

Bayesian Inference Methods for Univariate and Multivariate GARCH Models: a Survey

Audronė Virbickaitė

Universidad Carlos III de Madrid

Getafe (Madrid), Spain, 28903

audrone.virbickaite@uc3m.es

M. Concepción Ausín

Universidad Carlos III de Madrid

Getafe (Madrid), Spain, 28903

concepcion.ausin@uc3m.es

Pedro Galeano

Universidad Carlos III de Madrid

Getafe (Madrid), Spain, 28903

pedro.galeano@uc3m.es

Abstract

This survey reviews the existing literature on the most relevant Bayesian inference methods for univariate and multivariate GARCH models. The advantages and drawbacks of each procedure are outlined as well as the advantages of the Bayesian approach versus classical procedures. The paper makes emphasis on recent Bayesian non-parametric approaches for GARCH models that avoid imposing arbitrary parametric distributional assumptions. These novel approaches implicitly assume infinite mixture of Gaussian distributions on the standardized returns which have been shown to be more flexible and describe better the uncertainty about future volatilities. Finally, the survey presents an illustration using real data to show the flexibility and usefulness of the non-parametric approach.

Keywords: Bayesian inference; Dirichlet Process Mixture; Financial returns; GARCH models; Multivariate GARCH models; Volatility.

1 Introduction

Understanding, modeling and predicting the volatility of financial time series has been extensively researched for more than 30 years and the interest in the subject is far from decreasing. Volatility prediction has a very wide range of applications in finance, for example, in portfolio optimization, risk management, asset allocation, asset pricing, etc. The two most popular approaches to model volatility are based on the Autoregressive Conditional Heteroscedasticity (ARCH) type and Stochastic Volatility (SV) type models. The seminal paper of Engle (1982) proposed the primary ARCH model while Bollerslev (1986) generalized the purely autoregressive ARCH into an ARMA-type model, called the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. Since then, there has been a very large amount of research on the topic, stretching to various model extensions and generalizations. Meanwhile, the researchers have been addressing two important topics: looking for the best specification for the errors and selecting the most efficient approach for inference and prediction.

Besides selecting the best model for the data, distributional assumptions for the returns are equally important. It is well known, that every prediction, in order to be useful, has to come with a certain precision measurement. In this way the agent can know the risk she is facing, i.e. uncertainty. Distributional assumptions permit to quantify this uncertainty about the future. Traditionally, the errors have been assumed to be Gaussian, however, it has been widely acknowledged that financial returns display fat tails and are not conditionally Gaussian. Therefore, it is common to assume a Student-t distribution, see Bollerslev (1987), He & Teräsvirta (1999) and Bai et al. (2003), among others. However, the assumption of Gaussian or Student-t distributions is rather restrictive. An alternative approach is to use a mixture of distributions, which can approximate arbitrarily any distribution given a sufficient number of mixture components. A mixture of two Normals was used by Bai et al.

(2003), Ausín & Galeano (2007) and Giannikis et al. (2008), among others. These authors have shown that the models with the mixture distribution for the errors outperformed the Gaussian one and do not require additional restrictions on the degrees of freedom parameter as the Student-t one.

As for the inference and prediction, the Bayesian approach is especially well-suited for GARCH models and provides some advantages compared to classical estimation techniques, as outlined by Ardia & Hoogerheide (2010). Firstly, the positivity constraints on the parameters to ensure positive variance, may encumber some optimization procedures. In the Bayesian setting, constraints on the model parameters can be incorporated via priors. Secondly, in most of the cases we are more interested not in the model parameters directly, but in some non-linear functions of them. In the maximum likelihood (ML) setting, it is quite troublesome to perform inference on such quantities, while in the Bayesian setting it is usually straightforward to obtain the posterior distribution of any non-linear function of the model parameters. Furthermore, in the classical approach, models are usually compared by any other means than the likelihood. In the Bayesian setting, marginal likelihoods and Bayes factors allow for consistent comparison of non-nested models while incorporating Occam's razor for parsimony. Also, Bayesian estimation provides reliable results even for finite samples. Finally, Hall & Yao (2003) add that the ML approach presents some limitations when the errors are heavy tailed, also the convergence rate is slow and the estimators may not be asymptotically Gaussian.

This survey reviews the existing Bayesian inference methods for univariate and multivariate GARCH models while having in mind their error specifications. The main emphasis of the paper is on the recent development of an alternative inference approach for these models using Bayesian non-parametrics. The classical parametric modeling, relying on a finite number of parameters, although so widely used, has some certain drawbacks. Since the number of parameters for any model is fixed, one can encounter underfitting or overfitting, which

arises from the misfit between the data available and the parameters needed to estimate. Then, in order to avoid assuming wrong parametric distributions, which may lead to inconsistent estimators, it is better to consider a semi- or non-parametric approach. Bayesian non-parametrics may lead to less constrained models than classical parametric Bayesian statistics and provide an adequate description of the data, especially when the conditional return distribution is far away from Gaussian.

Up to our knowledge, there has been a very few papers using Bayesian non-parametrics for GARCH models. These are [Ausín et al. \(2014\)](#) for univariate GARCH, Jensen & Maheu (2013) and Virbickaite et al. (2013) for MGARCH. All of them have considered infinite mixtures of Gaussian distributions with a Dirichlet process (DP) prior over the mixing distribution, which results into DP mixture (DPM) models. This approach so far proves to be the most popular Bayesian non-parametric modeling procedure. The results over the papers have been consistent: The Bayesian non-parametric approach leads to more flexible models and is better in explaining heavy-tailed return distributions, which parametric models cannot fully capture.

The outline of this survey is as follows. Section 2 shortly introduces univariate GARCH models and different inference and prediction methods. Section 3 overviews the existing models for multivariate GARCH and different inference and prediction approaches. Section 4 introduces the Bayesian non-parametric modeling approach and reviews the limited literature of this area in time-varying volatility models. Section 5 presents a real data application. Finally, Section 6 concludes.

2 Univariate GARCH

As mentioned before, the two most popular approaches to model volatility are GARCH type and SV type models. In this survey we focus on GARCH models, therefore, SV models will

not be included thereafter. Also, we are not going to enter into the technical details of the Bayesian algorithms and refer to Robert & Casella (2004) for a more detailed description of Bayesian techniques.

2.1 Description of Models

The general structure of an asset return series modeled by a GARCH-type models can be written as:

$$r_t = \mu_t + a_t = \mu_t + \sqrt{h_t}\epsilon_t,$$

where $\mu_t = E[r_t|\mathcal{I}_{t-1}]$ is the conditional mean given \mathcal{I}_{t-1} , the information up to time $t - 1$, a_t is the mean corrected returns of the asset at time t , $h_t = \text{Var}[r_t|\mathcal{I}_{t-1}]$ is the conditional variance given \mathcal{I}_{t-1} and ϵ_t is the standard white noise shock. There are several ways to model the conditional mean, μ_t . The usual assumptions are to consider that the mean is either zero, equal to a constant ($\mu_t = \mu$), or follows an ARMA(p, q) process. However, sometimes the mean is also modeled as a function of the variance, say $g(h_t)$, which leads to the GARCH-in-Mean models. On the other hand, the conditional variance, h_t , is usually modeled using the GARCH-family models. In the basic GARCH model the conditional variance of the returns depends on a sum of three parts: a constant variance as the long-run average, a linear combination of the past conditional variances and a linear combination of the past mean squared returns. For instance, in the GARCH(1,1) model, the conditional variance at time t is given by $h_t = \omega + \alpha a_{t-1}^2 + \beta h_{t-1}$, for $t = 1, \dots, T$. There are some restrictions which have to be imposed such as $\omega > 0$, $\alpha, \beta \geq 0$ for positive variance, and $\alpha + \beta < 1$ for the covariance stationarity.

[Nelson \(1991\)](#) proposed the exponential GARCH (EGARCH) model that acknowledges the existence of asymmetry in the volatility response to the changes in the returns, sometimes also called the "leverage effect", introduced by Black (1976). Negative shocks to the

returns have a stronger effect on volatility than positive. Other ARCH extensions that try to incorporate the leverage effect are the GJR model by Glosten et al. (1993) and the TGARCH of Zakoian (1994), among many others. As Engle (2004) puts it, “there is now an alphabet soup” of ARCH family models, such as AARCH, APARCH, FIGARCH, STARCH etc, which try to incorporate such return features as fat tails, volatility clustering and volatility asymmetry. Papers by Bollerslev et al. (1992), Bollerslev et al. (1994), Engle (2002b), Ishida & Engle (2002) provide extensive reviews of the existing ARCH-type models. Bera & Higgins (1993) review ARCH type models, discuss their extensions, estimation and testing, also numerous applications. Also, one can find an explicit review with examples and applications concerning GARCH-family models in Tsay (2010) and Chapter 1 in Teräsvirta (2009).

2.2 Inference Methods

The main estimation approach for GARCH-family models is the classical maximum likelihood method. However, recently there has been a rapid development of Bayesian estimation techniques, which offer some advantages compared to the frequentist approach as already discussed in the Introduction. In addition, in the empirical finance setting, the frequentist approach presents an uncertainty problem. For instance, optimal allocation is greatly affected by the parameter uncertainty, which has been recognized in a number of papers, see [Jorion \(1986\)](#) and [Greyserman et al. \(2006\)](#), among others. These authors conclude that in the frequentist setting the estimated parameter values are considered to be the true ones, therefore, the optimal portfolio weights tend to inherit this estimation error. However, instead of solving the optimization problem on the basis of the choice of unique parameter values, the investor can choose the Bayesian approach, because it accounts for parameter uncertainty, as seen in [Kang \(2011\)](#) and [Jacquier & Polson \(2013\)](#), for example. A number of papers in this field have explored different Bayesian procedures for inference and predic-

tion and different approaches to modeling the fat-tailed errors and/or asymmetric volatility. The recent development of modern Bayesian computational methods, based on Monte Carlo approximations and MCMC methods have facilitated the usage of Bayesian techniques, see e.g. Robert & Casella (2004).

The standard Gibbs sampling procedure does not make the list because it cannot be used due to the recursive nature of the conditional variance: the conditional posterior distributions of the model parameters are not of a simple form. One of the alternatives is the *Griddy-Gibbs* sampler as in Bauwens & Lubrano (1998). They discuss that previously used importance sampling and Metropolis algorithms have certain drawbacks, such as that they require a careful choice of a good approximation of the posterior density. The authors propose a Griddy-Gibbs sampler which explores analytical properties of the posterior density as much as possible. In this paper the GARCH model has Student-t errors, which allows for fat tails. The authors choose to use flat (uniform) priors on parameters (ω, α, β) with whatever region is needed to ensure the positivity of variance, however, the flat prior for the degrees of freedom cannot be used, because then the posterior density is not integrable. Instead, they choose a half-right side of Cauchy. The posteriors of the parameters were found to be skewed, which is a disadvantage for the commonly used Gaussian approximation. On the other hand, Ausín & Galeano (2007) modeled the errors of a GARCH model with a mixture of two Gaussian distributions. The advantage of this approach compared to that of Student-t errors, is that if the number of the degrees of freedom is very small (less than 5), some moments may not exist. The authors have chosen flat priors for all the parameters, and discovered that there is little sensitivity to the change in the prior distributions (from uniform to Beta), unlike in Bauwens & Lubrano (1998), where the sensitivity for the prior choice for the degrees of freedom is high. More articles using a Griddy-Gibbs sampling approach are by Bauwens & Lubrano (2002), who have modeled asymmetric volatility with Gaussian innovations and have used uniform priors for all the parameters, and by Wago

(2004), who explored an asymmetric GARCH model with Student-t errors.

Another MCMC algorithm used in estimating GARCH model parameters, is the *Metropolis-Hastings* (MH) method, which samples from a candidate density and then accepts or rejects the draws depending on a certain acceptance probability. [Ardia \(2006\)](#) modeled the errors as Gaussian distributed with zero mean and unit variance while the priors are chosen as Gaussian and a MH algorithm is used to draw samples from the joint posterior distribution. The author has carried out a comparative analysis between ML and Bayesian approaches, finding, as in other papers, that some posterior distributions of the parameters were skewed, thus warning against the abusive use of the Gaussian approximation. Also, [Ardia \(2006\)](#) has performed a sensitivity analysis of the prior means and scale parameters and concluded that the initial priors in this case are vague enough. This approach has been also used by [Müller & Pole \(1998\)](#), [Nakatsuma \(2000\)](#) and [Vrontos et al. \(2000\)](#), among others. A special case of the MH method is the random walk Metropolis-Hastings (RWMH) where the proposal draws are generated by randomly perturbing the current value using a spherically symmetric distribution. A usual choice is to generate candidate values from a Gaussian distribution where the mean is the previous value of the parameter and the variance can be calibrated to achieve the desired acceptance probability. This procedure is repeated at each MCMC iteration. [Ausín & Galeano \(2007\)](#) have also carried out a comparison of estimation approaches, Griddy-Gibbs, RWMH and ML. Apparently, RWMH has difficulties in exploring the tails of the posterior distributions and ML estimates may be rather different for those parameters where posterior distributions are skewed.

In order to select one of the algorithms, one might consider some criteria, such as fast convergence for example. [Asai \(2006\)](#) numerically compares some of these approaches in the context of GARCH. The Griddy-Gibbs method is capable in handling the shape of the posterior by using smaller MCMC outputs comparing with other methods, also, it is flexible regarding parametric specification of a model. However, it can require a lot of computational

time. This author also investigates MH, adaptive rejection Metropolis sampling (ARMS), proposed by [Gilks et al. \(1995\)](#), and acceptance-rejection MH algorithms (ARMH), proposed by Tierney (1994). For more in detail about each method in GARCH models see [Nakatsuma \(2000\)](#) and Kim et al. (1998), among others. Using simulated data, Asai (2006) calculated geometric averages of inefficiency factors for each method. Inefficiency factor is just an inverse of [Geweke \(1992\)](#) efficiency factor. According to this, the ARMH algorithm performed the best. Also, computational time was taken into consideration, where ARMH clearly outperformed MH and ARMS, while Griddy-Gibbs stayed just a bit behind. The author observes that even though the ARMH method showed the best results, the posterior densities for each parameter did not quite explore the tails of the distributions, as desired. In this case Griddy-Gibbs performs better; also, it requires less draws than ARMH. Bauwens & Lubrano (1998) investigate one more convergence criteria, proposed by Yu & Mykland (1998), which is based on cumulative sum (cumsum) statistics. It basically shows that if MCMC is converging, the graph of a certain cumsum statistic against time should approach zero. Their employed Griddy-Gibbs algorithm converged in all four parameters quite fast. Then, the authors explored the advantages and disadvantages of alternative approaches: the importance sampling and the MH algorithm. Considering importance sampling, one of the main disadvantages, as mentioned before, is to find a good approximation of the posterior density (importance function). Also, comparing with Griddy-Gibbs algorithm, the importance sampling requires much more draws to get smooth graphs of the marginal densities. For the MH algorithm, same as in importance sampling, a good approximation needs to be found. Also, compared to Griddy-Gibbs, the MH algorithm did not fully explore the tails of the distribution, unless for a very big number of draws.

Another important aspect of the Bayesian approach, as commented before, are the advantages in model selection compared to the classical methods. Miazhyńska & Dorffner (2006) reviews some Bayesian model selection methods using MCMC for GARCH-type mod-

els, which allow for the estimation of either marginal model likelihoods, Bayes factors or posterior model probabilities. These are compared to the classical model selection criteria showing that the Bayesian approach clearly considers model complexity in a more unbiased way. Also, [Chen et al. \(2009\)](#) includes a revision of Bayesian selection methods for asymmetric GARCH models, such as the GJR-GARCH and threshold GARCH. They show how using the Bayesian approach it is possible to compare complex and non-nested models to choose for example between GARCH and stochastic volatility models, between symmetric or asymmetric GARCH models or to determine the number of regimes in threshold processes, among others.

An alternative approach to the previous parametric specifications is the use of Bayesian non-parametric methods, that allow to model the errors as an infinite mixture of normals, as seen in the paper by [Ausín et al. \(2014\)](#). The Bayesian non-parametric approach for time-varying volatility models will be discussed in detail in Section 4.

To sum up, considering the amount of articles published quite recently regarding the topic of estimating univariate GARCH models using MCMC methods indicates still growing interest in the area. Although numerous GARCH-family models have been investigated using different MCMC algorithms, there are still a lot of areas that need further research and development.

3 Multivariate GARCH

Returns and volatilities depend on each other, so multivariate analysis is a more natural and useful approach. The starting point of multivariate volatility models is a univariate GARCH, thus the most simple MGARCH models can be viewed as direct generalizations of their univariate counterparts. Consider a multivariate return series $\{r_t\}_{t=1}^T$ of size $K \times 1$.

Then

$$r_t = \mu_t + a_t = \mu_t + H_t^{1/2} \epsilon_t,$$

where $\mu_t = E[r_t | \mathcal{I}_{t-1}]$, a_t are mean-corrected returns, ϵ_t is a random vector, such that $E[\epsilon_t] = 0$ and $\text{Cov}[\epsilon_t] = I_K$ and $H_t^{1/2}$ is a positive definite matrix of dimensions $K \times K$, such that H_t is the conditional covariance matrix of r_t , i.e., $\text{Cov}[r_t | \mathcal{I}_{t-1}] = H_t^{1/2} \text{Cov}[\epsilon_t] (H_t^{1/2})' = H_t$. There is a wide range of MGARCH models, where most of them differ in specifying H_t . In the rest of this section we will review the most popular and widely used and the different Bayesian approaches to make inference and prediction. For general reviews on MGARCH models, see Bauwens et al. (2006), Silvennoinen & Teräsvirta (2009) and Tsay (2010) (Chapter 10), among others.

Regarding inference, one can also consider the same arguments provided in the univariate GARCH case above. Maximum likelihood estimation for MGARCH models can be obtained by using numerical optimization algorithms, such as Fisher scoring and Newton-Raphson. Vrontos et al. (2003b) have estimated several bivariate ARCH and GARCH models and found that some classical estimates of the parameters were quite different from their Bayesian counterparts. This was due to the non-Normality of the parameters. Thus, the authors suggest careful interpretation of the classical estimation approach. Also, Vrontos et al. (2003b) found it difficult to evaluate the classical estimates under the stationarity conditions, and consequently the resulting parameters, evaluated ignoring the stationarity constraints, produced non-stationary estimates. These difficulties can be overcome using the Bayesian approach.

3.1 VEC, DVEC and BEKK

The VEC model was proposed by Bollerslev et al. (1988), where every conditional variance and covariance (elements of the H_t matrix) is a function of all lagged conditional variances

and covariances, as well as lagged squared mean-corrected returns and cross-products of returns. Using this unrestricted VEC formulation, the number of parameters increases dramatically. For example, if $K = 3$, the number of parameters to estimate will be 78, and if $K = 4$, the number of parameters increases to 210, see Bauwens et al. (2006) for the explicit formula for the number of parameters in VEC models. To overcome this difficulty, Bollerslev et al. (1988) simplified the VEC model by proposing a diagonal VEC model, or DVEC, as follows:

$$H_t = \Omega + A \odot (a_{t-1}a'_{t-1}) + B \odot H_{t-1},$$

where \odot indicates the Hadamard product, Ω , A and B are symmetric $K \times K$ matrices. As noted in Bauwens et al. (2006), H_t is positive definite provided that Ω , A , B and the initial matrix H_0 are positive definite. However, these are quite strong restrictions on the parameters. Also, DVEC model does not allow for dynamic dependence between volatility series. In order to avoid such strong restrictions on the parameter matrices, Engle & Kroner (1995) propose the BEKK model, which is just a special case of a VEC and, consequently, less general. It has the attractive property that the conditional covariance matrices are positive definite by construction. The model looks as follows:

$$H_t = \Omega^* \Omega^{*'} + A^* (a_{t-1}a'_{t-1}) A^{*'} + B^* H_{t-1} B^{*'}, \quad (1)$$

where Ω^* is a lower triangular matrix and A^* and B^* are $K \times K$ matrices. In the BEKK model it is easy to impose the definite positiveness of the H_t matrix. However, the parameter matrices A^* and B^* do not have direct interpretations since they do not represent directly the size of the impact of the lagged values of volatilities and squared returns.

Osiewalski & Pipien (2004) present a paper that compares the performance of various bivariate ARCH and GARCH models, such as VEC, BEKK, etc, estimated using Bayesian techniques. As the authors observe, they are the first to perform model comparison using Bayes factors and posterior odds in the MGARCH setting. The algorithm used for parameter estimation and inference is Metropolis-Hastings, and to check for convergence they rely on the cumsum statistics, introduced by Yu & Mykland (1998), and used by Bauwens & Lubrano (1998) in the univariate GARCH setting. Using the real data the authors found that the t-BEKK models performed the best, leaving t-VEC not so far behind; t-VEC model, sometimes also called t-VECH, is a more general form of a DVEC, seen above, where the mean-corrected returns follow a Student-t distribution. The name comes from a function called *vech*, which reshapes the lower triangular portion of a symmetric variance-covariance matrix into a column vector. To sum up, the authors choose t-BEKK model as clearly better than the t-VEC, because it is relatively simple and has less parameters to estimate.

On the other hand, Hudson & Gerlach (2008) developed a prior distribution for a VECH specification that directly satisfy both necessary and sufficient conditions for positive definiteness and covariance stationarity, while remaining diffuse and non-informative over the allowable parameter space. These authors employed MCMC methods, including Metropolis-Hastings, to help enforce the conditions in this prior.

More recently, Burda & Maheu (2013) use the BEKK-GARCH model to show the usefulness of a new posterior sampler called the Adaptive Hamiltonian Monte Carlo (AHMC). Hamiltonian Monte Carlo (HMC) is a procedure to sample from complex distributions. The AHMC is an alternative inferential method based on HMC that is both fast and locally adaptive. The AHMC appears to work very well when the dimension of the parameter space is very high. Model selection based on marginal likelihood is used to show that full BEKK models are preferred to restricted diagonal specifications. Additionally, Burda (2013) suggests an approach called Constrained Hamiltonian Monte Carlo (CHMC) in order to deal

with high dimensional BEKK models with targeting, which allow for a parameter dimension reduction without compromising the model fit, unlike the diagonal BEKK. Model comparison of the full BEKK and the BEKK with targeting is performed indicating that the latter dominates the former in terms of marginal likelihood.

3.2 Factor-GARCH

Factor-GARCH was first proposed by Engle et al. (1990) to reduce the dimension of the multivariate model of interest using an accurate approximation of the multivariate volatility. The definition of the Factor-GARCH model, proposed by Lin (1992), says that BEKK model in (1) is a Factor-GARCH, if A^* and B^* have rank one and the same left and right eigenvalues: $A^* = \alpha w \lambda'$, $B^* = \beta w \lambda'$, where α and β are scalars and w and λ are eigenvectors. Several variants of the factor model have been proposed. One of them is the full-factor multivariate GARCH by Vrontos et al. (2003a):

$$\begin{aligned} r_t &= \mu + a_t, \\ a_t &= W X_t, \\ X_t | \mathcal{I}_{t-1} &\sim \mathcal{N}_K(0, \Sigma_t), \end{aligned}$$

where μ is a $K \times 1$ vector of constants, which is time invariant, W is a $K \times K$ parameter matrix, X_t is a $K \times 1$ vector of factors and $\Sigma_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{Kt}^2)$ is a $K \times K$ diagonal variance matrix such that $\sigma_{it}^2 = c_i + b_i x_{i,t-1}^2 + g_i \sigma_{i,t-1}^2$, where σ_{it}^2 is the conditional variance of the i th factor at time t such that $c_i > 0$, $b_i \geq 0$, $g_i \geq 0$. Then, the factors in the X_t vector are GARCH(1,1) processes and the vector a_t is a linear combination of such factors. It can be easily shown that H_t is always positive definite by construction. However, the structure of H_t depends on the order of the time series in r_t . Vrontos et al. (2003a) have considered this

problem to find the best ordering under the proposed model. Furthermore, Vrontos et al. (2003a) investigate a full-factor MGARCH model using the ML and Bayesian approaches. The authors compute maximum likelihood estimates using Fisher scoring algorithm. As for the Bayesian analysis, the authors have adopted a Metropolis-Hastings algorithm, and found that the algorithm is very time consuming, especially in high-dimensional data. To speed-up the convergence, Vrontos et al. (2003a) have proposed reparametrization of positive parameters and also a blocking sampling scheme, where the parameter vector is divided into three blocks: mean, variance and the matrix of constants W . As mentioned before, the ordering of the univariate time series in full-factor models is important, thus to select “the best” model one has to consider $K!$ possibilities for a multivariate dataset of dimension K . Instead of choosing one model and making inference (as if the selected model was the true one), the authors employ a Bayesian approach by calculating the posterior probabilities for all competing models and model averaging to provide “combined” predictions. The main contribution of this paper is that the authors were able to carry out an extensive Bayesian analysis of a full-factor MGARCH model considering not only parameter uncertainty, but model uncertainty as well.

As already discussed above, a very common stylized feature of financial time series is the asymmetric volatility. Dellaportas & Vrontos (2007) have proposed a new class of tree structured MGARCH models that explore the asymmetric volatility effect. Same as the paper by Vrontos et al. (2003a), the authors consider not only parameter-related uncertainty, but also uncertainty corresponding to model selection. Thus in this case Bayesian approach becomes particularly useful because an alternative method based on maximizing the pseudo-likelihood is only able to work after selecting a single model. The authors develop an MCMC stochastic search algorithm that generates candidate tree structures and their posterior probabilities. The proposed algorithm converged fast. Such modeling and inference approach leads to more reliable and more informative results concerning model-selection and individual parameter

inference.

There are more models that are nested in BEKK, such as the Orthogonal GARCH for example, see Alexander & Chibumba (1997) and Van der Weide (2002), among others. All of them fall into the class of direct generalizations of univariate GARCH or linear combinations of univariate GARCH models. Another class of models are the nonlinear combinations of univariate GARCH models, such as conditional correlation (CCC), dynamic condition correlation (DCC), general dynamic covariance (GDC) and Copula-GARCH models. A very recent alternative approach that also considers Bayesian estimation can be found in Jin & Maheu (2013) who proposes a new dynamic component models of returns and realized covariance (RCOV) matrices based on time-varying Wishart distributions. In particular, Bayesian estimation and model comparison is conducted with a existing range of multivariate GARCH models and RCOV models.

3.3 CCC

The CCC model, proposed by Bollerslev (1990) and the simplest in its class, is based on the decomposition of the conditional covariance matrix into conditional standard deviations and correlations. Then, the conditional covariance matrix H_t looks as follows:

$$H_t = D_t R D_t,$$

where D_t is diagonal matrix with the K conditional standard deviations and R is a time-invariant conditional correlation matrix such that $R = (\rho_{ij})$ and $\rho_{ij} = 1, \forall i = j$. The CCC approach can be applied to a wide range of univariate GARCH family models, such as exponential GARCH or GJR-GARCH, for example.

Vrontos et al. (2003b) have estimated some real data using a variety of bivariate ARCH and GARCH models in order to select the best model specification and to compare the

Bayesian parameter estimates to those of the ML. These authors have considered three ARCH and three GARCH models, all of them with constant conditional correlations (CCC). They have used a Metropolis-Hastings algorithm, which allows to simulate from the joint posterior distribution of the parameters. For model comparison and selection, Vrontos et al. (2003b) have obtained predictive distributions and assessed comparative validity of the analyzed models, according to which the CCC model with diagonal covariance matrix performed the best considering one-step-ahead predictions.

3.4 DCC

A natural extension of the simple CCC model are the dynamic conditional correlation (DCC) models, firstly proposed by Tse & Tsui (2002) and Engle (2002a). The DCC approach is more realistic, because the dependence between returns is likely to be time-varying.

The models proposed by Tse & Tsui (2002) and Engle (2002a) consider that the conditional covariance matrix H_t looks as $H_t = D_t R_t D_t$, where R_t is now a time varying correlation matrix at time t . The models differ in the specification of R_t . In the paper by Tse & Tsui (2002), the conditional correlation matrix is $R_t = (1 - \theta_1 - \theta_2)R + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1}$, where θ_1 and θ_2 are non-negative scalar parameters, such that $\theta_1 + \theta_2 < 1$, R is a positive definite matrix such that $\rho_{ii} = 1$ and Ψ_{t-1} is a $K \times K$ sample correlation matrix of the past m standardized mean-corrected returns $u_t = D_t^{-1}a_t$. On the other hand, in the paper by Engle (2002a), the specification of R_t is $R_t = (I \odot Q_t)^{-1/2} Q_t (I \odot Q_t)^{-1/2}$, where $Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha(u_{t-1}u'_{t-1}) + \beta Q_{t-1}$, $u_{i,t} = a_{i,t}/\sqrt{h_{ii,t}}$ is a mean-corrected standardized returns, α and β are non-negative scalar parameters, such that $\alpha + \beta < 1$ and \bar{Q} is unconditional covariance matrix of u_t . As noted in Bauwens et al. (2006), the model by Engle (2002a) does not formulate the conditional correlation as a weighted sum of past correlations, unlike in the DCC model by Tse & Tsui (2002), seen above. The drawback of both these

models is that θ_1 , θ_2 , α and β are scalar parameters, so all conditional correlations have the same dynamics. However, as Tsay (2010) notes it, the models are parsimonious.

Moreover, as financial returns display not only asymmetric volatility, but also excess kurtosis, previous research, as in univariate case, has mostly considered using a multivariate Student-t distribution for the errors. However, as already discussed above, this approach has several limitations. Galeano & Ausín (2010) propose a MGARCH-DCC model, where the standardized innovations follow a mixture of Gaussian distributions. This allows to capture long tails without being limited by the degrees of freedom constraint, which is necessary to impose in the Student-t distribution so that higher moments could exist. The authors estimate the proposed model using the classical ML and Bayesian approaches. In order to estimate model parameters, dynamics of single assets and dynamic correlations, and the parameters of the Gaussian mixture, Galeano & Ausín (2010) have relied on RWMH algorithm. BIC criteria was used for selecting the number of mixture components, which performed well in simulated data. Using real data, the authors provide an application to calculating the Value at Risk (VaR) and solving a portfolio selection problem. MLE and Bayesian approaches have performed similarly in point estimation, however, the Bayesian approach, besides giving just point estimates, allows the derivation of predictive distributions for the portfolio VaR.

An extension of the DCC model of Engle (2002a) is the Asymmetric DCC also proposed by Engle (2002a), which incorporates an asymmetric correlation effect. It means that correlations between asset returns decrease more in the bear market than they increase when the market performs well. Cappiello et al. (2006) generalizes the ADCC model into the AGDCC model, where the parameters of the correlation equation are vectors, and not scalars. This allows for asset-specific correlation dynamics. In the AGDCC model, the Q_t matrix in the

DCC model is replaced with:

$$Q_t = S(1 - \bar{\kappa}^2 - \bar{\lambda}^2 - \bar{\delta}^2/2) + \kappa\kappa' \odot u'_{t-1}u_{t-1} + \lambda\lambda' \odot Q_{t-1} + \delta\delta' \odot \eta'_{t-1}\eta_{t-1},$$

where $u_t = D_t^{-1}a_t$ are mean corrected standardized returns, $\eta_t = u_t \odot I(u_t < 0)$ selects just negative returns, "diag" stands for either taking just the diagonal elements from the matrix, or making a diagonal matrix from a vector, S is a sample correlation matrix of u_t , κ, λ and δ are $K \times 1$ vectors, $\bar{\kappa} = K^{-1} \sum_{i=1}^K \kappa_i$, $\bar{\lambda} = K^{-1} \sum_{i=1}^K \lambda_i$ and $\bar{\delta} = K^{-1} \sum_{i=1}^K \delta_i$. To ensure positivity and stationarity of Q_t , it is necessary to impose $\kappa_i, \lambda_i, \delta_i > 0$ and $\kappa_i^2 + \lambda_i^2 + \delta_i^2/2 < 1, \forall i = 1, \dots, K$. The AGDCC by Cappiello et al. (2006) is just a special case where $\kappa_1 = \dots = \kappa_K$, $\lambda_1 = \dots = \lambda_K$ and $\delta_1 = \dots = \delta_K$.

Up to our knowledge, the only paper that considers the AGDCC model in the Bayesian setting is Virbickaite et al. (2013) that propose to model the distribution of the standardized returns as an infinite scale mixture of Gaussian distributions by relying on Bayesian non-parametrics. This approach is presented in more detail in Section 4.

3.5 Copula-GARCH

The use of copulas is an alternative approach to study return time series and their volatilities. The main convenience of using copulas is that individual marginal densities of the returns can be defined separately from their dependence structure. Then, each marginal time series can be modeled using univariate specification and the dependence between the returns can be modeled by selecting an appropriate copula function. A K -dimensional copula $C(u_1, \dots, u_K)$, is a multivariate distribution function in the unit hypercube $[0, 1]^K$, with uniform $[0, 1]$ marginal distributions. Under certain conditions, the Sklar Theorem affirms that (see, Sklar, 1959), every joint distribution $F(x_1, \dots, x_K)$, whose marginals are given by $F_1(x_1), \dots, F_K(x_K)$, can be written as $F(x_1, \dots, x_K) = C(F_1(x_1), \dots, F_K(x_K))$, where C is

a copula function of F , which is unique if the marginal distributions are continuous.

The most popular approach to volatility modeling through copulas is called the Copula-GARCH model, where univariate GARCH models are specified for each marginal series and the dependence structure between them is described using a copula function. A very useful feature of copulas, as noted by Patton (2009), is that the marginal distributions of each random variable do not need to be similar to each other. This is very important in modeling time series, because each of them might be following different distributions. The choice of copulas can vary from a simple Gaussian copula to more flexible ones, such as Clayton, Gumbel, mixed Gaussian, etc. In the existing literature different parametric and non-parametric specifications can be used for the marginals and copula function C . Also, the copula function can be assumed to be constant or time varying, as seen in Ausín & Lopes (2010), among others.

The estimation for Copula-GARCH models can be performed in a variety of ways. Maximum likelihood is the obvious choice for fully parametric models. Estimation is generally based on a multi-stage method, where firstly the parameters of the marginal univariate distributions are estimated and then used to condition in estimating the parameters of the copula. Another approach is non- or semi-parametric estimation of the univariate marginal distributions followed by a parametric estimation of the copula parameters. As Patton (2006) has showed, the two-stage maximum likelihood approach lead to consistent, but not efficient estimators.

An alternative is to employ a Bayesian approach, as done by Ausín & Lopes (2010). The authors have developed a one-step Bayesian procedure where all parameters are estimated at the same time using the entire likelihood function and, provided the methodology, for obtaining optimal portfolio, calculating VaR and CVaR. Ausín & Lopes (2010) have used a Gibbs sampler to sample from a joint posterior, where each parameter is updated using a RWMH. In order to reduce computational cost, the model and copula parameters are

updated not one-by-one, but rather by blocks, that consist of highly correlated vectors of model parameters.

Arakelian & Dellaportas (2012) have also used Bayesian inference for copula-GARCH models. These authors have proposed a methodology for modelling dynamic dependence structure by allowing copula functions or copula parameters to change across time. The idea is to use a threshold approach so these changes, that are assumed to be unknown, do not evolve in time but occur in distinct points. These authors have also employed a RWMH for parameter estimation together with a Laplace approximation. The adoption of an MCMC algorithm allows the choice of different copula functions and/or different parameter values between two time thresholds. Bayesian model averaging is considered for predicting dependence measures such as the Kendall's correlation. They conclude that the new model performs well and offers a good insight into the time-varying dependencies between the financial returns.

Hofmann & Czado (2010) developed Bayesian inference of a multivariate GARCH model where the dependence is introduced by a D-vine copula on the innovations. A D-vine copula is a special case of vine copulas which are very flexible to construct multivariate copulas because it allows to model dependency between pairs of margins individually. Inference is carried out using a two-step MCMC method closely related with the usual two-step maximum likelihood procedure for estimating copula-GARCH models. The authors then focus on estimating the VaR of a portfolio that shows asymmetric dependencies between some pairs of assets and symmetric dependency between others.

All the previously introduced methods rely on parametric assumptions for the distribution of the errors. However, imposing a certain distribution can be rather restrictive and lead to underestimated uncertainty about future volatilities, as seen in Virbickaite et al. (2013). Therefore, Bayesian non-parametric methods become especially useful, since they do not impose any specific distribution on the standardized returns.

4 Bayesian Non-Parametrics for GARCH

Bayesian non-parametrics is an alternative approach to the classical parametric Bayesian statistics, where one usually gives some prior for the parameters of interest, whose distribution is unknown, and then observes the data and calculates the posterior. The priors come from the family of parametric distributions. Bayesian non-parametrics uses a prior over distributions with the support being the space of all distributions. Then, it can be viewed as a distribution over distributions.

4.1 DP and DPM

One of the most popular Bayesian non-parametric modeling approach is based on Dirichlet processes (DP) and mixtures of Dirichlet processes (DPM). DP was firstly introduced by Ferguson (1973). Suppose that we have a sequence of exchangeably distributed random variables X_1, X_2, \dots from an unknown distribution F , where the support for X_i is Ω . In order to perform Bayesian inference, we need to define the prior for F . This can be done by considering partitions of Ω , such as $\Omega = C_1 \cup C_2 \cup \dots \cup C_m$, and defining priors over all possible partitions. We say that F has a Dirichlet process prior, denoted as $F \sim \mathcal{DP}(\alpha, F_0)$, if the set of associated probabilities given F for any partition follows a Dirichlet distribution, $\{F(C_1), \dots, F(C_m)\} \sim \text{Dirichlet}(\alpha F_0(C_1), \dots, \alpha F_0(C_m))$, where $\alpha > 0$ is a precision parameter that represents our prior certainty of how concentrated the distribution is around F_0 , which is a known base distribution on Ω . The Dirichlet process is a conjugate prior. Thus, given n independent and identically distributed samples from F , the posterior distribution of F is also a Dirichlet process such that $F \sim \mathcal{DP}(\alpha + n, (\alpha F_0 + n F_n)(\alpha + n)^{-1})$, where F_n is the empirical distribution function.

There are two main ways for generating a sample from the marginal distribution of X , where $X|F \sim F$ and $F \sim \mathcal{DP}(\alpha, F_0)$: the Polya urn and stick breaking procedures. On the

one hand, the Polya urn scheme can be illustrated in terms of a urn with α black balls; when a non-black ball is drawn, it is placed back in the urn together with another ball of the same color. If the drawn ball is black, a new color is generated from F_0 and a ball of this new color is added to the urn together with the black ball we drew. This process gives a discrete marginal distribution for X since there is always a probability that a previously seen value is repeated. On the other hand, the stick-breaking procedure is based on the representation of the random distribution F as a countably infinite mixture:

$$F = \sum_{m=1}^{\infty} \omega_m \delta_{X_m},$$

where δ_X is a Dirac measure, $X_m \sim F_0$ and the weights are such that $\omega_1 = \beta_1$, $\omega_m = \beta_m \prod_{i=1}^{m-1} (1 - \beta_i)$, for $m = 1, \dots$, where $\beta_m \sim \text{Beta}(1, \alpha)$. This implies that the weights $\omega \rightarrow \text{Dirichlet}(\alpha/K, \dots, \alpha/K)$ as $K \rightarrow \infty$.

The discreteness of the Dirichlet process is clearly a disadvantage in practice. A solution was proposed by Antoniak (1974) by using DPM models where a DP prior is imposed over over the distribution of the model parameters, θ , as follows:

$$\begin{aligned} X_i | \theta_i &\sim F(X | \theta_i), \\ \theta_i | G &\sim G(\theta), \\ G | \alpha, G_0 &\sim \mathcal{DP}(\alpha, G_0). \end{aligned}$$

Observe that G is a random distribution drawn from the DP and because it is discrete, multiple θ_i 's can take the same value simultaneously, making it a mixture model. In fact, using the stick-breaking representation, the hierarchical model above can be seen as an

infinite mixture of distributions:

$$f(X|\theta, \omega) = \sum_{m=1}^{\infty} \omega_m f(X|\theta_m),$$

where the weights are obtained as before: $\omega_1 = \beta_1$, $\omega_m = \beta_m \prod_{i=1}^{m-1} (1 - \beta_i)$, for $m = 1, \dots$, and where $\beta_m \sim \text{Beta}(1, \alpha)$ and $\theta_m \sim G_0$.

Regarding inference algorithms, there are two main types of approaches. On the one hand, the marginal methods, such as those proposed by Escobar & West (1995), MacEachern (1994) and Neal (2000), which rely on the Polya urn representation. All these algorithms are based on integrating out the infinite dimensional part of the model. Recently, another class of algorithms, called conditional methods, have been proposed. These approaches, based on the stick-breaking scheme, leave the infinite part in the model and sample a finite number of variables. These include the procedure by Walker (2007), who introduces slice sampling schemes to deal with the infiniteness in DPM, and the retrospective MCMC method of Papaspiliopoulos & Roberts (2008), that is later combined by Papaspiliopoulos (2008) with slice sampling method by Walker (2007) to obtain a new composite algorithm, which is better, faster and easier to implement. Generally, the stick-breaking compared to the Polya urn procedures produce better mixing and simpler algorithms.

4.2 Volatility modeling using DPM

As mentioned above, so far there has been little research in modeling volatilities with MGARCH using the DPM models. Up to our knowledge, these only include: Ausín et al. (2014) for univariate GARCH, and Jensen & Maheu (2013) and Virbickaite et al. (2013), for MGARCH.

Ausín et al. (2014) have applied semi-parametric Bayesian techniques to estimate univariate GARCH-type models. These authors have used the class of scale mixtures of Gaussian

distributions, that allow for the variances to change over components, with a Dirichlet process prior on the mixing distribution to model innovations of the GARCH process. The resulting class of models is called DPM-GARCH type models. In order to perform Bayesian inference on the new model, the authors employ a stick-breaking sampling scheme and make use of the ideas proposed in Walker (2007), Papaspiliopoulos & Roberts (2008) and Papaspiliopoulos (2008). The new scale mixture model was compared to a simpler mixture of two Gaussians, Student-t and the usual Gaussian models. The estimation results in all three cases were quite similar, however, the scale mixture model is able to capture skewness as well as kurtosis and, based on the approximated Log Marginal Likelihood (LML) and DIC, provided the best performance in simulated and real data. Finally, Ausín et al. (2014) have applied the resulting model to perform one-step-ahead predictions for volatilities and VaR. In general, the non-parametric approach leads to wider Bayesian credible intervals and can better describe long tails.

Jensen & Maheu (2013) propose a Bayesian non-parametric modeling approach for the innovations in MGARCH models. They use a MGARCH specification, proposed by Ding & Engle (2001), which is a different representation of a well known DVEC model, introduced above. The innovations are modeled as an infinite mixture of multivariate Normals with a DP prior. The authors have employed Polya urn and stick-breaking schemes and, using two data sets, compared the three model specifications: parametric MGARCH with Student-t innovations (MGARCH-t), GARCH-DPM- Λ that allows for different covariances (scale mixture) and MGARCH-DPM, allowing for different means and covariances of each component (location-scale mixture). In general, both semi-parametric models produced wider density intervals. However, in MGARCH-t model a single degree of freedom parameter determines the tail thickness in all directions of the density, meanwhile the non-parametric models are able to capture various deviations from Normality by using a certain number of components. These results are consistent with the ones in Ausín et al. (2014). As for

predictions, both semi-parametric models performed equally good and outperformed the parametric MGARCH-t specification.

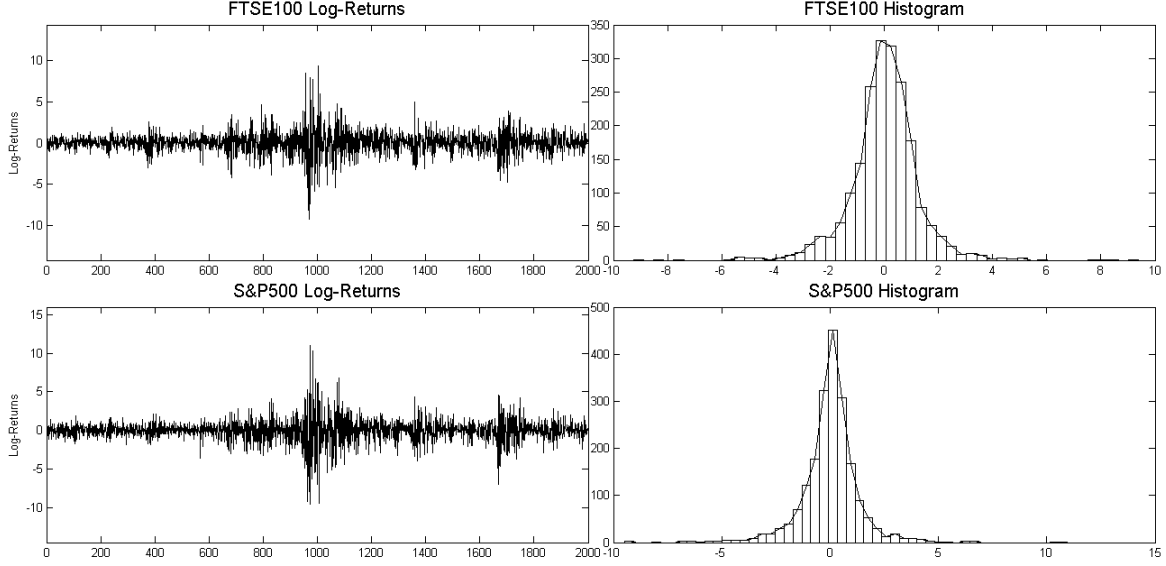
Finally, the paper by Virbickaite et al. (2013) can be seen as a direct generalization of the paper by Ausín et al. (2014) to the multivariate framework. Here, same as in Jensen & Maheu (2013), the authors have proposed using an infinite scale mixture of Normals for the standardized returns. For the MGARCH model a GJR-ADCC was chosen, allowing for asymmetric volatilities and asymmetric time-varying correlations. Moreover, the authors have carried-out a simulation study that illustrated the adaptability of the DPM model. Finally, the authors provided one real data application to portfolio decision problem concluding that DPM models are less restrictive and more adaptive to whatever distribution the data comes from, therefore, can better capture the uncertainty about financial decisions.

To sum up, the findings in the above papers are consistent: the Bayesian semi-parametric approach leads to more flexible models and is better in explaining heavy-tailed return distributions, which parametric models cannot fully capture. The parameters are less precise, i.e. wider Bayesian credible intervals are observed because the semi-parametric models are less restricted. This provides a more adequate measure of uncertainty. If in the Gaussian setting the credible intervals are very narrow and the real data is not Gaussian, this makes the agent overconfident about her decisions, and she takes more risk than she would like to assume. Steel (2008) observes that the combination of Bayesian methods and MCMC computational algorithms provide new modeling possibilities and calls for more research regarding non-parametric Bayesian time series modeling.

5 Illustration

This illustration study using real data has basically two goals: firstly, to show the advantages of the Bayesian approach, such as the ability to obtain posterior densities of quantities of

Figure 1. Log>Returns and Histograms of FTSE100 and S&P500 Indices



interest and the facility to incorporate various constraints on the parameters. Secondly, to illustrate the flexibility of the Bayesian non-parametric approach for GARCH modeling.

The data used for estimation are the log-returns (in percentages), obtained from close prices adjusted for dividends and splits, of two market indices: FTSE100 and S&P500 from November 10th, 2004 till December 10th, 2012, resulting into a sample size of 2000 observations. FTSE100 is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization. S&P500 is a stock market index based on the common stock prices of 500 top publicly traded American companies. The data was obtained from Yahoo Finance. Figure 1 and Table 1 present the basic plots and descriptive statistics of the two log-return series.

As seen from the plot and descriptive statistics, the data is slightly skewed and with high kurtosis, therefore, assuming a Gaussian distribution for the standardized returns would be inappropriate. Therefore, we estimate this bivariate time series using an ADCC model by Engle (2002a), presented in Section 3.4., which incorporates an asymmetric correlation effect. The univariate series are assumed to follow GJR-GARCH(1,1) models in order to

Table 1. Descriptive Statistics of the FTSE100 and S&P500 Log-Return Series

	FTSE100	S&P500
Mean	0.0112	0.0099
Median	0.0344	0.0779
Variance	1.7164	1.9617
Skewness	-0.0974	-0.3001
Kurtosis	10.5464	12.5674
Correlation	0.6060	

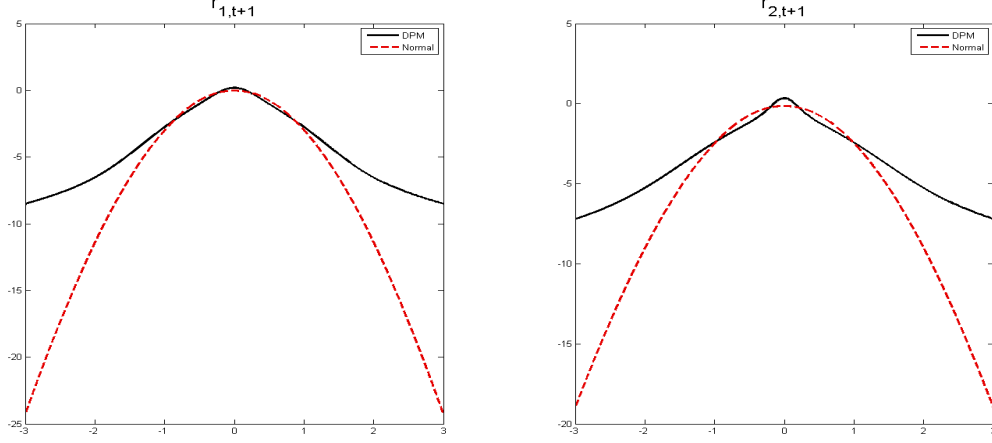
Table 2. Estimation Results for FTSE100 (subindex 1) and S&P500 (subindex 2) log-Returns, with 30,000 iterations plus 10,000 burn-in iterations

	ML Gaussian		Bayesian Gaussian		Bayesian DPM	
	Estimate	St. Dev.	Mean	95% CI	Mean	95% CI
ω_1	0.0166	0.0020	0.0192	(0.0130, 0.0258)	0.0181	(0.0104, 0.0264)
ω_2	0.0190	0.0016	0.0249	(0.0174, 0.0316)	0.0219	(0.0153, 0.0293)
α_1	$8.31 \cdot 10^{-8}$	$2.61 \cdot 10^{-4}$	0.0058	(0.0002, 0.0177)	0.0046	(0.0003, 0.0112)
α_2	$9.05 \cdot 10^{-9}$	$7.58 \cdot 10^{-5}$	0.0053	(0.0002, 0.0173)	0.0059	(0.0002, 0.0151)
β_1	0.9087	0.0045	0.9010	(0.8841, 0.9152)	0.8956	(0.8762, 0.9139)
β_2	0.9079	0.0050	0.8888	(0.8705, 0.9088)	0.8851	(0.8675, 0.9041)
ϕ_1	0.1535	0.0085	0.1587	(0.1351, 0.1871)	0.1586	(0.1057, 0.2089)
ϕ_2	0.1483	0.0092	0.1737	(0.1398, 0.2020)	0.1758	(0.1134, 0.2142)
κ	0.0075	0.0020	0.0071	(0.0014, 0.0145)	0.0095	(0.0040, 0.0156)
λ	0.9898	0.0029	0.9818	(0.9665, 0.9898)	0.9806	(0.9693, 0.9901)
δ	$5.50 \cdot 10^{-8}$	$1.33 \cdot 10^{-4}$	0.0076	(0.0002, 0.0153)	0.0039	(0.0001, 0.0114)

incorporate the leverage effect in volatilities. As for the errors, we use an infinite scale mixture of Gaussian distributions. Therefore, we call the final model GJR-ADCC-DPM. Inference and prediction is carried out using Bayesian non-parametric techniques, as seen in Virbickaite et al. (2013). The selection of the MGARCH specification is arbitrary and other models might work equally well. For the sake of comparison, we estimate a restricted GJR-ADCC-Gaussian model using Maximum Likelihood and Bayesian approaches. The estimation results are presented in Table 2.

The estimated parameters are very similar for all three approaches, except for α , the asymmetric correlation coefficient δ . Since α and δ are so close to zero, the ML has some

Figure 2. Log-Predictive Densities of the one-step-ahead Returns r_{t+1} for Bayesian Gaussian and DPM Models



trouble in estimating those parameters. Overall, the δ is small, indicating little evidence of asymmetrical behavior in correlations.

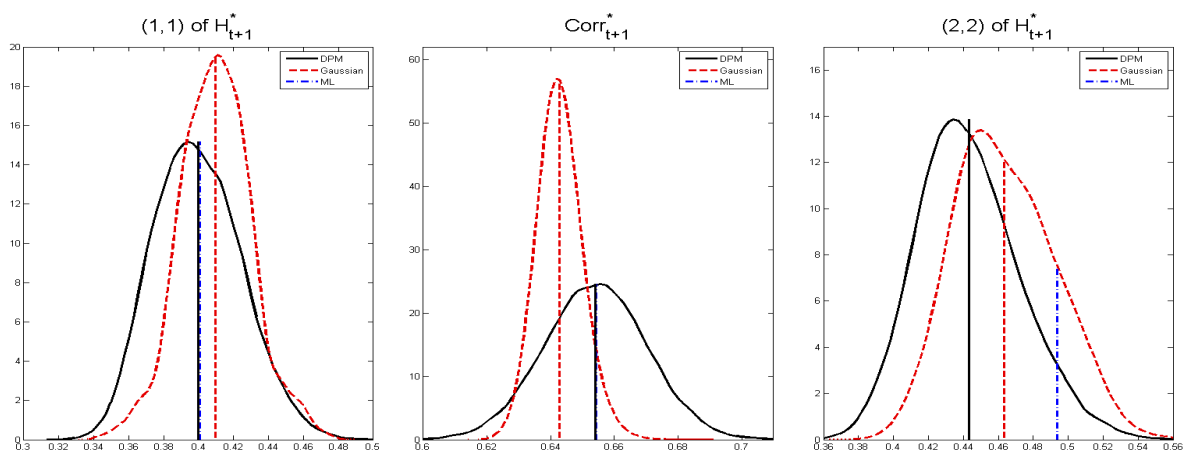
Figure 2 shows the estimated marginal predictive densities of the one-step-ahead returns in log scale using the Bayesian approach. We can observe the differences in tails arising from different specification of the errors. The DPM model allows for a more flexible distribution, therefore, for more extreme returns, i.e. fatter tails. The estimated densities were obtained using the procedure described in Virbickaite et al. (2013).

Table 3 presents the estimated mean, median and 95% credible intervals of one-step-ahead volatility matrices in Bayesian context. The matrix element (1,1) represents the volatility for the FTSE100 series, (2,2) for the S&P500, and the elements in the diagonal (1,2) and (2,1) represent the covariance of both financial returns. Figure 3 draws the posterior distributions for volatilities and correlation. The estimated mean volatilities for both, DPM and Gaussian approaches, are very similar, however, the main differences arise from the shape of the posterior distribution. 95% credible intervals for DPM model correlation are wider providing a more realistic measure of uncertainty about future correlations between two assets. This

Table 3. Estimated Means, Medians and 95% Credible Intervals of One-Step-Ahead Volatilities of FTSE100 and S&P500 log>Returns

	Constant	ML Gaussian	Bayesian Gaussian		Bayesian DPM	
			Mean Median	95% CI CI Length	Mean Median	95% CI CI Length
$H_{T+1}^{*(1,1)}$	1.7164	0.4007	0.4098 0.4099	(0.3681, 0.4538) 0.0857	0.3996 0.3983	(0.3550, 0.4512) 0.0962
$H_{T+1}^{*(1,2)}$	1.1120	0.2911	0.2800 0.2790	(0.2571, 0.3077) 0.0506	0.2751 0.2742	(0.2421, 0.3123) 0.0702
$H_{T+1}^{*(2,2)}$	1.9617	0.4939	0.4635 0.4606	(0.4159, 0.5193) 0.1034	0.4431 0.4408	(0.3912, 0.5059) 0.1146

Figure 3. Densities of One-Step-Ahead Volatilities of the Returns



is a very important implication in financial setting, because if an investor chooses to be Gaussian, she would be overconfident about her decision and unable to adequately measure the risk she is facing. See Virbickaite et al. (2013) for a more detailed comparison of DPM and alternative parametric approaches in portfolio decision problems.

To sum up, this illustration has shown the main differences between the standard estimation procedures and the new non-parametric approach. Even though the point estimates for the parameters and the one-step-ahead volatilities are very similar, the main differences arise from the thickness of tails of predictive distributions of one-step-ahead returns and the

shape of the posterior distribution for the one-step-ahead volatilities.

6 Conclusions

In this paper we reviewed univariate and multivariate GARCH models and inference methods, putting emphasis on the Bayesian approach. We have surveyed the existing literature that concerns various Bayesian inference methods for MGARCH models, outlining the advantages of the Bayesian approach versus the classical procedures. We have also discussed in more detail the recent Bayesian non-parametric method for GARCH models, which avoid imposing arbitrary parametric distributional assumptions. This new approach is more flexible and can describe better the uncertainty about future volatilities and returns, as has been illustrated using real data.

Acknowledgements

We are grateful to an anonymous referee for helpful comments. The first and second authors are grateful for the financial support from MEC grant ECO2011-25706. The third author acknowledges financial support from MEC grant ECO2012-38442.

References

- Alexander, C. O., & Chibumba, A. M. (1997). Multivariate Orthogonal Factor GARCH. *Mimeo, University of Sussex*, .
- Antoniak, C. E. (1974). Mixtures of Dirichlet Processes with Applications to Bayesian Non-parametric Problems. *Annals of Statistics*, 2, 1152–1174. doi:10.1214/aos/1176342871.

- Arakelian, V., & Dellaportas, P. (2012). Contagion Determination via Copula and Volatility Threshold Models. *Quantitative Finance*, 12, 295–310. doi:10.1080/14697680903410023.
- Ardia, D. (2006). Bayesian Estimation of the GARCH (1,1) Model with Normal Innovations. *Student*, 5, 1–13.
- Ardia, D., & Hoogerheide, L. F. (2010). Efficient Bayesian estimation and combination of GARCH-type models. In K. Bocker (Ed.), *Rethinking Risk Measurement and Reporting: Examples and Applications from Finance: Vol II* chapter 1. (pp. 1–22). London: RiskBooks.
- Asai, M. (2006). Comparison of MCMC Methods for Estimating GARCH Models. *Journal of the Japan Statistical Society*, 36, 199–212. doi:10.14490/jjss.36.199.
- Ausín, M. C., & Galeano, P. (2007). Bayesian Estimation of the Gaussian Mixture GARCH Model. *Computational Statistics & Data Analysis*, 51, 2636–2652. doi:10.1016/j.csda.2006.01.006.
- Ausín, M. C., Galeano, P., & Ghosh, P. (2014). A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *European Journal of Operational Research*, 232, 350–358. doi:10.1016/j.ejor.2013.07.008.
- Ausín, M. C., & Lopes, H. F. (2010). Time-Varying Joint Distribution Through Copulas. *Computational Statistics & Data Analysis*, 54, 2383–2399.
- Bai, X., Russell, J. R., & Tiao, G. C. (2003). Kurtosis of GARCH and Stochastic Volatility Models with Non-Normal Innovations. *Journal of Econometrics*, 114, 349–360.
- Bauwens, L., Laurent, S., & Rombouts, J. V. K. (2006). Multivariate GARCH Models: a Survey. *Journal of Applied Econometrics*, 21, 79–109.

- Bauwens, L., & Lubrano, M. (1998). Bayesian Inference on GARCH Models Using the Gibbs Sampler. *Econometrics Journal*, 1, 23–46.
- Bauwens, L., & Lubrano, M. (2002). Bayesian Option Pricing Using Asymmetric GARCH Models. *Journal of Empirical Finance*, 9, 321–342.
- Bera, A. K., & Higgins, M. L. (1993). ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, 7, 305–362.
- Black, F. (1976). Studies of Stock Market Volatility Changes. *Proceedings of the American Statistical Association; Business and Economic Statistics Section*, (pp. 177–181).
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307 – 327.
- Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, 69, 542–547.
- Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: a Multivariate Generalized ARCH Model. *The Review of Economics and Statistics*, 72, 498–505.
- Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence. *Journal of Econometrics*, 52, 5–59.
- Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH Models. In R. F. Engle, & D. McFadden (Eds.), *Handbook of Econometrics*, Vol. 4 November (pp. 2959–3038). Amsterdam: Elsevier.
- Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy*, 96, 116–131.

- Burda, M. (2013). Constrained Hamiltonian Monte Carlo in BEKK GARCH with Targeting. *Working Paper, University of Toronto*, .
- Burda, M., & Maheu, J. M. (2013). Bayesian Adaptively Updated Hamiltonian Monte Carlo with an Application to High-Dimensional BEKK GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, *17*, 345–372. doi:10.1515/snde-2013-0020.
- Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, *4*, 537–572.
- Chen, C. W. S., Gerlach, R., & So, M. K. P. (2009). Bayesian Model Selection for Heteroskedastic Models. *Advances in Econometrics*, *23*, 567–594. doi:10.1016/S0731-9053(08)23018-5.
- Dellaportas, P., & Vrontos, I. D. (2007). Modelling Volatility Asymmetries: a Bayesian Analysis of a Class of Tree Structured Multivariate GARCH Models. *The Econometrics Journal*, *10*, 503–520.
- Ding, Z., & Engle, R. F. (2001). Large Scale Conditional Covariance Matrix Modeling , Estimation and Testing. *Academia Economic Papers*, *29*, 157–184.
- Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*, 987–1008.
- Engle, R. F. (2002a). Dynamic Conditional Correlation. *Journal of Business & Economic Statistics*, *20*, 339–350.
- Engle, R. F. (2002b). New Frontiers for ARCH Models. *Journal of Applied Econometrics*, *17*, 425–446.
- Engle, R. F. (2004). Risk and Volatility : Econometric Models and Financial Practice. *The American Economic Review*, *94*, 405–420.

- Engle, R. F., & Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11, 122–150.
- Engle, R. F., Ng, V. K., & Rothschild, M. (1990). Asset Pricing With a Factor-ARCH Covariance Structure. *Journal of Econometrics*, 45, 213–237.
- Escobar, M. D., & West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, 90, 577–588.
- Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *Annals of Statistics*, 1, 209–230.
- Galeano, P., & Ausín, M. C. (2010). The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter Estimation, Value at Risk Calculation, and Portfolio Selection. *Journal of Business and Economic Statistics*, 28, 559–571.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics*, 4, 169–193.
- Giannikis, D., Vrontos, I. D., & Dellaportas, P. (2008). Modelling Nonlinearities and Heavy Tails via Threshold Normal Mixture GARCH Models. *Computational Statistics & Data Analysis*, 52, 1549–1571.
- Gilks, W. R., Best, N. G., & Tan, K. K. C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Journal of the Royal Statistical Society Series C*, 44, 455–472.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48, 1779–1801.

- 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.
- Hall, P., & Yao, Q. (2003). Inference in ARCH and GARCH Models with Heavy-Tailed Errors. *Econometrica*, *71*, 285–317.
- He, C., & Teräsvirta, T. (1999). Properties of Moments of a Family of GARCH Processes. *Journal of Econometrics*, *92*, 173–192.
- Hofmann, M., & Czado, C. (2010). Assessing the VaR of a portfolio using D-vine copula based multivariate GARCH models. *Working Paper, Technische Universität München Zentrum Mathematik.*, .
- Hudson, B. G., & Gerlach, R. H. (2008). A Bayesian approach to relaxing parameter restrictions in multivariate GARCH models. *Test*, *17*, 606–627.
- Ishida, I., & Engle, R. F. (2002). Modeling Variance of Variance: the Square-Root, the Affine, and the CEV GARCH Models. *Working Paper, New York University*, (pp. 1–47).
- Jacquier, E., & Polson, N. G. (2013). Asset Allocation in Finance: A Bayesian Perspective. In P. Demian, P. Dellaportas, N. G. Polson, & D. A. . Stephen (Eds.), *Bayesian Theory and Applications* chapter 25. (pp. 501–516). Oxford: Oxford University Press. (1st ed.).
- Jensen, M. J., & Maheu, J. M. (2013). Bayesian Semiparametric Multivariate GARCH Modeling. *Journal of Econometrics*, *176*, 3–17.
- Jin, X., & Maheu, J. M. (2013). Modeling Realized Covariances and Returns. *Journal of Financial Econometrics*, *11*, 335–369.
- Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis*, *21*, 279–292.

- Kang, L. (2011). Asset allocation in a Bayesian Copula-GARCH Framework: An Application to the "Passive Funds Versus Active Funds" Problem. *Journal of Asset Management*, 12, 45–66.
- Kim, S., Shephard, N., & Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65, 361–393.
- Lin, W.-L. (1992). Alternative Estimators for Factor GARCH Models - a Monte Carlo Comparison. *Journal of Applied Econometrics*, 7, 259–279.
- MacEachern, S. N. (1994). Estimating Normal Means with a Conjugate Style Dirichlet Process Prior. *Communications in Statistics - Simulation and Computation*, 23, 727–741.
- Miazhyńska, T., & Dorffner, G. (2006). A Comparison of Bayesian Model Selection Based on MCMC with Application to GARCH-type Models. *Statistical Papers*, 47, 525–549.
- Müller, P., & Pole, A. (1998). Monte Carlo Posterior Integration in GARCH Models. *Sankhya, Series B*, 60, 127–14.
- Nakatsuma, T. (2000). Bayesian Analysis of ARMA- GARCH Models: a Markov Chain Sampling Approach. *Journal of Econometrics*, 95, 57–69.
- Neal, R. M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9, 249–265.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: a New Approach. *Econometrica*, 59, 347–370.
- Osiewalski, J., & Pipien, M. (2004). Bayesian Comparison of Bivariate ARCH - Type Models for the Main Exchange Rates in Poland. *Journal of Econometrics*, 123, 371–391.

- Papaspiliopoulos, O. (2008). A Note on Posterior Sampling from Dirichlet Mixture Models. *Working Paper, University of Warwick. Centre for Research in Statistical Methodology, Coventry*, (pp. 1–8).
- Papaspiliopoulos, O., & Roberts, G. O. (2008). Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models. *Biometrika*, *95*, 169–186.
- Patton, A. J. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, *47*, 527–556.
- Patton, A. J. (2009). Copula - Based Models for Financial Time Series. In T. G. Andersen, T. Mikosch, J.-P. Kreiß, & R. A. Davis (Eds.), *Handbook of Financial Time Series* chapter 36. (pp. 767–785). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods*. (2nd ed.). Springer Texts in Statistics.
- Silvennoinen, A., & Teräsvirta, T. (2009). Multivariate GARCH models. In T. G. Andersen, T. Mikosch, J.-P. Kreiß, & R. A. Davis (Eds.), *Handbook of Financial Time Series* (pp. 201–229). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Sklar, A. (1959). Fonctions de répartition à n Dimensions et Leur Marges. *Publ. Inst. Statist. Univ. Paris*, *8*, 229–231.
- Steel, M. (2008). Bayesian Time Series Analysis. In S. Durlauf, & L. Blume (Eds.), *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan. (2nd ed.).
- Teräsvirta, T. (2009). An Introduction to Univariate GARCH Models. In T. G. Andersen, T. Mikosch, J.-P. Kreiß, & R. A. Davis (Eds.), *Handbook of Financial Time Series* (pp. 17–42). Berlin, Heidelberg: Springer Berlin Heidelberg.

- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22, 1701–1728.
- Tsay, R. S. (2010). *Analysis of Financial Time Series*. (3rd ed.). Hoboken: John Wiley & Sons, Inc.
- Tse, Y. K., & Tsui, A. K. C. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations. *Journal of Business & Economic Statistics*, 20, 351–362. doi:10.1198/073500102288618496.
- Van der Weide, R. (2002). GO-GARCH: a Multivariate Generalized Orthogonal GARCH Model. *Journal of Applied Econometrics*, 17, 549–564.
- Virbickaite, A., Ausín, M. C., & Galeano, P. (2013). A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model with Application to Portfolio Selection. *Working paper, arXiv:1301.5129v2*, .
- Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2000). Full Bayesian Inference for GARCH and EGARCH Models. *Journal of Business and Economic Statistics*, 18, 187–198. doi:10.2307/1392556.
- Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2003a). A Full-Factor Multivariate GARCH Model. *Econometrics Journal*, 6, 312–334. doi:10.1111/1368-423X.t01-1-00111.
- Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2003b). Inference for Some Multivariate ARCH and GARCH Models. *Journal of Forecasting*, 22, 427–446. doi:InferenceforSomeMultivariateARCHandGARCHModels.
- Wago, H. (2004). Bayesian Estimation of Smooth Transition GARCH Model Using Gibbs Sampling. *Mathematics and Computers in Simulation*, 64, 63–78. doi:10.1016/S0378-4754(03)00121-6.

- Walker, S. G. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation*, 36, 45–54. doi:10.1080/03610910601096262.
- Yu, B., & Mykland, P. (1998). Looking at Markov Samplers Through Cusum Path Plots: a Simple Diagnostic Idea. *Statistics and Computing*, 8, 275–286.
- Zakoian, J.-M. (1994). Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control*, 18, 931–955. doi:10.1016/0165-1889(94)90039-6.