Measuring uncertainty and disagreement about uncertainty from a large set of model predictions☆

David Ardia[a], Arnaud Dufays[b,c,*]

[a]*Department of Decision Sciences, HEC Montréal, Canada*

[b]*Information Systems, Decision Sciences and Statistics (IDS) Department, Essec Business School, France*

[c]*Department of Economics, Laval University, Canada*

IIF-SAS AWARD 2018
– FINAL REPORT –

Abstract

We construct measures of uncertainty and disagreement about uncertainty exploiting the heterogeneity of a large set of model predictions. The approach is forward-looking, can be computed in real-time, and can be applied at any frequency. We illustrate the methodology with expected shortfall predictions of worldwide equity indices generated from 71 risk models. We use the new measures in asset pricing, risk forecasting, and for explaining the aggregate trading volume of S&P 500 firms.

Keywords: Uncertainty, Uncertainty disagreement, Risk predictions, Model risk, Heterogeneity, Mixture models.

☆We acknowledge IIF/SAS (`https://forecasters.org`) for their financial support.

*Corresponding author. 2216 Pavillon J.-A.-DeSève, G1V 0A6, Quebec City, Canada. *Phone:* +1(418)6562131 - 7749.

Email addresses: `david.ardia@hec.ca` (David Ardia), `dufays@essec.edu` (Arnaud Dufays)

1. Introduction

Uncertainty is typically related to forecast uncertainty (e.g., Engle, 1983). When uncertainty increases, we expect that the predictive distribution of future outcomes becomes more diffuse. On this basis, many different estimators related to the variance of the predictive distribution have been used to measure uncertainty. Popular estimators based on the second moment are the volatility index (VIX, Bloom, 2009), the variance risk premium (VRP, Bollerslev et al., 2009; Conrad and Loch, 2015), the cross-sectional variance of returns (CSV, Bloom, 2009; Christiano et al., 2014), forecaster's disagreement (FD, Bomberger, 1996; Lahiri and Sheng, 2010) and the measure of Jurado et al. (2015). Each of these estimators is based on different information sets. The VIX and the VRP use options data. CSV is computed from a large panel of financial returns, while FD requires experts' predictions. Jurado et al. (2015) define uncertainty as the future conditional variance of the true data generating process (DGP). To be as close as the information set of the true DGP, they rely on a massive macroeconomic database to filter out the conditional expectation and compute the second moment (see also Carriero et al., 2018). In this paper, we use a large panel of series and model predictions to build a measure of uncertainty and uncertainty disagreement. Our two measures can be computed in any situation where multiple models exist to produce forecast uncertainty of the same series.

Uncertainty disagreement arises when forecasters disagree about the future level of uncertainty. In a period of disagreement, the market offers better prospects for investors as it increases their perceived trading profits (Gao et al., 2019). Multiple evidence exists about the impact of uncertainty on financial markets (Segal et al., 2015) and macroeconomic fundamentals (Bloom et al., 2018). Disagreement on the uncertainty level is one of the channels through which uncertainty could lead to such impacts (Carlin et al., 2014). We show that uncertainty disagreement increases in turbulent periods like any uncertainty measures.

Our framework relates the model risk literature to the literature on uncertainty. On the one hand, our uncertainty disagreement measure boils down to a model risk measure close to the model risk measure of Danielsson et al. (2016) when applied to a single time series. On the other hand, our uncertainty measure is similar to Jurado et al. (2015) when the measure is built upon many time series but with one model. In particular, given N time series, their h-step ahead uncertainty measure is defined as:

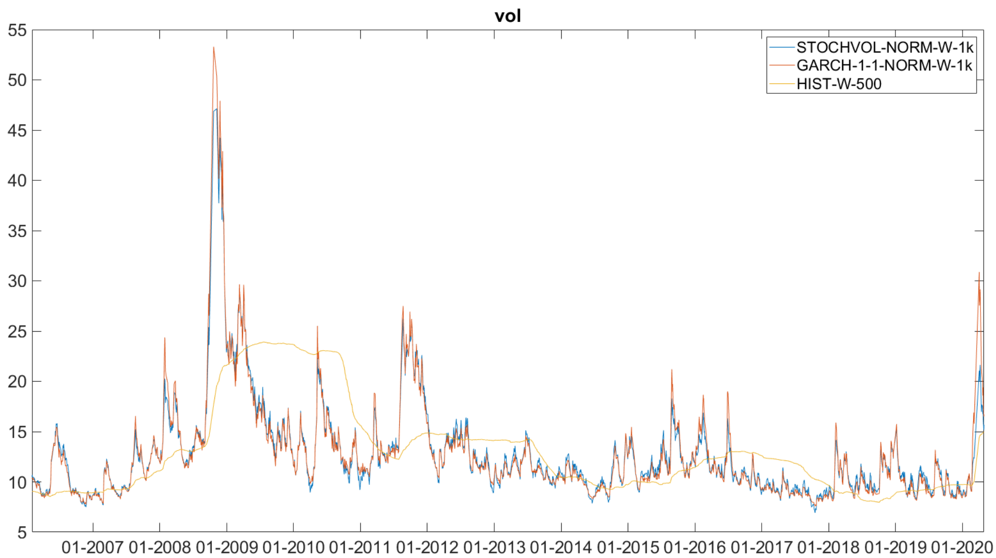
$$U_t(h) \equiv \frac{1}{N}\sum_{i=1}^{N} U_{t,i}(h)\,, \tag{1}$$

$$U_{t,i}(h) \equiv \sqrt{V[y_{t+h}^{(i)}|I_t]}\,, \tag{2}$$

where $i = 1, \ldots, N$ stands for the ith series and $V[y_{t+h}^{(i)}|I_t]$ denotes the h-step ahead conditional variance given the information set I_t up to time t. Jurado et al. (2015) use a flexible factor model with stochastic volatility to estimate (2). By doing so, their uncertainty measure is based on one volatility model estimated on multiple time series, but the uncertainty coming from other models is overlooked.[1] Figure 1 shows with the ten-step ahead uncertainty measure estimated using three different volatility models that the uncertainty dynamic can significantly differ among models. Recall that five new volatility models extending the GARCH equation are published on average every year, not mentioning volatility models with other types of innovation distributions and of dynamics such as the stochastic volatility (Taylor, 1994) or the multi-fractal (Calvet and Fisher, 2004).[2] Eventually and importantly, a two-step strategy consisting of estimating the uncertainty measure from one model and then using it for testing its impact on macro variables suffers from measurement errors (Carriero et al., 2018).

Figure 1: Measure of uncertainty for three risk models

This figure displays the daily evolution of uncertainty, expressed as the annualized volatility, captured by three risk models: 500-day historical standard deviation, GARCH(1,1) with normal innovations, and stochastic volatility with normal innovations. Models are estimated on rolling windows (of size 500 for the historical approach and 1,000 for the two other models).



To account for model risks, we could extend the measure of Jurado et al. (2015) as follows:

$$U_t(h, M) \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M} U_{t,i}(h, m), \tag{3}$$

$$U_{t,i}(h, m) \equiv \sqrt{V[y_{t+h}^{(i)}|I_t, m]}, \tag{4}$$

where $V[y_{t+h}^{(i)}|I_t, m]$ denotes the h-step ahead conditional variance given by model $m = 1, \ldots, M$. While this approach could limit the measurement errors, this new uncertainty measure is still

[1]The conditional variance is given by $V[y_{t+h}^{(i)}|I_t] = E[y_{t+h}^2|I_t] - E[y_{t+h}|I_t]$. Jurado et al. (2015) insist on estimating the conditional expectation $E[y_{t+h}|I_t]$ as accurately as possible; hence the use of a factor model. However, they do not use the most flexible volatility model for approaching $E[y_{t+h}^2|I_t]$.

[2]The glossary of Bollerslev (2010) lists over 140 volatility models solely based on the GARCH equation.

unsatisfactory because models contribute uniformly to the measure. If the panel of models is unbalanced, this could create a bias towards the most representative type of models. Also, volatility models are not equivalent as they differ in their performance in capturing financial stylized facts such as leverage effects (Glosten et al., 1993), skewed and fat tails (Verhoeven and McAleer, 2004), and structural breaks (Bauwens et al., 2013). We solve this issue by using a clustering technique. By doing so, our uncertainty measure exhibits time-varying model weights (instead of uniform weights as in (3)). Our measure of uncertainty disagreement is estimated from the heterogeneity of model predictions uncovered by the mixture model. Supported by empirical evidence (Carlin et al., 2014) and by the theoretical framework of Lahiri and Sheng (2010), we view each volatility model as a forecaster. The uncertainty disagreement measure is thus constructed like a measure of forecasters' disagreement. As many models provide similar predictions, our clustering technique helps group the model predictions. This way, only models (or forecasters) that provide different predictions contribute to the uncertainty disagreement.

Our framework relies on a clustering algorithm. While several methods could be used to group the model predictions, we use a recent mixture model proposed by Frühwirth-Schnatter et al. (2020). Instead of using a deterministic approach such as the K-means algorithm, mixture models have the advantage of clustering the data in a probabilistic way. However, standard mixture models require fixing the number of clusters. In practice, the number of groups is chosen by estimating and comparing $\bar{K}$ different mixture models (e.g., Chib, 1996) which is computationally costly. To tackle this problem, advanced mixture models either rely on Dirichlet processes (Lin et al., 2018), on Dirichlet distributions with shrinkage priors (Malsiner-Walli et al., 2016; Frühwirth-Schnatter and Malsiner-Walli, 2019) or on mixture of finite mixture (MFM) models to jointly estimate the model parameters and the number of clusters. The MFM approach encompasses the mixture processes based on a Dirichlet process and shrinkage priors. The reversible-jump MCMC algorithm has been the standard tool for estimating MFM models over the last decades (Green, 1995) but its computational cost has limited its use. Recently, Miller and Harrison (2018) have proposed a new MCMC scheme based on partitions that avoid the birth and death move of the reversible-jump algorithm. This breakthrough highly simplifies the estimation of MFM processes. Extending Miller and Harrison (2018), Frühwirth-Schnatter et al. (2020) introduce the dynamic MFM that we adapt for our clustering analysis.

Our measures exhibit several appealing features. As an average of predictions, our uncertainty measure is more robust to measurement errors than standard approaches based on a single volatility model. Second, our two measures are forward-looking because they are based on predictions, and they can be updated in real-time. As a final advantage, our two measures can be applied to other series such as macroeconomic variables or exchange rates. In particular, standard uncertainty measures require extra data such as options data for the VIX and big macroeconomic data for the measures of Jurado et al. (2015) and of Carriero et al. (2018). Our framework relies uniquely on computational resources. Like business expectation surveys (Altig et al., 2020b), our method could allow for rapid deployment of specific questions that target current developments and policy issues.

We compute the two measures based on eleven worldwide daily equity indices spanning from 1995 to 2021. We show that they spike at well-known uncertainty events such as the global financial crisis and the pandemic, and share commonalities with standard uncertainty measures. We investigate their asset pricing implication and show that stocks in the S&P 500 index are significantly exposed to the measures. In particular, using a Fama-French five-factor model, 20% expected returns of the financial returns are affected by our uncertainty measures even when controlling for other uncertainty variables. Also, in a predictive exercise, we highlight that our measures help predict daily realized variances. Finally, we investigate the link between uncertainty disagreement and trading volume. Carlin et al. (2014) put forward a positive impact of disagreement on future trading volumes of financial assets on a monthly basis. We show that our uncertainty disagreement measure also positively affects future trading volume at a daily frequency.

The paper is organized as follows. In Section 2, we motivate our two measures using empirical and theoretical facts. Section 3 introduces the clustering method to assess the model predictions' heterogeneity. We also detail the computation of the two measures. Section 4 is dedicated to the empirical analysis. In particular, subsection 4.1 documents the data and the models. Then, we study the heterogeneity of the model predictions in subsection 4.2. The dynamics of the two measures are analyzed and compared to existing alternatives in subsection 4.3. Section 5 provides empirical evidence of the two measures' relevance in asset pricing, risk forecasting, and in explaining the aggregate trading volume of S&P 500 firms. Section 6 concludes.

2. Model predictions and its use in measuring uncertainty

We build our uncertainty measures on expected shortfalls produced by 71 different volatility models (see section 4.1 for more information about the models). In this section, we discuss empirically and theoretically our motivations to use this database for measuring uncertainty.

2.1. Empirical facts

Over the last four decades, numerous volatility models have been proposed to capture stylized facts observed in financial returns. Each new volatility model is motivated by its ability to fit the dynamic of returns better. From a historical perspective, we can classify the volatility models into families capturing different stylized facts as follows:[3]

- **Volatility clustering.** Example: Historical variances, EWMA, GARCH and standard stochastic volatility models.

[3]We do not provide an exhaustive list of the volatility models and model families. The list is presented to illustrate the relevance of clustering the predictions.

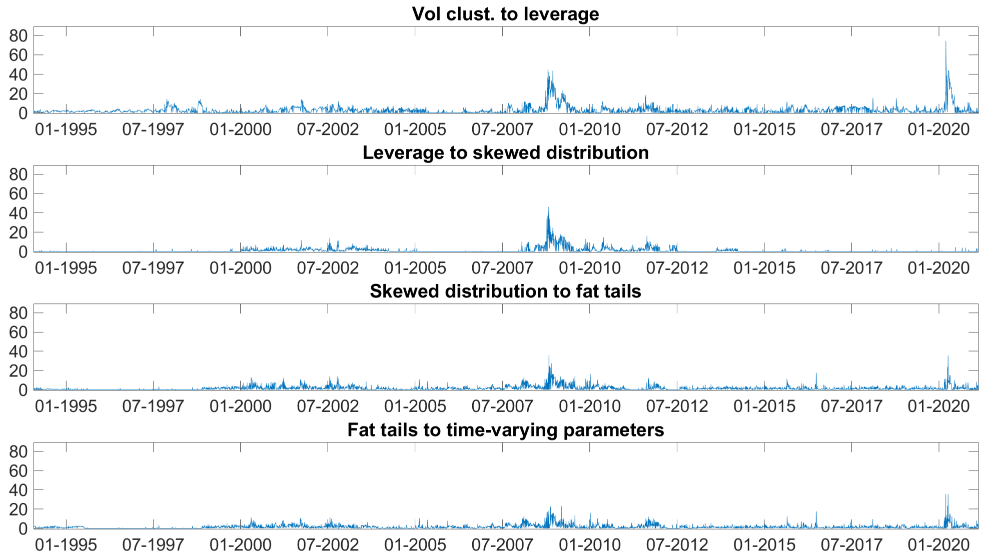
- **Leverage effect.** Example: GJR-GARCH and EGARCH models.
- **Asymmetry.** Example: GJR-GARCH and EGARCH models with skewed distributions.
- **Fat tails.** Example: GJR-GARCH and EGARCH models with fat tail distributions.
- **Structural breaks.** Example: Markov-switching (MS)-GARCH models.

We construct our two measures from the heterogeneity of these model families.

Measure of uncertainty. Generally speaking, we expect a bias in the estimation of the conditional variance when we use a standard GARCH model on returns exhibiting leverage effect (Glosten et al., 1993; Harvey and Sucarrat, 2014). However, a leverage model such as the GJR-GARCH process should unbiasedly estimate the conditional variance of returns exhibiting only volatility clustering. Logically, this argument extends to more flexible classes of models. Figure 2 show the absolute differences between the expected shortfall estimates of two consecutive volatility families. Clearly, we need a highly flexible volatility model to estimate uncertainty over time without bias.

Measure of uncertainty disagreement. Interestingly, Figure 2 highlights that volatility disagreement differs from class to class and also increases during turbulent periods. This is an evidence that agitated periods can be captured from the heterogeneity of the model predictions. This disagreement could potentially be the primary channel explaining why uncertainty is so damaging to the economy. Note that uncertainty disagreement does not depend on the level of uncertainty since it can be close to zero in periods of small and large uncertainties.

Figure 2: (absolute) Bias in the conditional variance from one family of volatility models to another This figure displays the absolute difference of ten-step ahead expected shortfalls between two classes of volatility models. The ten-step ahead expected shortfall is estimated as the sum over the eleven financial indices. For each model class, the expected shortfall is computed as follows. We first choose a benchmark model representing the volatility family. Then, the expected shortfall of the family at each time period is computed as the average over the expected shortfalls belonging to the same cluster as the benchmark model in which the clusters are inferred using our clustering method. We use the following benchmark models: **Volatility clustering:** EWMA-94 with Gaussian distribution, **Leverage effect:** GJR-GARCH with Gaussian distribution, **Asymmetry:** GJR-GARCH with skewed Gaussian distribution, **Fat tails:** GJR-GARCH with skewed Student-t distribution, **Structural breaks:** Markov-switching GJR-GARCH with skewed Student-t distribution.



2.2. Theoretical framework

Using a simple theoretical framework, we now motivate our two measures of uncertainty. Let us denote by $U_{t,h}$ the true uncertainty at time t over horizon h.[4] We get estimates of this uncertainty from M distinct models $\{\hat{U}_{m,t,h}\}_{m=1}^{M}$. The forecast error of model m is given by:

$$e_{m,t,h} \equiv U_{t,h} - \hat{U}_{m,t,h}\,. \tag{5}$$

As explained in section 2.1, volatility models are not equivalent, as they belong to different volatility families. Holding with this idea, we cannot assume unbiased forecast errors. We denote the bias and the variance forecast errors as $E[e_{m,t,h}] \equiv a_{m,t,h}$ and $V[e_{m,t,h}] \equiv \theta^2_{m,t,h}$, respectively.

Our first measure of uncertainty simply consists in averaging out the uncertainty estimates,

[4]In this paper, $U_{t,h}$ is understood as the expected shortfall at horizon h of the true data generating process as we believe that an uncertainty measure should depend on the full predictive distribution and not only on its second moment. Nevertheless, the framework also operates with the conditional variance.

that is $\bar{U}_{t,h} \equiv \sum_{m=1}^{M} \omega_m \hat{U}_{m,t,h}$ where $\sum_{m=1}^{M} \omega_m = 1$ and $\omega_m \geq 0$. One advantage of this average is that it limits the measurement error put forward by Carriero et al. (2018). In our framework, the sample average is obviously biased due to the model heterogeneity. We make the following assumption.

Assumption A.1. The bias $a_{m,t,h}$ decreases with the complexity of the volatility models and tends to zero for the most complex volatility models.

Assumption A.1 leads to multiple uncertainty estimators. The most straightforward estimator relies on one model (the most flexible one) and gives zero weights to the other models. This is the standard approach when uncertainty is estimated from volatility models (e.g., Jurado et al., 2015). To get a more efficient estimator, we propose to estimate uncertainty as the sample average of the uncertainty estimates of models exhibiting the same complexity. Since the model complexity is not observable, we rely on a clustering method to cluster the uncertainty estimates on a day-to-day basis. We then average the uncertainty estimates of the model belonging to the same cluster as a benchmark model (that is the MS-GJR-GARCH model with skewed Student-t distribution in our empirical exercise).

Our measure of uncertainty disagreement is equivalent to a forecaster's disagreement measure. In particular, forecast disagreement is typically estimated by the variance of the predictions (Bomberger, 1996):

$$\mathrm{UD}_{t,h} \equiv \frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{U}_{m,t,h} - \frac{1}{M} \sum_{j=1}^{M} \hat{U}_{j,t,h} \right)^2 = \frac{1}{M-1} \sum_{m=1}^{M} (e_{m,t,h} - \bar{e}_{t,h})^2 . \tag{6}$$

To illustrate our second measure, we make the following assumption.

Assumption A.2. The forecast error of two volatility models is not correlated: $E[e_{jth} e_{ith}] = 0$ for $i \neq j$.

While Assumption A.2 is not realistic for model predictions among the same volatility family, this assumption is more credible for predictions between families that are our focus when we estimate our uncertainty measure. Taking the expectation of the uncertainty disagreement leads to:

$$E[\mathrm{UD}_{t,h}] = \frac{1}{M-1} \sum_{m=1}^{M} E[(e_{m,t,h} - \bar{e}_{t,h})^2] = \frac{1}{M} \sum_{m=1}^{M} \left(\theta_{m,t,h}^2 + a_{m,t,h}^2 \right) . \tag{7}$$

Note that $a_{m,t,h}$ contributes positively to the model disagreement regardless of the bias's sign. In addition, more models will be biased in turbulent periods because high-volatile periods exacerbate the impact of the financial stylized facts (such as the leverage effect, asymmetries, fat tails, etc.) on the conditional variance. We conclude that uncertainty disagreement inflates during distress periods like a measure of uncertainty. However, this increase is related to different model "opinions" on the uncertainty level based on the same information set (Carlin et al., 2014).

To make assumption A.2 realistic, we cannot directly compute the variance of the uncertainty estimates produced by all volatility models. Using a clustering approach, we shall compute the uncertainty disagreement given by (6) over the uncertainty estimates from all volatility families.

2.3. Uncertainty measures from clustering algorithms

In Section 3, we lay down the details of our uncertainty measures based on the mixture of finite mixture models. Before going into that details, we stress that the measures can be estimated using any clustering approach. In particular, a standard classification algorithm provides the cluster partition $\boldsymbol{s} \equiv (s_1, \ldots, s_M)' \in [0, K]^M$ and the centroid location $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_K)' \in \mathbb{R}^K$ where K denotes the number of clusters. Only these two quantities are needed to compute our measures. Assuming without loss of generality that our most complex model is the first model in the model set, the two measures at time t are given by:

$$\bar{U}_{t,h} \equiv \frac{1}{\sum_{m=1}^{M} \mathbb{1}_{s_m = s_1}} \sum_{m=1}^{M} \mathbb{1}_{s_m = s_1} \hat{U}_{m,t,h}\,, \tag{8}$$

$$\bar{UD}_{t,h} \equiv \frac{1}{K-1} \sum_{k=1}^{K} (\mu_k - \bar{\mu})^2, \text{ with } \bar{\mu} \equiv \frac{1}{K} \sum_{k=1}^{K} \mu\,, \tag{9}$$

where $\mathbb{1}_{\bullet}$ denotes the indicator function which is equal to one if the condition is satisfied.

3. Mixture of finite mixture models and uncertainty measures

We use an unsupervised clustering technique to assess the heterogeneity in the model set's predictions. The MFM approach differentiates the number of possible clusters, K, to the number of active clusters, K^+, and sets a prior on these two quantities. At time t, we collect the predictions (e.g., volatilities or expected shortfalls) generated by the M models for the N series into matrix $\underset{\sim}{\mathrm{y}}^{(t)} \equiv (\underset{\sim}{\mathrm{y}}_1^{(t)}, \ldots, \underset{\sim}{\mathrm{y}}_M^{(t)}) \in \mathbb{R}^{N\times M}$, where $\underset{\sim}{\mathrm{y}}_m^{(t)} \equiv (y_{1,m}^{(t)}, \ldots y_{N,m}^{(t)})'$ is a vector containing the N series predictions of model m for time t. We assume that $N < M$. For each time period $t = 1, \ldots, T$, the MFM model is specified as:

$$\begin{aligned}
K &\sim \text{Beta-Negative-Binomial}(1, \underline{a}_k, \underline{b}_k)\,, \\
\boldsymbol{\mu}_k | K &\sim N(\boldsymbol{b}_0, \boldsymbol{R}) \quad \text{for } k = 1, \ldots, K\,, \\
\boldsymbol{\Sigma}_k | K &\sim \text{Inverse-Wishart}(\boldsymbol{V}, \underline{v}) \quad \text{for } k = 1, \ldots, K\,, \\
\underset{\sim}{\mathrm{p}}_{1:K} \equiv (p_1, \ldots, p_K)' | K &\sim \text{Dirichlet}\left(\frac{\alpha_{\text{Dir}}}{K}, \ldots, \frac{\alpha_{\text{Dir}}}{K}\right)\,, \\
s_m | \underset{\sim}{\mathrm{p}}_{1:K}, K &\sim \text{Multinomial}(\underset{\sim}{\mathrm{p}}_{1:K}) \quad \text{for } m = 1, \ldots, M\,, \\
\underset{\sim}{\mathrm{y}}_m^{(t)} | s_m &\sim N(\boldsymbol{\mu}_{s_m}, \boldsymbol{\Sigma}_{s_m})\,, \text{ with } s_m \in [1, K]\,.
\end{aligned} \tag{10}$$

To facilitate the birth of new clusters, we follow Song (2014) and assume hierarchical distributions on the mean and the variance of the parameters. In particular, we use the structure advocated by Malsiner-Walli et al. (2016):

$$\begin{aligned} \boldsymbol{b}_0 &\sim N(\text{median}(\underline{\mathrm{y}}^{(t)}), 100 I_k)\,, \\ \boldsymbol{R} &\sim \text{Inverse-Wishart}(\boldsymbol{R}_0, r_0)\,, \\ \boldsymbol{V} &\sim \text{Wishart}(S_0, s_0)\,, \end{aligned} \tag{11}$$

in which $\boldsymbol{R}_0 \equiv \text{diag}(V[\underline{\mathrm{y}}^{(t)}])(r_0 - N - 1)$, $r_0 \equiv N + 5$, $v_0 = N + 5$ and $S_0 \equiv \left(\frac{I_N}{1000} + V[\underline{\mathrm{y}}^{(t)}]\right)(v_0 - N - 1)$.

Importantly, the MFM model sets a prior on the number K of clusters. Nevertheless, note that the number K^+ of non-empty clusters can be smaller than K. The MFM model in (10)-(11) allows us to infer the number of clusters at a given time t but does not account for time-persistence in the number of clusters.[5] By doing so, we perform the estimation for all periods in parallel. This way, we easily handle a long time-span. It also makes the method feasible with unbalanced panels of model predictions. Moreover, when a new set of predictions is available for a given time period, we can estimate the model for that period. We use a Markov chain Monte Carlo (MCMC) algorithm to estimate the model parameters. For each time period, we run 10,000 MCMC iterations and keep the last 4,000 draws to build the posterior distribution. The algorithm is detailed in Appendix A.

3.1. Assessing models' predictions heterogeneity

Once models have been estimated on the predictions for all periods, we can explore the predictions' heterogeneity. Using the posterior draws $z = 1, \ldots, Z$ generated by the MCMC sampler, we compute the $M \times M$ heterogeneity matrix at time t as follows:

$$H^{(t)} \equiv \frac{1}{Z} \sum_{z=1}^{Z} \sum_{k=1}^{K_z^+} \mathbb{1}_{\boldsymbol{s}_z^{(t)}=k} (\mathbb{1}_{\boldsymbol{s}_z^{(t)}=k})'\,, \tag{12}$$

where K_z^+ denotes the number of active clusters at iteration z, $\boldsymbol{s}_z^{(t)} \equiv (s_{z,1}^{(t)}, \ldots, s_{z,M}^{(t)})'$ are the zth posterior draw of the states at time t (see distribution $s_m | \underline{\mathrm{p}}_{1:K}, K$ in (10)) and $\mathbb{1}_{\boldsymbol{s}_z^{(t)}=k} \equiv (\mathbb{1}_{s_{z,1}^{(t)}=k}, \ldots, \mathbb{1}_{s_{z,M}^{(t)}=k})'$. Let $H_{i,j}^{(t)}$ be the (i,j)-entry of $H^{(t)}$, then $H_{i,j}^{(t)}$ denotes the probability of similar predictions between models i and j at time t. By construction, elements on the diagonal of $H^{(t)}$ are equal to one and $H^{(t)}$ is symmetric. If all models produce similar predictions, the models belong to the same cluster and the matrix is full of ones. Averaging out these matrices over time, we end up with the unconditional heterogeneity matrix $\bar{H} \equiv \sum_{t=1}^{T} \frac{H^{(t)}}{T}$.

[5] To account for time-persistence, we explored a smoothing approach of the number of clusters based on a Markov-switching model. However, the results were qualitatively the same, so we did not include this extra procedure in the paper.

As explained in Section 2.3, we also rely on our mixture model to build our measure of uncertainty and of uncertainty disagreement. Setting our most flexible model as the first model in the model set, we define our measure of uncertainty (potentially) robust to measure errors as:

$$\mathcal{U}_{t,i} \equiv \frac{1}{Z} \sum_{z=1}^{Z} \left(\frac{1}{\sum_{m=1}^{M} \mathbb{1}_{s_{z,m}^{(t)}=s_{1,m}^{(t)}}} \sum_{m=1}^{M} \mathbb{1}_{s_{z,m}^{(t)}=s_{1,m}^{(t)}} \hat{U}_{t,m,i} \right) , \tag{13}$$

where $\hat{U}_{t,m,i}$ denotes the uncertainty estimation of model m for series i at time t. In addition, we measure the uncertainty disagreement as follows:

$$\mathcal{UD}_{t,i} \equiv \frac{1}{Z} \sum_{z=1}^{Z} \frac{1}{K_z^+} \sum_{k=1}^{K_z^+} (\mu_{z,i,k}^{(t)} - \bar{\mu}_{z,i}^{(t)})^2 \, , \text{ with } \bar{\mu}_{z,i}^{(t)} \equiv \frac{1}{K_z^+} \sum_{k=1}^{K_z^+} \mu_{z,i,k}^{(t)} \, , \tag{14}$$

where $\mu_{z,i,k}^{(t)}$ denotes the expectation of cluster k of the zth MCMC iteration at time t for series i. Recall that we simultaneously study the prediction heterogeneity of N series. The measures are thus defined at the series-level. Our global measures are defined by averaging these measures:

$$\mathcal{U}_t \equiv \frac{1}{N} \sum_{i=1}^{N} \mathcal{U}_{t,i} \, , \tag{15}$$

$$\mathcal{UD}_t \equiv \frac{1}{N} \sum_{i=1}^{N} \mathcal{UD}_{t,i} \, . \tag{16}$$

Note that the uncertainty disagreement of one series is closely related to one of the measures of Danielsson et al. (2016) and to model risk in general. In fact, our measure jointly captures the model risk on N series. One crucial difference with model risks lies in the clustering method that clusters the N *predictions* of each volatility model. Consequently, our uncertainty disagreement $\mathcal{UD}_{t,i}$ can be different from zero even though there is no disagreement in the predictions of series i among the models.

4. Application of the MFM model

We now turn to the estimation of the MFM model and the uncertainty measures. We begin by describing the data and the volatility models that we use in Section 4.1. Section 4.2 discusses the model heterogeneity and how it varies over time. We then devote Section 4.3 to the uncertainty measure.

4.1. Data and volatility models

We calibrate the volatility models on eleven worldwide equity indices: (1) S&P 500 (US; SPX), (2) FTSE 100 (UK; FTSE), (3) CAC 40 (France; FCHI), (4) DAX 30 (Germany; GDAXI), (5) Nikkei 225 (Japan; N225), (6) Hang Seng (China; HSI), (7) Dow Jones Industrial Average (US; DJI), (8) Euro Stoxx 50 (Europe; STOXX50), (9) KOSPI (South Korea; KS11), (10) S&P/TSX Composite (Canada; GSPTSE), and (11) Swiss Market Index (Switzerland; SSMI). Data are retrieved from Datastream. Each price series is expressed in local currency. We compute the daily percentage log-return series defined by $100 \times \ln(P_t/P_{t-1})$, where P_t is the adjusted closing value on day t. We then de-mean the returns using an AR(1)-filter, and use those filtered returns as our time series.

We consider 71 risk models:

1. Historical non-parametric approach with a window of size 500 and 1,000 days;
2. EWMA models with Gaussian, Student-t, and skewed Student-t conditional distribution, with memory parameter either set to 0.94 or estimated;
3. GARCH-type models with symmetric and asymmetric specifications (GJR and EGARCH) for the volatility and Gaussian, Student-t, GED, and skewed versions of these for the conditional distribution. We also include target variance versions of the models;
4. Log-GARCH model with a Gaussian conditional distribution (Sucarrat, 2015);
5. GAS model where only volatility is time-varying with Gaussian, Student-t, and skewed Student-t distribution (Ardia et al., 2018b);
6. Stochastic volatility models with Gaussian and Student-t conditional distributions, with and without leverage (Hosszejni and Kastner, 2020);
7. Two-state Markov-switching GARCH models with symmetric and asymmetric specification, Gaussian, Student-t, GED and skewed versions of those (Ardia et al., 2018a);
8. Smooth transition GARCH models with Gaussian, Student-t, and skewed Student-t conditional distributions (Gonzales-Rivera, 1998);
9. Multifractal models with five components (Calvet and Fisher, 2004).

We use 1,000 daily log-returns to estimate the models and run the backtest for a period ranging from January 27, 1994 to February 18, 2021. From the estimated density, we compute the expected shortfall risk measure at a ten-day horizon. Model parameters are estimated by maximum likelihood except for stochastic volatility models where we use an MCMC approach.

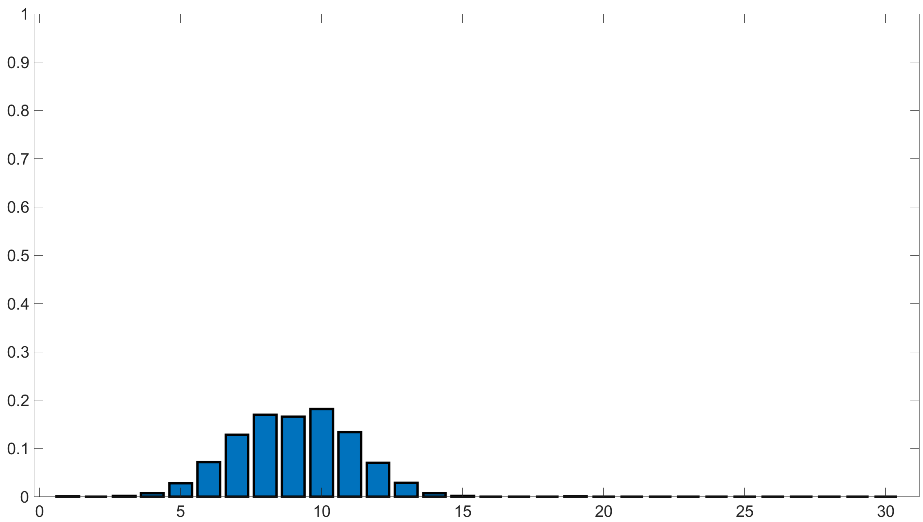
4.2. Prediction heterogeneity: volatility families

The MFM model has the advantage of providing information about predictive heterogeneity across models. Researchers in risk management get only an approximate idea of the models' predictive performance because the forecasting ability of new volatility models is typically assessed with respect to a few other processes (e.g., Conrad and Kleen, 2020). In addition, meta-analyzes on the predictive performance of volatility models use a limited number of financial returns (e.g., Chan and Grant, 2016) and/or models within the same class (e.g., Ardia et al., 2018a). Consequently, no volatility-modeling expert has a clear idea about the forecasting heterogeneity within and between classes of models. In this section, we explore the heterogeneity of model predictions uncovered by the clustering method.

The MFM model groups the model predictions one period at a time. A simple indicator of the number of model families stands as the number of clusters. Figure 3 shows the marginal distribution of the number of clusters over the entire period. We observe that the number of model families oscillates between four to fifteen with a mode at ten volatility classes. The number of clusters over time (not reported here) highlights that this range is relatively constant over time.

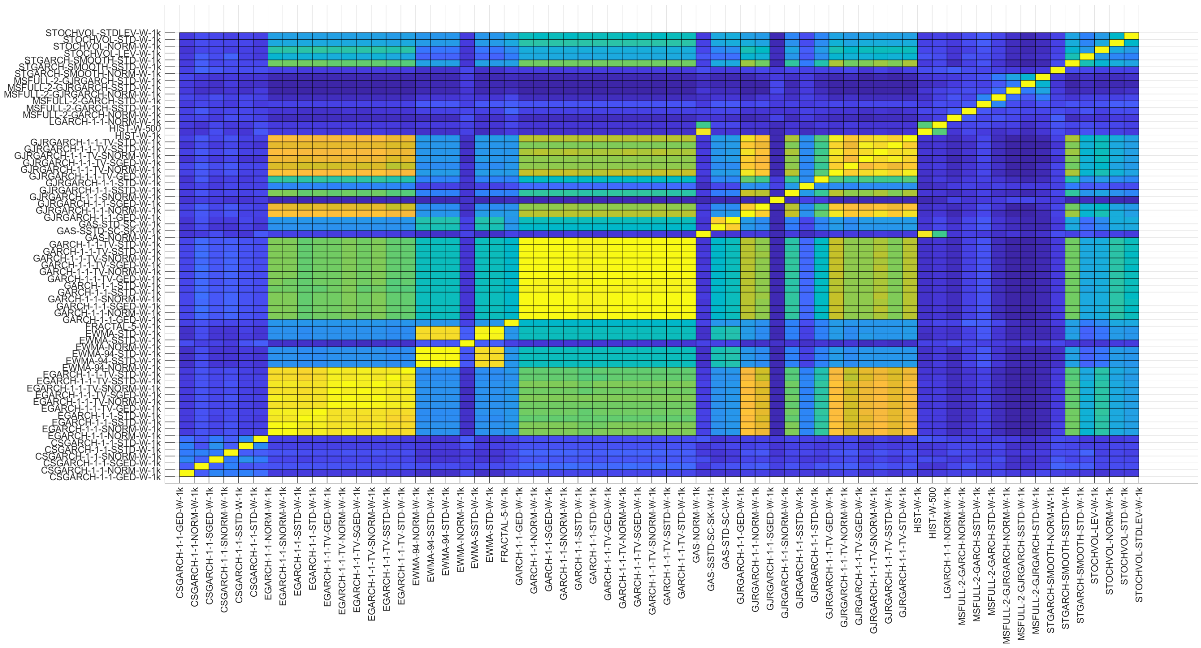
Figure 3: Marginal probability of the number of model families

The figure displays the marginal probability of the number of clusters averaged out over the full period.



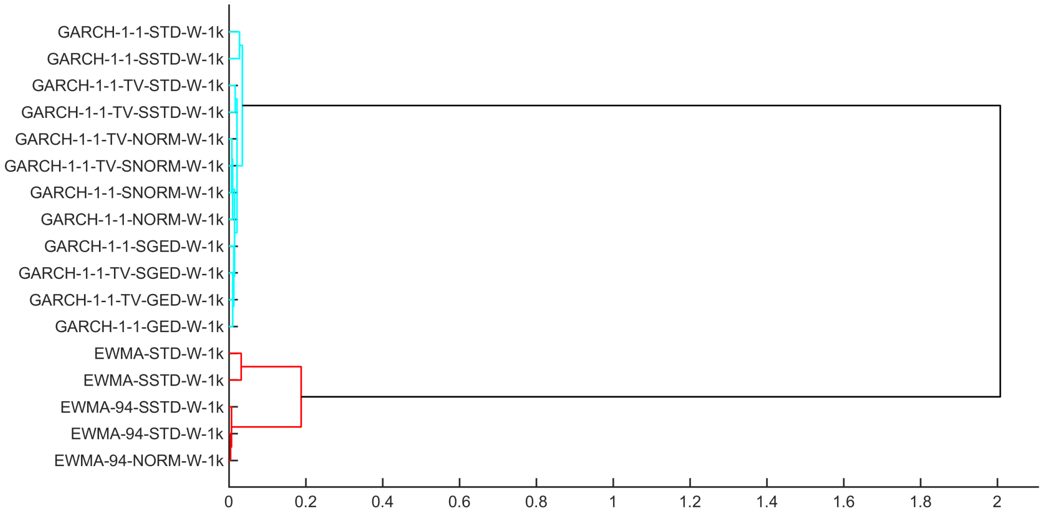
To identify the volatility families, we use the heterogeneity matrix in (12). Figure 4 shows the unconditional heterogeneity matrix (i.e., $\bar{H}$). The heterogeneity matrix makes clear that volatility models generate heterogeneous uncertainty predictions and that many of them cannot be easily classified into one volatility family. Nevertheless, four volatility families stand out. By order of complexity, several EWMA models deliver similar predictions. Then, GARCH models also constitute a homogeneous class in terms of expected shortfall predictions. The EGARCH processes are also very comparable. Finally, the GJR-GARCH models with targeted variance (TV) are relatively equivalent.

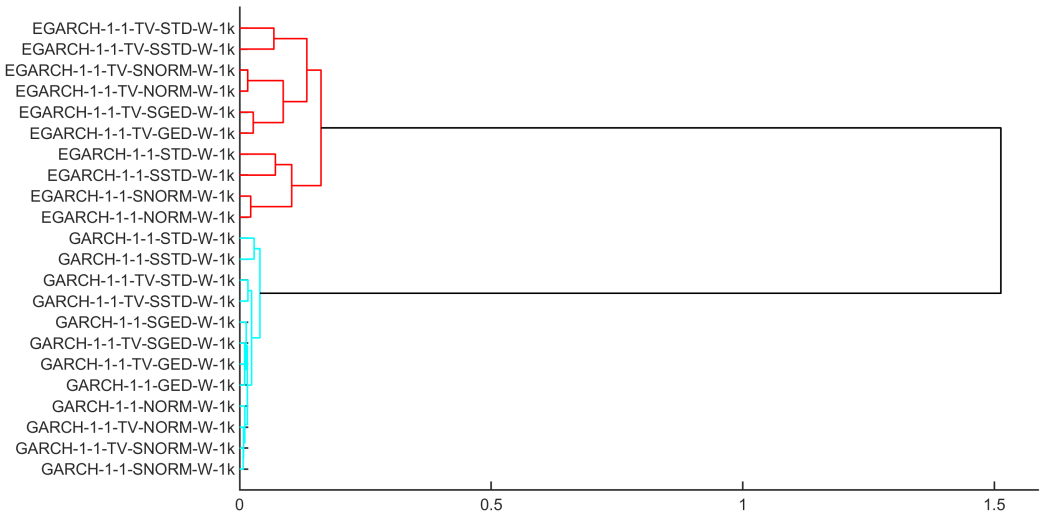
Figure 4: Heterogeneity matrix over the whole period

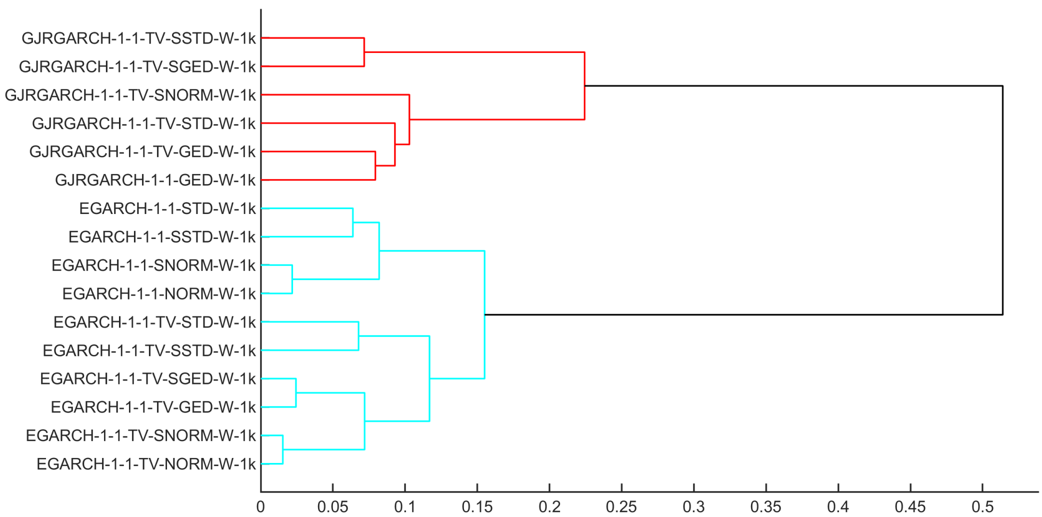


Using a hierarchical clustering technique, we zoom on these four families. Figure 5 shows the dendrogram of two consecutive volatility families based on the average linkage clustering. Focusing on the EWMA family, we observe that the most important difference in the family comes from estimating the memory parameter (instead of fixing it to 0.94 as suggested by Longerstaey and Spencer (1996)). In the EGARCH family, the TV component is more important than the innovation distribution as all the EGARCH with TV components are in the same cluster. Regarding the GJR-GARCH family, the skewed distribution is the most relevant component, followed by the target variance one.

Figure 5: Heterogeneity between two volatility families



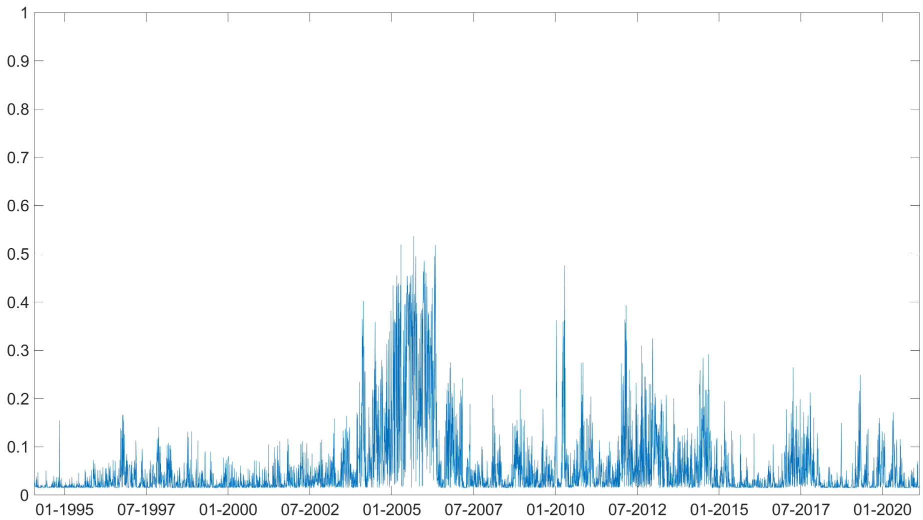




As expected, model heterogeneity varies over time. Figure 6 shows the probability that two models deliver the same prediction when there are randomly selected in the model set (with replacement). The probability at period t is computed as $\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{1}_{H_{i,j}^{(t)}=1}$. Interestingly, the probability surges during calm periods, especially over the period between the dot-com crisis and the global financial crisis. This time-varying heterogeneity favors Assumption A.1, as we expect that each volatility model's bias evolves over time. Figure 6 also highlights that the model families (in the sense of belonging to the same cluster) evolve over time. This is relevant for computing a robust uncertainty measure because we want to average out the uncertainties of models within the same cluster.

Figure 6: Model heterogeneity over time

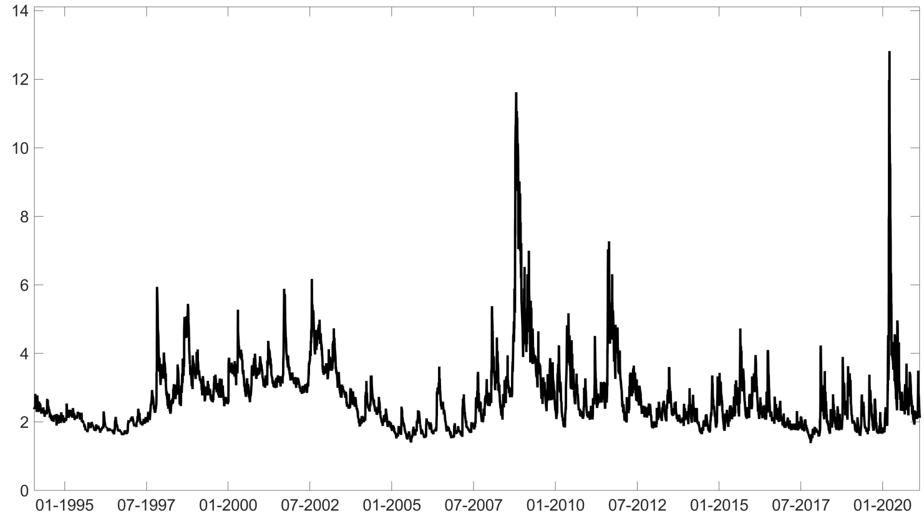
This figure displays the probability that two models in the model set give a similar prediction (with replacement). For each time period, we divide the sum of all the elements equal to one by the squares of the number of models (i.e., $p_t \equiv \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{1}_{H_{i,j}^{(t)}=1}$).



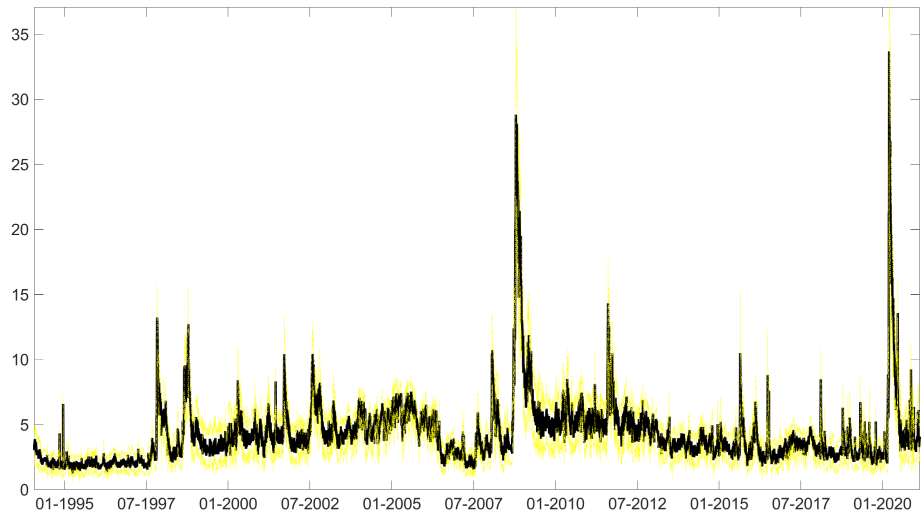
4.3. Uncertainty and uncertainty disagreement

Figure 7 shows the daily uncertainty measure and uncertainty disagreement over time. As expected, the two series strongly spike during the Global Financial Crisis and the COVID pandemic. The uncertainty measure and the disagreement one exhibit a first-order autocorrelation of 0.97 and 0.95, respectively. The correlation between the two series amounts to 0.77.

Figure 7: Uncertainty and uncertainty disagreement over time
This figure displays the evolution over time of the uncertainty $\mathcal{U}_t$ computed in (15) and the uncertainty disagreement $\mathcal{UD}_t$ in (16).



Uncertainty measure $\mathcal{U}_t$



Uncertainty Disagreement measure $\mathcal{UD}_t$ with 95% credible interval

Table 1 documents the correlation with other daily uncertainty measures that are also forward-looking and real-time. Our measures share similar dynamics with other uncertainty measures such as the one-month volatility index (VIX; see Bloom 2009) and the variance risk premium (VRP; Bollerslev et al. 2009). The correlation goes up to 0.89 for the uncertainty measure with the VIX. On the contrary, VRP and the VIX are strongly correlated with a correlation coefficient amounting to -0.96.

Table 1: Comparison with existing uncertainty measures

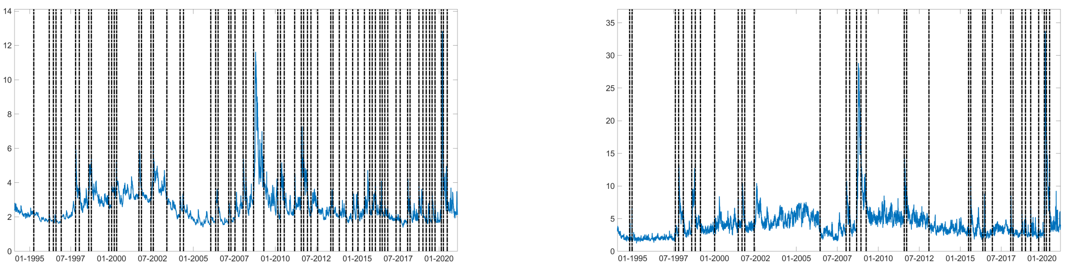
The table reports the correlation between our measures and alternative measures proposed in the literature such as the variance risk premium (VRP), the implied volatility index (VIX), and the cross-sectional volatility of S&P 500 returns (CSV, see Bloom 2009; Christiano et al. 2014).

Existing measure	$\mathcal{U}$	$\mathcal{UD}$
VRP	-0.88	-0.79
VIX	0.89	0.71
CSV	0.65	0.46

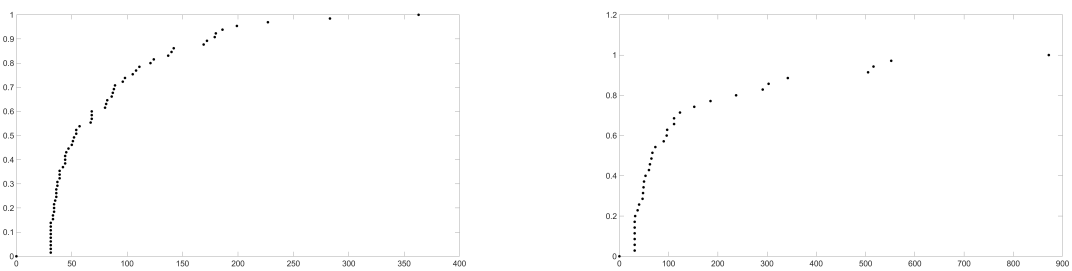
Our two measures are persistent and predictable. However, the spikes capturing the most critical crises cast doubts about the stationarity of the series. We thus fit the two series using a piecewise-stationary model. In particular, we use the standard change-point AR(2) model proposed by Ardia et al. (2019) to detect breaks in the series dynamics. Figure 8 shows the breakpoints locations and the cumulative density function of a segment duration in days. The change-point model captures all major events occurring on the global financial market. We detect 66 and 36 breakpoints in the uncertainty measure and the uncertainty disagreement, respectively. The duration medians of a segment amount to 54 and 70 days, respectively. It implies that the uncertainty dynamic of the financial market evolves every 2 to 3 months. In addition, the dynamic of the uncertainty measure changes more often than the uncertainty disagreement.

Figure 8: Detected breakpoints and duration of the segments

Uncertainty $\mathcal{U}_t$ Uncertainty disagreement $\mathcal{UD}_t$

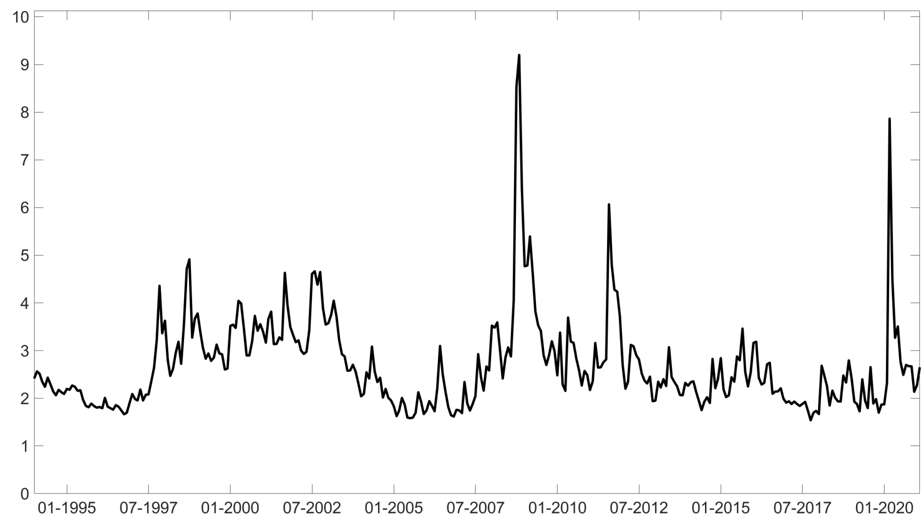


Cumulative density function of the segment durations

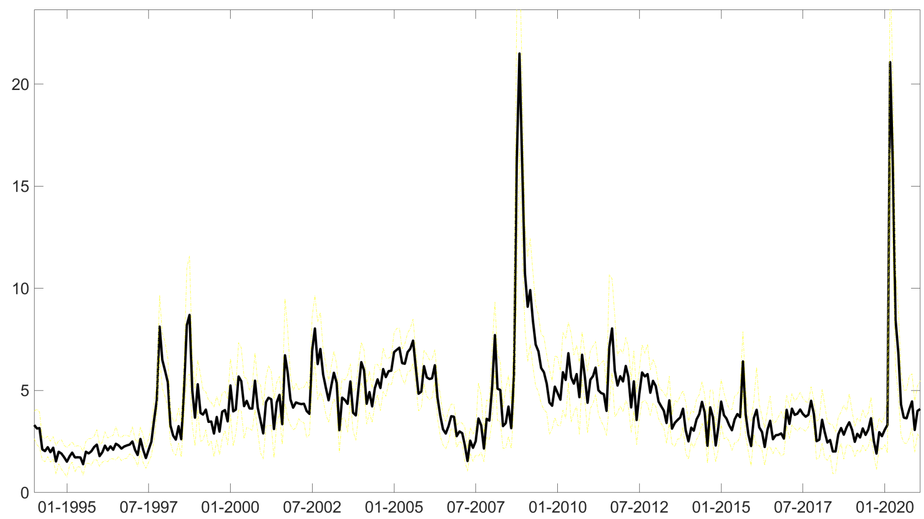


To compare with other recent uncertainty measures, we also compute our measures on a monthly basis. To do so, we average out the model predictions over each month and then apply the clustering method on the monthly data. Figure 9 shows the monthly measures. At several periods such as between 2005 and 2007, we observe that the uncertainty measure drops while the uncertainty disagreement increases. Regarding the COVID pandemic, both monthly measures reach a peak in March 2020. For the uncertainty measure, we observe an increase of 320% with respect to its value in January 2020. The uncertainty disagreement augments by 961% over the same period. These values are in line with the one-month VIX, the 24-month VIX, the news Economic Policy Uncertainty, and the Twitter Economic Uncertainty as documented by Altig et al. (2020a).

Figure 9: Monthly uncertainty and uncertainty disagreement
Using monthly data computed by averaging out the daily expected shortfall, the uncertainty measure is computed with (15). The uncertainty disagreement follows from (16).



Monthly uncertainty

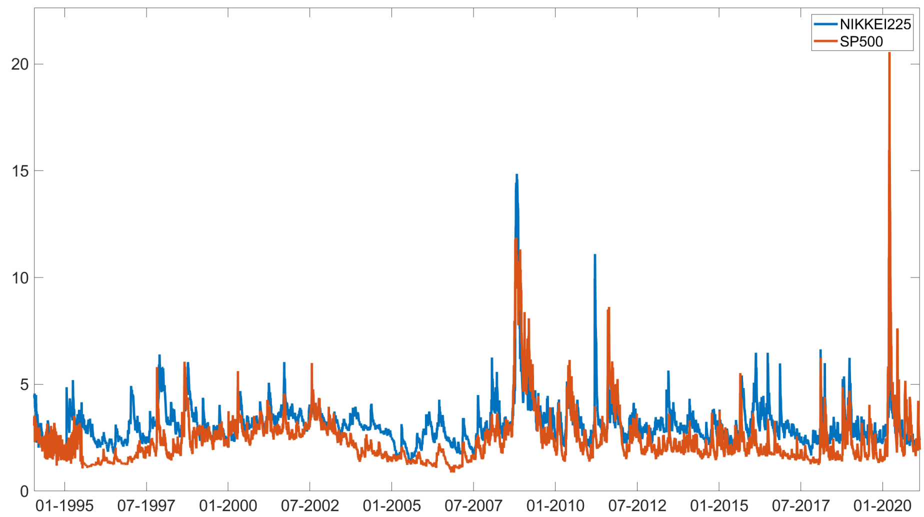


Monthly uncertainty disagreement with 95% credible interval

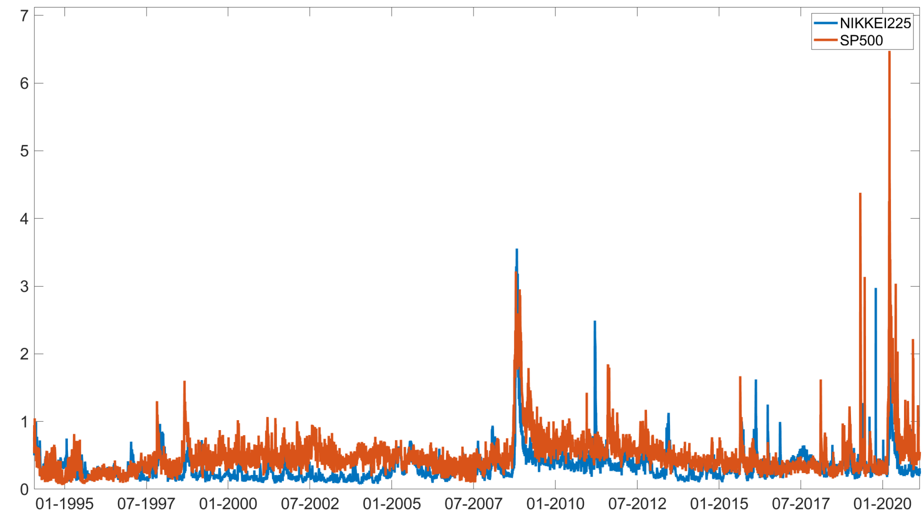
4.3.1. Decomposition of the uncertainty

Besides the global level of our measures, we can also investigate the uncertainty dynamic and the related disagreement per series. Figure 10 highlights our measures over time for two indices: the Nikkei 225 and the S&P 500. The two measures agree that the Japanese market has been even more affected during the financial crisis than the US market. On the other hand, Japan has limited the COVID pandemic impact compared to the US, as confirmed by the two measures. We also observe several crash events such as the Fukushima crisis and the US market's huge uncertainty in February 2018.

Figure 10: Uncertainty measures for two equity indices over time
Uncertainty and uncertainty disagreement over time for the Nikkei 225 and the S&P 500.



Uncertainty



Uncertainty disagreement

5. Applications of the measures

In this section, we explore three potential applications of the uncertainty measures. First, we look at factor models and check if returns of the firms in the S&P 500 are sensitive to the uncertainty captured by our measures. Second, we show that our measures help predict realized variance of equity indices. Finally, we investigate if there exists a link between trading volumes and the uncertainty disagreement.

5.1. Factor exposure

We test if our uncertainty measures stands as risk factors in asset pricing. We consider the following factor model:

$$r_{i,t} = \alpha_i + \beta_i \mathrm{Unc}_t + \boldsymbol{\theta}_i' \boldsymbol{x}_t + \boldsymbol{\gamma}_i' \mathbf{f}_t + \epsilon_{i,t}\,, \tag{17}$$

where $r_{i,t}$ is the log-return of stock i at time t, Unc_t is a selected measure of uncertainty at time t, $\boldsymbol{x}_t$ is a vector of control variables, $\mathbf{f}_t$ a vector containing the observations of the five factors of Fama and French (2015), both at time t, and $\epsilon_{i,t}$ is an error term. For each 1,237 stocks which entered the S&P 500 index for at least 30 days over the time period (from January 27, 1994 to February 18, 2021), we test if the OLS estimate of β_i is significantly different from zero at the 5% significance level. Table 2 reports the percentage of significant exposures for various specifications. For each measure, we observe that the level and the variable in first difference exhibit a significant exposure in about 20% of the stocks even when we control for the VIX and the cross-sectional volatility (in addition to the Fama-French five factors). This indicates that our measure captures a risk that could be hedged. Interestingly, the VIX partly looses its significance when we control for our measures. This is especially true when the measure on uncertainty disagreement kicks in the model.

Table 2: Exposures associated with the uncertainty measures
This table reports the percentage of significant exposures over the 1,237 stocks in our sample. Control stands for control variables in addition to the Fama-French five factors. VIX and CSV denote the one-month VIX and the cross-sectional volatility, respectively. Δ stands for the variable in first difference.

Measure	Controls ($\boldsymbol{x}_t$)	Significance
$\mathcal{U}$	None	19.0%
$\mathcal{U}$	VIX, CSV	18.2%
$\Delta\mathcal{U}$	None	24.6%
$\Delta\mathcal{U}$	VIX, CSV	24.3%
$\mathcal{UD}$	None	17.5%
$\mathcal{UD}$	VIX, CSV	14.1%
$\Delta\mathcal{UD}$	None	18.0%
$\Delta\mathcal{UD}$	VIX, CSV	17.3%
VIX	None	18.5%
VIX	$\mathcal{U}$, CSV	13.9%
VIX	$\mathcal{UD}$, CSV	10.4%

5.2. Forecasting realized variance

We explore now the predictive performance of our uncertainty measures when forecasting the realized variance of equity indices. We use an Heterogeneous Auto-Regressive (HAR) model (Corsi, 2009):

$$\ln \mathrm{RV}_t^{(i)} = \beta_0 + \beta_1 \ln \bar{\mathrm{RV}}_{t-1}^{(i)} + \beta_2 \ln \bar{\mathrm{RV}}_{t-5}^{(i)} + \beta_3 \ln \bar{\mathrm{RV}}_{t-22}^{(i)} + \boldsymbol{\theta}'\boldsymbol{x}_{t-1} + \epsilon_t, \tag{18}$$

where $\mathrm{RV}_t^{(i)}$ denotes the daily realized variance at time t of equity index i, $i = 1, \ldots, 11$, $\ln \bar{\mathrm{RV}}_{t-j}^{(i)} \equiv \frac{1}{j}\sum_{q=1}^{j} \ln \mathrm{RV}_{t-q}^{(i)}$, and $\boldsymbol{x}_{t-1}$ stand for additional variables such as the one-month VIX and our uncertainty measures. We apply the HAR model to eleven realized variance indices (10-min subsampled) retrieved from the Oxford-Man Institute website (Heber et al., 2009) and ranging from 2001 to 2021 (except the S&P/TSX Composite which starts in 2003).

We perform a predictive exercise to assess the forecasting performance of our uncertainty measures. At each time point, we use the past 250 days to estimate the parameters in (18). Then, we recursively add the observations one by one, and, at each iteration, we predict the one-step-ahead log realized variance using the estimated model. Finally, we compare the predictive performance using the root-mean-square forecast error (RMSFE) of various HAR specifications.

Results are reported in Table 3. In Panel A, we see that including an uncertainty variable, or a combination of them, does improve to a large extend the forecasting performance compared to a benchmark model without additional variable. In Panel B, we find mix evidence of outperformance of our measures when the benchmak model is with the VIX included. However, an extended model with VIX and our variables improves the performance for several series while it does not significantly deteriorates for the others.

Table 3: Forecasting results of the various HAR model specifications
This table reports the root-mean-square forecast error RMSFE ($\times 10^4$) of the benchmark HAR model and the relative performance, in percent, of the alternative specifications. A relative performance below 100 indicates an improvement of the alternative specification. In squared parenthesis, we report the p-value of a two-sided test of equal-performance between the benchmark and the alternative specifications. We use the Diebold-Mariano test statistic with robust HAC standard errors within the bootstrap approach of Clark and McCracken (2001). "None" indicates that no variable is included in the HAR model.

Panel A: Benchmark model is without uncertainty measures

		Set of variables in $\boldsymbol{x}_{t-1}$					
	RMSFE	$\mathcal{U}$	$\mathcal{UD}$	$\mathcal{U}, \mathcal{UD}$	$\mathcal{U}, \mathcal{UD}$,VIX	CSV	VIX
SP500	2.228	91.8 [0.00]	92.7 [0.00]	99.6 [0.03]	91.4 [0.00]	98.9 [0.65]	91.7 [0.01]
FTSE100	3.509	98.6 [0.47]	99.6 [0.84]	100.1 [0.53]	100.0 [0.98]	100.4 [0.52]	99.3 [0.87]
CAC40	2.077	92.1 [0.01]	93.3 [0.00]	99.7 [0.07]	92.7 [0.05]	99.0 [0.60]	92.8 [0.04]
DAX30	2.303	88.9 [0.00]	91.3 [0.00]	99.8 [0.38]	88.7 [0.00]	99.4 [0.47]	90.8 [0.02]
N225	2.032	93.8 [0.01]	93.9 [0.00]	99.9 [0.75]	94.4 [0.05]	99.7 [0.92]	97.2 [0.44]
HSI	1.408	95.5 [0.03]	95.2 [0.02]	99.5 [0.09]	98.0 [0.44]	100.0 [0.95]	97.8 [0.49]
DJI	2.274	91.8 [0.00]	93.1 [0.00]	99.6 [0.05]	91.6 [0.00]	99.1 [0.70]	92.0 [0.01]
STOXX50	3.073	97.7 [0.72]	98.1 [0.47]	100.1 [0.64]	98.6 [0.86]	99.4 [0.39]	97.6 [0.45]
KOSPI	1.597	92.0 [0.00]	91.1 [0.00]	100.0 [0.97]	91.4 [0.00]	99.6 [0.67]	95.0 [0.09]
TSX	4.753	104.1 [0.23]	102.2 [0.46]	99.7 [0.01]	104.9 [0.22]	99.3 [0.02]	103.6 [0.22]
SMI	1.749	94.3 [0.00]	96.4 [0.00]	100.3 [0.14]	96.0 [0.00]	100.2 [0.76]	97.2 [0.34]

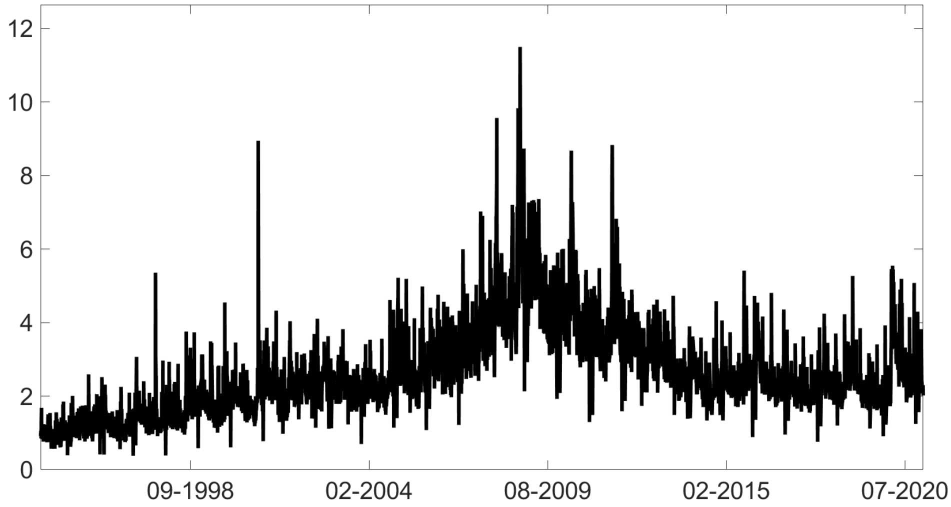
Panel B: Benchmark model is with VIX as uncertainty measure

		Set of variables in $\boldsymbol{x}_{t-1}$					
	RMSFE	$\mathcal{U}$	$\mathcal{UD}$	$\mathcal{U}, \mathcal{UD}$	$\mathcal{U}, \mathcal{UD}$,VIX	CSV	None
SP500	2.042	100.1 [0.00]	101.1 [0.00]	108.6 [0.03]	99.7 [0.00]	107.8 [0.65]	109.1 [0.01]
FTSE100	3.485	99.3 [0.47]	100.3 [0.84]	100.8 [0.53]	100.7 [0.98]	101.1 [0.52]	100.7 [0.87]
CAC40	1.929	99.2 [0.01]	100.5 [0.00]	107.4 [0.07]	99.8 [0.05]	106.6 [0.60]	107.7 [0.04]
DAX30	2.090	97.9 [0.00]	100.6 [0.00]	110.0 [0.38]	97.7 [0.00]	109.5 [0.47]	110.2 [0.02]
N225	1.974	96.6 [0.01]	96.6 [0.00]	102.9 [0.75]	97.2 [0.05]	102.7 [0.92]	102.9 [0.44]
HSI	1.377	97.6 [0.03]	97.3 [0.02]	101.7 [0.09]	100.1 [0.44]	102.2 [0.95]	102.2 [0.49]
DJI	2.091	99.8 [0.00]	101.2 [0.00]	108.3 [0.05]	99.6 [0.00]	107.7 [0.70]	108.8 [0.01]
STOXX50	2.999	100.2 [0.72]	100.5 [0.47]	102.6 [0.64]	101.1 [0.86]	101.9 [0.39]	102.5 [0.45]
KOSPI	1.517	96.8 [0.00]	95.9 [0.00]	105.3 [0.97]	96.3 [0.00]	104.9 [0.67]	105.3 [0.09]
TSX	4.923	100.5 [0.23]	98.7 [0.46]	96.2 [0.01]	101.3 [0.22]	95.8 [0.02]	96.5 [0.22]
SMI	1.700	97.0 [0.00]	99.2 [0.00]	103.2 [0.14]	98.7 [0.00]	103.1 [0.76]	102.9 [0.34]

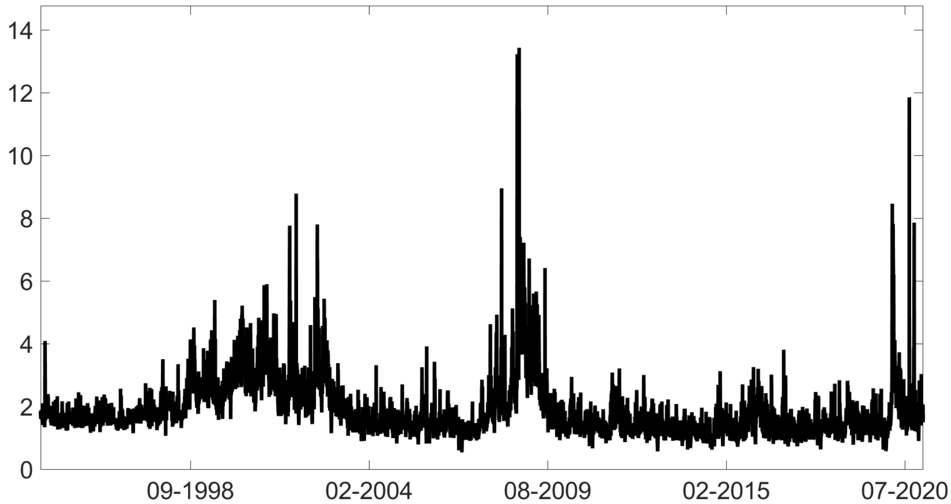
5.3. Uncertainty disagreement and trading volume

As emphasized by Carlin et al. (2014), an important channel through which uncertainty impacts the economy stands for opinion disagreement. As put forward by theoretical frameworks such as Harris and Raviv (1993), disagreement implies higher trading volumes and volatility. We check this relation with the aggregate trading volume of S&P 500 firms, the cross-sectional variance, and the uncertainty disagreement. We follow Carlin et al. (2014), and we use a Vector Autoregressive (VAR) model with three lags. Figure 11 shows the volume and the cross-sectional variance while Table 4 reports their summary statistics.

Figure 11: Trading volume and cross-sectional variance over time
The graphics show evolution of the aggregate volume of S&P 500 firms and the cross-sectional variance of the returns. To have similar scales, the trading volume is divided by 10^9.



Trading volume



Cross-sectional variance

Table 4: Summary statistics of the VAR variables
The table shows the summary statistics of the three variables used in the VAR model. The trading volume is divided by 10^9 to ensure a similar level as compared to the other two variables.

Variable	Mean	Std	Skewness	Kurtosis
$\mathcal{UD}$	4.31	2.65	4.33	32.38
CSV	1.88	0.90	3.06	22.63
Trading volume	2.62	1.21	1.41	6.42

Table 5 reports the VAR estimates. Results are in line with Carlin et al. (2014). Focusing on the disagreement equation, past variances do not significantly impact the current level of disagreement. However, and as expected, past trading volumes affect the disagreement variable. While the first lagged value increases the disagreement when trade augments, the third lag has a negative impact on the uncertainty disagreement. As explained in Carlin et al. (2014), this result is intuitive as investors learn about the trading volume and adapt their uncertainty accordingly. Zooming on the cross-sectional variance equation and besides the persistence of the variable, we observe a positive impact of the lagged uncertainty disagreement as well as the first-lagged trading volume. An increase in uncertainty disagreement tends to be followed by a higher level of variance. Regarding the trading volume equation, past variance and uncertainty disagreement affect the current level of trades. While the sign of the variance changes with the lags, the positive and significant estimates of the first and third lagged uncertainty disagreement indicate an increase in trading volumes when disagreement rises. This is additional evidence in favor of the theoretical framework of Harris and Raviv (1993) and supports the empirical evidence of Kandel and Pearson (1995).

Table 5: Relation between uncertainty disagreement and trading volume
The table reports the estimates of a VAR(3) model with the S&P 500 trading volume, the cross-sectional variance (CSV), and the uncertainty disagreement ($\mathcal{UD}$) as dependent variables. The dependent variables are in logarithm to limit the impact of extreme values. As shown in Table 4, the kurtosis is extremely large for the first two dependent variables. By taking the logarithm, the dependent variables' kurtoses become 4.54, 4.01, and 3.28, respectively.

	$\mathcal{UD}_t$	CSV_t	Volume_t
Intercept	0.03**	0.03**	0.03**
$\mathcal{UD}_{t-1}$	0.52**	0.04**	0.03*
CSV_{t-1}	0.01	0.41**	0.05**
Volume_{t-1}	0.08**	0.07**	0.54**
$\mathcal{UD}_{t-2}$	0.27**	-0.00	-0.01
CSV_{t-2}	0.01	0.23**	-0.07**
Volume_{t-2}	-0.01	-0.06**	0.15**
$\mathcal{UD}_{t-3}$	0.15**	0.00	0.04**
CSV_{t-3}	0.01	0.25**	-0.03**
Volume_{t-3}	-0.03**	-0.05**	0.22**

6. Conclusion

Uncertainty has dramatic impacts on economic conditions and is typically related to forecasting uncertainty. Many studies estimate uncertainty as the volatility of a predictive density (e.g., Bomberger, 1996; Bloom, 2009; Jurado et al., 2015). The existing uncertainty measures differ by the information set from which they have been estimated. Using a panel of financial indices and of model predictions, we propose a new uncertainty measure based on the average of the expected shortfall produced by multiple volatility processes belonging to the same model family.

To track down the model families' set over time, we apply the mixture of finite mixture model recently proposed by Frühwirth-Schnatter et al. (2020). This clustering approach highlights the prediction heterogeneity of the 71 volatility models used in this study and shows that the heterogeneity evolves over time. As an average, our uncertainty measure is less exposed to the measurement error put forward by Carriero et al. (2018) than other measures. In addition, the measure is forward-looking, real-time, and can be estimated on any financial series.

Uncertainty increases when future opinions about economic conditions differ. It constitutes an important channel through which uncertainty affects asset prices (see Harris and Raviv, 1993; Carlin et al., 2014). Investors with the same information set may have different opinions (Gao et al., 2019). We assume that some of these investors end up with different opinions because they do not use the same volatility model to predict forecast uncertainty. Therefore, we view the heterogeneity of volatility model predictions as forecast disagreement. Motivated by the theoretical framework of section 2.2, we build a measure of uncertainty disagreement and show that disagreement raises during recession periods.

Empirical-wise, our measures share commonalities with standard uncertainty measures such as the VIX, the variance risk premium, and the cross-sectional variance even though our approach is based on a smaller information set than those measures. We highlight that the uncertainty measure and, to a lesser extent, the uncertainty disagreement are associated with a risk premium. We also look at the predictive ability of the measures on daily realized variance. In a forecasting exercise, we show that the two measures substantially improve the root mean absolute forecast errors even when controlling for the VIX. Finally, using VAR model, we show with that the uncertainty disagreement measure positively impacts on future trading volume of financial assets.

Appendix A. MCMC algorithm for the MFM model

In this appendix, we detail the MCMC algorithm used to estimate the model parameters of the MFM model. For notational purposes, we skip the time subscript since the MCMC algorithm is applied in parallel to each time period. For a time period, we cluster the $N \times M$ model predictions: $\underset{\sim}{\mathrm{y}} \equiv (\underset{\sim}{\mathrm{y}}_1, \underset{\sim}{\mathrm{y}}_2, \ldots, \underset{\sim}{\mathrm{y}}_M) \in \mathbb{R}^{N\times M}$. We use the following notations: the number of elements in cluster k is denoted by $n_k \equiv \sum_{m=1}^{M} \mathbb{1}_{s_m=k}$, a vector of dimension p that is full of ones is written as $\mathbb{1}_p$ and the identity matrix of dimension N is given by $\boldsymbol{I}_N$. We initialize the MCMC algorithm using the K-means algorithm with 10 clusters. The active number of clusters is initialized at $K^+ = 10$ and the number of cluster is set to $K = K^+ + 2$. The MCMC scheme consists in nine steps that we detail below.

1. **Sampling the state vector $\boldsymbol{s} \equiv (s_1, \ldots, s_M)'$.** Given the K possible clusters, we sample each state from the discrete distribution as follows, for $m = 1, ..., M$:

$$P[s_m = k | \underset{\sim}{\mathrm{y}}, K, \underset{\sim}{\mathrm{p}}_{1:K}, \boldsymbol{\mu}, \boldsymbol{\Sigma}] \propto f_N(\underset{\sim}{\mathrm{y}}_m | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) p_k \,,$$

   where $f_N(\underset{\sim}{\mathrm{y}}_m | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ denotes the density of a multivariate normal distribution with expectation $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$.
   From the state vector $\boldsymbol{s}$, compute the number of models in each state: $n_k \equiv \sum_{m=1}^{M} \mathbb{1}_{s_m=k}$ and the number of active clusters: $K^+ \equiv \sum_{k=1}^{K} \mathbb{1}_{n_k>0}$.

2. **Sampling the covariance matrices $\boldsymbol{\Sigma}_k$ for $k = 1, \ldots, K$.** For each component k, the covariance matrix is sampled as:

$$\boldsymbol{\Sigma}_k^{-1} \sim \text{Inverse-Wishart}\left(\bar{\boldsymbol{V}}, \underline{v} + \sum_{m=1}^{M} \mathbb{1}_{s_m=k}\right) ,$$

   in which $\bar{\boldsymbol{V}} \equiv \left(\boldsymbol{V} + \sum_{m=1}^{M} \mathbb{1}_{s_m=k} (\underset{\sim}{\mathrm{y}}_m - \boldsymbol{\mu}_k)(\underset{\sim}{\mathrm{y}}_m - \boldsymbol{\mu}_k)'\right)^{-1}$.
3. **Sampling the expectation $\boldsymbol{\mu}_k$ for $k = 1, \ldots, K$.** For each component k, the expectation is drawn from:

$$\begin{aligned} \boldsymbol{\mu}_k &\sim N(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}), \\ \bar{\boldsymbol{\Sigma}} &\equiv \left(\boldsymbol{X}'\bar{\Omega}\boldsymbol{X} + \boldsymbol{R}^{-1}\right)^{-1}, \\ \bar{\boldsymbol{\mu}} &\equiv \bar{\boldsymbol{\Sigma}}(\boldsymbol{X}'\bar{\Omega}\tilde{\underset{\sim}{\mathrm{y}}} + \boldsymbol{R}^{-1}\boldsymbol{b}_0), \end{aligned}$$

   with $\boldsymbol{X} \equiv \boldsymbol{I}_N \otimes \mathbb{1}_{n_k}$, $\bar{\Omega} \equiv \boldsymbol{\Sigma}_k^{-1} \otimes I_{n_k}$ and $\tilde{\underset{\sim}{\mathrm{y}}} \equiv \vec{(}\boldsymbol{y}_{\mathbb{1}_{s=k}}) \in \mathbb{R}^{n_k N}$. Note that $\boldsymbol{y}_{\mathbb{1}_{s=k}} \in \mathbb{R}^{N\times n_k}$ denotes all the model predictions belonging to cluster k.
4. **Sampling S_0.** The hierarchical covariance matrix is sampled as follows:

$$S_0^{-1} \sim \text{Inverse-Wishart}(\bar{\boldsymbol{S}}, s + K^+\underline{v}) \,,$$

   where $\bar{\boldsymbol{S}} \equiv S_0^{-1} + \sum_{k=1}^{K^+} \boldsymbol{\Sigma}_k^{-1}$.
5. **Sampling the hierarchical expectation $\boldsymbol{b}_0$.** The posterior distribution of the expectation parameters given the other model parameters is a multivariate normal distribution

specified as:

$$
\begin{aligned}
\boldsymbol{b}_0 &\sim N(\bar{\boldsymbol{b}}, \bar{\boldsymbol{\Sigma}}_{\boldsymbol{b}_0}),\\
\bar{\boldsymbol{\Sigma}}_{\boldsymbol{b}_0} &\equiv \left(K^+ \boldsymbol{R}^{-1} + (I_k/100)\right)^{-1},\\
\bar{\boldsymbol{b}} &\equiv \bar{\boldsymbol{\Sigma}}_{\boldsymbol{b}_0}\left(\boldsymbol{R}^{-1}(\sum_{k=1}^{K^+} \boldsymbol{\mu}_k) + (I_k/100)\text{median}(\underline{\text{y}})\right),
\end{aligned}
$$

6. **Sampling $\boldsymbol{R}$.** The hierarchical covariance matrix of the expectation is given by:

$$
S_0^{-1} \sim \text{Inverse-Wishart}\left(\boldsymbol{R}_0 + \sum_{k=1}^{K^+} (\boldsymbol{\mu}_k - \boldsymbol{b}_0)(\boldsymbol{\mu}_k - \boldsymbol{b}_0)', r_0 + K^+\right).
$$

7. **Sampling the number of clusters K.** For $\bar{K} = K^+, K^+ + 1, \ldots$, sample K from:

$$
P[K = \bar{K}|n_1, \ldots, n_K, \alpha_{\text{Dir}}] \propto P[K = \bar{K}] \frac{(\alpha_{\text{Dir}})^{K^+} \bar{K}!}{\bar{K}^{K^+} (\bar{K} - K^+)!} \prod_{k=1}^{K^+} \frac{\Gamma(n_k + \frac{\alpha_{\text{Dir}}}{\bar{K}})}{\Gamma(1 + \frac{\alpha_{\text{Dir}}}{\bar{K}})},
$$

where $P[K = \bar{K}]$ stands for the density of the Beta-Negative-Binomial distribution evaluated at $\bar{K} - 1$ (i.e. $K_{BNB} \equiv K - 1 \sim BNB(1, \underline{a}_k = 4, \underline{b}_k = 3)$). The density is given by $P[K_{BNB} = k] = \frac{\Gamma(\underline{a}_k + \underline{b}_k)}{\Gamma(\underline{a}_k)\Gamma(\underline{b}_k)} \frac{\Gamma(\underline{a}_k + 1)\Gamma(\underline{b}_k + k)}{\Gamma(\underline{a}_k + \underline{b}_k + k + 1)}$ for $k = 0, 1, \ldots$. If the number of clusters sampled from the posterior distribution is greater than the previous number of clusters, sample the additional cluster parameters from their prior distributions.

8. **Sampling α_{Dir}.** Sample a proposal value from $\alpha_{\text{Dir}}^{\text{prop}} \sim N(\alpha_{\text{Dir}}, \sigma_R^2)$ where σ_R^2 is adapted using Atchade and Rosenthal (2005) as follows: $\sigma_R \equiv \min\big(\max(1e^{-6}, \sigma_R + (\text{acc}_{\text{rate}}^{(z)} - 0.44)/z^{0.6}), 10\big)$ in which $\text{acc}_{\text{rate}}^{(z)}$ denotes the acceptance rate up to iteration z of the MCMC algorithm. Then, accept the draw using the Metropolis ratio:

$$
\alpha_{\text{Metropolis}} \equiv \min\left(\frac{\bar{f}(\alpha_{\text{Dir}}^{\text{prop}})}{\bar{f}(\alpha_{\text{Dir}})}, 1\right),
$$

in which $\bar{f}(\alpha_{\text{Dir}}) \equiv f(\alpha_{\text{Dir}})(\alpha_{\text{Dir}})^{K^+} \frac{\Gamma(\alpha_{\text{Dir}})}{\Gamma(\alpha_{\text{Dir}} + M)} \prod_{k=1}^{K^+} \frac{\Gamma(n_k + \frac{\alpha_{\text{Dir}}}{K})}{\Gamma(1 + \frac{\alpha_{\text{Dir}}}{K})}$ and $f(\alpha_{\text{Dir}})$ denotes the prior distribution on α_{Dir} (i.e. $\alpha_{\text{Dir}} \sim \text{Fisher}(6, 3)$ as advocated by Frühwirth-Schnatter et al. (2020)).

9. **Sampling the probability of the clusters $\underline{\mathbf{p}}_{1:K}$.** The posterior distribution of the probabilities is given by:

$$
\underline{\text{p}}_{1:K} \sim \text{Dirichlet}\left(n_1 + \frac{\alpha_{\text{Dir}}}{K}, n_2 + \frac{\alpha_{\text{Dir}}}{K}, \ldots, n_K + \frac{\alpha_{\text{Dir}}}{K}\right).
$$

References

Altig, D., Baker, S., Barrero, J.M., Bloom, N., Bunn, P., Chen, S., Davis, S.J., Leather, J., Meyer, B., Mihaylov, E., et al., 2020a. Economic uncertainty before and during the COVID-19 pandemic. Journal of Public Economics 191, 104274.

Altig, D., Barrero, J.M., Bloom, N., Davis, S.J., Meyer, B., Parker, N., 2020b. Surveying business uncertainty. Forthcoming in Journal of Econometrics.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., 2018a. Forecasting risk with Markov-switching GARCH models: A large-scale performance study. International Journal of Forecasting 34, 733–747.

Ardia, D., Boudt, K., Catania, L., 2018b. Downside risk evaluation with the R package GAS. The R Journal 10, 410–421.

Ardia, D., Dufays, A., Ordas, C., 2019. Bayesian and classic change-point models: The missing link. Working paper.

Atchade, Y., Rosenthal, J., 2005. On adaptive Markov chain Monte Carlo algorithms. Bernoulli 11, 815–828.

Bauwens, L., Dufays, A., Rombouts, J., 2013. Marginal likelihood for Markov switching and change-point GARCH models. Journal of Econometrics 178, 508–522.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77, 623–685.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., Terry, S.J., 2018. Really uncertain business cycles. Econometrica 86, 1031–1065.

Bollerslev, T., 2010. Glossary to ARCH (GARCH), in: Volatility and Time Series Econometrics. Oxford University Press, 137–163.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Review of Financial Studies 22, 4463–4492.

Bomberger, W.A., 1996. Disagreement as a measure of uncertainty. Journal of Money, Credit and Banking 28, 381–392.

Calvet, L.E., Fisher, A.J., 2004. How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. Journal of Financial Econometrics 2, 49–83.

Carlin, B.I., Longstaff, F.A., Matoba, K., 2014. Disagreement and asset prices. Journal of Financial Economics 114, 226–238.

Carriero, A., Clark, T.E., Marcellino, M., 2018. Measuring uncertainty and its impact on the economy. Review of Economics and Statistics 100, 799–815.

Chan, J.C., Grant, A.L., 2016. Modeling energy price dynamics: GARCH versus stochastic volatility. Energy Economics 54, 182–189.

Chib, S., 1996. Calculating posterior distributions and modal estimates in markov mixture models. Journal of Econometrics 75, 79–97.

Christiano, L.J., Motto, R., Rostagno, M., 2014. Risk shocks. American Economic Review 104, 27–65.

Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85–110.

Conrad, C., Kleen, O., 2020. Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models. Journal of Applied Econometrics 35, 19–45.

Conrad, C., Loch, K., 2015. The variance risk premium and fundamental uncertainty. Economics Letters 132, 56–60.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Danielsson, J., James, K.R., Valenzuela, M., Zer, I., 2016. Model risk of risk models. Journal of Financial Stability 23, 79–91.

Engle, R.F., 1983. Estimates of the variance of US inflation based upon the ARCH model. Journal of Money, Credit and Banking 15, 286–301.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Frühwirth-Schnatter, S., Malsiner-Walli, G., 2019. From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. Advances in Data Analysis and Classification 13, 33–64.

Frühwirth-Schnatter, S., Malsiner-Walli, G., Grün, B., 2020. Dynamic mixtures of finite mixtures and telescoping sampling. Working paper.

Gao, G.P., Lu, X., Song, Z., Yan, H., 2019. Disagreement beta. Journal of Monetary Economics 107, 96–113.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gonzales-Rivera, G., 1998. Smooth transition GARCH models. Studies in Nonlinear Dynamics and Econometrics 3, 61–78.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. Biometrika 82, 711–732.

Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. Review of Financial Studies 6, 473–506.
Harvey, A., Sucarrat, G., 2014. EGARCH models with fat tails, skewness and leverage. Computational Statistics & Data Analysis 76, 320–338.
Heber, G., Lunde, A., Shephard, N., Sheppard, K.K., 2009. Oxford-Man Institute's realized library. University of Oxford.
Hosszejni, D., Kastner, G., 2020. Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. Forthcoming inJournal of Statistical Software.
Jurado, K., Ludvigson, S.C., Ng, S., 2015. Measuring uncertainty. American Economic Review 105, 1177–1216.
Kandel, E., Pearson, N.D., 1995. Differential interpretation of public signals and trade in speculative markets. Journal of Political Economy 103, 831–872.
Lahiri, K., Sheng, X., 2010. Measuring forecast uncertainty by disagreement: The missing link. Journal of Applied Econometrics 25, 514–538.
Lin, A., Zhang, Y., Heng, J., Allsop, S., Tye, K.M., Jacob, P., Ba, D., 2018. Clustering time series with nonlinear dynamics: A Bayesian non-parametric and particle-based approach. Working paper.
Longerstaey, J., Spencer, M., 1996. Riskmetrics — Technical document. Morgan Guaranty Trust Company of New York: New York 51, 54.
Malsiner-Walli, G., Frühwirth-Schnatter, S., Grün, B., 2016. Model-based clustering based on sparse finite Gaussian mixtures. Statistics and Computing 26, 303–324.
Miller, J.W., Harrison, M.T., 2018. Mixture models with a prior on the number of components. Journal of the American Statistical Association 113, 340–356.
Segal, G., Shaliastovich, I., Yaron, A., 2015. Good and bad uncertainty: Macroeconomic and financial market implications. Journal of Financial Economics 117, 369–397.
Song, Y., 2014. Modelling regime switching and structural breaks with an infinite hidden Markov Model. Journal of Applied Econometrics 25, 825–842.
Sucarrat, G., 2015. lgarch: Simulation and estimation of Log-GARCH models. R package.
Taylor, S., 1994. Modeling stochastic volatility: A review and comparative study. Mathematical Finance 4, 183–204.
Verhoeven, P., McAleer, M., 2004. Fat tails and asymmetry in financial volatility models. Mathematics and Computers in Simulation 64, 351–361.