coull (1998) showed that in a certain context Bayesian interval estimation for the mean is much more exact than the standard frequentist Wald interval: for example, for a sample size of 20, a Bayesian procedure provides an extremely accurate real coverage for a 95% nominal probability interval, while the frequentist procedure yields a real coverage of around 80%. This study suggests that the same conclusion might be drawn for a wider range of problems.

A substantial body of literature subsequently adopted Bayesian methods for market risk modeling, primarily focusing on univariate return series. So et al. (2007) evaluated Normal, Student-t, and Generalized Error distributions within a Bayesian GARCH(1,1) framework to model fat tails in financial time series. Their results showed that the Student-t GARCH(1,1) model outperformed competing specifications based on Mean Squared Error (MSE) and Mean Absolute Error (MAE), which was confirmed by Diebold–Mariano forecast comparison tests.

Similarly, Geweke and Amisano (2008), in a study conducted at the European Central Bank, compared Bayesian predictive distributions from five conditional volatility models, including Normal and Student-t GARCH variants. Using predictive ex post likelihood and probability integral transform diagnostics, they found that the Student-t Bayesian GARCH(1,1) model dominated alternative models and performed comparably to a hierarchical Markov normal mixture model.

More recently, Das (2025) applied a Bayesian Generalized Pareto regression model with Cauchy and g-priors, benchmarking its predictive performance against frequentist Lasso and Ridge regressions. Using RMSE for out-of-sample evaluation and AIC and BIC for in-sample fit, the study identified the Cauchy prior as superior across all metrics.

A comprehensive Bayesian treatment of Value-at-Risk (VaR) and Expected Shortfall (ES) was developed by Hoogerheide and van Dijk (2008) at the Tinbergen Institute. Using a Student-t GARCH(1,1) model and Adaptive Importance Sampling, they analyzed 10-day logarithmic returns of the S&P 500 over several years. Their results demonstrated that Adaptive Importance Sampling substantially outperformed direct sampling methods by achieving higher precision in VaR and ES estimation with significantly fewer simulated draws.

Despite the robustness of these findings, the studies share notable limitations. All rely on a fixed (G)ARCH(1,1) specification without addressing model order selection within the broader (G)ARCH(p,q) family. Their evaluation focuses exclusively on predictive accuracy, omitting formal backtesting procedures for VaR and ES estimates. Furthermore, they neglect posterior predictive checks via test quantities, a standard Bayesian diagnostic for assessing model adequacy. Hierarchical data structures, a core strength of Bayesian analysis, are not incorporated, and the implications of different return horizons (e.g., daily, weekly, monthly) remain unexplored. Another pitfall of the mentioned works is that the conditional mean modeling is completely omitted.

A distinct strand of literature emphasizes Bayesian hierarchical modeling for asset returns. Greyserman, Jones, and Strawderman (2006) provided one of the earliest contributions by comparing four return estimation approaches—equally weighted, classical, James–Stein, and hierarchical Bayesian—within the context of Markowitz mean–variance portfolio optimization. Using daily stock index data from 11 countries (US, UK, Canada, Belgium, Australia, France, Japan, Austria, Spain, Germany, and Hong Kong) over 1975–2002, they found that the Bayesian hierarchical model delivered superior portfolio returns and utility.

More recently, Feng and He (2021) investigated factor investing using a Bayesian hierarchical framework. They implemented a hierarchical multivariate regression model to forecast one-step-ahead asset returns. Across portfolios sorted by 10, 20, and 100 characteristics, the Bayesian hierarchical model outperformed Lasso regression, Random Forests, and Principal Component Regression in terms of out-of-sample R-squared, predictive interval coverage, and interval length. Backtesting results further showed that the model achieved the highest returns and the greatest Jensen's alpha.

While these studies demonstrate empirical advantages of jointly modeling multiple asset returns, both from a practical perspective and in light of Stein's paradox (Stein, 1956), they remain limited to a two-level structure. Consequently, they are not able to jointly model stocks and

markets, so they have to choose between modeling different stocks from the same market or grouping individual stocks in some pre-determined way to model them, which constrains their applicability in a broader systemic risk context.

Thesis description

This thesis is primarily empirical. The dataset consists of daily stock prices from 19 major indices: AEX (the Netherlands), ATX (Austria), BEL20 (Belgium), BUX (Hungary), FTSE (the UK), GDAX (Germany), HSI (Hong Kong), IT30 (Italy), IBEX35 (Spain), OMXB10 (Baltic), OMXC25 (Denmark), OMXH25 (Finland), OMXS30 (Sweden), OSBEX (Norway), PSI20 (Portugal), PX (the Czech Republic), S&P500 (the USA), SMIC (Switzerland) and WIG30 (Poland); comprising 1003 individual stock time series over 3,827 trading days, spanning the period from 2010-05-18 to 2026-01-12. All market data were obtained via Eikon under an academic license, and the overall data volume exceeds that used in any of the referenced studies.

From the outset, Bayesian methods provide a principled framework for handling missing data. A standard Bayesian multivariate normal model is used to draw posterior samples of missing observations, allowing their direct inclusion in subsequent analyses without discarding data or resorting to ad hoc imputations such as unconditional or conditional means.

The methodological component details the theoretical and practical foundations required to implement and assess Bayesian models for market risk. It begins with Bayes' rule and follows the standard modeling workflow: specifying the likelihood, assigning prior distributions, deriving posterior distributions, and conducting posterior predictive checks to assess model adequacy. Key theoretical results are presented, including asymptotic normality of posterior distributions and Bayesian point estimators, ensuring frequentist validity in large samples, as well as optimality properties in small samples. Markov Chain Monte Carlo (MCMC) methods are discussed in depth due to their central role in posterior approximation. Definitions for VaR and ES, their formal backtesting procedures and predictive performance metrics are also presented.

Two main classes of models are proposed: one-dimensional models and a Bayesian hierarchical model. The first class includes Bayesian normal and a two-component mixture of normals model, alongside comparisons with frequentist counterparts. As a key original contribution, the thesis develops a joint mean-variance model that combines a factor structure for the conditional mean with a GARCH(1,1) specification for the conditional variance with Student-t errors. Bayesian variable selection is applied to the conditional mean driven by the factor model, providing a way to take advantage of Bayesian model averaging. Unlike much of the existing literature, posterior predictive checks and test quantities are systematically employed to evaluate goodness of fit in this thesis.

Each model produces Bayesian Value-at-Risk (VaR) and Expected Shortfall (ES) measures, which are explicitly backtested, addressing a notable gap in the literature. The Bayesian factor-driven t-GARCH(1,1) model is applied to a large sample of 1003 individual stock time series of: 1. Daily returns obtained from 22, 66 and 252 daily consecutive observations; 2. Weekly returns obtained from 26 and 52 weekly consecutive observations; 3. Monthly returns obtained from 12 and 24 monthly consecutive observations. It allowed for a systematic comparison of model performance across different timescales and with the frequentist equivalent models.

The result of this empirical application shows that the Bayesian approach provides highly precise estimates of VaR and ES, which generally pass the backtesting procedures. It significantly outperforms the frequentist equivalent models for timescales and training sample sizes. The predictive performance of the conditional mean model is also higher under the Bayesian approach than under the frequentist one. Furthermore, the Bayesian variable selection latent variable method recognizes the pertinence of an individual stock to a particular market in a more consistent way, which explains the higher precision. Moreover, the Bayesian approach yields many possible point estimates, whereas the frequentist approach generally provides just one.

The Bayesian hierarchical normal model is a generalization of a one-dimensional normal model. It is developed in full and represents both an empirical and conceptual advance with respect to the reviewed references. Empirically, it exploits an unusually large dataset. Conceptually, it introduces an additional layer of hierarchy beyond the standard intra-asset and intra-market levels: an inter-market level. At this highest level, return means of different markets are modeled jointly as normal variables, independent conditionally on a latent intra-market location variable. This structure enables novel efficient market hypothesis testing on three levels: individually, at a market level and globally. This model also provides an insight in structures of marker-level and global-level expected returns, while retaining optimal modeling at the individual stock level.

An empirical application of the hierarchical model for different timescales and training sample sizes leads to the same conclusion as before: the Bayesian approach provides consistently more accurate VaR and ES estimates than the frequentist estimations under a normal model, although the factor-driven t-GARCH(1,1) model outperforms the normal hierarchical model, as expected. The model provides a way of exploring latent market structures: it reveals the effects of cross-market diversification. It also offers a unique way of testing the efficient market hypothesis: on an individual, market and global level without making any other assumptions or grouping stocks in a particular way. The hypothesis is found to hold on the global level for a large sample of 1003 stocks from 19 different markets for all types of returns (daily, weekly and monthly) and different training sample sizes. The hypothesis also holds for most markets and individual stocks, although there are particular markets and stocks for which the hypothesis is rejected in all cases.

Document structure

This thesis is organized into five parts:

I. Introduction motivates the use of Bayesian methods as an alternative to frequentist statistics, outlines the research context, objectives, resources, and the approach, and summarizes the methodology, results, and conclusions, with a review of relevant recent literature.

II. Methodology presents the theoretical foundations of Bayesian statistics and introduces models for market risk inference.

III. Data describes the data sources, the employed dataset and the data processing.

IV. Results analyze a set of market risk problems using Bayesian methods, with detailed interpretation of the findings.

V. Conclusion summarizes the main results and discusses directions for future research.

Methodology

This chapter starts with basic definitions and theoretical results concerning Bayesian statistical methods, and steps to follow in order to use them in practice. The first part of the chapter includes some useful asymptotical and finite samples theoretical results; a treatment of numerical approximation techniques (Monte Carlo Markov Chain) and Bayesian point estimators. All these details do not depend on a specific model, so they are presented before modeling. A description of Value-at-Risk, Expected Shortfall and their formal backtesting procedures is provided along with measures of predictive accuracy for point estimates. One-dimensional case is treated for the sake of briefness and clarity.

The second part of the chapter develops concrete statistical models of market risk: one-dimensional models (a normal, a two-component mixture of normals and a factor-driven t-GARCH(1,1) model) for a stock's returns and a hierarchical Bayesian model. The chapter also describes a Bayesian model for treatment of missing values.

A slight abuse of notation is employed for briefness: a probability distribution function (p.d.f.) is denoted by $p(\dots)$. For example, $p(x, y)$ denotes a joint p.d.f. of random variables x and y. Consider the following decomposition: $p(x, y) = \mathrm{p}(x|y)p(y)$. Although the three p.d.f. are denoted by $p(\dots)$, they don't have to be the same, and generally they will be different: $p_1(x, y) = \mathrm{p}_2(x|y)p_3(y)$.

Part I. General framework.

The conditional probability is the essential part of the Bayes' theorem.

Definition 2.1. Conditional probability distribution function.

Let $p_Y(y)$ be a marginal probability density (mass) function of a continuous (discrete) random variable Y and $p(y, \theta)$ is a joint probability density function of Y and another continuous or discrete random variable Θ. It is not necessary that they are defined on the same probability space Ω. Then the conditional probability distribution function of a random variable Θ given $Y = y$ is given by the above expression:

$$p(\theta|Y = y) = \frac{p(y, \theta)}{p_Y(y)}$$

With $p(\theta|Y = y) = 0$ when $p_Y(y) = 0$.

The Bayes' rule originated the Bayesian statistical approach:

Theorem 2.2. Bayes' rule.

The conditional probability distribution function from Definition 2.1 is a proper probability density function. Moreover, it can be expressed as follows:

$$p(\theta|Y = y) = \frac{p(y|\theta)p(\theta)}{\int_{\theta\in\Theta} p(y|\theta)p(\theta)\, d\Theta}$$

Where $\int_{\theta\in\Theta} p(y|\theta)p(\theta)\, d\Theta$ is to be understood as an integral if Θ is a continuous random variable or as a sum if Θ is a discrete random variable.

For a proof see Annex 1.1.

A usual modeling assumption within the Bayesian framework is unconditional exchangeability of the observed data, although the actual modeling is often carried out assuming that the observations are independent conditionally on the model parameters. As the following theorem shows, the two assumptions are equivalent.

Theorem 2.3. De Finetti.

Let random variables $Y_i \in \mathcal{Y}$ for all $i \in \{1,2,\dots\}$. Suppose that, for any n, the joint probability function of $Y_1, \dots, Y_n$ is exchangeable:

$$p(y_1, \dots, y_n) = p\left(y_{\pi_1}, \dots, y_{\pi_n}\right)$$

for all permutations π of $\{1, \dots, n\}$. Then this joint probability function of $Y_1, \dots, Y_n$ can be written as

$$p(y_1, \dots, y_n) = \int_\Theta \left\{\prod_{i=1}^{n} p(y_i|\theta)\right\} p(\theta) d\theta$$

for some parameter θ, some prior distribution on θ and some sampling model $p(y|\theta)$. The prior and sampling model depend on the form of the belief model $p(y_1, \dots, y_n)$.

For a sketch of proof see Annex 1.2.

Asymptotic results

Although Bayesian statistical methods are usually employed for and function well in finite sample contexts, there are some important asymptotic statistical properties that deserve a mention. As Hartigan (1966) shows, the Bayesian coverage is asymptotically approximately equivalent to the frequentist coverage. However, it is possible to establish an even more general result that connects the Bayesian approach with the frequentist one:

Theorem 2.4. Bernstein-von Mises.

Let $p(Y = y|\theta)$ be differentiable in quadratic mean at θ_0 with nonsingular Fisher information matrix I_{θ_0}, and suppose that for every $\epsilon > 0$ there exists a sequence of tests ϕ_n such that $p(Y = y|\theta_0)\phi_n \to 0$ and $\sup_{||\theta - \theta_0|| \geq \epsilon} p(Y = y|\theta)(1 - \phi_n) \to 0$

Furthermore, let the prior measure be absolutely continuous in a neighborhood of θ_0 with a continuous positive density at θ_0. Then the corresponding posterior distributions satisfy, with $||\cdot||$ the total variation norm.

$$\left|\left|p\left(\sqrt{n}(\bar{\Theta}_n - \theta_0)\middle|Y = y\right) - \mathcal{N}\left(\Delta_{n,\theta_0}, I_{\theta_0}^{-1}\right)\right|\right| \xrightarrow{p(Y=y|\theta_0)} 0$$

Where Δ_{n,θ_0} is any estimator satisfying $\sqrt{n}\left(\Delta_{n,\theta_0} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, I_{\theta_0}^{-1}\right)$.
The proof can be found in van der Vaart (2012).

This theorem asserts that under some regularity conditions, the posterior distribution of a model's parameters converges to a normal random variable centered at θ_0 with variance $I_{\theta_0}^{-1}$ in total variation distance. This type of convergence is a strong one and entails, for example, a convergence in distribution. A similar result can be obtained for Bayesian point estimators:

Theorem 2.5. Convergence of Bayesian point estimators.

Let the conditions of the Theorem 2.4 hold, and let $l(h)$ be a loss function that satisfy the following conditions

$$\sup_{||h||<M} l(h) \leq \inf_{||h||\geq 2M} l(h)$$

For every $M > 0$ and

$$l(h) \leq 1 + \left||h|\right|^p$$

Then the sequence $\sqrt{n}(T_n - \theta_0)$ converges under θ_0 in distribution to the minimizer of $t \to \int l(t-h)d\mathcal{N}(X, I_{\theta_0}^{-1})(h)$, for X possessing the $\mathcal{N}\left(0, I_{\theta_0}^{-1}\right)$ distribution, provided that any two minimizers of this process coincide almost surely. In particular, for every nonzero, subconvex loss function it converges to X.
The proof can be found in van der Vaart (2012).

The idea behind this theorem is simple: all reasonable Bayesian point estimators are asymptotically equivalent in distribution, so any election of a loss function (which entails a particular point estimator) leads to the same result in an asymptotic sense. Both theorems constitute a frequentist asymptotic validity of the Bayesian statistical methods: in general, the larger a sample size is, the fewer differences there will be between the posterior distribution of a parameter and the frequentist distribution of the parameter estimator.

Bayesian posterior point estimator

A Bayesian posterior point estimator is an estimator that minimizes a risk function with respect to the posterior parameter distribution. As the following proposition shows, a posterior mean and median minimize the squared and absolute value error risk functions respectively.

Proposition 2.6. Bayesian estimator for square and absolute value risk functions.

Let θ be a parameter (scalar) and let a be an estimator based on the data. The squared error risk function is defined as follows:

$$L(\theta, a) = (a - \theta)^2$$

Meanwhile the absolute value error risk function has the following definition:

$$L(\theta, a) = |a - \theta|$$

Then, after having observed data $Y = y$, the Bayesian posterior point estimator which minimizes the expected squared error risk function is the posterior mean:

$$a_{SE} = \mathbb{E}[\theta|Y = y]$$

Whereas the Bayesian posterior point estimator which minimizes the expected absolute value error risk function is the posterior median:

$$a_{AE} = \mathbb{F}_{1/2}\,[\theta|Y = y]$$

For a proof see Annex 1.3.

Since the Theorem 2.5 guarantees the convergence of Bayesian point estimators, as a general convention, in this thesis a posterior mean will be used as a Bayesian point estimator.

Bayesian framework in finite samples

Generally, the Bayesian posterior point estimators are biased. It can be shown that for a family of exponential family distributions (which includes a normal distribution) the posterior mean of θ can be decomposed in the following way (Hoff, 2009):

$$\hat{\theta}_b(Y = y) = \mathbb{E}[\theta|Y = y] = \frac{n}{k_0 + n}\bar{y} + \frac{k_0}{k_0 + n}\mu_0 = \omega\bar{y} + (1 - \omega)\mu_0$$

Where $\bar{y}$ is the sample mean of the observed values of Y, n is the sample size, k_0 is a prior parameter ("prior sample size") and μ_0 is the prior mean.

It can be shown that $\hat{\theta}_b(Y = y)$ can be more efficient than an unbiased estimator $\hat{\theta}_e(Y = y) = \bar{y}$ (the sample mean) in frequentists terms, i.e. assuming that the true value of the population mean is fixed at θ_0. For instance, the Mean Squared Error (MSE) can be used to evaluate the average distance between the estimator and the value of the true parameter:

$$MSE\left[\hat{\theta}|\theta_0\right] = \mathbb{E}\left[\left(\hat{\theta} - \theta_0\right)^2|\theta_0\right]$$

It is a frequentist version of the mean squared error risk function. The following proposition shows that the posterior mean is a more efficient point estimator than an unbiased maximum likelihood estimator (a sample average).

Proposition 2.7. Efficiency of a Bayesian point estimator.

Define a posterior-mean Bayesian point estimator: $\hat{\theta}_b(Y = y) = \omega\bar{y} + (1 - \omega)\mu_0$ and an unbiased estimator $\hat{\theta}_e(Y = y) = \bar{y}$. Then, the following statements hold for any sample size:

1. $\hat{\theta}_b(Y = y)$ has lower variance than $\hat{\theta}_e(Y = y)$ for any $\omega \in (0,1)$.

2. Under some conditions, $MSE\left[\hat{\theta}_b(Y = y)|\theta_0\right] < MSE\left[\widehat{\hat{\theta}_e(Y = y)}|\theta_0\right]$.

For a proof see Annex 1.4.

As it follows from the proof of the proposition, it is not hard to satisfy the conditions for the second statement to hold, providing a disperse prior with a mean that is reasonably near the true value.

Another key statistical property of Bayesian point estimators is their stability, especially for very small samples (Hoff, 2009). This fact turns out to be crucial in dealing with very short observation periods of stock prices, which is the case of recently listed companies or time series aggregated in months or even years. Real examples will be provided in the Results.

Apart from more efficient point estimators, Bayesian statistical methods also provide more accurate probability interval estimations. According to a study conducted by Alan Agresti and Brent Coull (1998) in which different estimation procedures of a real converge of nominal probability intervals were investigated via Monte Carlo simulations. Bayesian estimations were found to be extremely accurate for all sample sizes, and for a range of different distributions.

Posterior numerical approximation techniques

Numerical techniques provide the most common, general and feasible methods for approximation of a posterior distribution. This section is dedicated to an overview of the Markov Chain Monte Carlo (MCMC) algorithm due to its utmost importance both in Bayesian statistical

methods in general and in this thesis in particular. The most general version of MCMC is the Metropolis-Hastings algorithm presented for a bivariate problem as follows:

Algorithm 2.8. Metropolis-Hastings.

Let the target distribution be $p_0(u, v)$, where U and V are random variables, and suppose the vector of initial values $x^0 = (u^0, v^0)$ is known. The Metropolis-Hastings algorithm follows the next scheme for every step $s = 0, 1, 2, \dots, N$:

1. Update U:
    a. Sample a proposal value for u: $u^* \sim J_u\left(u \middle| u^{(s)}, v^{(s)}\right)$
    b. Compute the acceptance ratio:
$$r_u = \frac{p_0\left(u^*, v^{(s)}\right)}{p_0(u^{(s)}, v^{(s)})} \frac{J_u\left(u^{(s)} \middle| u^*, v^{(s)}\right)}{J_u\left(u^* \middle| u^{(s)}, v^{(s)}\right)}$$
    c. Set $u^{(s+1)} = u^*$ with probability $min\ (1, r_u)$ or $u^{(s+1)} = u^{(s)}$ with probability $max\ (0, 1 - r_u)$
2. Update V:
    a. Sample a proposal value for v: $v^* \sim J_v\left(v \middle| u^{(s+1)}, v^{(s)}\right)$
    b. Compute the acceptance ratio:
$$r_v = \frac{p_0\left(u^{(s+1)}, v^*\right)}{p_0(u^{(s+1)}, v^{(s)})} \frac{J_v\left(v^{(s)} \middle| u^{(s+1)}, v^*\right)}{J_v\left(v^* \middle| u^{(s+1)}, v^{(s)}\right)}$$
    c. Set $v^{(s+1)} = v^*$ with probability $min(1, r_v)$ or $v^{(s+1)} = v^{(s)}$ with probability $max(0, 1 - r_v)$
3. Update the parameter vector $x^{(s+1)} = \left(u^{(s+1)}, v^{(s+1)}\right)$

This algorithm approximates the target distribution $p_0(u, v)$ with a sequence of numerically simulated pseudo-random variables x. The order of updating of the parameters does not matter, i.e. it is possible to first update V and then update U, or vice versa, because the resulting approximation will not change. The vector of initial values may contain arbitrary values, although it is convenient to use values from a high posterior density region to facilitate a fast convergence. The proposal distribution can be any reasonable distribution. These facts stem from the convergence properties of the Metropolis-Hastings algorithm. All the mentioned facts and further details can be found in Hoff (2009) and Gelman et al. (2013).

There are two associated algorithms which are instantiations of the Algorithm 2.8 that are worth mentioning: the Metropolis and the Gibbs algorithms. The former suggests using symmetric proposal distributions, and the latter relies on fully conditional distributions for proposals. The details of both algorithms can be found in Annexes 1.5 and 1.6 respectively. In this thesis standard practices associated with the MCMC algorithm are applied: the first 20% of the simulated values are discarded as a *burn-in period*, then only every 10th simulated value is considered to ensure that the value are not autocorrelated, which is a process called *leaning*.

Value-at-Risk and Expected Shortfall

A posterior predictive distribution can be used to obtain Bayesian estimates of VaR and ES, which are risk metrics of an utmost importance in practice (Alexander, 2009):

Definition 2.9. Value-at-Risk and Expected Shortfall.

Let Y be a vector of random variables representing market returns consisting of components indexed by time, i.e. $Y = (Y_1, Y_2, \dots, Y_t, \dots)$. Further, let $\alpha \in (0,1)$ be a probability (confidence or belief) level. Assume that Y_t follows an unknown distribution F_t. Then, the Value-at-Risk (VaR) and the Expected Shortfall (ES) at time t are defined as follows (Acerbi & Szekely, 2014):

$$VaR_{\alpha,t}(Y) = -F_t^{-1}(\alpha); ES_{a,t} = -\frac{1}{\alpha}\int_0^{\alpha} F_t^{-1}(q)dq$$

The VaR can be defined using an alternative notation:

$$VaR_{\alpha,t}(Y) = -inf\{\mathrm{x} \in \mathbb{R}: \mathbb{P}(Y_t \leq \mathrm{x}) > \alpha\}$$

The ES can be defined as the tail conditional expectation if the subjacent distribution is assumed to be continuous and strictly increasing:

$$ES_{\alpha,t}(Y) = -\mathbb{E}[\mathrm{Y_t} \mid Y_t + VaR_{\alpha,t}(Y) < 0]$$

The $VaR_{\alpha,t}(Y)$ can be interpreted as a α-qunatile of returns at time t and the $ES_{\alpha,t}(Y)$ as an average return (loss) which drops below the $VaR_{\alpha,t}(Y)$.

The Bayesian version of the VaR and ES differs from the frequentist one precisely in the interpretation of Y, a random variable representing market returns. A natural choice in the Bayesian framework is to use the posterior predictive distribution of Y, while the frequentist only choice is to use a model distribution with fitted parameters. Nevertheless, there exist other approaches, for example, a historical simulation approach, which entirely depends on the observed data and is essentially model-free.

Nevertheless, posterior approximation (Bayesian approach) or model fitting (frequentist approach) and the subsequent estimation of risk metrics is only the first step in building a market risk model. The second step consists in ensuring that the estimation is viable, and it is as important as the first step. This thesis considers the Koupiec's (1995) and Christoffersen's (1998) tests for VaR backtesting; and conditional and unconditional tests for ES backtesting developed by Carlo Acerbi and Balazs Szekely (2014). Their descriptions can be found in Annex 1.7 and Annex 1.8.

It is also possible to assess the model's ability to predict the observed returns, so in this work the following predictive accuracy metrics will be considered:

Definition 2.10. Mean Absolute Error and Mean Squared Error.

Let y be a vector of length N representing observed market returns. Further, let $\hat{y}$ be a vector of the same length representing a point prediction of y_n. The mean absolute error (MAE), the mean squared error (MSE) and the coefficient of determination (R^2) are defined as follows:

$$MAE = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N}; MSE = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}; R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Obviously, the lower the MAE and MSE and the higher the R^2, the better are the point predictions. It should be noted that the values of the coefficient of determination are not guaranteed to be $R^2 \in [0,1]$ when applied to out-of-sample data

The Diebold-Mariano test (Diebold & Mariano, 1995) for MAE and MSE will also be applied, the details are provided in Annex 1.9.

Part II. Bayesian approach to market risk modeling

A Bayesian statistical modeling can be resumed in the following scheme:

1. Identification of the conditional model $p(y|\theta)$ and the prior parameter $p(\theta)$ distributions.
2. Calculation of the posterior distribution of the model parameters $p(\theta|Y = y)$.
3. Calculation of the posterior predictive distribution $p(\tilde{y}|Y = y)$.
4. Statistical inference on both the posterior parameter and posterior predictive distributions.

This chapter is dedicated to the first two steps in a context of different models. The third step follows as an immediate consequence of the second step:

$$p(\tilde{y}|Y = y) = \int_{\Theta} p(\tilde{y}, \theta|Y = y)d\theta = \int_{\Theta} p(\tilde{y}|Y = y, \theta)p(\theta|Y = y)d\theta$$
$$= \int_{\Theta} p(\tilde{y}|\theta)p(\theta|Y = y)d\theta$$

The last equation assumes independence of observations conditional on model parameters, which is assumed throughout this work. The posterior distribution can be obtained analytically or approximated numerically.

The fourth step requires having calculated $p(\theta|Y = y)$ and $p(\tilde{y}|Y = y)$ and consists of conducting analyses of these distributions, such as a posterior description, model adequacy checks, a predictive capacity assessment and Bayesian point estimators.

One-dimensional normal model

A one-dimensional normal model is perhaps the most widely used model both in statistical modeling and in quantitative finance. Besides it is of value per se, it also serves as a budling block for more complex hierarchical models, which will be discussed later.

A continuous random variable $Y \in \mathbb{R}$ is called a normal random variable with a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma^2 \in \mathbb{R}_+$ if its probability density function (p.d.f.) has the following form:

$$p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$

The assumption of conditional independence of observations makes it possible to represent the likelihood function of the observed data as a product of normal densities:

$$p(Y = y \mid \mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)$$

The following factorization is useful to derive the posterior distributions of both parameters:

$$f(Y = y, \mu, \sigma^2) = f(Y = Y|\, \mu, \sigma^2) f(\mu|\sigma^2) f(\sigma^2)$$

Location parameter

A single most important prior conjugate distribution for the location parameter μ is also normal: $\mu \sim \mathcal{N}(\mu_0, \tau_0^2)$, where μ_0 and τ_0^2 are fixed values often called *hyperparameters*. They can either represent a prior knowledge or be conveniently chosen to be disperse enough, so that there will be very little to practically no influence on the posterior distribution. Under this prior, the posterior distribution of μ takes the following form (see Annex 1.10).

$$f(\mu|Y = y\,, \sigma^2) \sim \mathcal{N}\left(\mu_{\mathrm{n}} = \left(\mu_0 \frac{1}{\tau_0^2} + \bar{y}\frac{N}{\sigma^2}\right)\left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right)^{-1}; \tau_{\mathrm{n}}^2 = \left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right)^{-1}\right)$$

The posterior mean is a weighted average of the prior mean μ_0 and the sample average $\bar{y}$: $\mu_{\mathrm{n}} = \mu_0 \frac{\omega_1}{\omega_1+\omega_2} + \bar{y}\frac{\omega_2}{\omega_1+\omega_2}$, where $\omega_1 = \frac{1}{\tau_0^2}$ (prior precision), $\omega_2 = \frac{N}{\sigma^2}$ (sample precision). The scale parameter (variance) is the inverse of prior and sample precisions: $\tau_{\mathrm{n}}^2 = \left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right)^{-1}$. It should be noted that as the sample size becomes larger, $\mu_{\mathrm{n}} \xrightarrow[N\to\infty]{} \bar{y}$ and $\tau_{\mathrm{n}}^2 \xrightarrow[N\to\infty]{} 0$.

The choice of a normal prior distribution is motivated by several factors: the Central Limit Theorem states that under quite general conditions the distribution of a sample mean is normal; it facilitates the posterior calculations, and it is also widely used in the literature.

Scale parameter

The posterior joint distribution of the location and scale parameters can be decomposed:

$$f(\mu, \sigma^2|Y = y) = f(\mu|Y = y\,, \sigma^2) f(\sigma^2|Y = y)$$

The posterior distribution of the location parameter conditional on the scale parameter has already been obtained, so now it's the turn of the scale parameter. Given a prior distribution $f(\sigma^2) \sim InvGamma\left(\frac{v_0}{2}, \frac{v_0}{2}\sigma_0^2\right)$, the posterior distribution is $f(\sigma^2|Y = y) \sim InvGamma\left(\frac{N+v_0}{2}, \frac{1}{N+v_0}\left[v_0\sigma_0^2 + \sum_{i=1}^{N}(y_i - \bar{y})^2 + \frac{Nk_0}{N+k_0}(\bar{y} - \mu_0)^2\right]\right)$ (see Annex 1.11 for details). A usual interpretation of v_0 is as a prior sample size and $v_0\sigma_0^2$ as a prior sum of squares. Additionally, $N + v_0$ is a posterior sample size, $\sum_{i=1}^{N}(y_i - \bar{y})^2$ is a sample sum of squares and $\frac{Nk_0}{N+k_0}(\bar{y} - \mu_0)^2$ is an additional term correcting for a square error between the sample mean and the prior expectation, which can be called *a prior bias*.

With this result, it is possible to obtain a posterior distribution for the location parameter unconditional on the scale parameter by calculating the following integral: $f(\mu|Y = y\,) = \int_0^\infty f(\mu|Y = y\,, \sigma^2) f(\sigma^2|Y = y) d\sigma$. However, it is not necessary because it can be easily approximated by the Gibbs algorithm (the full numerical scheme is presented in Annex 1.12).

Multidimensional normal model

The multidimensional normal model is a direct generalization of a one-dimensional normal model. The conditional multivariate normal model distribution has the following form:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp(-(\boldsymbol{Y} - \boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\theta})/2\,)$$

Where $\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \cdots \\ y_p \end{pmatrix}$, $\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \cdots \\ \theta_p \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{p,1} \\ \vdots & \ddots & \vdots \\ \sigma_{1,p} & \cdots & \sigma_p^2 \end{pmatrix}$.

The multivariate distribution is completely characterized by the mean vector and the covariance matrix. As in the one-dimensional case, it is possible to show that the following choice of the prior distribution of $\boldsymbol{\theta}$ leads to a semi-conjugate posterior distribution:

$$p(\boldsymbol{\theta}) \sim MN(\boldsymbol{\mu_0}, \Lambda_0)$$

Where $\boldsymbol{\mu_0}$ is a vector of prior means, $\boldsymbol{\Lambda}_0$ is a prior covariance matrix of the means and MN denotes a multivariate normal distribution. The posterior distribution of $\boldsymbol{\theta}$ follows:

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\Sigma}) \sim MN(\boldsymbol{\mu_n}, \boldsymbol{\Lambda}_n)$$

Where $\boldsymbol{\mu_n} = (\boldsymbol{\Lambda}_0^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1}(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu_0} + n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{Y}})$ and $\boldsymbol{\Lambda}_n = (\boldsymbol{\Lambda}_0^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1}$. This posterior specification is completely analogous to the one-dimensional case. The following facts follow readily from the properties of multivariate normal distribution:

$$\mathbb{E}[\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\Sigma}] = \boldsymbol{\mu_n}$$
$$\mathbb{V}[\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\Sigma}] = \boldsymbol{\Lambda}_n$$

The situation with the model covariance matrix $\boldsymbol{\Sigma}$ is also completely analogous to the one-dimensional case: it can be shown that an inverse Wishart prior distribution of $\boldsymbol{\Sigma}$ leads to a conjugate posterior distribution:

$$p(\boldsymbol{\Sigma}) \sim InvWishart(v_0, \boldsymbol{S}_0^{-1})$$

Just like a multivariate normal distribution, an inverse Wishart distribution is a generalization of an inverse gamma distribution. The posterior distribution of Σ follows:

$$p(\Sigma|\boldsymbol{Y}, \boldsymbol{\theta}) \sim InverseWishart(v_0 + n, (\boldsymbol{S}_0 + \boldsymbol{S}_\theta)^{-1})$$

Where $\boldsymbol{S}_\theta = \sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{\theta})(\boldsymbol{Y}_i - \boldsymbol{\theta})^T$.

A detailed derivation is analogous to the one-dimensional normal model and can be found in Hoff (2009). The numerical implementation consists in a straightforward use of a Gibbs sampler in a manner similar to a one-dimensional normal model (the full numerical scheme is presented in Annex 1.13). The application of the multidimensional normal model to a Bayesian treatment of missing values can be found in Annex 1.14.

One-dimensional two-component normal mixture model

Bayesian normal mixture models were studied in detail by Stephens (1997). A recent study (Yan & Han, 2019) suggests that, under the frequentist framework, a three-component normal mixture provides the best empirical fit of the model distribution to the observed market data. However, the results for a mixture consisting of two components are comparably good and remarkably better than those of a single-component normal model.

In this thesis a two-component specification has been chosen because of its high interpretability (a two-regime market: low volatile and high volatile) and clarity. A two-component mixture of normal distributions is the following generalization of a one-dimensional normal distribution function:

$$f(y \mid \mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p) = \mathrm{p}\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}(y-\mu_1)^2\right) + (1-\mathrm{p})\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2\sigma_2^2}(y-\mu_2)^2\right)$$

The assumption of conditional independence of observations makes it possible to represent the likelihood function of the observed data as a product of normal mixture densities:

$$f(Y = y \mid \mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p) = \prod_{i=1}^{N}\left[\mathrm{p}\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2\sigma_1^2}(y_i-\mu_1)^2\right) + (1-\mathrm{p})\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2\sigma_2^2}(y_i-\mu_2)^2\right)\right].$$

A two-component mixture of normal distributions has an advantage of being as analytically tractable as a normal distribution, but it also has a greater flexibility of being able to cope with bimodal data and a two-regime volatility structure. It is highly relevant to market returns data which usually present periods of high and low volatility.

Even though this model is also implemented with the MCMC algorithm, there is one conceptual distinction. First, there will be two groups of values, the ones assigned to one cluster and the ones assigned to another cluster. Second, the means and the variances of both components are updated via Gibbs sampler treating the two groups individually. Third, the mixture proportion parameter p is updated.

A $Beta(\alpha_0, \beta_0)$ distribution is assumed as a prior: $h(p) \sim Beta(\alpha_0, \beta_0)$. Given the observations, the proportion parameter is updated for each observation:

$$c_{1,i} = f(Y = y_i \mid \mu_1, \sigma_1^2, p)h(p)$$
$$c_{2,i} = f(Y = y_i \mid \mu_2, \sigma_2^2, 1-p)h(1-p)$$
$$k_i = \frac{c_{1,i}}{c_{1,i} + c_{2,i}}$$

An auxiliary variable k_i measures the probability of each observation to pertain to the first group. Another auxiliary binary random variable z_i, $p(z_i) \sim Bin(k_i)$, is drawn to update each of the component of the vector $\boldsymbol{z} = (z_1, z_2, \dots, z_N)$. Finally, the posterior distribution of a probability of an i-th observation to pertain to the first group is calculated in the following way:

$$p_i \mid Y = y_i \sim Beta\left(\alpha_0 + \sum_{i=1}^{N}(z_i = 1), \beta_0 + \sum_{i=1}^{N}(z_i = 0)\right)$$

For the model to be correctly specified, it needs an ordering constraint, which can be an ordering of means or variances. Since the difference in variances is suspected to be more relevant than the differences in means for the components in a context of logarithmic returns, the variance ordering constraint is chosen to be enforced. It means that in each iteration of the MCMC algorithm the simulated variance component with higher value will always be assigned to one group (e.g., "a high variance cluster") and the simulated variance component with lower value will always be assigned to the other group (e.g., "a low variance cluster"). A detailed treatment can be found in Gelman et al. (2013) and Stephens (1997). The full numerical scheme is presented in Annex 1.15.

Factor-driven t-GARCH(1,1) model with variable selection

On one hand, one of the first factor-driven models for a conditional mean of stock returns was the CAPM proposed by William Sharpe (1964) who won the 1990 Nobel Memorial Prize in Economic Sciences for its development. While the objective of their use can be quite different, e.g. a risk decomposition into a systematic and idiosyncratic components, or a prediction of returns for investment purposes; they are essentially regression-based models that are still in use (see Das, 2025; Feng & Ge, 2022).

On the other hand, a family of conditional variance models called GARCH (general autoregressive conditional heterogeneity) was developed by Tim Bollerslev (1986). Although he didn't win a Nobel Prize himself, his doctoral supervisor Robert Engle won the 2003 Nobel Memorial Prize in Economic Sciences for the development of the ARCH model. These models are still considered to be the gold standard for variance modeling (Hoogerheide & van Dijk, 2008).

A Bayesian factor-driven t-GARCH(1,1) model is a unique contribution of this work. It consists of a joint model for the conditional mean governed by a Bayesian factor-driven linear regression model with a variable selection procedure and a Bayesian t-GARCH(1,1) model for the conditional variance.

Let Y be an $N \times 1$ response vector, X be a $N \times p$ matrix of regressors, β be a $p \times 1$ regression coefficient vector and z is a $p \times 1$ binary vector associated with the regressor coefficients. Furthermore, $\sigma_t^2, \alpha_0, \alpha_1, \beta, \delta, v$ are scalar values.

The model adopts the following functional form:

$$\begin{cases} y_t = z\beta X_t + \epsilon_t \\ \epsilon_t \sim Student(0, \sigma_t^2, v) \\ \sigma_t^2 = \gamma + \alpha\epsilon_{t-1}^2 + \theta\sigma_{t-1}^2 \end{cases}$$

Therefore, assuming conditional independence, the conditional probability model for the observations takes the following form:

$$p(Y = y\,|\boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, v) = \prod_{t=1}^{N} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)\sigma_t^2}} \left(1 + \frac{(y_i - (\boldsymbol{Z\beta})_t)^2}{(v-2)\sigma_t^2}\right)^{-\frac{v+1}{2}}$$

The model has the following parameters: $\boldsymbol{Z} \in \{0,1\}^p, \boldsymbol{\beta} \in \mathbb{R}^p; \gamma > 0, \alpha \geq 0, \theta \geq 0; v > 0$. A wide-sense stationarity is guaranteed when $\alpha + \theta < 1$ is satisfied (Bollerslev, 1986).

The particular election of prior distributions for the parameters is subject to either expert knowledge or the absence of it, which can be represented by uninformative or weakly informative prior specifications. For instance, each element of the vector $\boldsymbol{Z}$ can have a Binary(0.5) prior distribution which is a weakly informative prior. The prior distribution for $\boldsymbol{\beta}$ can be a multidimensional normal distribution centered at maximum-likelihood estimates (an Empirical Bayes prior) or any prior values, for instance, they can be centered at 0 to represent a prior hypothesis of weak relevance of all the regressors. The covariance matrix associated with $\boldsymbol{\beta}$ can have a large determinant and thus be a *diffuse prior*, permitting for a large range of values. A prior distribution for γ can be any distribution generating positive random variable, for example, an inverse-gamma distribution with a particular election of hyperparameters making it a diffuse prior. A prior distribution for α and θ can be a Beta(1,1) distribution, which is equivalent to a Uniform(0,1) distribution and thus it would be an uninformative prior, with a restriction of $\alpha + \theta < 1$. Finally, a prior distribution for v can be any distribution suitable for positive random variables, either discrete or continuous. It should be kept in mind that a Student-t distribution with around 30 degrees of freedom is practically indistinguishable from a normal distribution. For instance, it can be a Poisson distribution centered at 15, for instance, or a discrete uniform distribution for some reasonably large range of values, e.g. from 1 to 100.

The Bayesian conditional mean linear regression model is qualitatively different form the frequentist linear regression model in the following senses:

First, it provides a posterior distribution for each model parameter including a posterior predictive distribution for observations. Second, it incorporates an implicit variable selection procedure via the variable z whose posterior distribution indicates the inclusion or exclusion of each variable in the model. This fact implies that the model performs a variable selection procedure without further assumptions (e.g., a common frequentist approach to a variable selection procedure is based on forward, backward or stepwise algorithm which adds or eliminates variables depending on their p-value or F-statistics change), so the model is naturally prone to overparametrisation. Third, in comparison with a standard frequentist approach, this model does not assume that the errors satisfy standard hypotheses of having zero-mean and being heteroscedastic, non-autocorrelated and normally distributed. Furthermore, it does not assume that the regressor matrix is deterministic, which is a common implicit assumption of a frequentist linear regression analysis. That is because the Bayesian framework provides posterior distributions for every parameter as an innate characteristic, so it does not generally need to rely on either analytical or asymptotical results to carry out statistical inference because it is possible to numerically approximate the posterior distributions.

The conditional variance model is represented by a Bayesian version of a GARCH(1,1) model with a Student-t distribution for the errors, which has been found to be one of the best-

performing GARCH specifications (see Literature review). In conjunction with the conditional mean model, the performance of the conditional variance model is expected to be even higher and the variance estimation to be more accurate, which could provide a more precise estimation of individual market risk. A full description of a MCMC numerical scheme for this model is presented in Annex 1.15.

Bayesian hierarchical three-level normal model

The previously reviewed models concerned unidimensional time series of returns. The factor-driven model decomposes the market risk of an individual stock's returns into a systematic (which can be attributed to the factors) and an idiosyncratic (unexplained by the conditional mean model, thus specific to that stock) components. The idea behind the risk decomposition is to quantify the risk that can be hedged away (the idiosyncratic) and the one that must be assumed as non-hedgeable (the systematic). Nevertheless, this approach ignores the existence of different markets, which is an undeniable fact, in a sense that it is not possible to further decompose the systematic risk into more basic elements.

Stock returns from different markets can be represented by a hierarchical multilevel structure: individual stocks represent the unit level of this structure, then they are grouped into different markets which is the market level, and finally markets are grouped together to make up the global level. This structure is a completely natural way of modeling stock returns, and it is easily handled as a Bayesian hierarchical model. It should be noted that this model is impossible to implement under the frequentist approach because it requires to assume that the model parameters are random variables, which is the core of the Bayesian approach.

A Bayesian hierarchical modeling offers a number of theoretical and practical advantages over individual modeling, either Bayesian or frequentist: it permits to take advantage of the Stein effect (Stein, 1955) by conducting a joint parameter modeling, so that the resulting estimators will have a lower mean squared error. Another key feature is a so-called *pulling effect* (Hoff, 2009), which consists in "pulling" together the estimates around a group average, and the force of this effect generally depends on a sample size (smaller the sample size, higher the pulling effect). Furthermore, the observed data is allowed to differ in size, i.e. vectors of stocks' returns don't have to be of the same length. Perhaps the most important feature of this approach is a possibility of conducting statistical inference on nested latent variables (a market level and a global level).

Assume that there are M markets and N_m stocks in a market $m \in \{1, \dots, M\}$, so there are $N = \sum_{m=1}^{M} N_m$ stocks in total. The model is represented by the following structure:

1. A within-stock conditional model: $\phi_i = \{\mu_i, \sigma_i^2\}: p(y|\phi_i) = Normal(\mu_i, \sigma_i^2)$.
2. A within-market conditional model: $\psi_{\mathrm{m}} = \{\delta_m, \tau_{\mathrm{m}}^2\}: p(\mu|\psi_{\mathrm{m}}) = Normal(\delta_m, \tau_{\mathrm{m}}^2)$.
3. A global conditional model: $\omega = \{\theta, \gamma^2\}: p(\delta|\omega) = Normal(\theta, \gamma^2)$.

There are three conditional models, namely the within-stock, within-market and global conditional models, each of the models represent the unit, intra-market and inter-market levels respectively. The key difference between a standard Bayesian model and a Bayesian hierarchical model setting is considering some of the parameters (ϕ_i and ψ_{m} in this case) as random variables conditioned on other random parameters (ψ_{m} in case of ϕ_i, ω in case of ψ_{m}). That is why this type of modelling is beyond the reach under the frequentist approach. A full description of a MCMC numerical scheme for this model is presented in Annex 1.16.

Data

All the data for this thesis has been retrieved from Refinitiv Eikon using an academic license. The time series of prices associated with the following indexes have been retrieved: S&P500 (502 stocks), AEX (31 stocks), ATX (21 stocks), BEL20 (27 stocks), BUX (17 stocks), FTSE (101 stocks), GDAX (41 stocks), HSI (90 stocks), IBEX35 (36 stocks), IT30 (32 stocks), OMBX (10 stocks), OMXC25 (26 stocks), OMXH25 (26 stocks), OMXS30 (31 stocks), OSEBX (61 stocks), PSI20 (17 stocks), PX (13 stocks), SMIC (21 stocks), WIG30 (31 stocks). The observation period spans from 18.05.2010 to 12.01.2026.

The retrieved data has been treated in the following way: any time series consisting of missing values only was eliminated. The remaining number of stocks are 470, 21, 20, 24, 14, 93, 37, 65, 33, 29, 6, 25, 23, 28, 45, 16, 8, 18, 27, respectively for the above-mentioned indexes. The are time series of 1003 stocks remaining after the treatment. The next step of the data treatment concerns the rows: any row consisting of more than 90% of missing values has been eliminated. The observation period has shrunk to be from 21.11.2014 to 12.01.2026.

The final dataset contains 1004 columns, 1003 corresponding to each of the stock's price series index by an additional column representing time variable. The number of rows is 2640, each of them corresponds to one business day. The dataset has a descending order with respect to the time variable by rows, e.g. the first row corresponds to the oldest observations (21.11.2014) and the last row correspond to the most recent observations (12.01.2026).

A specific treatment of outliers was not carried out because of a high quality of the data provider. Nevertheless, an outlier detection system has been implemented to inform in case of any normalized value associated with a given vector being higher than 10 sample standard deviations of the values of that vector. No outlier has been detected in the whole dataset.

A treatment of missing value, possibly due to trading holidays in some markets which do not overlap with trading holidays in other markets, is a distinctive characteristic of a data treatment done in this thesis: the posterior distribution of the missing values has been numerically approximated by the procedure described in Annex 1.14. The draws from this posterior distribution allow to use the observations as if they were available, and at the same time to account for the uncertainty associated with the missing values, which is never achieved by frequentist methods of missing values treatment which generally consist in either discarding of missing observations or an imputation of a fixed value (e.g., an unconditional or conditional mean).

Since the retrieved information contained price information only, logarithmic returns have been defined in the following way: $r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$, where p_t is a price at time t.

Results

This chapter is dedicated to an application of the Bayesian statistical models described in this thesis to a real market data. Unidimensional models will be applied and discussed first, followed by the Bayesian three-level hierarchical normal model.

Unidimensional models

For the sake of briefness and clarity, a 22-day observation window for a time series of logarithmic returns of an example stock (Apple) will be presented in detail. Estimation and backtesting of VaR and ES is provided for every model, as well as posterior predictive checks and predictive accuracy assessment. The frequentist results are calculated using maximum likelihood estimation (MLE) and are provided for comparison with the Bayesian results

Consider the following logarithmic returns distribution of Apple stock prices between 09.12.2024 and 10.01.2026. The period between 09.12.2024 and 10.01.2025 (22 business days) comprises the train set and the period between 11.01.2025 and 10.01.2026 (250 business days) comprises the test set:

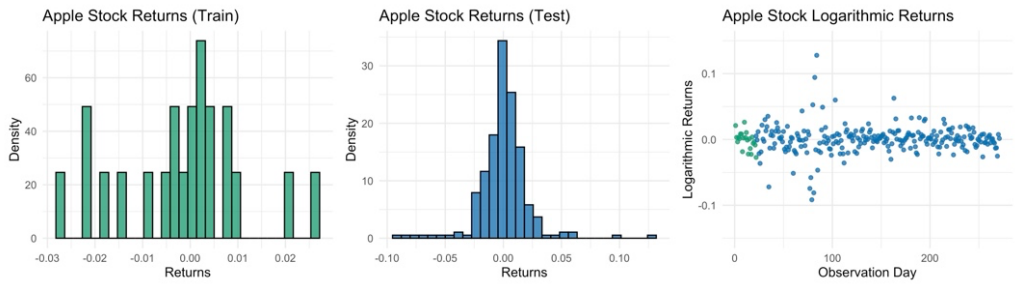

Figure 1. Apple stock's logarithmic returns between 09.12.2024 and 10.01.2026 (Eikon). Histograms of the train set (09.12.2024 - 10.01.2025) and the test set (11.01.2025 - 10.01.2026), and a joint plot. Green points correspond to the train set and the blue points correspond to the test set. Proper elaboration.

As the Figure 1 shows, the train period is quite steady and comparable with the rest of the time series. Nevertheless, there is at least one period of pronouncedly higher volatility and at least one period of lower volatility. Some sample statistics are shown in the table below (see Annex 1.17 for details), along with a 10%, 5% and 1% samples estimate of VaR and ES:

Sample statistics	Mean	Standard Deviation	Kurtosis	Skewness	10% VaR (ES)	5% VaR (ES)	1% VaR (ES)
Train	-0.0016	0.0134	2.80	-0.09	-0.0218 (-0.0240)	-0.0225 (-0.0250)	-0.0263 (-0.0273)
Test	0.0004	0.0213	11.80	0.48	-0.0193 (-0.0364)	-0.0237 (-0.0502)	-0.0733 (-0.0824)

Table 1. Apple stock logarithmic returns between 09.12.2024 and 10.01.2026 (Eikon). Sample statistics of the train set (09.12.2024 - 10.01.2025) and the test set (11.01.2025 - 10.01.2026). Proper elaboration.

Normal model

Empirical Bayes prior.

An empirical Bayes ($\mu_0 = \overline{y_{train}}$ and $\sigma_0^2 = \hat{\sigma}_{train}^2$) diffuse ($k_0 = v_0 = 1$) specification described in Casella (1985) has been employed to set up the Bayesian normal model:

Parameter	μ_0	σ_0^2	k_0	v_0
Value	$\overline{y_{train}}$	$\hat{\sigma}^2_{train}$	1	1

Table 2. An empirical Bayes prior specification for the Bayesian normal unidimensional model. Proper creation.

This specification does not introduce bias in the posterior point estimator of the location parameter and allows to show a pure effect of introducing uncertainty into the parameter estimation in comparison with a usual frequentist approach. The posterior distribution is obtained via MCMC simulations in a straightforward way according to the Methodology. Posterior predictive checks provide the best way to visualize the results:

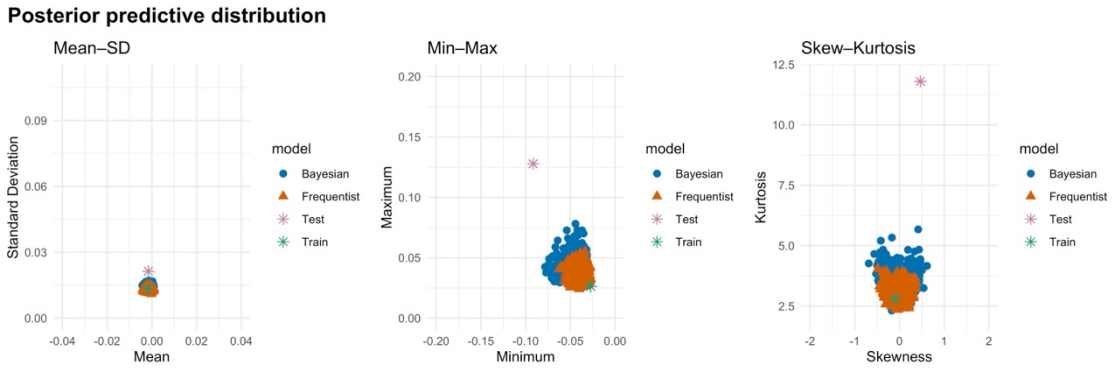

Figure 2. The unidimensional normal model with empirical Bayes prior: posterior predictive distributions checks and comparison with the frequentist equivalent model in three plots: 1. Mean – Standard deviation; 2. Maximum – Minimum; 3. Kurtosis – Skewness. Proper creation.

The plots presented above represent the distribution of the following test values: 1. Mean and standard deviation; 2. Maximum and minimum; 3. Coefficients of kurtosis and skewness; a Bayesian posterior predictive distribution is depicted in blue, a frequentist fitted model distribution is depicted in brown; observed statistics in the train and the test datasets are depicted in green and pink respectively. Generally, both the Bayesian and the frequentist approaches show a reasonably good capacity to approximate the test statistics observed in the train set. Nevertheless, the test statistics from the test set suggest that both models fail to model returns distributions adequately: the sample standard deviation, coefficient of kurtosis and maximum value from the test set are higher than usually predicted values, and the minimum value is lower than the values predicted by both models.

The Kolmogorov-Smirnov test has been applied to find out whether it is possible to reject the hypothesis that the Bayesian posterior predictive distribution (and the frequentist fitted model distribution) and the observed data from the test set come from the same distribution. The associated p-values are 0.01 for the Bayesian and 0.02 for the frequentist distributions. Therefore, at a 5% level of significance the null hypothesis is rejected, concluding that the modelled distributions do not come from the same distribution as the observed returns.

The Bayesian posterior point estimate (mean) and the frequentist estimate for the model parameters are presented in the table below:

Parameter \ Estimate	Bayesian	Frequentist	Train	Test
$\hat{\mu}$	-0.0016	-0.0016	-0.0016	0.0004
$\hat{\sigma}$	0.0135	0.0131	0.0134	0.0213

Table 3. The unidimensional normal model with an empirical Bayes prior: the posterior point estimate (mean) and the frequentist estimate; sample values observed in the train and test datasets. Proper creation.

Both point estimates for the location parameter are equal because of the empirical Bayes choice of the prior distribution hyperparameters. A slightly higher value of the Bayesian estimate for the scale parameter represents a greater uncertainty associated with the posterior which was presented in Methodology.

The estimates of VaR and ES (Bayesian and frequentist) are in a table below:

Risk metrics \ quantiles	10%	5%	1%
Bayesian VaR	-0.0193 (9.9%) (0.97, 1.00)	-0.0246 (4.4%) (0.65, 1.00)	-0.0349 (2.9%) (0.01, 0.37)
Frequentist VaR	-0.0184 (12.5%) (0.18, 1.00)	-0.0231 (5.5%) (0.70, 1.00)	-0.0321 (3.3%) (0.00, 0.15)
Bayesian ES	-0.0264 (0.00, 0.07)	-0.0310 (0.00, 0.05)	-0.0406 (0.03, 0.00)
Frequentist ES	-0.0246 (0.00, 0.00)	-0.0287 (0.00, 0.00)	-0.0366 (0.00, 0.00)

Table 4. The VaR and ES estimates within a normal model: Bayesian posterior predictive distribution with empirical Bayes priors (Bayesian VaR and ES) and a fitted model distribution (Frequentist VaR and ES). The VaR estimates include the proportion of breaches in brackets, then the p-values of unconditional coverage and conditional coverage tests are presented below. The ES estimates include the p-values of unconditional and direct tests below. Proper creation.

As expected, the Bayesian approach has provided results that are quite close to the frequentist ones, but the risk metrics estimations are a bit more severe: the difference is slightest for 10% metrics, but it becomes more pronounced for 1% metrics because. The backtesting results are also comparable, with a slight advantage in favor of the Bayesian results: the Bayesian normal model provides generally more precise estimates for VaR, but neither model has been able to estimate ES well because they have underestimated it.

An informative prior.

Just as an example of what the Bayesian approach is capable of in terms of incorporating prior knowledge, let us consider the following prior specifications for the unidimensional Bayesian normal model:

Parameter	μ_0	σ_0^2	k_0	v_0
Value	$\overline{y_{test}}$	$\hat{\sigma}_{test}^2$	88	88

Table 5. An informative prior specification for the Bayesian normal unidimensional model. Proper creation.

In other words, if the test sample were known in advance or if it was possible to infer these values in any reasonable way, i.e. by the means of another model or as an expert opinion, it would be possible to set the prior parameters to be equal to them ($\mu_0 = \overline{y_{test}}$ and $\sigma_0^2 = \hat{\sigma}_{test}^2$). Furthermore, let us assume that this prior knowledge is quite solid ($k_0 = v_0 = 88$), i.e. the "prior sample size is four times the actually available sample size (train set). The posterior checks analysis shows that the model fits the data observed in the test set to a higher extent, but it still lacks to capture maximum, minimum and skewness observed values, which can safely be attributed to an inadequacy of a normal model:

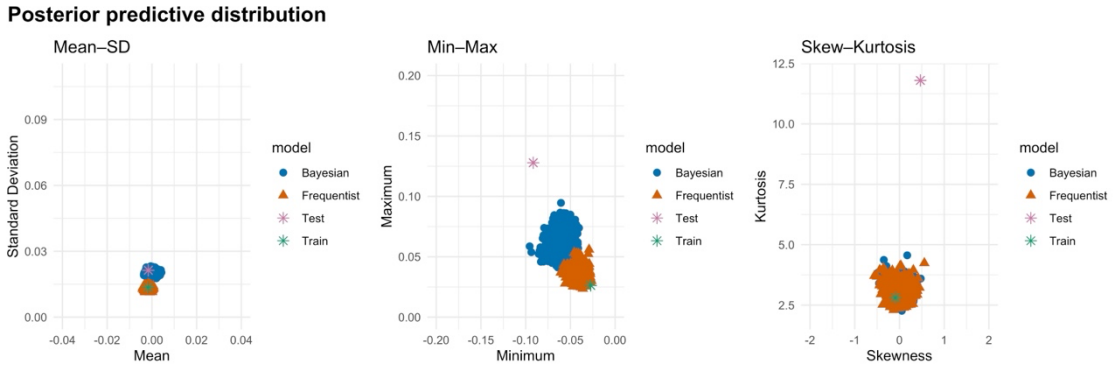

Figure 3. The unidimensional normal model with informative prior: posterior predictive distributions checks and comparison with the frequentist equivalent model in three plots: 1. Mean – Standard deviation; 2. Maximum – Minimum; 3. Kurtosis – Skewness. Proper creation.

The Kolmogorov-Smirnov test has been applied both to the Bayesian posterior predictive distribution and the frequentist fitted model distribution with respect to the observed test data. The associated p-values are again 0.01 and 0.02 respectively, with the same conclusion as before: neither method within a normal model has been able to provide a distribution that could be the same distribution as the observed test data.

The Bayesian posterior point estimates (mean) and the frequentist estimates for the model parameters are presented in the table below:

Parameter \ Estimate	Bayesian	Frequentist	Train	Test
$\hat{\mu}$	0.0000	-0.0016	-0.0016	0.0004
$\hat{\sigma}$	0.0201	0.0131	0.0134	0.0213

Table 6. The unidimensional normal model with an informative prior: the posterior point estimate (mean) and the frequentist estimate; sample values observed in the train and test datasets. Proper creation.

An informative prior has accounted for distinctions between the Bayesian posterior point estimates and the frequentist estimates for the model parameters. The most noticeable change has been accosted with the scale parameter incorporating a prior knowledge of more volatile character of the returns distribution.

The estimates of VaR and ES (Bayesian and frequentist) are in a table below:

Risk metrics \ quantiles	10%	5%	1%
Bayesian VaR	-0.0256 (4,4%) (0.00, 0.07)	-0.0332 (3,3%) (0.17, 1.00)	-0.0472 (2,2%) (0.09, 1.00)
Bayesian ES	-0.0355 (0.00, 0.97)	-0.0418 (0.00, 0.54)	-0.0543 (0.01, 0.01)

Table 7. The VaR and ES estimates within a normal model: Bayesian posterior predictive distribution with an informative prior (Bayesian VaR and ES) and a fitted model distribution (Frequentist VaR and ES). The VaR estimates include the proportion of breaches in brackets, then the p-values of unconditional coverage and conditional coverage tests are presented below. The ES estimates include the p-values of conditional and direct tests below. Proper creation.

The 10% and 5% ES estimates unconditional on the VaR estimates are much more precise than before. The 1% VaR estimate has improved a bit, but 10% and 5% VaR have much worse backtesting results than before. In other words, even if it were possible to guess in advance the test dataset with a high precision and be very sure about this guess, which was encoded into the

prior distribution, it would still be insufficient for the normal model to adequately estimate the market risk. A need in changing the model is a logical conclusion.

Mixture of two normals model

A mixture of two normal distributions model can potentially provide better estimates for risk metrics and represent the returns in a better way. As before, the model starts with the following election of the prior hyperparameters according to the empirical Bayes principle:

Empirical Bayes prior.

Parameter vector	μ_0	σ_0^2	k_0	v_0	(α_0, β_0)
Value	$\overline{(y_{train}, y_{train})}$	$(\hat{\sigma}^2_{train}, \hat{\sigma}^2_{train})$	(1,1)	(1,1)	(0.5, 0.5)

Table 8. An empirical Bayes prior specification for the Bayesian mixture of normals model. Proper creation.

Posterior predictive checks show that the two-component mixture model provides a better fit, both for the train and the test dataset, than both the frequentist equivalent model and the previously described normal model:

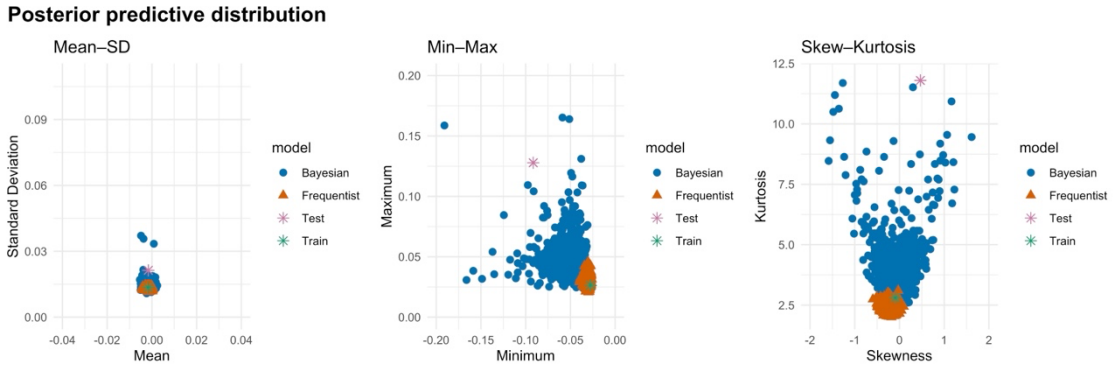

Figure 4. The unidimensional mixture of two normals model with an empirical Bayes prior: posterior predictive distributions checks and comparison with the frequentist equivalent model in three plots: 1. Mean – Standard deviation; 2. Maximum – Minimum; 3. Kurtosis – Skewness. Proper creation.

The posterior predictive distribution seems to account much better for the mean-variance sample values from both the train, and the test data sets, without incorporating any out-of-sample information. The posterior distribution of minimum and maximum values still generally fails to fit the observed values, but they have gotten much closer than the one-component normal model, even with an informative prior. However, the two-component normal mixture model hasn't coped with predicting a slight skewness and a high kurtosis of the test dataset.

The Kolmogorov-Smirnov test renders a p-value of 0.2 and 0.1 for the Bayesian posterior predictive distribution and the frequentist model fitted distribution respectively, so the null hypothesis of being drawn from the same distribution as the observed returns in the test set cannot be rejected. It indicates that the two-component normal mixture model provides a reasonably good fit to the stock's returns distribution.

The Bayesian posterior point estimates (mean) and the frequentist estimates for the model parameters differ significantly:

Parameter \ Estimate	μ	σ	ρ
Bayesian	(-0.0007, -0.0026)	(0.0103, 0.0157)	0.6564
Frequentist	(0.0024, -0.0229)	(0.0031, 0.0101)	0.8411

Table 9. The mixture of normals model with an empirical Bayes prior: the posterior point estimates (mean) and frequentist estimates. The vectors are ordered component-wise. The ρ parameter refers to the weight of the first component. Proper creation.

The first distinction of the results is that the Bayesian model proposes a more equibalanced mixture with the posterior mean of the mixture parameter being nearer to 0.5 than to 1, which is the case of the estimate provided by the frequentist equivalent model. Bayesian posterior point estimates for the variance parameters are higher than those provided by the frequentist approach, whereas the mean estimates are closer to 0.

The VaR and ES estimates are represented in the following table:

Risk metrics \ quantiles	10%	5%	1%
Bayesian VaR	-0.0195 (10,3%) (0.87, 1.00)	-0.0253 (4,4%) (0.65, 1.00)	-0.0380 (2,6%) (0.03, 0.79)
Frequentist VaR	-0.0222 (6,6%) (0.05, 1.00)	-0.0246 (4,8%) (0.87, 1.00)	-0.0278 (3,7%) (0.00, 0.05)
Bayesian ES	-0.0278 (0.01, 0.13)	-0.0335 (0.00, 0.14)	-0.0475 (0.02, 0.01)
Frequentist ES	-0.0249 (0.00, 0.24)	-0.0266 (0.00, 0.00)	-0.0293 (0.00, 0.00)

Table 10. The VaR and ES estimates within a two-component mixture of normals model: Bayesian posterior predictive distribution with an empirical Bayes prior (Bayesian VaR and ES) and a fitted model distribution (Frequentist VaR and ES). The VaR estimates include the proportion of breaches in brackets, then the p-values of unconditional coverage and conditional coverage tests are presented below. The ES estimates include the p-values of conditional and direct tests below. Proper creation.

The Bayesian point estimates of the risk metrics within a two-component normal mixture model are generally more accurate than the frequentist ones. It provides a more moderate 10% VaR, a comparable 5% VaR and a steeper 1% VaR corresponding to a more heavy-tailed posterior predictive distribution, which seems to be a more precise description of the test data. All three Bayesian estimates of ES are higher than the frequentist ones, but both approaches fail to estimate the 1% ES. Changing from a one-component to a two-component normal mixture model, both with empirical Bayes priors, has provided a significant improvement in market risk quantification. Nevertheless, the two-component normal mixture model is still far from predicting the ES in a precise way.

Factor-driven t-GARCH(1,1) model

The factor-driven t-GARCH(1,1) model consists of two models: the factor-driven conditional mean model and the t-GARCH(1,1) model for the conditional variance.

Conditional mean part.

For the conditional mean model, the mean returns of the following markets have been considered: S&P500, FTSE, IBEX35 and globally. The intercept term has also been included. A prior distribution for the conditional model coefficients is assumed to be a multivariate normal distribution. A choice for the hyperparameters has been the following:

Parameter vector	$\boldsymbol{\mu}_0$	$\boldsymbol{\Sigma}_0$	$\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0$
Value	$\widehat{\boldsymbol{\beta}_{MLE}}$	$\widehat{\boldsymbol{\Sigma}_{MLE}}$	**(1,1)**

Table 11. An empirical Bayes prior specification for the factor-driven part of the model. Proper creation.

In case of the hyperparameters associated with the model coefficients, an empirical Bayes prior has been chosen: the maximum likelihood estimates by the means of a frequentist linear regression model minimizing the sum of squared errors. The hyperparameters for the vectors associated with the variable selection parameter are unitary values, in which case the associated Beta(1,1) distribution converts itself into a Uniform(0,1) distribution representing an absence of prior knowledge, so this prior can be called uninformative.

The Bayesian posterior point estimates (mean | median) and the frequentist estimates for the model parameters are presented in the table below:

Model \ Estimate	$\mu_{Intercept}$	μ_{Global}	$\mu_{S\&P500}$	μ_{FTSE}	μ_{IBEX35}
Bayesian regression ($\neq 0$)	0.0007 | 0 (0.26)	0.2885 | 0 (0.38)	0.9185 | 0.9738 (0.78)	-0.1741 | 0 (0.29)	-0.4409 | 0 (0.48)
Frequentist linear regression (p-value)	0.0026 (0.402)	0.6143 (0.744)	1.1419 (0.268)	-0.2361 (0.752)	-0.9088 (0.152)
Frequentist ridge regression	0.0018	0.6335	0.6487	-0.1597	-0.4398
Frequentist lasso regression	0.0021	0	0.927	0	-0.2132

Table 12. The factor-driven part of the model: point estimates from the Bayesian posterior distribution, frequentist standard linear regression, the frequentist ridge regression and the frequentist lasso regression. A posterior expectation of the estimated coefficient of being different from exactly zero is shown below the estimated parameter value by the Bayesian regression. In case of the frequentist linear regression, the associated p-value is shown below the estimate. Proper creation.

From a purely logical point of view, the Bayesian regression is an evident winner: it has been able to recognize that the stock's returns (Apple) are associated primarily with the S&P500's average returns, so the draws from the posterior distribution of other coefficients have been exactly zero in more than a half of posterior draws. The frequentist linear regression has not been able to deliver the same result. Instead, it has provided estimates that are greater in absolute value, which can be attributed to a multicollinearity problem. There isn't a single coefficient with a p-value lower than any reasonable threshold (e.g., 0.01, 0.05, 0.1), so the null hypothesis of these coefficients being equal to zero individually cannot be rejected. The frequentist ridge regression has apparently coped with the problem of multicollinearity since all the estimates (except for the μ_{Global}) that it provides are much smaller in absolute value. It should be noted that the estimates for μ_{FTSE} and μ_{IBEX35} are quite close to the ones provided by the Bayesian posterior mean. However, the ridge regression has the same drawback as the usual linear regression: it has provided a non-zero estimates for all the regressors, which does not seem reasonable for the problem at hand. The frequentist lasso regression has provided the most convincing results among the three frequentist regression models since it has estimated the coefficients of μ_{Global} and μ_{FTSE} to be exactly zero. Moreover, its estimate for $\mu_{S\&P500}$ is very close to the Bayesian posterior mean estimate. However, it hasn't discarded a relation between the stock's returns and the IBEX35 returns, although the absolute value of the estimated coefficient is a half of the Bayesian posterior mean estimate.

All-in-all, key virtues of the Bayesian factor-driven model set-up are the following: a flexibility to incorporate a prior knowledge or its absence; a possibility to choose a Bayesian point estimator or use the posterior distribution directly without the need in choosing any point

estimator at all; a greater ability in identifying the regressors that are of a higher subjacent importance. It should also be noted that, apart from the information encoded into the prior distributions, the Bayesian model does not require any further assumptions to carry out a variable selection process, yet it provides a much more accurate result in terms of underlying logic, e.g. that the stock's returns primarily depend on the S&P500's returns.

The next step is to examine the performance these models using the test dataset:

Method \ Metrics	Bayesian posterior mean	Bayesian posterior median	Pseudo Bayesian posterior mean	Pseudo Bayesian posterior median	Frequentist linear regression	Frequentist ridge regression	Frequentist lasso regression
MSE	0.030	0.0287	0.030	0.0243*	0.0347	0.0286	0.0277
MAE	1.136	1.1170	1.137	1.0678*	1.2342	1.1255	1.1195
R2	0.3018	0.3341	0.3040	0.4352*	0.1929	0.3366	0.3556

Table 13. The factor-driven part of the model: point estimates from the Bayesian posterior distribution, frequentist standard linear regression, the frequentist ridge regression and the frequentist lasso regression. Predictive performance metrics. MSE and MAE are multiplied by 100. Proper elaboration.

The Bayesian framework offers a great flexibility of point estimators depending on a decision rule (loss function) and an approach: a purely Bayesian one or a pseudo-Bayesian one. The purely Bayesian approach provides point estimates based on the posterior predictive distribution which accounts for both sample uncertainty and parameter uncertainty. The pseudo-Bayesian approach is an interchange between the purely Bayesian and the frequentist approach: it consists in deriving the point estimates from the posterior distribution of the parameters and using this as a fixed input into the conditional model, eliminating the parameter uncertainty but taking advantage of using Bayesian statistical methods. The Pseudo Bayesian posterior median point estimator has provided the best predictive performance in all three considered metrics (MAE, MSE and R2). The differences between the point estimates provided by this method and the rest are statistically significant at 5% level. Nevertheless, there differences between the point estimates provided by the purely Bayesian and the frequentist ridge and lasso regressions are not statistically significant. The frequentist linear regression has been the single worst-performing statistically significant method. For the p-values associated with the Diebold-Mariano (DM) tests, see Annex 1.19.

As a conclusion, the Bayesian regression model provides a much more flexible modeling framework for the conditional expectation than the standard frequentist models providing posterior distribution for all the model parameters and point estimates that are at least as good as the ones obtained by the means of the frequentist models in term of predictive accuracy.

Conditional variance part.

The conditional variance part has been modeled in a joint way with the conditional mean part, and the posterior distributions have been approximated within the same MCMC algorithm, so they share the same conditional model. The only part that remains to be determined is the prior distributions for the t-GARCH(1,1) model.

The choice of the prior distribution for the t-GARCH(1,1) model parameters is the following: a diffuse inverse-gamma distributions with the vector of parameters (10,1) for the unconditional variance γ; a Uniform(0,1) distribution for α and θ with a restrictions of $\alpha + \theta < 1$ ensured by rescaling (see Annex 1.15 for details); and a Uniform(1,100) distribution for the Student-t distribution v. It is a well-known fact that a Student-t distribution almost converges to

a normal distribution when the degrees of freedom parameter is above 30, so limiting the degrees of freedom by 100 can hardly be considered a real restriction. These choices are summarized in the following table:

Parameter vector	γ_0	α_0	θ_0	v_0
Value	IG(10,1)	U(0,1)	U(0,1)	U(1, 100)

Table 14. Prior specification for the t-GARCH(1,1) part of the model. Proper creation.

The Bayesian posterior point estimates (mean) and the frequentist estimates for the model parameters are presented in the table below:

Model \ Estimate	γ	α	θ	v
Bayesian (train)	0.0001	0.2712	0.3567	50.52
Frequentist (train)	0.0000	0.05	0.90	4.00
Bayesian (train + test)	0.0001	0.2140	0.1700	5.00
Frequentist (train + test)	0.0001	0.4390	0.3320	3.64

Table 15. The t-GARCH(1,1) part of the model: point estimates from the Bayesian and the frequentist models applied to the train dataset and the complete (train + test) dataset. Proper elaboration.

The Bayesian and the frequentist point estimates for the t-GARCH(1,1) model conditional on the previously discussed factor model for the conditional mean clearly show both methods provide completely different results. Nevertheless, if the frequentist model is estimated considering the whole dataset (train and test), then the estimates get very close to the Bayesian ones estimated with train data only. It is a remarkable result because it confirms both consistency and stability of Bayesian point estimators, even when the sample size is small. When the same is done for the Bayesian model, the point estimate for v approaches the frequentist estimate, the point estimates for γ and α stay reasonably stable, but the estimate for θ is halved. A possible explanation for this fact is the influence of the conditional mean model which has also been updated in light of the new data (the test dataset). The differences between the Bayesian and the frequentist point estimates, when the complete dataset is provided, are believed to stem from the fact that the associated conditional mean models provide different results.

Complete model

From now one, the rest of the discussion and comparison between the Bayesian factor-driven t-GARCH(1,1) model and its frequentist equivalent will be carried out for the complete model, i.e. considering both parts as a whole. The lasso regression in conjunction with a maximum likelihood estimation for the t-GARCH(1,1) model represents the frequentist equivalent model. The MCMC convergence checks can be found in Annex 1.21.

Posterior predictive checks show that both the Bayesian and the frequentist models provide a reasonably good fit, both for the train and the test data:

Figure 5. The factor-drive t-GARCH(1,1) model: posterior predictive distributions checks and comparison with the frequentist equivalent model in three plots: 1. Mean – Standard deviation; 2. Maximum – Minimum; 3. Kurtosis – Skewness. Proper creation.

The Bayesian factor-driven t-GARCH(1,1) model has provided a substantial improvement in terms of posterior checks in comparison with both a normal model and a two-component mixture of normals model. It also appears to be as good as the frequentist equivalent model in terms of posterior checks. However, it seems that the Bayesian approach account for the characteristics observed both in the train and in the test data, whereas the frequentist approach overestimates the kurtosis, so it doesn't fit the train data well in this sense. A high value of kurtosis might be a proper characteristic of the stock's returns or a result of a sample uncertainty.

The Kolmogorov-Smirnov test renders a p-value of 0.63 and 0.913 for the Bayesian posterior predictive distribution and the frequentist model fitted distribution respectively, so the null hypothesis of being drawn from the same distribution as the observed returns in the test set cannot be rejected. It indicates that the factor-driven t-GARCH(1,1) model provides a reasonably good fit to the stock's returns distribution.

The % of breaches and the associated p-values with the VaR and ES estimates for the Bayesian and frequentist approaches are represented in the following table:

Risk metrics \ quantiles	10%	5%	1%
Bayesian VaR	5.9% (0.015, 1.00)	3.3% (0.174, 1.00)	0.7% (0.65, 1.00)
Frequentist VaR	11.4% (0.45, 1.00)	5.9% (0.52, 1.00)	2.2% (0.08, 1.00)
Bayesian ES	(0.27, 0.80)	(0.66, 0.07)	(0.01, 0.44)
Frequentist ES	(0.71, 0.00)	(0.11, 0.31)	(0.92, 0.00)

Table 16. The VaR and ES estimates within a factor-driven t-GARCH(1,1) model: Bayesian posterior predictive distribution with an empirical Bayes prior (Bayesian VaR and ES) and a fitted model distribution (Frequentist VaR and ES). The VaR estimates include the proportion of breaches in brackets, then the p-values of unconditional coverage and conditional coverage tests are presented below. The ES estimates include the p-values of conditional and direct tests below. Proper creation.

The risk metrics estimation results clearly show that factor-based t-GARCH(1,1) model is the best among the three models considered for the unidimensional market risk modeling. The Bayesian approach estimates all risk metrics with a reasonably high precision, except for the 10% VaR. The frequentist estimation has provided quite accurate estimates for the 10% and 5% VaR, but it failed to estimate the 10% and 1% ES and the 1% VaR unconditionally. These results may seem a bit confusing, but it should be noted that the observations window contained 22 consecutive daily observations for the stock's logarithmic returns only. Therefore, in order to draw a conclusion about the factor-based t-GARCH(1,1) model ability to estimate the market

risk metrics in general and whether there is a substantial difference between the Bayesian and the frequentist approach to its implementation, a larger sample of stocks must be considered under different timescales and types of returns.

Massive application

Without a doubt, the factor-driven t-GARCH(1,1) model has provided the best fit to the observed data in the test set and the most accurate estimates for VaR and ES. However, it is unclear if this fact is due to a chance or it reflects a general characteristic. It can be found out by applying it to a large sample of stock returns under different conditions.

In this section a large sample of 1003 stocks is considered. The train data consists of consecutive observations for daily (22, 66 and 252 days), weekly (26 and 52 weeks) and monthly (12 and 24 months) returns. The test dataset comprises 252 daily, 400 weekly and 60 monthly consecutive returns, which immediately follow the train datasets. For each of these conditions the model will be applied individually to each of the 1003 time series of stock returns. The frequentist alternative model will be fitted in all cases.

First, the goodness-of-fit and the predictive performance metrics are shown:

Risk metrics \ quantiles	KS test	MSE	MAE	R2
Bayesian	0.92	0.0415 (75%)	0.0133 (78%)	0.1094 (75%)
Frequentist	0.60	0.0441 (25%)	0.0138 (22%)	0.0515 (25%)

Table 17. The Kolmogorov-Smirnov (KS) test, MAE, MSE and R2 for the factor-driven t-GARCH(1,1) model under the Bayesian and the frequentist approaches. For KS test, the proportion of stocks for which the null hypothesis cannot be rejected at 5% of confidence level is shown. For MSE (multiplied by 100), MAE and R2, an average value is presented first, then a % of stocks for which the method's metrics is lower than the one provided by the competing method. Daily returns, 22 days observation window. Proper creation.

Bayesian point estimates of returns are consistently more precise than the frequentist ones. It should be noted that the estimation has been carried out with a train dataset, while these metrics have been obtained with a test dataset. Therefore, there is no guarantee that the point estimates for the conditional mean should provide accurate results in the test dataset.

The test quantities include an average of % of VaR breaches, the percentage of VaR and ES estimates that are considered accurate with a 5% of significance according to the associated backtesting procedures for the sample of 1003 stocks:

Risk metrics \ quantiles	10%	5%	1%
Bayesian VaR	9,0% (0.68, 0.89)	5,0% (0.72, 0.90)	1,7% (0.75, 0.88)
Frequentist VaR	13,3% (0.53, 0.88)	8,4% (0.53, 0.87)	3,8% (0.50, 0.76)
Bayesian ES	1.28 (0.91, 0.33)	1.20 (0.94, 0.33)	1.19 (0.95, 0.14)
Frequentist ES	1.94 (0.67, 0.01)	6.52 (0.66, 0.01)	1.46 (0.64, 0.05)

Table 18. The VaR and ES estimates obtained via Monte Carlo simulations from the Bayesian posterior predictive distribution (Bayesian VaR) within the factor-driven t-GARCH(1,1) and from the frequentist fitted model distribution (Frequentist VaR). Average values for a sample of 1003 stocks are shown. For VaR, % of

breaches is presented first, and the p-values for backtesting is presented below them in the following order: (unconditional coverage, conditional coverage). For ES, the sample average ES ratio (real/estimated) is provided, and the proportion of stocks for which the null hypothesis of the conditional and the direct tests is not rejected is shown below. Daily returns, 22 days observation window. Proper creation.

Bayesian VaR estimates are highly precise on average, passing the unconditional VaR backtesting for more than 68% of stocks and the conditional VaR backtesting for more than 88% of stocks. The frequentist results are more moderate: 50% and 76% respectively. The Bayesian estimate of ES is also accurate: for more than 91% of stocks the ES test conditional on VaR accepts its estimation as a valid one, whereas the frequentist result is around 64%. Moreover, the unconditional ES test ranges from 14% (1% ES) to 33% (10% and 5% ES) of stocks for which the null hypothesis cannot be rejected, i.e. the estimate is accurate. In the case of the frequentist approach, the results range from 1% (10% and 5% ES) to 5% (1% ES), indicating that it fails to model the market risk in a consistent way with a limited number of observations (22 consecutive observations of daily logarithmic returns).

This massive application of the factor-based t-GARCH(1,1) model to the daily logarithmic returns of 1003 stocks under a 22-day window train set shows that the Bayesian approach provides consistently more accurate estimation of VaR and ES, as well as point estimators of the returns. The rest of the results is presented in Annex 1.19, which analysis is provided below:

Viewing the results as a function of a train dataset sample size, the Bayesian point estimates of VaR, ES and conditional mean of returns consistently outperform the frequentist estimates, although they converge steadily, which is quite expected due to the Theorem 2.4 and Theorem 2.5. The application of the model to weekly and monthly returns renders the same result: the Bayesian point estimates of the VaR and the conditional mean not only consistently outperform the frequentist estimates but also have a high degree of accuracy. The frequentist VaR underestimates the real risk, which is reflected by an elevated percentage of breaches, generally much higher than the nominal ones. It indicates that the frequentist estimation of the model couldn't adequately fit the model parameters, a typical problem of this method that is more pronounced in small sample cases (monthly returns). The Bayesian approach tackles with small samples as successfully as big samples, so that the variability of the VaR breaches across different testing conditions are minimal and are practically centered at the nominal values. It should be noted that a steady convergence of the Bayesian point estimates to the frequentist ones takes place in all types of returns. The major differences are found for monthly logarithmic returns data under 12-month observation window, which represents the smallest size of a training set (12 observations). Nevertheless, both methods generally underestimate the 1% VaR because the percentage of observed breaches is always higher than the nominal level, although it passes the backtesting procedures. It can indicate that the conditional model itself isn't perfect and other options should be explored to estimate the tail risk better.

Therefore, the following conclusion can be drawn: a factor-drive t-GARCH(1,1) model provides a reasonably good fitting for the logarithmic returns data for a range of timescales (daily, weekly and monthly returns). The Bayesian version of the model outperforms the frequentist equivalent model. The smaller the training sample size, the more the Bayesian approach outperforms the frequentist one.

Hierarchical normal model

The last model to apply is the Bayesian hierarchical normal model. It can be viewed as a direct generalization of a one-dimensional normal model and can be used to assess market risk at an individual, market and even global level. This model will be applied directly to the large sample of daily logarithmic returns of 1003 stocks with 252 daily consecutive observations.

A set-up of the model is identical to the one explained earlier for the one-dimensional normal model.

Empirical Bayes prior.

An empirical Bayes ($\mu_0 = \overline{y_{train}}$ and $\sigma_0^2 = \hat{\sigma}_{train}^2$) diffuse ($k_0 = v_0 = 1$) has been employed to set up the Bayesian hierarchical normal model:

Parameter	θ_0	$\boldsymbol{\sigma_0^2}$	$k_{0,\theta}$	$\boldsymbol{v}_{0,\sigma^2}$	γ_0^2	$\boldsymbol{v}_{0,\gamma_0^2}$
Value	$\overline{y_{train}}$	$\hat{\sigma}_{train}^2$	1	1	$\overline{\hat{\sigma}_{train}^2}$	1

Table 19. An empirical Bayes prior specification for the Bayesian normal unidimensional model. Proper creation.

The difference between a one-dimensional set-up and the hierarchical normal model is that there are 1003 prior distributions for the individual scale parameters σ^2 and degrees of freedom parameters $\boldsymbol{v}_{0,\sigma^2}$. The global mean ($\theta$) and its variance ($\gamma^2$) are taken to be the average and a standard deviation of the returns of all 1003 stocks. The prior distributions are chosen to be the most diffuse possible: $k_{0,\theta} = \boldsymbol{v}_{0,\sigma^2} = \boldsymbol{v}_{0,\gamma_0^2} = 1$.

First, the posterior distributions are approximated numerically according to the Annex 1.16. Then, all the tests provided in previous sections are applied to the observed data from the test dataset:

Risk metrics \ quantiles	KS test	MSE	MAE	R2
Bayesian	0.66	0.047 (83%)	0.0143 (75%)	0.00 (83%)
Frequentist	0.65	0.048 (17%)	0.0144 (25%)	-0.02 (17%)

Table 20. The Kolmogorov-Smirnov (KS) test, MAE, MSE and R2 for the Bayesian hierarchical model and frequentist individual fitting of normal distributions. For KS test, the proportion of stocks for which the null hypothesis cannot be rejected at 5% of confidence level is shown. For MSE (multiplied by 100), MAE and R2, an average value is presented first, then a % of stocks for which the method's metrics is lower than the one provided by the competing method. Daily returns, 252 days observation window. Proper creation.

Although the differences are very small, Bayesian point estimates of returns are consistently more precise than the frequentist ones. It should be noted that the estimation has been carried out with a train dataset, while these metrics have been obtained with a test dataset. Therefore, there is no guarantee that the point estimates for the conditional mean should provide accurate results in the test dataset.

The test quantities include an average of % of VaR breaches, the percentage of VaR and ES estimates that are considered accurate with a 5% of significance according to the associated backtesting procedures for the sample of 1003 stocks:

Risk metrics \ quantiles	10%	5%	1%
Bayesian VaR	8,8% (0.67, 0.86)	5,4% (0.75, 0.92)	2,5% (0.57, 0.75)
Frequentist VaR	9,0% (0.68, 0.86)	5,6% (0.73, 0.91)	2,7% (0.53, 0.73)
Bayesian ES	1.28 (0.17, 0.80)	1.34 (0.15, 0.69)	1.40 (0.22, 0.42)
Frequentist ES	1.29 (0.15, 0.77)	1.34 (0.15, 0.65)	1.40 (0.20, 0.39)

Table 21. The VaR and ES estimates obtained via Monte Carlo simulations from the Bayesian posterior predictive distribution (Bayesian VaR) within the Bayesian hierarchical normal model and from the frequentist fitted model distribution (Frequentist VaR). Average values for a sample of 1003 stocks are shown. For VaR, % of breaches is presented first, and the p-values for backtesting is presented below them in the following order: (unconditional coverage, conditional coverage). For ES, the sample average ES ratio (real/estimated) is provided, and the proportion of stocks for which the null hypothesis of the conditional and the direct tests is not rejected is shown below. Daily returns, 252 days observation window. Proper creation.

There is hardly any difference between the estimates of VaR and ES provided by the Bayesian hierarchical normal model and the frequentist individual fitting of normal distributions. Moreover, a large training sample size should also have practically neglected the influence of prior distributions. However, there is a slight advantage in favor of the Bayesian hierarchical model in estimating VaR and ES.

One of the well-known effects of a Bayesian hierarchical model a *pulling effect* on the means which consists in pulling the estimates of the means of individual units (stocks' returns) around a common mean (market):

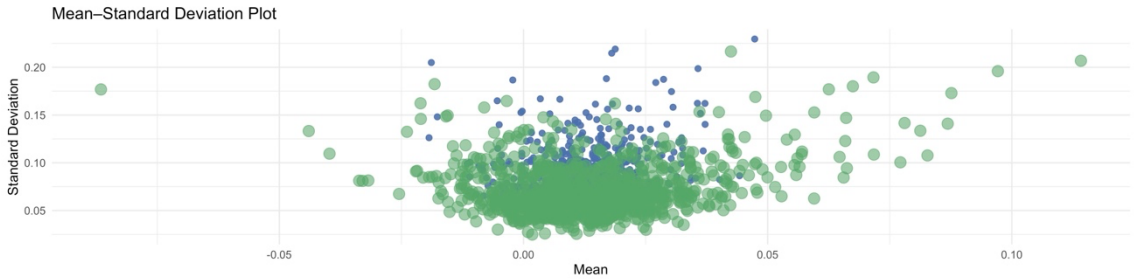

Figure 6. The Bayesian hierarchical normal model: the means of the posterior distribution of means and variances of each of the 1003 individual stocks (blue) along with the frequentist sample estimates (green). The pooling effect of the means and higher posterior variance. Proper creation.

It is also the case that the averages of the posterior variances are a little higher than the sample unbiased estimates of the variances. In turn, the means of the market returns are pulled around the global mean. The force of the pulling effect depends directly on the sample size: if a stock has few observations, it will be pulled with a greater force than if it has many observations. The same hold for markets: if there are very few markets included into the analysis, their means will be pulled together with a greater force than it there were many different markets.

The real power of the Bayesian three-level hierarchical normal model consists in an inference of market-level and global-level distribution of returns:

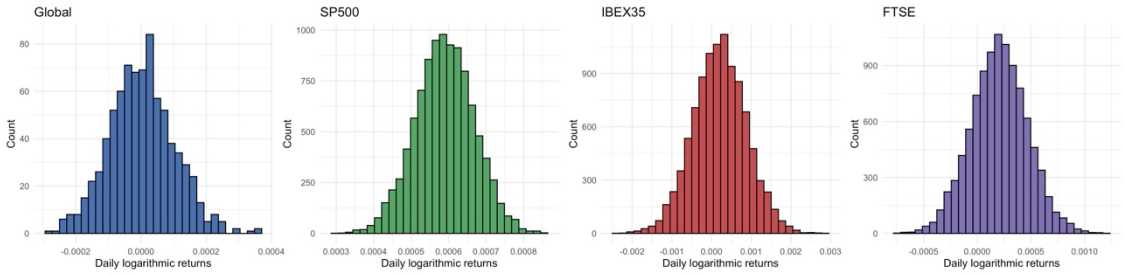

Figure 7. The Bayesian hierarchical normal model: the posterior distributions of the location of global, S&P50, IBEX35 and FTSE daily logarithmic returns. Proper creation.

The global mean returns are practically centered at 0 and have less variability than any market-mean parameter. It highlights the fact that the diversification isn't limited to a distribution of a portfolio inside one particular market: there is a specific market (or country) risk that is associated with each market, which can be hedged away by diversifying the portfolio with stocks from different markets. Another finding is that the posterior mean of market means of 16 out of 19 markets are positive (3 out of 19 markets are shown in Figure 7 above: S&P500, IBEX35 and FTSE). This may indicate that there are possibly arbitrage opportunities present in most market, which besides being important to market risk assessment, is a relevant question in quantitative finance.

One of the unique uses that can be given to the Bayesian hierarchical normal model is testing a simplified version of the efficient market hypothesis: if the value of 0 enters the α-probability posterior credible interval. It is essentially a frequentist formulation because under the pure Bayesian approach it would be more natural to put the question in a form of the posterior odds of a positive return to a negative return, for example. Nevertheless, even the frequentist formulation can be successfully addressed by the Bayesian hierarchical model at individual, market and global level:

Level \ Credible interval	90%	95%	99%
Individual	91,63%	96,61%	99,4%
Market	18/19	18/19	18/19
Global	1/1	1/1	1/1

Table 22. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Daily logarithmic returns, 252 consecutive daily observations in training set. Proper creation.

The efficient market hypothesis cannot be rejected for 96,61% of daily logarithmic returns of 1003 analyzed stocks on an individual level. A very similar conclusion can be drawn on a market level where the hypothesis is not rejected for 18 out of 19 analyzed markets. It should be noted that the hypothesis is rejected for the market composed by the companies comprising the S&P500 index, which is the largest market-group unit (470 stocks out of 1003). The implication of these findings is that on a global level there are no unconditional arbitrage opportunities. However, at 5% significance level, there is a market (S&P500) and some stocks (around 3,4%) that may admit arbitrage opportunities.

The results of the same analysis applied to the daily, weekly and monthly returns under different training observation windows can be found in Annex 1.20. All-in-all, similar conclusions hold for all timescales and training sample sizes. However, the hypothesis is rejected for more stocks and markets considering weekly returns, and in even more cases taking into account monthly returns. More observations also lead to a higher percentage of the hypothesis rejection.

Conclusions and future investigation

This thesis conducts a study of three Bayesian statistical models for unidimensional time series of market returns (a normal, a two-component mixture of normals and a factor-driven t-GARCH(1,1) models); and a Bayesian hierarchical model for multidimensional modeling considering three levels of hierarchy: a stock level, a market level and a global level. The missing data is treated by the means of a Bayesian statistical approach.

The unidimensional models are developed and applied to an example stock's returns (Apple). The example is carried out with a small train sample of 22 consecutive price observations and a large test sample of 252 immediately posterior consecutive observations. The VaR and ES are estimated by the Bayesian models and the frequentist equivalent models, a detailed backtesting procedures are provided. In case of a conditional mean model, the predictive accuracy is examined. Posterior predictive checks are provided for each model.

The factor-driven t-GARCH(1,1) model has been found to be the most accurate in estimating the market risk since its Bayesian version has passed all the backtesting procedures for 10%, 5% and 1% VaR and ES estimates. The model includes a variable selection latent variable which has permitted to identify the single most important regressor (S&P500 returns). For the example at hand, it is a logical and expected result. Moreover, the Bayesian estimates for the conditional mean provide the best predictive performance in comparison with the frequentist ones.

A massive application of the factor-driven t-GARCH(1,1) model has been carried out on a large sample of 1003 stocks from 19 different markets for 22, 66 and 252 daily observations; 26 and 52 weekly observations; and 12 and 24 monthly observations. The Bayesian version of the factor-driven t-GARCH(1,1) model provides consistently better results than the frequentist version. The Bayesian VaR point estimates are highly precise, so that the percentage of observed breaches is near the nominal values and their variation across both sample sizes and types of returns isn't statistically significant, so that in a vast majority of cases the backtesting procedures confirm the accuracy of the risk metrics on a test dataset.

The Bayesian hierarchical model has been applied to the same large sample of stocks under the same conditions. The model provides a unique way of inferring into the global market structure: the market-specific (country) risk can be measured for global and market means of returns. The efficient market hypothesis has been shown to hold on a global level under all considered conditions (daily, weekly and monthly logarithmic returns and different training sample sizes), for a vast majority of individual markets and stocks. However, there are particular stocks and markets that are susceptible of rejecting the efficient market hypothesis. It should be noted that the Bayesian hierarchical model also provides estimates for market risk metrics not only on an individual level, but also on a market and a global level. Therefore, it provides a unique tool of inferring into latent market risk structures of logarithmic returns.

Future investigation

There are several future investigation opportunities that may stem from this work. First, the factor-driven GARCH(1,1) model may be extended to carry out variable selection of the GARCH model specification, and not only of the factors. An effect of different elections of prior distributions is also a promising investigation line.

The Bayesian hierarchical model presented in this thesis can be extended in a number of ways: for instance, the individual stocks may be modeled as a multivariate normal distribution across markets. Another option is to include a conditional variance model for each stock, e.g. a GARCH model. It is also possible to account for a Bayesian regression within the hierarchical model permitting for a conditional mean modeling of the logarithmic returns of individual stocks. This extension could include the factor-driven t-GARCH(1,1) model within the Bayesian hierarchical model.

References

1. Acerbi, C., & Szekely, B. (2014). Backtesting expected shortfall. MSCI Inc.
2. Alexander, C. (2009). Market risk analysis, volume IV: Value at risk models (1st ed.). Wiley.
3. Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3), 637–654.
4. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327.
5. Casella, G. (1985). An introduction to empirical Bayes data analysis (Technical Report No. 85–11). Cornell University, Department of Statistics.
6. Christoffersen, P. F. (1998). Evaluating interval forecasts. International Economic Review, 39(4), 841–862.
7. Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223–236.
8. Das, S. (2025). Predicting stock market crash with Bayesian generalised Pareto regression. In Special proceedings of the 27th Annual Conference. 121–134. Chennai Mathematical Institute.
9. Diebold, F. X. (2012). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests (NBER Working Paper No. 18391). National Bureau of Economic Research.
10. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.
11. Duan, J., Liu, H., Zeng, J. (2013). Posterior probability model for stock return prediction base don analyst's recommendation behaviour. Knowledge-Based Systems. Volume 50, 151-158.
12. Feng, G., & He, J. (2022). Factor investing: A Bayesian hierarchical approach. Journal of Econometrics, 230(1), 183–200.
13. Fornacon-Wood, I., Mistry, H., Johnson-Hart, C., Faivre-Finn, C., O'Connor, J. P. B., & Price, G. J. (2022). Understanding the differences between Bayesian and frequentist statistics. International Journal of Radiation Oncology, Biology and Physics. 112(5), 1076-1082.
14. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC.
15. Geweke, J., & Amisano, G. (2008). Comparing and evaluating Bayesian predictive distributions of asset returns. Working Paper 969. European Central Bank.
16. Greyserman, A., Jones, D. H., & Strawderman, W. E. (2006). Portfolio selection using hierarchical Bayesian analysis and MCMC methods. Journal of Banking & Finance, 30, 669–678.
17. Hartigan, J. A. (1966). Estimation by ranking parameters. Journal of the Royal Statistical Society. Series B (Methodological), 28(1), 32–44.
18. Hartmann, S., & Sprenger, J. (2013). Bayesian Epistemology. The Routledge Companion to Epistemology. 609-620.
19. Hoff, P. D. (2009). A First Course in Bayesian Statistical Methods. Springer.
20. Hoogerheide, L. & van Dijk, H.K. (2008). Bayesian Forecasting of Value at Risk and Expected Shortfall using Adaptive Importance Sampling. Tinbergen Institute Discussion Paper. 08-092/4, Tinbergen Institute.

21. Karami, H., Luo, R., Sanaei, P., Chowell, G. (2025). Comparative study of Bayesian and Frequentist methods for epidemic forecasting: Insights from simulated and historical data. Statistical Methods in Medical Research. 1-42.
22. Klenke, A. (2020). Probability theory: A comprehensive course (3rd ed.). Springer.
23. Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives, 3(2), 73–84.
24. Lehmann, E. L., & Casella, G. (1998). Theory of Point Estimation (2nd ed.). Springer.
25. Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), 77–91.
26. Savage, L. J. (1954). The Foundations of Statistics. John Wiley & Sons.
27. So, M. K. P., Chen, C. W. S., Lee, J. Y., & Chang, Y. P. (2008). An empirical evaluation of fat-tailed distributions in modeling financial time series. Mathematics and Computers in Simulation, 77(1), 96–108.
28. Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal of Finance, 19(3), 425–442.
29. Stein, C. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1, 197–206. University of California Press.
30. Stephens, M. (1997). Bayesian Methods for Mixtures of Normal Distributions. PhD Thesis. University of Oxford.
31. van der Vaart, A. W. (2012). Asymptotic Statistics. Cambridge University Press.
32. Yan, H., & Han, L. (2019). Empirical distributions of stock returns: Mixed normal or kernel density? Physica A: Statistical Mechanics and its Applications, 514, 473–486.

Annex

Annex 1.1. Proof of the Theorem 2.2.

First, assume that $p(\theta)$ is a continuous function, i.e. Θ is a continuous random variable. It is sufficient to check the following two conditions for $p(\theta|y)$ to be a probability density function:

1. $p(\theta|Y = y) \geq 0$. It follows from the fact that $p(y, \theta)$ and $p_Y(y)$ are proper probability distribution functions themselves, so they are nonnegative, and their quotient is nonnegative as well.
2. $\int_\Theta p(\theta|Y = y)d\theta = 1$. Without loss of generality it may be assumed that $Y, \Theta \subset \mathbb{R}$, so $\int_\Theta p(\theta|Y = y)d\theta = \int_{-\infty}^{+\infty} p(\theta|Y = y)d\theta = \int_{-\infty}^{+\infty} \frac{p(y,\theta)}{p_Y(y)} d\theta = \frac{1}{p_Y(y)} \int_{-\infty}^{+\infty} p(y, \theta)d\theta = \frac{1}{p_Y(y)} p_Y(y) = 1$.

Therefore, $p(\theta|Y = y)$ is a proper probability density function. Moreover, since $p(y, \theta) = p(y|\theta)p(\theta)$ by Definition 2.1 and $p_Y(y) = \int_{\theta\in\Theta} p(y|\theta)p(\theta)\, d\Theta$ by the rule of total probability, the second part of the theorem follows immediately.

Second, assume that Θ is a discrete random variable with a probability mass function $f(\theta)$. This case is treated in the exact same way with the same result.

Annex 1.2. Sketch of proof of the Theorem 2.3.

$\begin{cases} Y_1, \dots, Y_n|\theta \sim i.i.d. \\ \theta \sim p(\theta) \end{cases} \Rightarrow Y_1, \dots, Y_n$ are exchangeable (the "if" part).

$$p(y_1, \dots, y_n) = \int_\Theta p(y_1, \dots, y_n|\theta)\, p(\theta)d\theta = \int_\Theta \left\{\prod_{i=1}^{n} p(y_i|\theta)\right\} p(\theta)d\theta$$
$$= \int_\Theta \left\{\prod_{i=1}^{n} p(y_{\pi_i}|\theta)\right\} p(\theta)d\theta = \int_\Theta p(y_{\pi_1}, \dots, y_{\pi_n}|\theta)\, p(\theta)d\theta = p(y_{\pi_1}, \dots, y_{\pi_n})$$

The first equality is the law of total probability, the second equality follows from the definition of conditional independence, and the third equality is a consequence that a product is commutative. Therefore, the conditional independence implies exchangeability.

The proof of the inverse implication (the "only if" part) can be found in (Klenke, 2008) and will not be presented here because it is lengthy and is not used in this work.

Annex 1.3. Proof of the Proposition 2.6.

Part I. The expected squared error risk function is defined as follows:

$$p(a) = \mathbb{E}[L(\theta, a)|Y = y] = \mathbb{E}[(a - \theta)^2|Y = y] = \int_\Theta (a - \theta)^2 p(\theta|Y = y)d\theta$$

It is possible to differentiate $p(a)$ applying the Leibnitz's rule:

$$\frac{\partial p(a)}{\partial a} = \int_\Theta 2(a - \theta)p(\theta|Y = y)d\theta = 2\left(a - \int_\Theta \theta p(\theta|Y = y)d\theta\right) = 0 \rightarrow a = \mathbb{E}[\theta|Y = y]$$

The second-order derivative can be obtained in the same way:

$$\frac{\partial^2 p(a)}{\partial a^2} = 2 > 0$$

Which shows that $a = \mathbb{E}[\theta|Y = y]$ is a local minimum of the expected risk function $p(a)$. Since the problem is a convex optimization problem, it is also a global minimum. The existence is guaranteed if $\mathbb{E}[\theta|Y = y] < \infty$ and uniqueness is guaranteed provided that $\mathbb{E}[\theta^2|Y = y] < \infty$.

Part II. The expected absolute value error risk function is defined as follows:

$$p(a) = \mathbb{E}[L(\theta, a)|Y = y] = \mathbb{E}[|a - \theta| \mid Y = y] = \int_{\Theta} |a - \theta| p(\theta|Y = y) d\theta$$
$$= \int_{-\infty}^{a} (a - \theta) p(\theta|Y = y) d\theta + \int_{a}^{\infty} (\theta - a) p(\theta|Y = y) d\theta$$

It is possible to differentiate $p(a)$ applying the Leibnitz's rule:

$$\frac{\partial p(a)}{\partial a} = \int_{-\infty}^{a} p(\theta|Y = y) d\theta - \int_{a}^{\infty} p(\theta|Y = y) d\theta = \mathbb{F}_a[\theta|Y = y] - (1 - \mathbb{F}_a[\theta|Y = y])$$
$$= 2\mathbb{F}_a[\theta|Y = y] - 1 = 0 \rightarrow \mathbb{F}_a[\theta|Y = y] = 1/2$$

By definition, $\mathbb{F}_a[\theta|Y = y] = 1/2$ is a posterior median of θ, so $a = \mathbb{F}_{1/2}[\theta|Y = y]$. The second-order derivative can be obtained in the same way:

$$\frac{\partial^2 p(a)}{\partial a^2} = 2\frac{\partial \mathbb{F}_a[\theta|Y = y]}{\partial a} = 2\frac{\int_{-\infty}^{a} p(\theta|Y = y) d\theta}{\partial a} = 2p(a|Y = y) \geq 0$$

The last inequality follows from the fact that $p(a|Y = y)$ is a probability density function. This shows that $a = \mathbb{F}_{1/2}[\theta|Y = y]$ is a local minimum of the expected risk function $p(a)$. The existence is guaranteed if $\mathbb{F}_{1/2}[\theta|Y = y] < \infty$. The uniqueness is guaranteed provided that the median is unique and $p(a|Y = y) > 0$ (i.e., the posterior distribution is continuous at that point).

Annex 1.4. Proof of Proposition 2.7.

The first part follows immediately from the variance of a sample mean:

$$\mathbb{V}(\widehat{\theta_e}|\theta_0, \sigma^2) = \frac{\sigma^2}{n}$$

$$\mathbb{V}(\widehat{\theta_b}|\theta_0, \sigma^2) = \omega^2 \frac{\sigma^2}{n}$$

Then, for any $\omega \in (0,1)$ the following holds:

$$\omega^2 \frac{\sigma^2}{n} < \frac{\sigma^2}{n}$$

For the second part, let $\mu = \mathbb{E}[\hat{\theta}|\theta_0]$ and $\text{Bias}^2[\hat{\theta}|\theta_0] = \mathbb{E}[(\mu - \theta_0)^2|\theta_0]$, then:

$$\text{MSE}[\hat{\theta}|\theta_0] = \mathbb{E}\left[(\hat{\theta} - \theta_0)^2|\theta_0\right] = \mathbb{E}\left[(\hat{\theta} - \mu + \mu - \theta_0)^2|\theta_0\right]$$
$$= \mathbb{E}\left[(\hat{\theta} - \mu)^2|\theta_0\right] + \mathbb{E}[(\mu - \theta_0)^2|\theta_0] + 2\mathbb{E}[(\hat{\theta} - \mu)|\theta_0]\mathbb{E}[(\mu - \theta_0)|\theta_0]$$
$$= \mathbb{V}(\hat{\theta}|\theta_0) + \text{Bias}^2[\hat{\theta}|\theta_0]$$

For an unbiased estimator $\widehat{\theta_e}$, $\text{Bias}^2[\hat{\theta}|\theta_0] = 0$, so the formula simplifies:

$$\mathrm{MSE}\left[\widehat{\theta_e}|\theta_0\right] = \mathbb{V}\left(\widehat{\theta_e}|\theta_0\right) = \frac{\sigma^2}{n}$$

Nevertheless, a Bayesian point estimator generally has $\mathrm{Bias}^2\left[\hat{\theta}|\theta_0\right] \neq 0$:

$$\mathrm{MSE}\left[\widehat{\theta_b}|\theta_0\right] = \omega^2\frac{\sigma^2}{n} + (1-\omega)^2(\mu-\theta_0)^2$$

Therefore, $\mathrm{MSE}\left[\widehat{\theta_b}|\theta_0\right] < \mathrm{MSE}\left[\widehat{\theta_e}|\theta_0\right]$ holds if the following condition is satisfied:

$$(\mu-\theta_0)^2 < \frac{\sigma^2}{n}\left(\frac{1+\omega}{1-\omega}\right) = \sigma^2\left(\frac{1}{n}+\frac{2}{k_0}\right)$$

Annex 1.5. Metropolis algorithm.

The Metropolis algorithm is the Metropolis-Hastings algorithm where the proposal distributions $J_u\left(u\middle|u^{(s)},v^{(s)}\right)$ and $J_v\left(v\middle|u^{(s+1)},v^{(s)}\right)$ are symmetric, i.e. $J_u\left(u^{(s)}\middle|u^*,v^{(s)}\right) = J_u\left(u^*\middle|u^{(s)},v^{(s)}\right)$ and $J_v\left(v^{(s)}\middle|u^{(s)},v^*\right) = J_v\left(v^*\middle|u^{(s)},v^{(s)}\right)$. This implies that the acceptance ratio simplifies as follows:

$$r_u = \frac{p_0\left(u^*,v^{(s)}\right)}{p_0\left(u^{(s)},v^{(s)}\right)}$$
$$r_v = \frac{p_0\left(u^{(s)},v^*\right)}{p_0\left(u^{(s)},v^{(s)}\right)}$$

Annex 1.6. Gibbs algorithm.

The Gibbs algorithm is the Metropolis-Hastings algorithm where the proposal distributions are the fully conditional target distributions:

$$J_u\left(u\middle|u^*,v^{(s)}\right) = p_0\left(u^*\middle|v^{(s)}\right)$$
$$J_u\left(v\middle|u^{(s)},v^*\right) = p_0\left(v^*\middle|u^{(s)}\right)$$

This election of the proposal distributions implies that they will be accepted with probability one:

$$r_u = \frac{p_0\left(u^*,v^{(s)}\right)}{p_0\left(u^{(s)},v^{(s)}\right)}\frac{p_0\left(u^{(s)}\middle|v^{(s)}\right)}{p_0\left(u^*\middle|v^{(s)}\right)} = \frac{p_0\left(u^*\middle|v^{(s)}\right)p_0\left(v^{(s)}\right)}{p_0\left(u^{(s)}\middle|v^{(s)}\right)p_0\left(v^{(s)}\right)}\frac{p_0\left(u^{(s)}\middle|v^{(s)}\right)}{p_0\left(u^*\middle|v^{(s)}\right)} = 1$$
$$r_v = \frac{p_0\left(u^{(s)},v^*\right)}{p_0\left(u^{(s)},v^{(s)}\right)} = \frac{p_0\left(v^*\middle|u^{(s)}\right)p_0\left(u^{(s)}\right)}{p_0\left(v^{(s)}\middle|u^{(s)}\right)p_0\left(u^{(s)}\right)}\frac{p_0\left(v^{(s)}\middle|u^{(s)}\right)}{p_0\left(v^*\middle|u^{(s)}\right)} = 1$$

Annex 1.7. VaR backtesting procedures (unconditional and conditional coverage tests).

Let Y be a vector of length T such that each observation is indexed by time: $Y = \{y_1, y_2, \dots, y_T\}$. Let $\widehat{VaR_{\alpha,t}}$ be an estimate of the Value-at-Risk corresponding to an α-probability (confidence or belief) level at time t. Define the following auxiliary variables:

$$X_t = \begin{cases} 1, if\ y_t < \widehat{VaR_{\alpha,t}} \\ 0, otherwise \end{cases}; N_T = \sum_{t=1}^{T} X_t; \hat{\pi} = \frac{N_T}{T}; \pi_0 = \alpha$$

With the following interpretation: X_t is an indicator variable which takes value of 1 if there has been observed a breach of an estimated α-VaR at time t, and 0 otherwise; N is a number of observed breaches of an estimated α-VaR within the observation window; $\hat{\pi}$ is the observed proportion (frequency) of breaches and π_0 is a nominal proportion of breaches.

A. The Koupiec test (Koupiec, 1995) is also called an unconditional coverage test which evaluates whether the observed frequency of VaR breaches is consistent with the nominal α-VaR. The test hypotheses are the following:

$$H_0: \pi_0 = \hat{\pi}; H_1: \pi_0 \neq \hat{\pi}$$

The test is based on likelihood ratios:

$$\ln L(\hat{\pi}) = N_T \cdot ln(\hat{\pi}) + (T - N_T) \cdot ln(1 - \hat{\pi})$$

$$\ln L(\pi_0) = N_T \cdot ln(\pi_0) + (T - N_T) \cdot ln(1 - \pi_0)$$

$$LR_{UC} = -2 \cdot (\ln L(\pi_0) - \ln L(\hat{\pi}))$$

The asymptotic distribution of the test statistics is as follows: $LR_{UC} \sim \chi_1^2$

B. The Christoffersen test (Christoffersen, 1998) consists of two steps, an independence test, which evaluates the independence of the VaR breaches over time, and a conditional coverage test, which evaluates the consistency of α-VaR estimation and temporal independence of the breaches. It is based on a first-order Markov chain, so it requires additional variables to be defined:

$$n_{ij} = \sum_{t=2}^{T} \mathbf{1}(X_{t-1} = i, X_t = j);\ i, j \in \{0,1\}$$

Where $\mathbf{1}$ is an indicator function. There are four additional variables that are created in this way: n_{00} (a no-breach followed by a no-breach), n_{01} (a no-breach followed by a breach), n_{10} (a breach followed by a no-breach), n_{11} (a breach followed by another breach). The associated proportion variables must also be defined:

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}, \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}$$

The hypotheses of the independence test are the following:

$$H_0: \hat{\pi}_{01} = \hat{\pi}_{11} = \hat{\pi} = \frac{N}{T}; H_1: \hat{\pi}_{01} \neq \hat{\pi}_{11}$$

The test is based on likelihood ratios:

$$\ln L_1 = n_{00} \cdot ln(1 - \widehat{\pi_{01}}) + n_{01} \cdot ln(\hat{\pi}_{01}) + n_{10} \cdot ln(1 - \hat{\pi}_{11}) + \mathrm{n}_{11} \cdot ln(\hat{\pi}_{11})$$

$$\ln L_0 = n_{00} \cdot ln(1 - \hat{\pi}) + n_{01} \cdot ln(\hat{\pi}) + n_{10} \cdot ln(1 - \hat{\pi}) + \mathrm{n}_{11} \cdot ln(\hat{\pi})$$

$$LR_{IND} = -2(\ln L_0 - \ln L_1)$$

The asymptotic distribution of the test statistics is as follows: $LR_{IND} \sim \chi_1^2$

The conditional coverage test combines both unconditional coverage Koupiec's test and the Christoffersen's independence test with the following test statistic:

$$LR_{CC} = LR_{UC} + LR_{IND}$$

The asymptotic distribution of the test statistics is as follows: $LR_{CC} \sim \chi_2^2$.

Annex 1.8. ES Backtesting procedures (unconditional and conditional coverage tests).

Let Y, $\widehat{VaR_{\alpha,t}}$, X_t and N_T be defined as in Annex 1.7. Assume that Y_t, elements of the returns vector Y, follow an unknown distribution F_t and are forecasted by a model predictive distribution P_t. Let $\widehat{ES}_{a,t}$ be an estimate of the Expected Exposure corresponding to the same α-probability level at time t: $\widehat{ES}_{a,t} = -\frac{1}{\alpha}\int_0^{\alpha} P_t^{-1}(q)dq$, where P_t is a predictive probability distribution function of Y_t.

Acerbi and Szekely (2014) proposed a novel methodology of Expected Shortfall backtesting consisting in three statistical tests. In this thesis two of them are considered.

A. The ES test conditional on VaR consists in calculating the following test statistics:

$$Z_1(Y) = \frac{1}{N_T}\sum_{t=1}^{T}\frac{Y_t \cdot X_t}{\widehat{ES}_{a,t}} + 1$$

The test hypotheses are the following:

$$H_0: P_t^{[\alpha]} = F_t^{[\alpha]}, \forall t$$

$$H_1: \begin{cases} ES_{a,t} \geq \widehat{ES}_{a,t}\ \forall t, \text{with } > \text{ for some } t; \\ VaR_{\alpha,t} = \widehat{VaR_{\alpha,t}}, \forall t \end{cases}$$

Where $P_t^{[\alpha]}(x) = \min\left(1, \frac{P_t(x)}{\alpha}\right)$ is the distribution tail for a VaR breach $x < -\widehat{VaR_{\alpha,t}}$. Under these conditions, $\mathbb{E}_{H_0}[Z_1|N_T > 0] = 0$ and $\mathbb{E}_{H_1}[Z_1|N_T > 0] < 0$. It should be noted that these results assume independence of Y_t, but it doesn't have to be identically distributed.

B. ES direct test consists in calculating the following test statistics:

$$Z_2(Y) = \frac{1}{T \cdot \alpha}\sum_{t=1}^{T}\frac{Y_t \cdot X_t}{\widehat{ES}_{a,t}} + 1$$

The test hypotheses are the following:

$$H_0: P_t^{[\alpha]} = F_t^{[\alpha]}, \forall t$$

$$H_1: \begin{cases} ES_{a,t} \geq \widehat{ES}_{a,t}\ \forall t, \text{with } > \text{ for some } t; \\ VaR_{\alpha,t} \geq \widehat{VaR_{\alpha,t}}, \forall t \end{cases}$$

Under these conditions, $\mathbb{E}_{H_0}[Z_1|N_T > 0] = 0$ and $\mathbb{E}_{H_1}[Z_1|N_T > 0] < 0$. These results do not require an assumption of either independence or identical distribution of Y_t.

The distribution of the test statistics is unknown. The authors of the tests propose to simulate such a distribution following the next scheme:

1. Simulate independent $Y_t^i \sim P_t, \forall t, \forall i = \{1,2,\dots,M\}$, where M is a large number of simulations.
2. Compute the test statistics $Z^i = Z\left(Y^i\right)$.
3. Estimate the p-value $p = \sum_{i=1}^{M}\left(Z_i < Z\left(\tilde{Y}\right)\right)/M$, where $\tilde{Y}$ is an observed Y.

Annex 1.9. Diebold-Mariano test.

The Diebold-Mariano (DM) test is due to Diebold and Mariano (1995). A more recent publication of Diebold (2012) describes the DM test in the following way:

$$DM_{12} = \frac{\overline{d_{12}}}{\widehat{\sigma_{\overline{d_{12}}}}} \xrightarrow{d} \mathcal{N}(0,1)$$

Where $d_{12,t} = L(e_{1,t}) - L(e_{2,t})$; $\overline{d_{12}} = \frac{1}{T}\sum_{t=1}^{T} d_{12,t}$; and $L(e_{i,t})$ is a loss function. In case of a squared loss function $L(e_{i,t}) = e_{i,t}^2$, and in case of an absolute value loss function $L(e_{i,t}) = |e_{i,t}|$.

The assumptions of Diebold-Mariano test are the following:

$$\begin{cases} \mathbb{E}(d_{12,t}) = \mu, \forall t \\ \mathbb{C}(d_{12,t}, d_{12,(t-\tau)}) = \gamma(\tau), \forall t \\ 0 < \mathbb{V}(d_{12,t}) = \sigma^2 < \infty \end{cases}$$

There are several well-known statistical procedures to test these hypotheses, but since it goes beyond the reach of the thesis, they are assumed to hold.

Annex 1.10. Univariate normal model: posterior location parameter distribution conditional on scale parameter.

$$p(\mu|Y = y, \sigma^2) \propto p(Y = y \mid \mu, \sigma^2) p(\mu|\sigma^2)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{1}{2\tau_0^2}(\mu - \mu_0)^2\right) \propto \exp\{\}$$

$$\propto \exp\left\{\sum_{i=1}^{N}\left(-\frac{1}{2\sigma^2}(y_i^2 - 2y_i\mu + \mu^2) - \frac{1}{2\tau_0^2}(\mu^2 - 2\mu_0\mu + \mu_0^2)\right)\right\}$$

$$\propto \exp\left(-\mu^2\frac{N}{2\sigma^2} + \mu\sum_{i=1}^{N}\frac{y_i}{\sigma^2} - \mu^2\frac{1}{2\tau_0^2} + \mu\frac{\mu_0}{\tau_0^2}\right)$$

$$\propto \exp\left(-\frac{\mu^2}{2}\left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right) + \mu\left(\frac{\mu_0}{\tau_0^2} + \frac{N\bar{y}}{\sigma^2}\right)\right) = \exp\left(-\frac{1}{2a}\left(\mu^2 - 2\mu\frac{b}{a}\right)\right)$$

$$\propto \exp\left(-\frac{1}{2a}\left(\mu^2 - 2\mu\frac{b}{a} + \frac{b^2}{a^2}\right)\right) = \exp\left(-\frac{1}{2a}\left(\mu - \frac{b}{a}\right)^2\right)$$

Where $a = \frac{1}{\tau_0^2} + \frac{N}{\sigma^2}$ and $b = \frac{\mu_0}{\tau_0^2} + \frac{N\bar{y}}{\sigma^2}$. Therefore, the posterior distribution of a conditional model location parameter $f(\mu|y, \sigma^2)$ is Normal with the location parameter $\mu_{\mathrm{n}} = \frac{b}{a} = \left(\mu_0\frac{1}{\tau_0^2} + \bar{y}\frac{N}{\sigma^2}\right)\left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right)^{-1}$ and the scale parameter $\tau_1^2 = a^{-1} = \left(\frac{1}{\tau_0^2} + \frac{N}{\sigma^2}\right)^{-1}$.

Annex 1.11. Univariate normal model: posterior scale parameter distribution.

$$p(\sigma^2|Y = y) \propto p(Y = y\,|\sigma^2)p(\sigma^2) = p(\sigma^2)\int_{-\infty}^{\infty} p(Y = y\,|\mu, \sigma^2)p(\mu|\sigma^2)\,d\mu$$

$$p(Y=y\,|\sigma^2)=\int_{-\infty}^{\infty}p(Y=y\,|\mu,\sigma^2)p(\mu|\sigma^2)\,d\mu$$
$$=\int_{-\infty}^{\infty}\prod_{i=1}^{N}\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y_i-\mu)^2\right)\right\}\frac{1}{\sqrt{2\pi\tau_0^2}}\exp\left(-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2\right)d\mu$$
$$=\left(\frac{1}{2\pi}\right)^{\frac{N+1}{2}}\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\frac{1}{\tau_0}\int_{-\infty}^{\infty}\exp\left(\sum_{i=1}^{N}\left\{-\frac{1}{2\sigma^2}(y_i-\mu)^2\right\}-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2\right)d\mu$$
$$=\left(\frac{1}{2\pi}\right)^{\frac{N+1}{2}}\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\frac{1}{\tau_0}\sqrt{2\pi a}\exp\left(-\frac{1}{2}\left(\frac{\sum_{i=1}^{N}y_i^2}{\sigma^2}+\frac{\mu_0^2}{\tau_0^2}+\frac{b^2}{a}\right)\right)$$
$$=\left(\frac{1}{2\pi}\right)^{\frac{N}{2}}\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\frac{1}{\tau_0}\sqrt{\frac{1}{\tau_0^2}+\frac{N}{\sigma^2}}\exp\left(-\frac{1}{2}\left(\frac{\sum_{i=1}^{N}y_i^2}{\sigma^2}+\frac{\mu_0^2}{\tau_0^2}+\frac{b^2}{a}\right)\right)$$

Which can be simplified as follows if $\tau_0^2=\frac{\sigma^2}{k_0}$ is assumed:

$$p(Y=y\,|\sigma^2)\propto\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}y_i^2+\mu_0^2k_0+\frac{(k_0\mu_0+N\bar{y})^2}{k_0+N}\right)\right)=\cdots$$
$$=\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}(y_i-\bar{y})^2+\frac{Nk_0}{N+k_0}(\bar{y}-\mu_0)^2\right)\right)$$

The conjugate family of distributions for the scale parameter is an inverse-gamma distribution:

$$p(\sigma^2|a,b)=\frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1}\exp\left(-\frac{b}{\sigma^2}\right);\ a,b,\sigma^2>0.$$

Therefore, the scale parameter posterior distribution in given by the following expression:

$$p(\sigma^2|Y=y)\propto p(Y=y\,|\sigma^2)p(\,\sigma^2)$$
$$=(\sigma^2)^{-\frac{N}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}(y_i-\bar{y})^2\right.\right.$$
$$\left.\left.+\frac{Nk_0}{N+k_0}(\bar{y}-\mu_0)^2\right)\right)\frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1}\exp\left(-\frac{b}{\sigma^2}\right)$$
$$\propto(\sigma^2)^{-\frac{N}{2}-a-1}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}(y_i-\bar{y})^2+\frac{Nk_0}{N+k_0}(\bar{y}-\mu_0)^2+2b\right)\right)$$

It is clear that $f(\sigma^2|Y=y)$ has the same shape as $f(\sigma^2|a,b)$, that is, $f(\sigma^2|Y=y)$ is also an inverse gamma distribution. The result will be reparametrized: given a prior distribution $f(\sigma^2)\sim InverseGamma\left(\frac{v_0}{2},\frac{v_0}{2}\sigma_0^2\right)$, the posterior distribution is $f(\sigma^2|Y=y)\sim InverseGamma\left(\frac{N+v_0}{2},\frac{1}{N+v_0}\left[v_0\sigma_0^2+\sum_{i=1}^{N}(y_i-\bar{y})^2+\frac{Nk_0}{N+k_0}(\bar{y}-\mu_0)^2\right]\right)$.

Annex 1.12. Numerical approximation scheme for the normal model.

Specify the prior hyperparameters: $\mu_0,\tau_0^2,v_0,\sigma_0^2$. Provide an initial value for μ and σ^2. Then, apply the following Gibbs sampler loop (for a sufficiently large number of iterations N):

For s = 1, 2…, N:

1. Sample $\mu^{(s+1)}$ to approximate $p\left(\mu\middle|Y=y,{\sigma^2}^{(s)}\right)$ from the following distribution: $Normal\left(\mu=\left(\mu_0\frac{1}{\tau_0^2}+\bar{y}\frac{N}{\sigma^2}\right)\left(\frac{1}{\tau_0^2}+\frac{N}{\sigma^2}\right)^{-1},\sigma^2=\left(\frac{1}{\tau_0^2}+\frac{N}{{\sigma^2}^{(s)}}\right)^{-1}\right)$.
2. Sample ${\sigma^2}^{(s+1)}$ to approximate $p(\sigma^2|Y=y)$ from the following distribution: $InverseGamma\left(\alpha=\frac{N+v_0}{2},\beta=\frac{1}{N+v_0}\left[v_0\sigma_0^2+\sum_{i=1}^{N}(y_i-\bar{y})^2+\frac{Nk_0}{N+k_0}(\bar{y}-\mu_0)^2\right]\right)$.

End the loop.

Annex 1.13. Numerical approximation scheme for a multivariate normal model.

Specify the prior hyperparameters: $\boldsymbol{\mu_0},\boldsymbol{\Lambda}_0,v_0,\mathbf{S}_0$. Provide an initial value for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. Then, apply the following Gibbs sampler loop (for a sufficiently large number of iterations N):
For s = 1, 2…, N:

1. Sample $\boldsymbol{\theta}^{(s+1)}$ to approximate $p\left(\boldsymbol{\theta}|\mathbf{Y}=\boldsymbol{y},{\boldsymbol{\Sigma}^{(s)}}^{-1}\right)$ from the following distribution: $MultivariateNormal\left(\left(\boldsymbol{\Lambda}_0^{-1}+n{\boldsymbol{\Sigma}^{(s)}}^{-1}\right)^{-1}\left(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu_0}+n{\boldsymbol{\Sigma}^{(s)}}^{-1}\bar{\boldsymbol{Y}}\right),\left(\boldsymbol{\Lambda}_0^{-1}+n{\boldsymbol{\Sigma}^{(s)}}^{-1}\right)^{-1}\right)$
2. Sample $\boldsymbol{\Sigma}^{(s+1)}$ to approximate $p(\boldsymbol{\Sigma}|\boldsymbol{Y},\boldsymbol{\theta})$ from the following distribution: $InverseWishart(v_0+n,(\boldsymbol{S}_0+\sum_{i=1}^{n}(\boldsymbol{Y}_i-\boldsymbol{\theta})(\boldsymbol{Y}_i-\boldsymbol{\theta})^T)^{-1})$.

End the loop.

Annex 1.14. Application of multidimensional normal model to a treatment of missing values.

The multidimensional normal model provides an easy and highly valuable way of treating missing values. In this case, the model is expanded in the following way:

Let $\boldsymbol{Y}$ be a n × m matrix of all the data, observed and unobserved. Let $\boldsymbol{O}$ be a n × m matrix in which $o_{i,j}=1$ if $Y_{i,j}$ is observed and $o_{i,j}=0$ if $Y_{i,j}$ is missing. Then $\boldsymbol{Y}$ can be partitioned into two matrixes: $\mathrm{Y_{obs}}=\{y_{i,j}:o_{i,j}=1\}$ (the observed data), $\mathrm{Y_{obs}}=\{y_{i,j}:o_{i,j}=0\}$ (the unobserved data). In this case, assuming conditional independence of observations, the conditional model for the missing data can be defined as follows:

$$p(\boldsymbol{Y_{miss}}|\boldsymbol{Y_{obs}},\boldsymbol{\theta},\Sigma)\propto p(\boldsymbol{Y_{miss}},\boldsymbol{Y_{obs}}|\boldsymbol{\theta},\Sigma)=\prod_{i=1}^{n}p(y_{i,miss},y_{i,obs}\,|\boldsymbol{\theta},\Sigma)$$

$$\propto\prod_{i=1}^{n}p(y_{i,miss}\,|y_{i,obs},\boldsymbol{\theta},\Sigma)$$

This model specification renders the following result:

$$\boldsymbol{Y_{miss}}|\boldsymbol{Y_{obs}},\boldsymbol{\theta},\boldsymbol{\Sigma}\sim\mathrm{MultivariateNormal}\left(\boldsymbol{\theta_{miss|obs}},\boldsymbol{\Sigma_{miss|obs}}\right)$$

Where $\boldsymbol{\theta_{miss|obs}}=\boldsymbol{\theta_{[miss]}}+\boldsymbol{\Sigma_{[miss,obs]}}\left(\boldsymbol{\Sigma_{[obs,obs]}}\right)^{-1}\left(\boldsymbol{y_{[obs]}}-\boldsymbol{\theta_{[obs]}}\right)$ and $\boldsymbol{\Sigma_{miss|obs}}=\boldsymbol{\Sigma_{[miss,miss]}}-\boldsymbol{\Sigma_{[miss,obs]}}\left(\boldsymbol{\Sigma_{[obs,obs]}}\right)^{-1}\boldsymbol{\Sigma_{[obs,miss]}}$. Subindexes [miss] and [obs] correspond to the elements of vectors y and θ, while their combination indicates the rows (the first position) and the columns (the second position) of the matrix $\boldsymbol{\Sigma}$.

The posterior distributions are then easily approximated by the Gibbs sampler. First, specify the prior hyperparameters: $\boldsymbol{\mu_0},\boldsymbol{\Lambda}_0,v_0,\mathbf{S}_0$. Provide initial values for $\boldsymbol{\theta},\boldsymbol{\Sigma}$ and $\boldsymbol{Y_{miss}}$. Then, apply the following Gibbs sampler loop (for a sufficiently large number of iterations N):

For s = 1, 2…, N:

1. Sample $\boldsymbol{\theta}^{(s+1)}$ from $p\left(\boldsymbol{\theta}\middle|\boldsymbol{Y_{obs}}, \boldsymbol{Y_{miss}^{(s)}}, \boldsymbol{\Sigma}^{(s)}\right)$.
2. Sample $\boldsymbol{\Sigma}^{(s+1)}$ from $p\left(\boldsymbol{\Sigma}\middle|\boldsymbol{Y_{obs}}, \boldsymbol{Y_{miss}^{(s)}}, \boldsymbol{\theta}^{(s+1)}\right)$.
3. Sample $\boldsymbol{Y_{miss}^{(s+1)}}$ from $p\left(\boldsymbol{Y_{miss}}\middle|\boldsymbol{Y_{obs}}, \boldsymbol{\Sigma}^{(s+1)}, \boldsymbol{\theta}^{(s+1)}\right)$.

End the loop.

The first two fully conditional distributions have already been presented earlier, and the third fully conditional distribution is the one presented above. A detailed treatment can be found in Hoff (2009).

Annex 1.14. Numerical approximation scheme for a mixture of normals model.

Specify the following two-component vectors of prior hyperparameters: $\boldsymbol{\mu_0}, \boldsymbol{\tau_0^2}, \boldsymbol{\nu_0}, \boldsymbol{\sigma_0^2}$ and the following scalar prior hyperparameters α_0, β_0. Provide an initial value for $\boldsymbol{\mu}$, $\boldsymbol{\sigma^2}$ and $\boldsymbol{Z}$. Then, apply the following Gibbs sampler loop (for a sufficiently large number of iterations N):
For s = 1, 2…, N:

1. Sample $\boldsymbol{Z}^{(s+1)}$ from a Binomial($p^{(s)}$) distribution, where $p = \frac{p_1}{p_1+p_2}$ and $p_1 = p^{(s)} \cdot \prod_{i=1}^{n} f\left(y_i \middle| \mu_1^{(s)}, \sigma_1^{2\,(s)}, \boldsymbol{Z}^{(s)}\right)$ and $p_2 = \left(1-p^{(s)}\right) \cdot \prod_{i=1}^{n} f\left(y_i \middle| \mu_2^{(s)}, \sigma_2^{2\,(s)}, \boldsymbol{Z}^{(s)}\right)$.
2. Sample $p^{(s+1)}$ to approximate $f(p \mid Y = y)$ from the following distribution: $Beta(\alpha_0 + n_1^{(s+1)}, \beta_0 + n_2^{(s+1)})$, where $n_1^{(s+1)} = \#\{\boldsymbol{Z}^{(s+1)} = 1\}, n_2^{(s+1)} = \#\{\boldsymbol{Z}^{(s+1)} = 0\}$
3. Sample $\boldsymbol{\mu}^{(s+1)}$ to approximate $p\left(\boldsymbol{\mu}\middle| Y = y, \boldsymbol{\sigma}^{2\,(s)}, \boldsymbol{Z}^{(s+1)}\right)$ as explained in Annex 1.12, applying it to each vector component separately.
4. Sample $\boldsymbol{\sigma}^{2\,(s+1)}$ to approximate $p\left(\boldsymbol{\sigma}^2\middle| Y = y, \boldsymbol{Z}^{(s+1)}\right)$ as explained in Annex 1.12, applying it to each vector component separately.

End the loop.

Annex 1.15. Numerical approximation scheme for the factor-driven t-GARCH(1,1) model.

Specify prior distributions and initial values for the following parameters: $\boldsymbol{Z}, \boldsymbol{\beta}, v, \gamma, \alpha, \theta$. Initial values can be arbitrary, but they are usually set in a convenient way to facilitate convergence, e.g. maximum-likelihood estimates. Assume there are N observed values and p regressors. The prior distributions may be arbitrary probability distributions which respect the support of the variables: $\boldsymbol{Z} \in \{0,1\}^p, \boldsymbol{\beta} \in \mathbb{R}^p; \gamma > 0, \alpha \geq 0, \theta \geq 0; v > 0$. Furthermore, the condition $\alpha + \theta < 1$ must be satisfied to ensure the wide sense stationarity. The factor-driven t-GARCH(1,1) model has the following form:

$$\begin{cases} y_t = \boldsymbol{Z\beta} X_t + \epsilon_t \\ \epsilon_t \sim Student(0, \sigma_t^2, v) \\ \sigma_t^2 = \gamma + \alpha\epsilon_{t-1}^2 + \theta\sigma_{t-1}^2 \end{cases}$$

Therefore, assuming conditional independence, the conditional probability model for the observations takes the following form:

$$p(Y = y \mid \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, v) = \prod_{t=1}^{N} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)\sigma_t^2}} \left(1 + \frac{(y_i - (\boldsymbol{Z\beta})_t)^2}{(v-2)\sigma_t^2}\right)^{-\frac{v+1}{2}}$$

The model can naturally be divided into three parts: a variable selection procedure, a conditional mean factor-driven regression model and a conditional variance model. Each step can be performed following the Metropolis-Hastings, Metropolis or Gibbs algorithm conditioned on the most recent value of all other variables (for a sufficiently large number of iterations N):

For $s \in \{1, 2, \dots, S\}$:

1. Update the vector of variable selection variable $\boldsymbol{Z} = (z_1, z_2, \dots, z_p)$, either individually for each component of the vector (each regressor) or the whole vector at once. Since z_i takes only two values, 0 and 1, the proposal distribution must account for that.
2. Update the vector of regressor coefficients $\boldsymbol{\beta}$ conditioned on just updated variable $\boldsymbol{Z}$.
3. Update the scalar GARCH model coefficients: $\gamma, \alpha, \theta, v$. Since γ and v are positive and the other variables are nonnegative, the prior and proposal distributions must account for that. In order to ensure a wide-sense stationarity, the prior and proposal distributions should also account for the stationarity condition: $\alpha + \theta < 1$. One of the ways of taking this condition into account is to discard iterations in which this condition is not satisfied considering them as those having zero probability. Another way of treating iterations in which $\alpha + \theta \geq 1$ is to rescale the parameters: $\alpha^{(s+1)} = \frac{\alpha^{(s+1)}}{\alpha^{(s+1)} + \theta^{(s+1)} + \delta}$; $\theta^{(s+1)} = \frac{\theta^{(s+1)}}{\alpha^{(s+1)} + \theta^{(s+1)} + \delta}$, adding reasonably a small value $\delta > 0$ to the denominator to ensure the stationarity.

Annex 1.16. Numerical approximation scheme for the three-level hierarchical model.

Assume that there are M markets and N_m stocks in a market $m \in \{1, \dots, M\}$, so there are $N = \sum_{m=1}^{M} N_m$ stocks in total. The model is represented by the following structure:

1. A within-stock conditional model: $\phi_i = \{\mu_i, \sigma_i^2\}: p(y|\phi_i) = Normal(\mu_i, \sigma_i^2)$.
2. A within-market conditional model: $\psi_{\mathrm{m}} = \{\delta_m, \tau_{\mathrm{m}}^2\}: p(\mu|\psi_{\mathrm{m}}) = Normal(\delta_m, \tau_{\mathrm{m}}^2)$.
3. A global conditional model: $\omega = \{\theta, \gamma^2\}: p(\delta|\omega) = Normal(\theta, \gamma^2)$.

Apply the following Gibbs sampler loop (for a sufficiently large number of iterations N) to approximate the posterior distributions (initial values can be arbitrary):

For s = 1, 2…, N:

1. **Unit level**.

A conditionally normal model for individual stocks' returns is assumed:

$$f\left(y_{i_m,t_{i_m}} \mid \mu_{i_m}, \sigma_{i_m}^2\right) = \frac{1}{\sqrt{2\pi \cdot \sigma_{i_m}^2}} \cdot \exp\left(-\frac{1}{2\sigma_{i_m}^2} \cdot \left(y_{i_m,t_{i_m}} - \mu_{i_m}\right)^2\right)$$

Where $i_m \in \{1, \dots, N_m\}$ indicates the i-th series of stock's returns from a market m and t is a time index of observations $t_{i_m} \in \{1, \dots, T_{i_m}\}$ of that series. In particular, $y_{i_m,t_{i_m}}$ is a return of the i-th stock from the m-th market at time t_{i_m}.

Since each stock's return distribution has a location and a scale parameter, there are $2N$ parameters at this level. The μ_{i_m} parameter will be treated at the next step. The $\sigma_{i_m}^2$ parameter is assumed to have a prior inverse-gamma distribution discussed earlier in the one-dimensional

normal model section: $f\left(\sigma_{i_m}^2\right) \sim InverseGamma\left(\frac{v_{0,\mathrm{i_m}}}{2}, \frac{v_{0,\mathrm{i_m}}}{2} \cdot \sigma_{0,\mathrm{i_m}}^2\right)$, yielding a conjugate inverse-gamma posterior distribution:

$$f\left(\sigma_{i_m}^2 \middle| Y_{i_m} = y_{i_m}\right) \sim InverseGamma\left(\frac{T_{i_m} + v_{0,\mathrm{i_m}}}{2}, \frac{1}{T_{i_m} + v_{0,\mathrm{i_m}}} \cdot \left[v_{0,\mathrm{i_m}} \cdot \sigma_{0,\mathrm{i_m}}^2 + \sum_{i_m=1}^{T_{i_m}} \left(y_{i_m,t_{i_m}} - \overline{y_{i_m}}\right)^2 + \frac{T_{i_m} \cdot k_{0,\mathrm{i_m}}}{T_{i_m} + k_{0,\mathrm{i_m}}} \cdot \left(\overline{y_{i_m}} - \mu_{0,\mathrm{i_m}}\right)^2\right]\right)$$

where $\overline{y_{i_m}} = \sum_{i_m=1}^{T_{i_m}} \frac{y_{i_m,t_{i_m}}}{T_{i_m}}$. The prior hyperparameters that need to be specified for each stock are $\mu_{0,\mathrm{i_m}}, k_{0,\mathrm{i_m}}, v_{0,\mathrm{i_m}}, \sigma_{0,\mathrm{i_m}}^2$ with the usual interpretation previously given, $4N$ in total.

2. **Inter-market level**.

Individual stock returns series comprise a market. There may be one or several different market. Just as an individual stock's returns series, each location parameter is treated as a draw from the following conditionally normal model:

$$f\left(\mu_{i_m} \middle| \delta_m, \tau_{\mathrm{m}}^2\right) = \frac{1}{\sqrt{2\pi\tau_{\mathrm{m}}^2}} \cdot \exp\left(-\frac{1}{2\tau_{\mathrm{m}}^2} \cdot \left(\mu_{i_m} - \delta_m\right)^2\right)$$

For all $i_m \in \{1, \dots, N_m\}$ that pertain to a m-th market, $m \in \{1, \dots, M\}$. Since each stock's location parameter distribution has a location and a scale parameter, there are $2M$ parameters at this level.

Analogously as in the case of the unit level, δ_m will be treated at the next step and the τ_{m}^2 parameters are assumed to have a prior inverse-gamma distribution:
$f(\tau_{\mathrm{m}}^2) \sim InverseGamma\left(\frac{v_{0,\mathrm{m}}}{2}, \frac{v_{0,\mathrm{m}}}{2} \cdot \sigma_{0,\mathrm{m}}^2\right)$, yielding a conjugate inverse-gamma posterior distribution:

$$f(\tau_{\mathrm{m}}^2 | \mathrm{M_i} = \mu_i) \sim InvGamma\left(\frac{N_m + v_{0,\mathrm{m}}}{2}, \frac{1}{N_m + v_{0,\mathrm{m}}} \cdot \left[v_{0,\mathrm{m}} \cdot \sigma_{0,\mathrm{m}}^2 + \sum_{i=1}^{N_m} \left(\mu_{i_m} - \overline{\mu_m}\right)^2 + \frac{N_m \cdot k_{0,\mathrm{m}}}{N_m + k_{0,\mathrm{m}}} \cdot \left(\overline{\mu_m} - \mu_{0,\mathrm{m}}\right)^2\right]\right)$$

where $\overline{\mu_m} = \sum_{i=1}^{N_m} \frac{\mu_{i_m}}{N_m}$. The prior hyperparameters that need to be specified for each market are $\mu_{0,\mathrm{m}}, k_{0,\mathrm{m}}, v_{0,\mathrm{m}}, \sigma_{0,\mathrm{m}}^2$ with the usual interpretation previously given, $4M$ in total.

3. **Intra-market level**.

Markets make up a global level. A market location parameter is treated as a draw from the following conditionally normal model:

$$f(\delta_m | \theta, \gamma^2) = \frac{1}{\sqrt{2\pi\gamma^2}} \cdot \exp\left(-\frac{1}{2\gamma^2} \cdot (\delta_m - \theta)^2\right)$$

There are just two parameters at this level. Assuming a normal prior for the global location parameter: $\theta \sim N\left(\mu_{0,\theta}, \tau_{0,\theta}^2\right)$, the posterior distribution of a global location parameter adopts the following form:

$$f(\theta | \Delta_{\mathrm{m}} = \delta_m, \gamma^2) \sim Normal\left(\left(\mu_{0,\theta} \cdot \frac{1}{\tau_{0,\theta}^2} + \bar{\delta} \cdot \frac{M}{\gamma^2}\right)\left(\frac{1}{\tau_{0,\theta}^2} + \frac{M}{\gamma^2}\right)^{-1}, \left(\frac{1}{\tau_{0,\theta}^2} + \frac{M}{\gamma^2}\right)^{-1}\right)$$

Making a usual assumption of an inverse-gamma prior distribution for the scale parameter: $f(\gamma^2) \sim InvGamma\left(\frac{v_{0,\gamma^2}}{2}, \frac{v_{0,\gamma^2}}{2} \cdot \sigma^2_{0,\gamma^2}\right)$, this yields the following posterior distribution:

$$f(\gamma^2|\Delta_{\mathrm{m}} = \delta_m) \sim InvGamma\left(\frac{M + v_{0,\gamma^2}}{2}, \frac{1}{M + v_{0,\gamma^2}}\left[v_{0,\gamma^2} \cdot \sigma^2_{0,\gamma^2} + \sum_{i=1}^{M}\left(\delta_m - \bar{\delta}\right)^2 + \frac{M \cdot k_{0,\gamma^2}}{M + k_{0,\gamma^2}} \cdot \left(\bar{\delta} - \mu_{0,\theta}\right)^2\right]\right)$$

where $\bar{\delta} = \sum_{i=1}^{M} \frac{\delta_m}{M}$. The prior hyperparameters that need to be specified are $\mu_{0,\theta}, \tau^2_{0,\theta}, v_{0,\gamma^2}, \sigma^2_{0,\gamma^2}$ with the usual interpretation previously given, 4 in total.

All-in-all, the model consists of three hierarchical levels, provides a posterior distribution for $2N + 2M + 2$ parameters and requires $4N + 4M + 4$ prior hyperparameters, so there are $6(N + M + 1)$ parameters included in the model.

Annex 1.17. Sample statistics formulas.

Suppose there is a vector y consisting of N observations. Define the following sample statistics:

Sample mean: $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$; sample variance (unbiased): $\widehat{\sigma^2} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$; sample variance (biased): $\widehat{\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$; sample coefficient of skewness: $Skew = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2\right)^{\frac{3}{2}}}$; sample coefficient of kurtosis: $Kurt = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2\right)^2} - 3$.

Annex 1.18. Diebold-Mariano test p-values for a factor-driven t-GARCH(1,1) model.

Method \ Metrics	Bayesian posterior mean	Bayesian posterior median	Pseudo Bayesian posterior mean	Pseudo Bayesian posterior median	Frequentist linear regression	Frequentist ridge regression	Frequentist lasso regression
Bayesian posterior mean	-	0.02 | 0.00	0.39 | 0.78	0.04 | 0.02	0.04 | 0.00	0.09 | 0.36	0.09 | 0.34
Bayesian posterior median	0.02 | 0.00	-	0.03 | 0.00	0.04 | 0.06	0.03 | 0.00	0.80 | 0.42	0.28 | 0.87
Pseudo Bayesian posterior mean	0.39 | 0.78	0.04 | 0.06	-	0.04 | 0.02	0.04 | 0.00	0.10 | 0.31	0.10 | 0.31
Pseudo Bayesian posterior median	0.04 | 0.02	0.04 | 0.06	0.04 | 0.02	-	0.03 | 0.00	0.03 | 0.02	0.02 | 0.02
Frequentist linear regression	0.04 | 0.00	0.03 | 0.00	0.04 | 0.00	0.03 | 0.00	-	0.04 | 0.00	0.05 | 0.00

Table A.1. P-value for the Diebold-Mariano test for the point predictions with respect to the squared mean error | the absolute value error, carried out for a factor-driven t-GARCH(1,1) model. Proper creation.

Annex 1.19. Massive application of the factor-driven t-GARCH(1,1) model.

Tables below show the following results of a massive application of the factor-driven t-GARCH(1,1) model to a large sample of logarithmic returns of 1003 stocks under different conditions: 1. VaR estimate (the percentage of observed breaches is indicated above, the associated unconditional and conditional coverage tests are provided below); 2. ES estimate (the observed ES ratio (real/estimated) is indicated above, the associated unconditional and direct tests are provided below); 3. Kolmogorov-Smirnov (KS) test; 4. MSE, MAE and R2: average values. In the case of tests (VaR, ES, KS), the proportion of stocks for which the null hypothesis at 5% of significance cannot be rejected is shown. All tables are of proper elaboration.

Risk metrics \ quantiles	10% VaR	5% VaR	1% VaR	10% ES	5% ES	1% ES	KS test	MSE	MAE	R2
Bayesian	9,3% (0.77, 0.95)	5,2% (0.80, 0.96)	1,3% (0.76, 0.88)	1.37 (0.96, 0.37)	1.21 (0.98, 0.26)	1.24 (0.99, 0.10)	0.94	0.0403 (62%)	0.0130 (65%)	0.1448 (62%)
Frequentist	12.7% (0.61, 0.96)	7.6% (0.59, 0.94)	2.8% (0.54, 0.79)	1.42 (0.68, 0.00)	1.46 (0.71, 0.00)	1.49 (0.99, 0.03)	0.80	0.0410 (38%)	0.0131 (35%)	0.1341 (38%)

Table A.2. Massive application of the factor-driven t-GARCH(1,1) model. Daily returns, 66-day observation window for a train dataset. Proper creation.

Risk metrics \ quantiles	10% VaR	5% VaR	1% VaR	10% ES	5% ES	1% ES	KS test	MSE	MAE	R2
Bayesian	10,8% (0.79, 0.98)	5,9% (0.79, 0.97)	2,0% (0.74, 0.87)	1.41 (0.97, 0.46)	1.16 (0.99, 0.37)	1.17 (1.00, 0.13)	0.97	0.03535 (55%)	0.01236 (55%)	0.2426 (62%)
Frequentist	12.6% (0.68, 0.99)	7.3% (0.63, 0.98)	2.5% (0.55, 0.81)	1.82 (0.77, 0.00)	1.62 (0.80, 0.00)	1.35 (0.98, 0.02)	0.97	0.03537 (45%)	0.01237 (45%)	0.2420 (38%)

Table A.3. Massive application of the factor-driven t-GARCH(1,1) model. Daily returns, 252-day observation window for a train dataset. Proper creation.

Risk metrics \ quantiles	10% VaR	5% VaR	1% VaR	10% ES	5% ES	1% ES	KS test	MSE	MAE	R2
Bayesian	9,6% (0.59, 0.88)	5,3% (0.60, 0.85)	1,7% (0.64, 0.77)	1.24 (0.93, 0.44)	1.20 (0.94, 0.30)	1.08 (0.95, 0.13)	0.92	0.2207 (69%)	0.03157 (72%)	0.0955 (69%)
Frequentist	14,2% (0.36, 0.92)	8,7% (0.35, 0.88)	3,6% (0.33, 0.65)	2.40 (0.77, 0.01)	2.29 (0.76, 0.00)	2.33 (0.72, 0.03)	0.58	0.2346 (31%)	0.03287 (28%)	0.0213 (31%)

Table A.4. Massive application of the factor-driven t-GARCH(1,1) model. Weekly returns, 26-week observation window for a train dataset. Proper creation.

Risk metrics \ quantiles	10% VaR	5% VaR	1% VaR	10% ES	5% ES	1% ES	KS test	MSE	MAE	R2
Bayesian	10,8%	6,3%	2,4%	1.32	1.29	1.17	0.95	0.1796	0.0288	0.2469

	(0.67, 0.96)	(0.62, 0.92)	(0.46, 0.65)	(0.95, 0.38)	(0.96, 0.24)	(0.97, 0.08)		(71%)	(72%)	(71%)
Frequentist	13,3% (0.50, 0.96)	7,9% (0.49, 0.94)	3,0% (0.42, 0.75)	1.52 (0.40, 0.01)	1.42 (0.52, 0.00)	1.19 (0.63, 0.01)	0.77	0.1850 (29%)	0.0293 (28%)	0.2267 (29%)

Table A.5. Massive application of the factor-driven t-GARCH(1,1) model. Weekly returns, 52-week observation window for a train dataset. Proper creation.

Risk metrics \ quantiles	10% VaR	5% VaR	1% VaR	10% ES	5% ES	1% ES	KS test	MSE	MAE	R2
Bayesian	8,3% (0.74, 0.87)	4,2% (0.77, 0.81)	1,1% (0.95, 0.98)	0.98 (0.86, 0.26)	0.81 (0.92, 0.25)	0.45 (0.96, 0.58)	0.99	0.8648 (80%)	0.0665 (79%)	0.057 (80%)
Frequentist	17,0% (0.51, 0.76)	11,2% (0.44, 0.69)	5,1% (0.53, 0.71)	2.04 (0.65, 0.07)	2.37 (0.68, 0.19)	2.50 (0.64, 0.28)	0.79	1.0123 (20%)	0.0720 (21%)	-0.115 (20%)

Table A.6. Massive application of the factor-driven t-GARCH(1,1) model. Monthly returns, 12-month observation window for a train dataset. Proper creation.

Risk metrics \ quantiles	10% VaR	5% VaR	1% VaR	10% ES	5% ES	1% ES	KS test	MSE	MAE	R2
Bayesian	9,4% (0.81, 0.92)	5,0% (0.83, 0.88)	1,4% (0.92, 0.97)	1.34 (0.88, 0.23)	0.92 (0.94, 0.19)	0.57 (0.97, 0.48)	0.99	0.7732 (75%)	0.0636 (73%)	0.151 (75%)
Frequentist	16,4% (0.58, 0.83)	10,3% (0.50, 0.76)	4,4% (0.59, 0.77)	2.18 (0.55, 0.03)	1.52 (0.62, 0.11)	1.94 (0.63, 0.25)	0.89	0.8365 (25%)	0.0729 (27%)	0.077 (25%)

Table A.7. Massive application of the factor-driven t-GARCH(1,1) model. Monthly returns, 24-month observation window for a train dataset. Proper creation.

Annex 1.20. Massive application of the Bayesian three-level hierarchical normal model.

Level \ Credible interval	90%	95%	99%
Individual	97,21%	99,6%	99,9%
Market	17/19	18/19	18/19
Global	1/1	1/1	1/1

Table A.8. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Daily returns for 22 consecutive days training set. Proper creation.

Level \ Credible interval	90%	95%	99%
Individual	96,01%	98,6%	99,8%
Market	18/19	19/19	19/19
Global	1/1	1/1	1/1

Table A.9. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Daily returns for 66 consecutive days training set. Proper creation.

Level \ Credible interval	90%	95%	99%
Individual	81,26%	90,93%	99,3%
Market	10/19	16/19	16/19

Global	1/1	1/1	1/1

Table A.10. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Weekly returns for 26 consecutive weeks training set. Proper creation.

Level \ Credible interval	90%	95%	99%
Individual	69,89%	84,05%	96,71%
Market	6/19	10/19	14/19
Global	1/1	1/1	1/1

Table A.11. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Weekly returns for 26 consecutive weeks training set. Proper creation.

Level \ Credible interval	90%	95%	99%
Individual	93,02%	98,5%	100%
Market	15/19	16/19	17/19
Global	1/1	1/1	1/1

Table A.12. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Monthly returns for 12 consecutive months training set. Proper creation.

Level \ Credible interval	90%	95%	99%
Individual	58,42%	75,17%	95,31%
Market	7/19	10/19	14/19
Global	1/1	1/1	1/1

Table A.13. Bayesian three-level hierarchical model: the efficient market hypothesis testing. The proportion of integrands of each level for which the hypothesis cannot be rejected is shown. Monthly returns for 24 consecutive months training set. Proper creation.

Annex 1.21. MCMC convergence checks for the factor-driven t-GARCH(1,1) model.

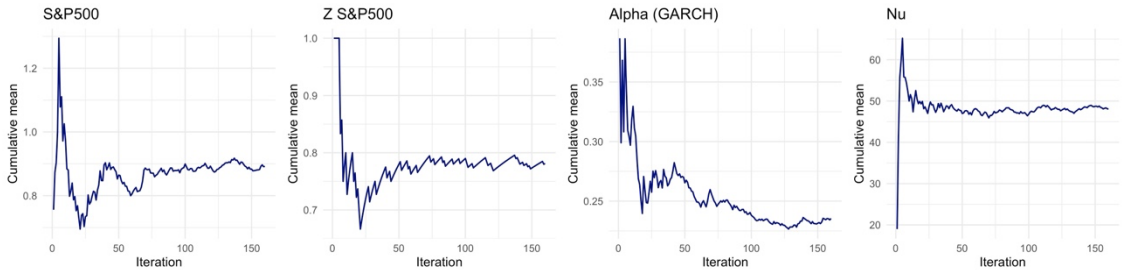

Figure A.1. MCMC convergence checks for the factor-driven t-GARCH(1,1) model. Posterior cumulative mean evolution by iteration. Proper creation.