



# Modeling intraday volatility of European bond markets: A data filtering application<sup>☆</sup>

Hanyu Zhang<sup>a,\*</sup>, Alfonso Dufour<sup>b</sup>

<sup>a</sup> Smurfit Business School, University College Dublin, Ireland

<sup>b</sup> ICMA Centre, Henley Business School, University of Reading, Whiteknights, Reading RG6 6BA, UK



## ARTICLE INFO

### Keywords:

Intraday GARCH  
European Bond Markets  
Data Filters

### JEL classification:

C22  
G10  
G15

## ABSTRACT

This paper studies the intraday volatility of European government bonds under the framework of the multiplicative component GARCH model (Engle and Sokalska, 2012). Intraday return volatility is specified as the product of daily volatility, intraday seasonality, and a unit GARCH process. The model is applied to 10-year European government bonds during the sovereign debt crisis. We observe large transitory intraday volatility often due to illiquidity effects and outliers. We suggest a flexible and effective procedure for jointly filtering mid-quote prices and estimating volatility models. Finally, we show that intraday data contain relevant information for daily volatility forecasts.

## 1. Introduction

With the advent of the sovereign debt crisis raging through Europe, government bond volatility becomes a greater concern to researchers, regulators and practitioners. The study of interest rate volatility which is important for bond volatility dates back to the earlier studies of affine models. Longstaff and Schwartz (1992) are among the first to suggest that yield change volatility is an important factor in explaining the term structure of interest rates. The roles and features of bond market volatility have been explored in numerous papers. Blume, Keim, and Patel (1991) investigate volatility risk of junk bonds relative to long-term government securities. Jones, Lamont, and Lumsdaine (1998) examine macroeconomic news effect on daily volatility and find different responses to a broad range of news using a GARCH(1,1) model (Bollerslev, 1986). De Goeij and Marquering (2004) estimate a multivariate model for bond and stock conditional variance using weekly data. Christiansen (2007) uses a GARCH model to study European bond markets before and after the introduction of the Euro and observes a substantial volatility spill-over effect from the aggregate European bond market to national markets.

High-frequency volatility remains less studied in contrast to the vast literature on daily and weekly volatility models (see Bollerslev, Chou, & Kroner, 1992; Poon & Granger, 2003). Taylor and Xu (1997) build a general ARCH model using hourly option returns and subsequently

compare the information content of conditional variance, realized variance and implied volatility. Fleming and Lopez (1999) estimate a multivariate GARCH model on hourly returns for the US Treasury bond interdealer market. Bollerslev, Cai, and Song (2000) adopt the flexible Fourier form (FFF) to model intraday seasonality and explicitly account for the macroeconomic news impact on 5-minute US Treasury bond futures volatility. They find long-memory effects and estimate an MA(1)-FIGARCH(1,d,1) model (Baillie, Bollerslev, & Mikkelsen, 1996) to forecast the daily variance. Deo, Hurvich, and Lu (2006) propose a long-memory stochastic volatility model and evaluate its forecasting performance against the component GARCH and ARFIMA (1,d,0) models. They introduce a gradually changing seasonal pattern to improve the forecasting performance of the model. Engle and Sokalska (2012) focus on the forecasting performance of a multiplicative component intraday GARCH model estimated for a large universe of US stocks.

The literature on the European debt crisis focuses on the spillover and contagion effects (see e.g., De Bruyckere, Gerhardt, Schepens, & Vander Vennet, 2013; Ehrmann & Fratzscher, 2017). Dimitriou, Kenourgios, and Simos (2013) use the Dynamic Conditional Correlation model (DCC) developed by Engle (2002a) to estimate time-varying volatilities and correlations of emerging stock markets. Although in this paper we estimate individual bond volatilities, a multivariate extension of our approach could be employed to explore and test contagion

<sup>☆</sup> We thank Professor Michael Clements, the editor Professor Brian Lucey, and two anonymous referees for their helpful comments, which greatly improve the article.

\* Corresponding author.

E-mail addresses: [hanyu.zhang@ucd.ie](mailto:hanyu.zhang@ucd.ie) (H. Zhang), [a.dufour@icmacentre.ac.uk](mailto:a.dufour@icmacentre.ac.uk) (A. Dufour).

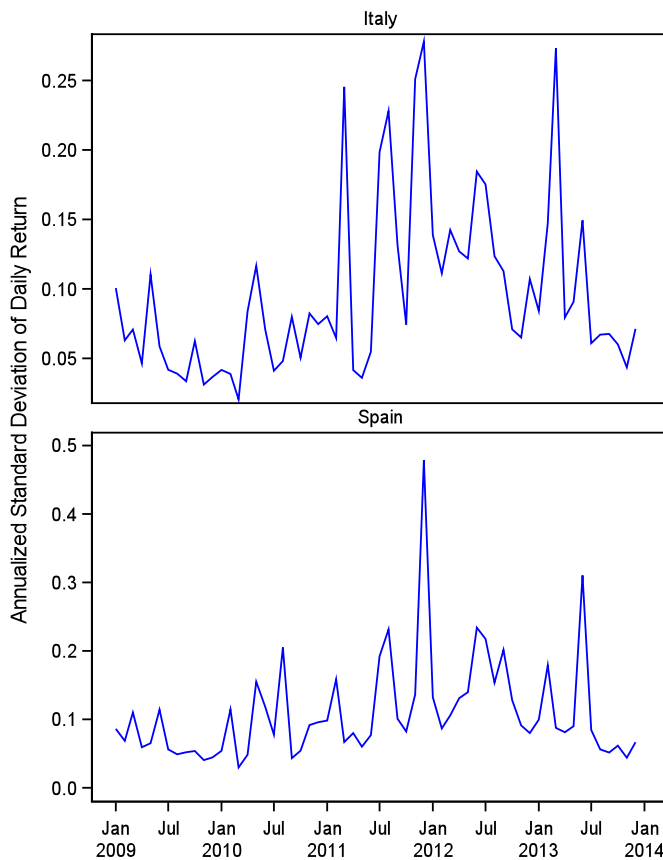


Fig. 1. Annualized standard deviations of daily returns for 10-year Italian and Spanish government bonds.

effects.

Almgren and Chriss (2001) develop a framework for computing optimal trade execution strategies which balance trade timing risk and impact cost. In their model the variance of asset returns affects the variance of trade execution costs. Engle and Ferstenberg (2007) extend this approach and show how to apply it to both optimal portfolio selection and portfolio liquidation. Again, asset volatility is an essential determinant of the joint asset selection and liquidation problems. Asset purchases and sales are performed throughout the trading day and thus, it is important to develop an intraday volatility model which can be used to inform optimal trading decisions. A recent paper by Bollerslev, Hood, Huss, and Pedersen (2018) emphasizes the importance of being able to predict and control volatility for portfolio managers who implement risk targeting strategies. The authors utilize an extensive range of high-frequency datasets and apply a series of data filters to clean out outliers induced by illiquidity of financial assets.

As shown in Fig. 1, European sovereign debt markets exhibited high volatility levels during the sovereign debt crisis of 2010–12 when the European Central Bank (ECB) was forced to intervene with several ad-hoc measures for stabilizing treasury yields. In particular, the volatilities of Italian and Spanish government bond returns sharply increased in May 2010 and showed annualized daily volatilities of 20–25 % with a peak of 50% for Spanish bonds towards the end of 2011. Government bond volatility remained at high levels for large part of 2013. These are volatility levels commonly observed for blue-chip stocks but they are unusual for government bonds.

The selected sample provides the justification for developing an appropriate bond volatility model and it offers the ideal conditions for testing its effectiveness. The three fundamental questions we want to address in this study are: How can we accurately quantify the short-term fluctuations in bond returns? How can we properly filter out the

temporary effects of liquidity dynamics on volatility models? Is intraday volatility important for predicting future daily volatility? Our study considers the European bond crisis period because it offers a unique opportunity for developing and testing a volatility model which can be used by portfolio managers and traders to control for and manage extreme, intraday bond risk.

We adopt the framework of Engle and Sokalska (2012) and develop a volatility modeling approach which simultaneously addresses the problems of filtering transitory liquidity effects, modeling intraday periodicity and estimating fundamental intraday volatility. We model the intraday periodicity as a piecewise linear structure in the spirit of the Spline-GARCH (Engle & Rangel, 2008) model. The daily volatility dynamics are captured by a simple GARCH(1,1) model. Our findings further improve our understanding of the European bond market during the sovereign debt crisis and lay the foundation of further work on modeling the time-varying contagion effect when the debt of distressed countries is no longer a safe asset, with serious repercussions for the whole economic environment. We study the volatility of benchmark 10-year bonds for 7 Euro area countries. With our sample, the dynamics of liquidity are paramount for understanding the short-term volatility of quoted prices and this poses a challenge to researchers trying to disentangle transitory versus fundamental volatility. It is well documented in the literature that high frequency data often contain noise and short-term effects due to frictions and liquidity imbalances (Fleming, 2001; Bandi & Russell, 2008), which make proper data cleaning both necessary and challenging. Obviously, the data cleaning/sampling process will affect the computation of fundamental volatility (see Bandi & Russell, 2008). Hence it is important to jointly address the data filtering and the volatility estimation problems. We consider several alternative data cleaning techniques and develop a procedure for choosing the filter which provides the best estimates of fundamental volatility. Finally, many papers from the realized volatility literature (see Barndorff-Nielsen & Shephard, 2002a; Barndorff-Nielsen & Shephard, 2002b) suggest that intraday data may contain information that is helpful for estimating volatility at longer horizons. Our results provide further evidence that intraday data can help in improving the forecasts of daily volatility.

The rest of the paper is organized as follows. Section 2 introduces the motivations and properties of our econometric high frequency model. Section 3 explains our method for cleaning the time series of bond data. The Appendix section explains the details of how we construct the return series. Section 4 presents the estimation results and interpretations. Section 5 carries out the forecasting comparison between the intraday GARCH and the daily GARCH(1,1). Finally Section 6 summarizes our findings.

## 2. A multiplicative error model of intraday volatility

We denote the intraday log return by  $r_{t,n}$  and the daily return by  $r_t$ .  $t$  represents the daily index ( $t = 1, 2, \dots, T$ ) and  $n$  is the intraday index ( $n = 1, 2, \dots, N$ ). Each intraday time interval  $n$  is referred to as “bin”  $n$ . The log return  $r_{t,n}$  is calculated as the difference in log mid-quote prices in a limit order market with designated market makers.

The multiplicative error model introduced by Engle (2002b) and adopted by Engle and Sokalska (2012) suggests that

$$r_{t,n} = \sqrt{h_t s_n q_{t,n}} \varepsilon_{t,n} \quad \text{and} \quad \varepsilon_{t,n} | \Phi_{t,n-1} \sim N(0, 1) \quad (1)$$

where  $h_t$  is the daily variance component

$s_n$  is the intraday periodicity or diurnal component

$q_{t,n}$  is the intraday variance component with  $E(q_{t,n}) = 1$

$\varepsilon_{t,n}$  is an error term

$\Phi_{t,n-1}$  denotes the set containing all the available information up to the bin preceding the current time interval. To avoid any confusion, we will refer in the subsequent analysis to the volatility of  $r_{t,n}$  as intraday return volatility and  $q_{t,n}$  as intraday volatility. Here we assume that the

conditional distribution of the error term is standard normal, but this does not imply a normal distribution of returns. The overnight return  $r_{t,0}$  is neither used for estimating the diurnal component nor the intraday variance component. However, the overnight return affects daily returns and thus the daily variance component.

## 2.1. Daily model

Andersen and Bollerslev (1998) find a close relationship between intraday volatility, computed as the cumulative absolute intraday return, and the one-step ahead daily volatility forecast, computed using an MA(1)-GARCH(1,1). The daily conditional variance forecast, which is not affected by short-term intraday volatility dynamics, represents a certain amount of *anticipated* intraday return variation. Failing to capture this lower-frequency component would distort the overall volatility computation. Hansen and Lunde (2005) confirm the superior predictive ability of the GARCH(1,1) model against more than 300 specifications for the daily conditional variance of foreign exchange rates. As the forex market has a very similar structure to the sovereign bond market we study, we choose the GARCH(1,1) model as our forecast model for daily conditional variance. Our model can be seen as an extension of the GJR GARCH model which is a simple and tractable type of asymmetric GARCH model. More sophisticated asymmetric volatility models include Nelson (1991)'s EGARCH, and Creal, Koopman, and Lucas (2013)'s GAS model which was recently applied in Apergis, Barunfk, and Lau (2017) and Wu, Lau, and Vigne (2017).

During the crisis, sovereign bond volatility was affected by the ECB's actions through a series of interventions. The Security Market Programme (SMP) was announced on May 10, 2010 along with several long-term refinancing operations (LTRO) aimed at alleviating the heightened market tension. The programme was described as “*interventions in the euro area public and private debt securities markets to ensure depth and liquidity in those segments which are dysfunctional*”.<sup>1</sup> With the first SMP the ECB purchased the government bonds of Greece, Ireland and Portugal and a second SMP was implemented to buy Italian and Spanish government bonds. The second SMP was announced on August 7, 2011<sup>2</sup> and on the following day, the price of the Italian 10-year bond jumped by €5.7 to €96.32. In a press release on February 21, 2013, the ECB disclosed the total amount of bonds acquired under the SMP and Italian and Spanish bonds accounted for two-thirds of those purchases.<sup>3</sup>

As mentioned above, Ghysels, Idier, Manganelli, and Vergote (2017) VAR model with errors following a GARCH(1,1) process to assess the effect of the ECB bond purchases during the SMP interventions. We adopt the same approach to control for the SMP effects when estimating volatility during the SMP window.

Only the first two lags of returns are included in the conditional mean equations, as indicated by the t-test on the coefficients and by the Schwartz information criterion (BIC). The daily model is estimated via maximum likelihood. Specifically,

$$r_t = c_1 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \sum_{p=1}^4 d_p * dummy_p + v_t \quad v_t | \mathcal{F}_{t-1} \sim N(0, h_t) \quad (2)$$

$$h_t = w + (a_1 + a_2 * I(SMP_{t-1} > 0)) v_{t-1}^2 + (b_1 + b_2 * I(SMP_{t-1} > 0)) h_{t-1} \quad (3)$$

$$I(SMP_{t-1}) = \begin{cases} 1 & \text{if purchasing amount} > 0 \text{ at } t-1 \\ 0 & \text{if purchasing amount} = 0 \end{cases}$$

The dummy  $I(SMP_{t-1})$  controls for the high level of volatility for Italian and Spanish bonds during the period from August 08, 2011 to March

09, 2012 which prompted the ECB intervention with the second round of the SMP. We assume that the SMP is active for the entire week as long as the weekly amount purchased by the ECB is positive. We also adopt a series of dummy variables in the conditional mean equations to control for specific news corresponding to four dates with large daily returns caused by institutional announcements.<sup>4</sup> The first 3 dummy variables are for Italian and Spanish government bonds, while  $dummy_4$  is for Italian bonds only.  $dummy_1$  controls for the big drop in returns on May 06, 2010 when the ECB maintained its base rate unchanged with no action with respect to the Greek debt crisis.<sup>5</sup>  $dummy_2$  and  $dummy_3$  capture the two jumps in bond prices due to the activation of the SMP (see above).  $dummy_4$  controls for the return of December 05, 2011 when former Italian Prime Minister Monti announced budget cut plans and Italian markets witnessed a big rally.

## 2.2. Intraday seasonal pattern and volatility

Obviously, we expect all bond returns of a particular day to have the same daily variance component. We then assume that intraday volatility has a seasonal or diurnal component, which captures the periodic time of day effect, and intraday innovations around this diurnal component. A spline model is used for the diurnal component and a unit GARCH(1,1) model (i.e. the unconditional variance is 1) for the intraday innovations. Our intraday model is implemented as follows:

$$s_n = \delta_0 * \exp \left( \sum_{j=1}^m \delta_j (\Delta_n - k_{j-1})_+ \right) \quad (4)$$

$$q_{t,n} = 1 - \alpha - \beta + \alpha \left( \frac{(r_{t,n-1})^2}{s_{n-1} h_t} \right) + \beta q_{t,n-1} \quad (5)$$

$$(\Delta_n - k_{j-1})_+ = \begin{cases} (\Delta_n - k_{j-1}) & \text{if } \Delta_n > k_{j-1} \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta_n = \frac{n}{N} \quad n = 0, 1, \dots, N.$$

This specification has the advantage of estimating the intraday volatility and the diurnal component jointly and eliminates the need for a two-step estimation. In the original framework of Engle and Sokalska (2012), intraday seasonality is estimated in a separate step with a simple average of returns for every interval of the trading day. It can be shown that the statistical properties of a two-step estimator can be derived from the Generalized Method of Moment (GMM) by Newey and McFadden (1994). But there is an efficiency loss in the parameter estimation of the second step. Furthermore, the linear spline of Eq. (4) has reduced the number of parameters substantially as compared to the original model. On the other hand, while Engle and Sokalska (2012) utilize a commercial forecast of daily volatility, we need to make a one-step-ahead forecast of daily conditional variance first. The consistency of the estimators in Eqs. (4) and (5) still holds according to the argument in the Appendix of Andersen and Bollerslev (1998) while the possible autocorrelations and heteroskedasticity caused by including the daily GARCH volatility forecast should be adjusted. The autocorrelations can be alleviated by a longer sampling interval and heteroskedasticity is naturally controlled by the unit GARCH.

The exponential form in Eq. (4) guarantees the positivity of the diurnal component.  $k_0 = 0$ ,  $k_1 = \frac{6}{N}$ ,  $k_2 = \frac{12}{N}$ ,  $k_j = \frac{j*6}{N}$ , ...,  $k_m = 1$  denote knots in the linear spline. The knots are set respectively at 9:00, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 16:00, 17:00 and 17:30 (official closing time) for Belgium, Germany, Italy and Spain. Three nodes at 11:00, 12:00 and 13:00 are omitted for Austria, France and the

<sup>1</sup> See, [www.ecb.europa.eu/press/pr/date/2010/html/pr100510.en.html](http://www.ecb.europa.eu/press/pr/date/2010/html/pr100510.en.html).

<sup>2</sup> See, [www.ecb.europa.eu/press/pr/date/2011/html/pr110807.en.html](http://www.ecb.europa.eu/press/pr/date/2011/html/pr110807.en.html).

<sup>3</sup> See, [www.ecb.europa.eu/press/pr/date/2013/html/pr130221\\_1.en.html](http://www.ecb.europa.eu/press/pr/date/2013/html/pr130221_1.en.html).

<sup>4</sup> Controlling for one-time event with dummy variables is a common approach in volatility analysis, e.g. Andersen and Bollerslev (1998).

<sup>5</sup> See, [www.ecb.europa.eu/press/pr/date/2010/html/pr100506.en.html](http://www.ecb.europa.eu/press/pr/date/2010/html/pr100506.en.html).

Netherlands because the estimation of the exponential spline makes the optimization algorithm difficult to converge for these three countries' data. As can be seen in Fig. 7 in Section 4, volatility stays low in the middle of the day for all major European countries. So we choose to remove the knots during the 11:00–13:00 interval when the return does not vary significantly. The spline we use is different from Engle and Rangel (2008) in terms of functional form and purpose. Their quadratic spline coupled with exogenous variables aims to incorporate the low-frequency volatility related to the macroeconomic environment. While our linear spline has the same frequency as the intraday volatility and we assume it is not affected by exogenous variables. Obviously, this could easily be extended to allow exogenous factors to affect the diurnal pattern, for example if we wanted to distinguish information days, with relatively higher trading intensity, from normal days. Notice that  $E(q_{t,n}) = 1$  implies that the unconditional variance of the stochastic intraday volatility component is one. Hence, the unconditional variance of high frequency returns is entirely dependent on the unconditional daily variance and the diurnal component, i.e.

$$E[(r_{t,n})^2] = s_n E(h_t). \quad (6)$$

### 3. Data and cleaning procedures

Our high frequency data contain 10-minute log returns constructed from the quote mid-points for ten-year benchmark government bonds from the MTS interdealer market. The intraday data runs from April 02, 2012 to December 30, 2013. We rely on a longer time series of daily data from 02 January, 2009 through December 30, 2013 to estimate the daily volatility component.<sup>6</sup>

We focus on benchmark ten-year bonds issued by major Euro-zone countries including Austria, Belgium, France, Germany, Italy, the Netherlands and Spain. Since we concentrate on one maturity category, we choose on-the-run 10-year bonds defined as long-term bonds with a remaining time to maturity ranging from 8.5 years to 11.5 years. The details of variable construction are described in the Appendix.

#### 3.1. Institutional details

MTS is an electronic trading platform where unique counterparties trade various fixed-income securities including European government bonds, quasi-government bonds, corporate bonds and repurchase agreements. Here we describe the market features that are most relevant for our analysis. Detailed information on the MTS market structure is given in Darbha and Dufour (2013). There are two parallel platforms for benchmark bonds: the MTS domestic markets devoted to trading domestic bonds and the EuroMTS market where all benchmark securities across countries can be traded. Each platform has its own features in terms of trading rules, market participants, and market makers. The database has information on all changes in the best three quotes either in the ask side or in the bid side of the order book. Quote changes are due to either changes in the quote prices or in the quote sizes. The quote changes may reflect revisions to existing quotes, the arrival of new limit orders or trades. Price discrepancies for the same bond due to the parallel trading structure can be eliminated by traders with access to both markets. Cheung, de Jong, and Rindi (2005) find that the liquidity conditions on domestic markets are very similar to those observed on the EuroMTS market. Market makers are obliged to post two-way quotes called “proposals” for the securities which are assigned to them by MTS. The limit orders they submit must satisfy a series of conditions including a minimum volume varying from €2.5 to €10 million, and a minimum tick value. Before 2007, market makers were required to post quotes for a minimum period of the trading day

**Table 1**

Data preparation. This Table presents the various steps followed in the data preparation process and provides the number of raw observations lost at each step. The quote updates recorded outside the trading hours (8:15–17:30 CET) are deleted. All ticks on 22 Oct 2009 of all countries are excluded because multiple transactions were recorded at various prices after the last quote was recorded. Simultaneous ticks due to parallel tradings are identified and adjusted. The best available bid and ask quotes are selected. The observations with negative spreads are also dropped. Finally, we only keep the changes to the best bid and ask quotes.

Operation	No. of obs.	Deleted obs. (%)
Number of raw observations	13,772,614	0
Quotes outside trading hours	184,407	1.3389
Quotes on 22 Oct 2009	9467	0.0687
Simultaneous quotes	4,241,931	30.7998
Negative spreads	2194	0.0159
Unchanged bid and ask prices	3,705,491	26.9048
Final sample size	5,629,291	40.8731

and for a maximum spread. During the 2007 financial crisis, MTS relaxed market makers' dealing obligations recognizing that market makers were facing higher liquidity and credit risks. Now MTS tracks the average length of quoting time and the average spreads pertaining to an individual market maker and requires that these averages are consistent with the market averages derived from all market makers. Trading is possible from 8:15 to 17:30 CET.

#### 3.2. Data preparation and filter evaluation

We follow a series of steps to prepare the sample data, which covers the period from the start of 2009 to the end of 2013, for the analysis.<sup>7</sup> Table 1 below presents the number of observations deleted at each step of the preparation process before applying any filtering method. Firstly, we remove the quotes recorded outside the trading hours.<sup>8</sup> Following Fleming (2001), all quotes on October 22, 2009 are excluded from our dataset because the last quote update on that day was recorded at 15:26 and there were multiple transactions happening at different prices afterwards. Secondly, we compute the global best bid and offer prices across the two platforms for each country. Due to the parallel status of domestic MTS and the EuroMTS platforms, quotes are often updated simultaneously on both platforms with recorded time stamps differing by a few milliseconds. The adjustment is made for the delay and the overall best available quotes are computed from the simultaneous ticks. We also remove quotes with negative spreads and only keep the changes in the best bid and ask prices. Thirdly, we apply a range of filters to remove temporary illiquidity effects and choose the best filter for each country. The procedure for the selection of the best filter is explained below. Finally, the longer daily sample and the 10-minute sample are generated from the prepared data. The daily return is calculated as 100 times the log difference of 5 PM quote midpoints extracted from the intraday data. The use of quote mid-points is discussed in Hasbrouck (1991) and can alleviate the temporary autocorrelation induced by any bid-ask bounce. The reason we select the 5 PM mid-quote price instead of the closing one at 5:30 PM is that the quoting activity is less intensive for some days towards the end of the trading day. The final quote updates sometimes appear considerably earlier than 5:30 PM and thus the closing prices are often stale. The 10-minute returns are dropped between 8:15 and 8:50 AM. The first 10-minute return is computed as the log difference between quotes from 8:50 AM to 9:00 AM because daily quote submission is often delayed at the

<sup>7</sup> The cleaning covers all the daily and intraday sample as it helps us to estimate both models more accurately.

<sup>8</sup> Some pre-market quotes and post-market settlements are stored in the data set.

<sup>6</sup> In the overlapping period of intraday and daily data, the daily volatility is computed as a one-step-ahead forecast.



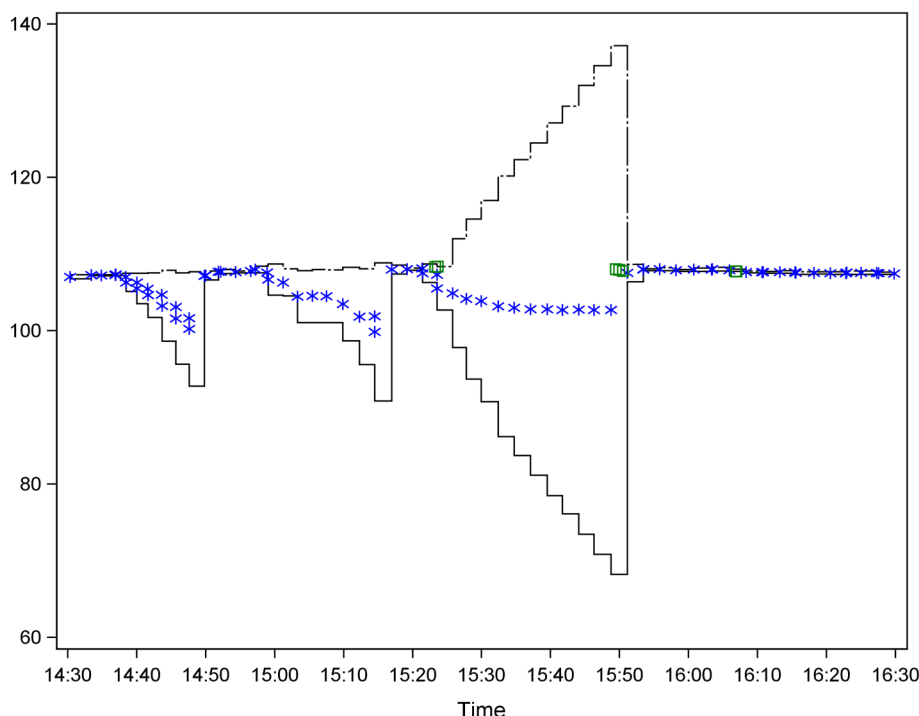


Fig. 2. Plot of the best quotes for a 10-year benchmark French government bond (ISIN code: FR0011196856) on June 01, 2012 from 14:30:00 to 16:30:00. Tick-by-tick mid-quote prices (stars), transaction prices (squares), best available bid prices (solid line), best available ask prices (dashed line).

beginning of the trading day.

With the increasing frequency of financial data production and recording, researchers face the challenging task of separating relevant data from noise and odd entries due to specific market regulations and frictions. For example, market makers operating on MTS platforms are obliged to keep their quotes on the system even when they have satisfied their quoting obligations. At times, this results in very large spreads which simply indicate to the market that dealers have temporarily withdrawn their competitive quotes. No rational traders would trade at these quotes. Possible causes include macro news announcements, unscheduled ECB interventions on debt markets, human errors and holiday effects (Fleming, 2001). Attention must be paid to distinguishing transitory volatility due to illiquidity effects. A 3% jump in log returns is plausible if some macro news is released. A large jump would be suspicious in the absence of any observable information, especially when there are no trades. In fact, a temporary shortage of liquidity on one side of the order book may lead to asymmetric changes in the best quotes which would result in changes in mid-quote prices. Filtering is a way of identifying and controlling for abnormal outliers. The temporary volatility caused by illiquidity is best illustrated by Fig. 2.

Transactions are unlikely to occur when liquidity evaporates and the quoted price may be extreme (the literature often refers to these extreme quotes as stub quotes).<sup>9</sup> We provide two examples of odd quotes on the MTS platforms and their effect on mid-quote prices. In Fig. 2 we show that the dynamics of bid, ask and mid-quote prices for the 10-year French government bond on June 01, 2012. From 14:38:28 to 15:16:57, the bid price gradually moved away from the offer and reached a low level of 90 twice while the offer price barely changed. From 15:21:24 until 15:51:08, the two sides of the order book deviated from each other. In both cases, the mid-quote was affected by the temporary illiquidity and the volatility of the mid-quote price was artificially increased by the sudden adjustment in the bid and ask quotes at the end of the two periods. A preliminary analysis which involves

sampling the data from every 2 s to every 2 h<sup>10</sup> for this particular day reveals that the daily realized variance can reach an annualized level as high as 200%! Even when we reduce the sampling frequency to every 50 minutes, the realized variance is still above 50%. Obviously, using intraday data without filtering is not an appropriate way to study volatility for this market. The asymmetric dynamics in bid and ask quotes are also observed in other markets (see for example, Hasbrouck, 2018; Engle & Russell, 1998 for the stock market). Interestingly, in the MTS markets we often observe a gradual deterioration in liquidity on one side of the market which is then followed by a prompt recovery. This is the opposite of what the literature describes as the effect of a trade on prices (see for example Hasbrouck (1991) for the equity market). The immediate effect of a trade is then gradually, although partially, reverted. Many markets implement trading halts when there are periods of extreme uncertainty or illiquidity and traders are not allowed to trade. The MTS market, however, is not halted when there is scarce liquidity. Normally, MTS traders will not trade on the side of the order book displaying extreme quotes. Similarly to these traders, the empirical researcher trying to assess the fundamental volatility of an asset has to choose an optimal sampling procedure which filters out temporary illiquidity effects (see Bandi & Russell, 2008).

To establish the benchmark when evaluating various cleaning procedures, we resort to the study of Bandi and Russell (2008). They prove that the microstructure noise, which causes transitory volatility, heavily influences the estimation of the fundamental volatility. The optimal sampling frequency should minimize the mean squared error (MSE) of the realized variance against the true variance tick-by-tick returns, which in their case are assumed to follow an MA(1) process. The true daily integrated variance is approximated by the realized variance of 15-minute squared returns. In the spirit of Bandi and Russell (2008) and in the interest of studying fundamental government bond volatility, we propose a method to find the best filtering procedure, which minimizes

<sup>9</sup> See, Kirilenko, Kyle, Samadi, and Tuzun (2017).

<sup>10</sup> We increment the sampling interval by 1 s every time and recompute the daily realized variance as the sum of squared log returns.

the effect of short-term frictions and noise when modeling and estimating bond volatility. We utilize the concept of the MSE and try to minimize the average difference between the daily summation of conditional variance of 10-minute return and realized variance derived from 2-hour returns. Specifically, let  $V_t$  denote the true daily integrated variance. The MSE  $E(\sum_{n=1}^N h_t s_n q_{t,n} - V_t)^2$  is estimated by  $\frac{1}{T} \sum_{t=1}^T (\sum_{n=1}^N \hat{h}_t \hat{s}_n \hat{q}_{t,n} - \hat{V}_t)^2$  and the best filter should minimize this criterion. Our benchmark realized variance is a model-free measure of fundamental volatility. The 2-hour sampling interval is conservative enough to avoid including liquidity effects in the realized variance as those observed in Fig. 2. We assume that the returns of daily and intraday frequency follow a GARCH process. It should be emphasized that the benchmark realized variance is computed from unfiltered data.

### 3.3. Choosing filters

We now turn to describing the filtering methods. The methods can be categorized into three main groups. (A) Filters based on “maximum tradable spreads”; (B) filters using “spread quantiles” and (C) “local window” filters. The filters belonging to group (A) use both quotes and trades to determine the largest acceptable spread which is defined as the largest spread when trades occur. This filter was developed for a market like MTS where market makers are obliged to maintain their quotes on the system even when they are not required to make the market. At times, the spread is so wide that no traders are willing to execute transactions at the posted quotes (stub quotes). The market is open, but the bid-ask spread is so large that it is de facto halted. The filters in group (B) simply compute quantiles of quoted percentage spread and, again, discard quotes with spread above a certain threshold. Trimming extreme observations is often implemented to check the robustness of estimation results to outliers. Group (C) comprises filters based on the local properties of the observations. All the quotes identified as outliers are replaced with the most recent valid ones. The filtered series have the same number of observations across filters in order to compare them fairly. For the first two group of filters, we do not discard any observations with percentage spreads less than

50 basis points regardless of the corresponding threshold.

A. The “maximum tradable spread” approach matches trades with their immediately preceding quotes in order to determine the maximum percentage spread prevailing when trades occur. Due to the sparsity of bond transactions, we filter the data based on maximum tradable spreads of the entire sample for 7 countries. The percentage spreads associated with actual transactions can be reasonably expected to be tradable. The percentage spread, which is computed as the bid-ask spread divided by the mid-quote price, facilitates the comparison of different filters across assets. Filtering based on bid-ask spreads seems a natural choice, given it is a measure of the liquidity and quality of the market and market data (Hasbrouck, 1993). This approach brings trades and quotes together and relies on the economic meaning of percentage spread. Harris (2002) illustrates that the posted spread represents a measure of transaction cost, which traders tend to minimize by searching for smaller spreads. Venues with high frequency of trades and quote revisions have often tight bid-ask spreads because of the fierce competition among liquidity providers. Moreover, matching trades with quotes can be a way of identifying erroneous trades (Barndorff-Nielsen, Hansen, Lunde, & Shephard, 2009). Hence using additional information about MTS trades could be appropriate for filtering the time series of quote midpoints.

On the other hand, this method may have some drawbacks. Sometimes, trades are executed at the midpoint of the spread even when the quoted spread is large. See for example in Fig. 2, trades executed at about 15:50. In this case the filtering method may choose a maximum tradable bid-ask spread that would be too large and hence it would not be very effective. Perhaps, a solution to this problem would be to set the maximum tradable spread only using trades that can be matched to prevailing quotes. However, this may lead to a loss of relevant information. Another potential issue with this method is that trades could be executed when liquidity is scarce only on one side of the order book. Notice, for example, the last execution in Fig. 3. This is a buyer-initiated trade which is executed at the ask

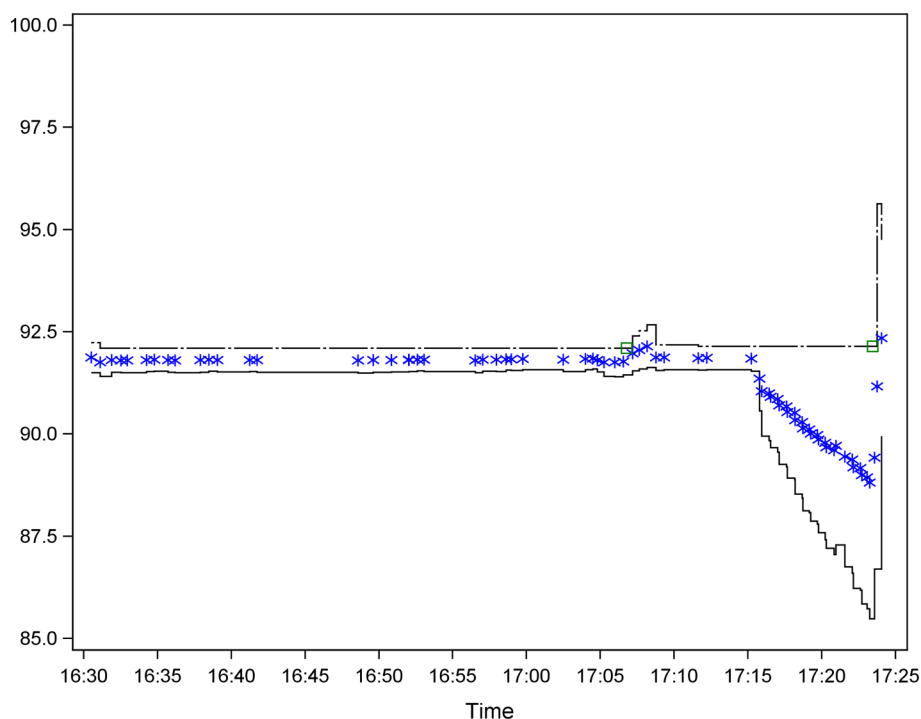


Fig. 3. Plot of the best quotes for a 10-year benchmark Spanish government bond (ISIN code: ES00000123B9) on November 25, 2011 from 16:30:00 to 17:25:00. Tick-by-tick mid-quote prices (stars), transaction prices (squares), best available bid prices (solid line), best available ask prices (dashed line).

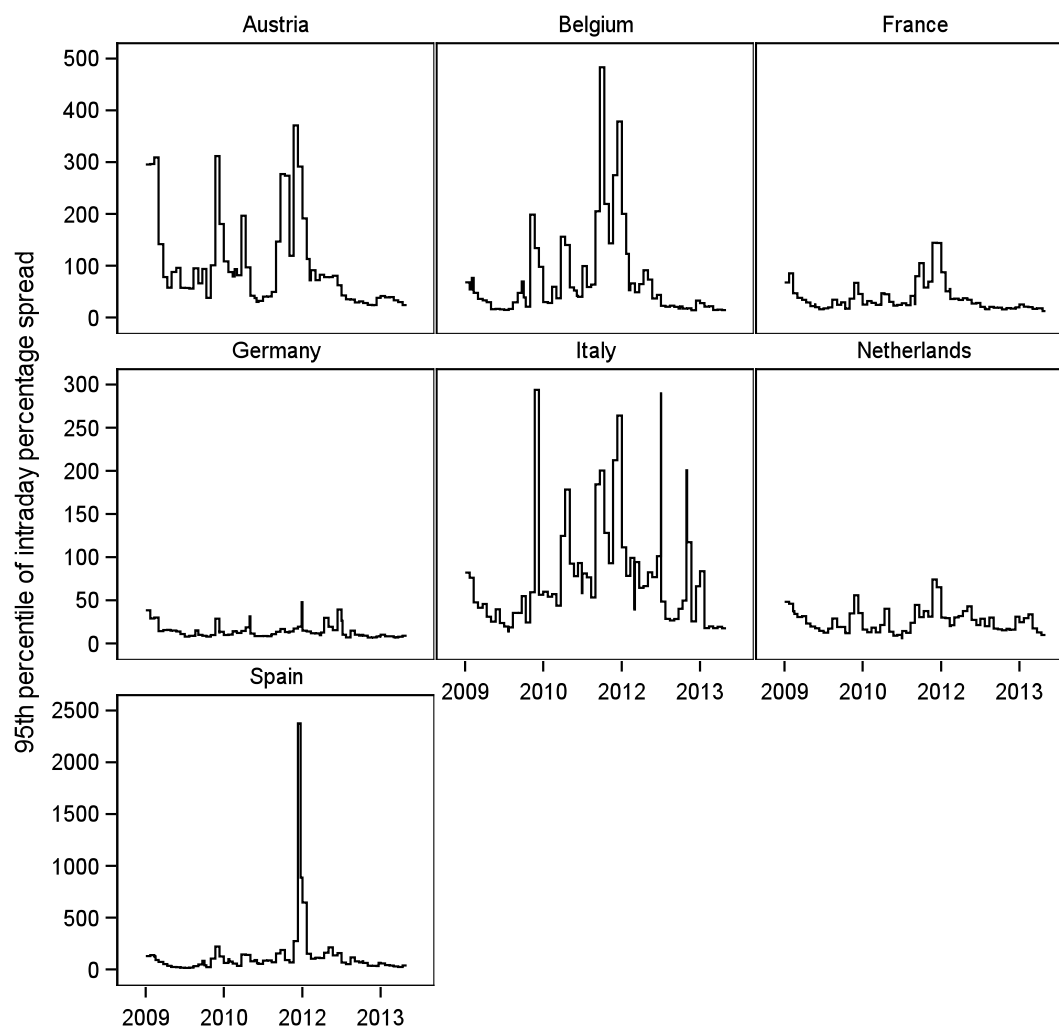


Fig. 4. 95th percentile of intraday percentage spread by countries. The percentile is drawn from the empirical distribution of the intraday percentage spread measured in basis points from the consolidated order book. The percentiles are real observations and not interpolated values. Notice the different scale of each row of each panel.

quote when the bid price is very low thus making the spread large. In this case, the presence of a trade does not support the assumption that both bid and ask quotes are tradable. Obviously, the ask quote offers a competitive price and it is tradable, but the bid quote offers conditions that traders are unlikely to accept, namely, it is not tradable. These examples cast some doubts on the reliability of the “maximum tradable spread” approach without further adjustments and prompts us to look at alternative statistics such as percentiles. The 99th percentile of all traded percentage spreads may potentially give a more reliable estimate of the threshold within which trades will probably be executed.

- B. Filters in the second group identify quote outliers using quantiles for percentage spreads. However, percentage spreads are not stable during the sovereign debt crisis. According to [Darbha and Dufour \(2013\)](#), the spreads of European government bonds have significantly increased during the crisis period. When defining the threshold for removing extreme percentage spreads, a successful filter needs to reflect the development of the liquidity conditions. Specifically, we first compute the 90–99th percentile of the empirical distribution of percentage spreads belonging to one bond. We then remove the quotes with a percentage spread larger than the percentile. To accommodate the dynamics in the liquidity conditions, the computation of the percentage spread quantiles and the filtering of outliers are performed each month. Arguably this approach is ad hoc but it is very simple to implement. Dropping any

predetermined amount of data is purely mechanical and has no economic significance. Additionally, it is unlikely that any particular percentile filter uniformly outperforms the others for all countries. Due to the limited space we cannot present the detailed figure of each percentile each month but the patterns of percentiles would be the same. [Fig. 4](#) shows the 95th percentile by countries. We can see that liquidity is very volatile during the sample period. Germany has Treasury markets with the lowest spreads: roughly 95% of the spreads are below 50 basis points. For other countries, the outliers of the percentage spread are present in many months. Surprisingly, even French bonds have nearly 5% of their quoted percentage spread well above 100 basis points in late 2011. Austria, Belgium and Italy all have large spreads for a considerable time from 2009 to 2012. Spanish bonds experience the worst liquidity during December 2011, when the 95th percentile reaches 2000 basis points.

- C. The third approach is applied directly to price series. Some of the bid prices or offer prices deviate substantially from the quotes around them. [Gençay, Dacorogna, Müller, Pictet, and Olsen \(2001\)](#) propose a technique of detecting outliers, called “adaptive filtering”. They suggest that a filter should learn from the series and develop its standard with a consideration of local properties. The same idea is also applied in two other papers, namely [Brownlees and Gallo \(2006\)](#) and [Barndorff-Nielsen et al. \(2009\)](#). [Brownlees and Gallo \(2006\)](#) devise a filter based on changes in transaction prices. The filter examines a local window of  $k$  trades near the current trade and

**Table 2**

Number of observations deleted by various methods. We apply 12 alternative filtering methods to the final sample data. These can be classified into three groups. The first group, which contains the maximum tradable spreads and 99th percentile of all tradable spreads, attempts to find a reliable threshold with the aid of transaction records. The second group, gathering the 7 percentiles of all percentage spreads, simply runs through the data month by month in order to ascertain outliers according to the empirical distributions of spreads. The third group, following the concept of local filtering, consists of two established methods which have been applied to stock data.

Filtering method	Number of observations deleted	Percentage removed (%)
Maximum tradable spread	219,246	3.8947
99th Percentile of tradable spreads	333,351	5.9271
97th Percentile of percentage spreads	98,597	1.7515
96th Percentile of percentage spreads	121,806	2.1638
95th Percentile of percentage spreads	142,513	2.5316
94th Percentile of percentage spreads	161,729	2.8730
93th Percentile of percentage spreads	179,681	3.1919
92th Percentile of percentage spreads	196,836	3.4966
91th Percentile of percentage spreads	212,915	3.7823
90th Percentile of percentage spreads	227,838	4.0474
Brownlees and Gallo (2006)'s method	11,046	0.20
Barndorff-Nielsen et al. (2009)'s method	5338	0.0948

computes the mean and variance of those trades after trimming the 10% tail values. We apply Brownlees and Gallo (2006)'s core method to mid-quote prices  $p_n$ . That is

$$(|p_n - \bar{p}_{-n}(k)| < 3\sigma_{-n}(k) + \gamma) \\ = \begin{cases} \text{True} & \text{observation } n \text{ is kept} \\ \text{False} & \text{observation } n \text{ is removed} \end{cases}$$

where  $\bar{p}_{-n}(k)$  and  $\sigma_{-n}(k)$  are, respectively, the  $\delta$ -trimmed mean and standard deviation of a length of  $k$  quotes around the current quote. The  $-n$  subscript indicates that we exclude the current observation from calculating the mean and standard deviation.  $\delta$  is kept as 10% and the  $k$  observations should belong to the same day as the current observation. Specifically, as in Brownlees and Gallo (2006), the local window of the first mid-quote price of a day should be the  $k$  quotes following it; the neighborhood of the last observation of a day is chosen as the  $k$  data points preceding it. In the middle of the day, we select the  $k/2$  points before and after the current observation.  $k$  and  $\gamma$  are set to 60 and 0.02 as in the original paper, respectively. Barndorff-Nielsen et al. (2009) apply a similar filtering procedure to the quotes of stocks. Barndorff-Nielsen et al. (2009) remove trades if the price change is larger than 50 times the median price change for the same day. In addition, the algorithm considers the average distance between the trade price and the median of the 50 trade prices in the neighborhood of the current price. It classifies as outliers observations where the distance between the trade price and the median of the 50 trade prices is greater than 10 times the average distance.

$$|p_n - \text{median}(p_{-n})| < 10 * \frac{1}{50} \sum_{j=1}^{50} |p_j - \text{median}(p_j)| \\ = \begin{cases} \text{True} & \text{observation } n \text{ is kept} \\ \text{False} & \text{observation } n \text{ is removed} \end{cases}$$

Intuitively, these two methods do well when there are only “a few” quotes heavily deviating from others. However, it is difficult for this approach to filter out outliers similar to those in Fig. 2 because the local property of current observation is distorted due to the persistent enlargement of bid-ask spread. Also the parameters for identifying outliers rely on the discretion of econometricians. Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009) choose parameters values related to the filters either through visual inspections or intensive experimentation, without evaluating them against a benchmark. More examples can be found from other microstructure papers. For instance, Fleming and Lopez, 1999, delete ticks with spreads larger than 50 times the median

spread for the day. Engle and Russell (1998) filter the bid and ask of the IBM stock based on a simple threshold. They observe some disassociation of the bid and ask changes, which causes the mid-price to vary temporarily. They decide 4 ticks to be the minimum amount of change for bid or offer price to trigger a genuine price movement. There are no apparent reasons why 50 or 4 is a proper choice for filtering. This further underlies the need for a systematic evaluation of all filters based on an explicit benchmark.

### 3.4. Cleaning result based on the benchmark

We attempt to remove the illiquidity effect by choosing the best filter which minimizes the distance between the fitted volatility and  $V_t$ , which is estimated with the realized variance of 2-hour returns. We assume that both daily and intraday returns follow a GARCH process. Note that the realized variance is computed from unfiltered data.

In general, we do not see any danger that over-cleaning would be suggested by our benchmark in Table 3. Barndorff-Nielsen et al. (2009)'s method appears to be suitable for Germany and Italy. Given that this procedure was originally designed to filter stock data, we can infer that the dynamics of German and Italian bond prices are fairly smooth and resembles stock price fluctuation. For other countries, the two local window filtering methods have the worst performance. The distinctive characteristic of these two “local window” filters is that they eliminate very few outliers compared to the other methods (see Table 2). Due to the inadequate filtering, the fitted intraday return volatility for Spain diverges from the model-free daily realized variance. The first 10 filters, which are based on the properties of the percentage spread, yield similar result. The 95th percentile of the percentage spread turns out to be the best filter for Austria, Belgium and Spain. The 97th percentile wins in the Netherlands while the 92nd is preferred for French bonds.

For each country, the estimated MSE across models are often very close. This prompts us to examine whether the performance of the various filters is statistically significantly different. Table 4 shows the Wilcoxon rank sum test on the equality of every MSE against the lowest one. Not surprisingly, the difference in most of the filters is not statistically significant. In particular, it makes very little difference to choose one particular filter for German and Italian bonds. However, the two local window filters performance significantly worse than the other filters. The results for France are surprising. Although the mean square error of Brownlees and Gallo (2006)'s method is thirty times larger than the MSE of the 92nd percentile method, the MSEs are not statistically different. A further investigation of the squared error series for France reveals that the large numerical difference arises from only a few observations. Therefore a rank sum test which is robust to outliers cannot



**Table 3**

MSE of various filters. The sample performance of the conditional volatility estimation is measured using the mean squared deviations from the benchmark realized volatility. The Mean Squared Errors for the various filters are computed as the  $E(\sum_{n=1}^N h_t s_n q_{t,n} - V_t)^2$ , which is estimated by  $\frac{1}{T} \sum_{t=1}^T (\sum_{n=1}^N \hat{h}_t \hat{s}_n \hat{q}_{t,n} - \hat{V}_t)^2$ .  $h_t$  is estimated using the GARCH(1,1) daily volatility model  $r_t = c_1 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + v_t$ ,  $h_t = w + a_1 v_{t-1}^2 + b_1 h_{t-1}$ .  $s_n$  is fitted by  $\delta_0 \exp(\sum_{j=1}^m \delta_j^* (\Delta_n - k_{j-1})_+)$  where  $(\Delta_n - k_{j-1})_+ > 0$  when  $\Delta_n > k_{j-1}$  and  $(\Delta_n - k_{j-1})_+ = 0$  otherwise,  $t = 1, 2, \dots, N$ .  $q_{t,n}$  is specified as  $1 - \alpha - \beta + \alpha \left( \frac{r_{t,n-1}^2}{s_{n-1} h_t} \right) + \beta q_{t,n-1}$ .

Filtering method	Austria	Belgium	France	Germany	Italy	Netherlands	Spain
Maximum tradable spread	0.020533	0.049768	0.130702	0.025478	0.33235	0.067315	4.326815
99th percentile of tradable spread	0.021192	0.049413	0.042893	0.025478	0.40676	0.060009	1.772955
90th percentile of percentage spread	0.019372	0.049347	0.042902	0.025472	0.40128	0.059920	1.750791
91th percentile of percentage spread	0.019122	0.049835	0.042744	0.025472	0.39348	0.059920	1.759567
92th percentile of percentage spread	0.019014	0.049667	<b>0.042453</b>	0.025472	0.38263	0.059920	1.735467
93th percentile of percentage spread	0.018788	0.049481	0.042872	0.025472	0.37201	0.059920	1.659913
94th percentile of percentage spread	0.018771	0.049092	0.042862	0.025472	0.36575	0.059920	1.653448
95th percentile of percentage spread	<b>0.018630</b>	<b>0.048796</b>	0.042871	0.025472	0.38935	0.060061	<b>1.506469</b>
96th percentile of percentage spread	0.018641	0.049311	0.042897	0.025472	0.38484	0.059973	1.679483
97th percentile of percentage spread	0.018691	0.050138	0.042896	0.025487	0.37845	<b>0.058922</b>	1.720455
Brownlees and Gallo (2006)'s method	12.011837	2.559600	12.601892	0.025435	0.30209	12.292285	44.393817
Barndorff-Nielsen et al. (2009)'s method	14.623279	3.245954	20.015327	<b>0.025369</b>	<b>0.30065</b>	10.185809	59.904429

**Table 4**

P-value of Wilcoxon rank sum test on the equality of a MSE and the lowest MSE for each country. The null hypothesis is that the median of the difference in the squared errors generated by two methods is zero.

Filtering method	Austria	Belgium	France	Germany	Italy	Netherlands	Spain
Maximum tradable spread	0.66944	0.59580	0.02855**	0.88823	0.45793	0.00765***	0.32792
99th percentile of tradable spread	0.60441	0.88053	0.96039	0.88823	0.17381	0.44494	0.95368
90th percentile of percentage spread	0.62422	0.83688	0.97825	0.87015	0.13896	0.41328	0.53680
91th percentile of percentage spread	0.72121	0.85049	0.97404	0.87015	0.13514	0.41328	0.54080
92th percentile of percentage spread	0.71825	0.90573		0.87015	0.14533	0.41328	0.66829
93th percentile of percentage spread	0.75560	0.90469	0.96816	0.87015	0.16077	0.41328	0.59124
94th percentile of percentage spread	0.71314	0.98182	0.96312	0.87015	0.20074	0.41328	0.57102
95th percentile of percentage spread			0.95074	0.87015	0.25316	0.44087	
96th percentile of percentage spread	0.90490	0.95368	0.97636	0.87015	0.35711	0.41478	0.95032
97th percentile of percentage spread	0.97720	0.84223	0.96648	0.85813	0.36445		0.83647
Brownlees and Gallo (2006)'s method	0.15389	0.00004***	0.17331	0.98770	0.93879	0.00000***	0.00000***
Barndorff-Nielsen et al. (2009)'s method	0.00000***	0.00000***	0.01565**			0.00000***	0.00000***

reject the null hypothesis. On the other hand, the test suggests that the maximum tradable spread method is statistically worse than the best method identified for French and Dutch data. Judging from Fig. 2, the presence of trades executed at prices away from the prevailing quotes may reduce the effectiveness of the filter by setting very large maximum tradable spreads which may lead to the inappropriate inclusion of some outliers.

For robustness check we compute the benchmark realized volatility by sampling using a range of alternative sampling frequencies: 15, 30, 45, 60, 75, 90, and 105 minutes, respectively. The relative performance of the filters remains qualitatively the same when using a sampling frequency of at least 90 minutes. However, the results significantly change for all countries (except for Germany and Italy) when the sampling window is smaller than 75 minutes, which suggests relatively long periods of quote distortions due to illiquidity for these countries.

#### 4. Model estimation result

##### 4.1. Daily model result and evaluation

The subsequent results are all based on the best filters obtained for each country. The summary statistics for the sample series of daily returns are presented in Table 5.<sup>11</sup>

Average daily returns are generally positive and relatively small. Spain is the only country with a negative, although not significant,

**Table 5**

Summary statistics of daily series. The daily log returns are computed from 17:00 mid-quote price of cleaned series. The mean and standard deviation are in percentage point. The daily sample covers the period from January 02, 2009 to March 30, 2012.

Country	N	Mean	St.D.	Skew.	Kurt. (excess)
Austria	827	0.0132	0.450	−0.393	3.560
Belgium	826	0.0096	0.489	−0.238	5.875
France	827	0.0116	0.411	−0.018	2.592
Germany	827	0.0167	0.462	0.165	1.559
Italy	827	0.0013	0.646	1.453	22.210
Netherlands	827	0.0180	0.402	0.198	1.701
Spain	827	−0.0051	0.757	−0.184	50.320

average daily return. Italy and Spain exhibit higher standard deviation. All daily return series show excess kurtosis. The estimated coefficients of the daily GARCH models are presented in Table 6. Some of the first order autoregressive coefficients are not significantly different from zero and therefore are not reported. In the conditional variance equation,  $w$  is significantly different from zero except for Italy. Given that we control for the persistent increase in volatility at the time of the implementation of the SMP with a slightly more complicated structure, the significance of  $w$  is of lesser importance.<sup>12</sup> We note the high persistence of volatility for France, Germany and the Netherlands (with

<sup>11</sup> We deleted one day of Belgian data because some of the filters eliminate January 02, 2009 entirely.

<sup>12</sup> We tested the change in  $w$  during the SMP period. The change turns out to be insignificant.

**Table 6**

Estimation of the daily GARCH(1,1) model. The GARCH(1,1) model for Austrian, Belgian, French, German and Dutch government bonds is  $r_t = c_1 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \nu_t$ ,  $h_t = w + a_1 \nu_{t-1}^2 + b_1 h_{t-1}$ . The GARCH(1,1) model for Italian and Spanish government bonds is specified as  $r_t = c_1 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \sum_{p=1}^4 d_p \text{dummy}_p + \nu_t$ ,  $h_t = w + (a_1 + a_2 I(SMP_{t-1} > 0)) \nu_{t-1}^2 + (b_1 + b_2 I(SMP_{t-1} > 0)) h_{t-1}$ . The conditional distribution of error  $\nu_t | \mathcal{F}_{t-1}$  follows a normal distribution  $N(0, h_t)$ . T-statistics are reported in parentheses. The insignificant parameters are omitted. The sample covers the period from January 02, 2009 to March 30, 2012. The daily return is generated using the prevailing mid-quotes at 17:00:00.

Country	c	$\phi_1$	$\phi_2$	w	$a_1$	$b_1$	$a_2$	$b_2$	$d_1$	$d_2$	$d_3$	$d_4$
Austria	0.0233 (1.8928)		-0.0847 (-2.3031)	0.0038 (2.4496)	0.0914 (4.8150)	0.8879 (39.7472)						
Belgium	0.0130 (1.0166)	0.1437 (3.7280)	-0.0933 (-2.4822)	0.0067 (2.8869)	0.0972 (4.3066)	0.8686 (29.7712)						
France	0.0188 (1.5808)		-0.0983 (-2.7008)	0.0026 (2.1455)	0.0659 (4.4179)	0.9161 (49.5816)						
Germany	0.0137 (0.9857)	0.0773 (2.1916)	-0.1085 (-3.0450)	0.0036 (1.8888)	0.0623 (3.5788)	0.9202 (40.9629)						
Italy	0.015769 (1.4350)	0.1426 (3.8034)	-0.1306 (-3.5829)	0.0041 (1.5359)	0.1736 (4.0766)	0.8182 (16.3389)	0.4125 (2.3154)	-0.2213 (-2.2692)	-1.716 (-5.2562)	2.1789 (5.8314)	6.8597 (7.7226)	5.5181 (2.1629)
Netherlands	0.0197 (1.6213)		-0.0832 (-2.3055)	0.0022 (1.8271)	0.0583 (3.6269)	0.9277 (46.0359)						
Spain	-0.0084 (-0.5908)	0.1338 (4.2407)	-0.0803 (-2.6381)	0.0114 (2.1584)	0.1271 (3.9550)	0.8232 (17.0000)	0.4945 (2.9308)	-0.2459 (-3.2145)	-1.9745 (-5.0229)	3.4307 (9.2132)	6.5602 (14.2098)	

estimated coefficients above 0.9) compared to the volatility of the other countries. The high  $a_1$ s of Italian and Spanish bonds clearly indicates that investors attach relatively more importance to volatility shocks. The low persistence ( $b_1$ ) of the two distressed countries is consistent with Chou (1988) who examines the US stock market during the period 1967–1973 and finds that a low persistence coefficient characterizes this period of high uncertainty. During the period when the SMP was launched, we do find a 30% reduction in persistence for both Italy and Spain, which is confirmed by a significant and negative  $b_2$ . Despite the ECB intervention aimed at dissipating the pressure on yields of distressed countries we estimate a significant increase in the coefficients measuring the effect of shocks ( $a_1 + a_2$ ) to around 0.6. This shows that the ECB intervention was not enough to maintain the conditional volatility in line with the values estimated over the first part of the sample. Note that our model was developed for estimating bond volatility and not for assessing the effectiveness of the ECB intervention. However, we do control for the shift in volatility that triggered the ECB intervention. The sum of  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  exceeds 1 and thus this implies a non-stationary daily conditional variance during the turbulent period of the ECB intervention.

We also want to examine the correlation of daily volatility forecasts with intraday activity. Theoretically, different types of traders and market makers may be exposed to and concerned about risk with different time horizons. Active fund managers and market makers attribute greater importance to short-term volatility, whereas pension and passive fund managers are mainly concerned with long-term fluctuations. In addition, the increasing uncertainty about the macro environment and country credit risk may produce greater short-term bond price fluctuations which may affect intraday returns relatively more than daily returns. It is therefore always important to compare daily volatility with volatility computed from intraday returns, and assess whether it is necessary to include the daily variance component.

To study the relation between daily volatility forecasts and intraday volatility, we compute the *ex post* correlation, as in Andersen and Bollerslev (1998) between the daily volatility forecast and the cumulative squared intraday returns for the period from April 02, 2012 to December 30, 2013. Traditionally, the  $R^2$  of a Mincer-Zarnowitz (MZ),  $r_t^2 = a + b h_t + u_t$ , regression is used to evaluate the out-of-sample forecast performance of a GARCH type model. The  $R^2$  is simply the square of the correlation between the regressor and the regressand. As noted by Engle and Patton (2001), squared daily returns are a noisy measure of the latent  $h_t$ . The noise could mask the true relationship of the forecast and the “real” volatility. On the other hand, realized variance, which is the cumulative squared intraday return, proves to be

**Table 7**

Ex-post correlations between forecasted daily volatility with cumulative squared 10-minute returns.

Austria	Belgium	France	Germany	Italy	Netherlands	Spain
0.345	0.401	0.466	0.404	0.507	0.437	0.514

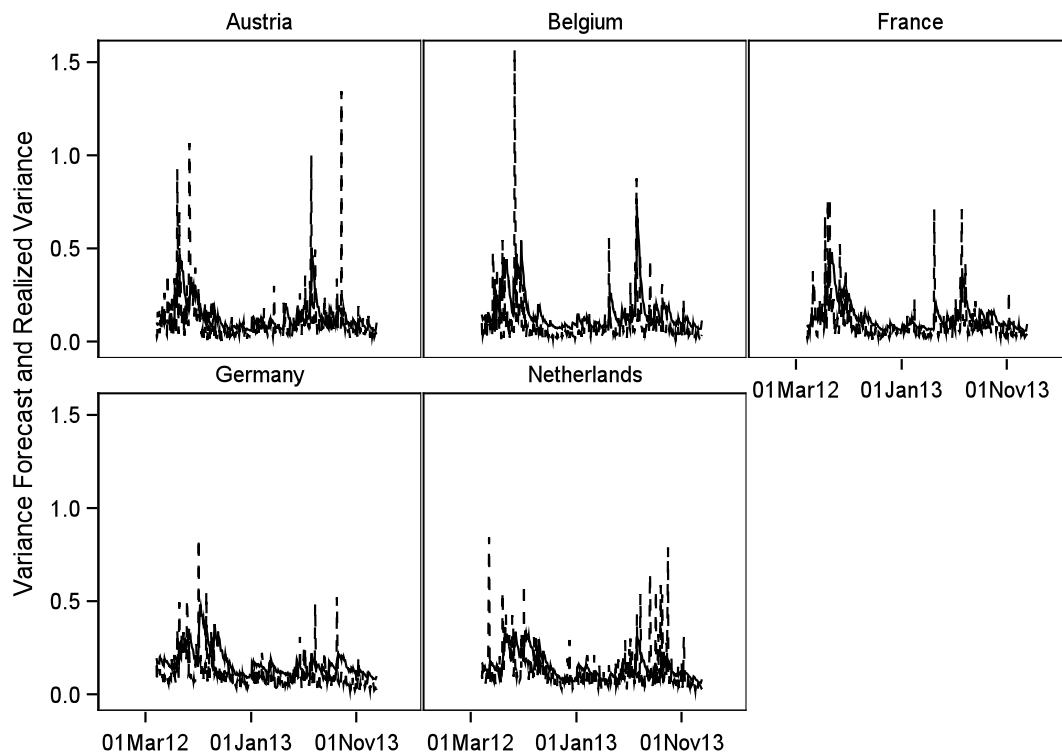
able to provide a more efficient benchmark for the valuation of the volatility forecast.<sup>13</sup> Hence, we use the same approach for assessing the forecasting ability of our model.

The correlation ranges from as low as 0.345 for Austria to as high as roughly 0.5 for Italy and Spain. A simple regression of cumulative squared returns on forecast conditional variances indicates that the forecast explains at least  $0.345^2 = 0.12 = 12\%$  of the total intraday variation for the Austrian market. The Spanish and Italian markets show a relatively high correlation between the volatility computed using intraday returns and the volatility predicted using daily returns (see Table 7). Apart from the big jump of daily volatility on August 02, 2012, we generally see that the two lines closely follow each other in Figs. 5 and 6. As the daily volatility is independent of the two intraday components, it does embody some degree of predictability, which could be explained by investors' risk preferences. Ignoring this daily effect would mistakenly attribute this part either to intraday periodicity or intraday volatility. However, high frequency fluctuations (see peaks in each panel of Figs. 5 and 6) have certainly become a primary concern for investors. Instead of being subordinated as a secondary source of risk, intraday volatility is sometimes the prevailing source of risk.

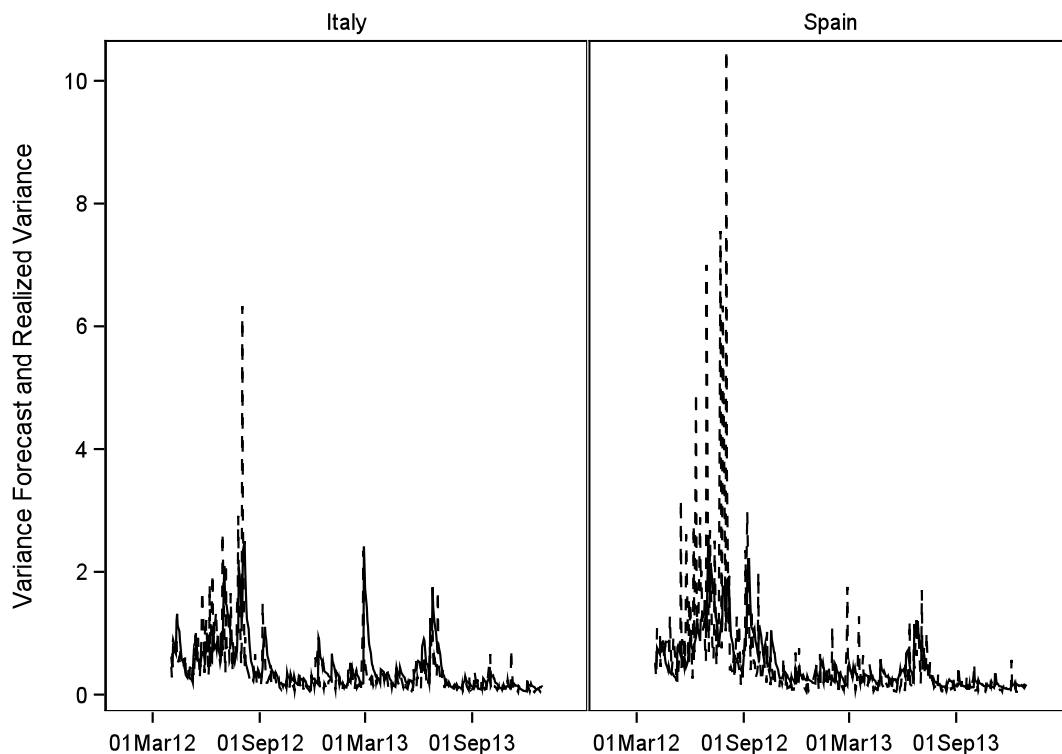
#### 4.2. Intraday result

As expected, in the intraday data, Italy and Spain still have higher standard deviations, with twice the magnitude of the standard deviations of the other countries (see Table 8). The higher average of intraday returns tends to compensate the higher risk of Italian and Spanish government bonds. The signs of skewness seem not consistent with daily returns based on Tables 5 and 8. The skewness of returns for Austria, Belgium, France and Italy shows the opposite sign to the skewness obtained for the daily series. Nonetheless the kurtosis tells a consistent story in both daily and intraday data. Spain and Italy still

<sup>13</sup> Hansen and Lunde (2006) show a significant increase in  $R^2$  when the realized variance is used in a MZ regression.



**Fig. 5.** Daily volatility forecast and realized variance of the five safer countries. The solid line represents one-day-ahead daily conditional variance forecast. The dashed line is the realized variance which is computed using squared 10-minute returns. The one-day-ahead daily conditional variance is generated from the GARCH (1,1) model  $r_t = c_1 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + v_t$ ,  $v_t | \mathcal{F}_{t-1} \sim N(0, h_t)$ ,  $h_t = w + a_1 v_{t-1}^2 + b_1 h_{t-1}$ . The realized variance is computed as  $\sum_{n=1}^N r_{t,n}^2$ . The forecast period is from April 02, 2012 to December 30, 2013.



**Fig. 6.** Daily volatility forecast and realized variance for Italy and Spain. The solid line represents one-day-ahead daily conditional variance forecast. The dashed line is the realized variance which is computed using squared 10-minute returns. The one-day-ahead forecast of daily conditional variance is generated from the GARCH (1,1) model  $r_t = c_1 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + v_t$ ,  $v_t | \mathcal{F}_{t-1} \sim N(0, h_t)$ ,  $h_t = w + a_1 v_{t-1}^2 + b_1 h_{t-1}$ . The realized variance is computed as  $\sum_{n=1}^N r_{t,n}^2$ . The forecast model of Italy and Spain has no dummy variable. The forecast period is from April 02, 2012 to December 30, 2013.

have the most extreme kurtosis, with Austria and Belgium following them. Overall the kurtosis of the 10-minute returns is larger than that of the daily returns.

#### 4.2.1. Diurnal component

The intraday periodicity estimation consistently highlights the distinctive risk of Italian and Spanish government bonds. The results can be categorized into two groups. The typical patterns of Austria, Belgium, Germany, France constitute of one group. Even though there are three knots omitted in the model specification of Austrian, French and Dutch government bonds (as explained in Section 2.2), the estimated intraday patterns are similar. Since we remove the first 45 minutes of returns (see Section 3.2), the seasonal pattern is estimated for the intraday period from 9:00 to 17:30. The market volatility decreases rapidly in the first hour until 10:00, after which the decrease in volatility continues but at a reduced rate. The diurnal volatility starts to pick up from around 14:00 and peaks at 15:00, which is probably due to the opening of the US market and to the volatility spill-over effect. The markets then adjust calmly towards the end of trading day without any further increase in volatility.

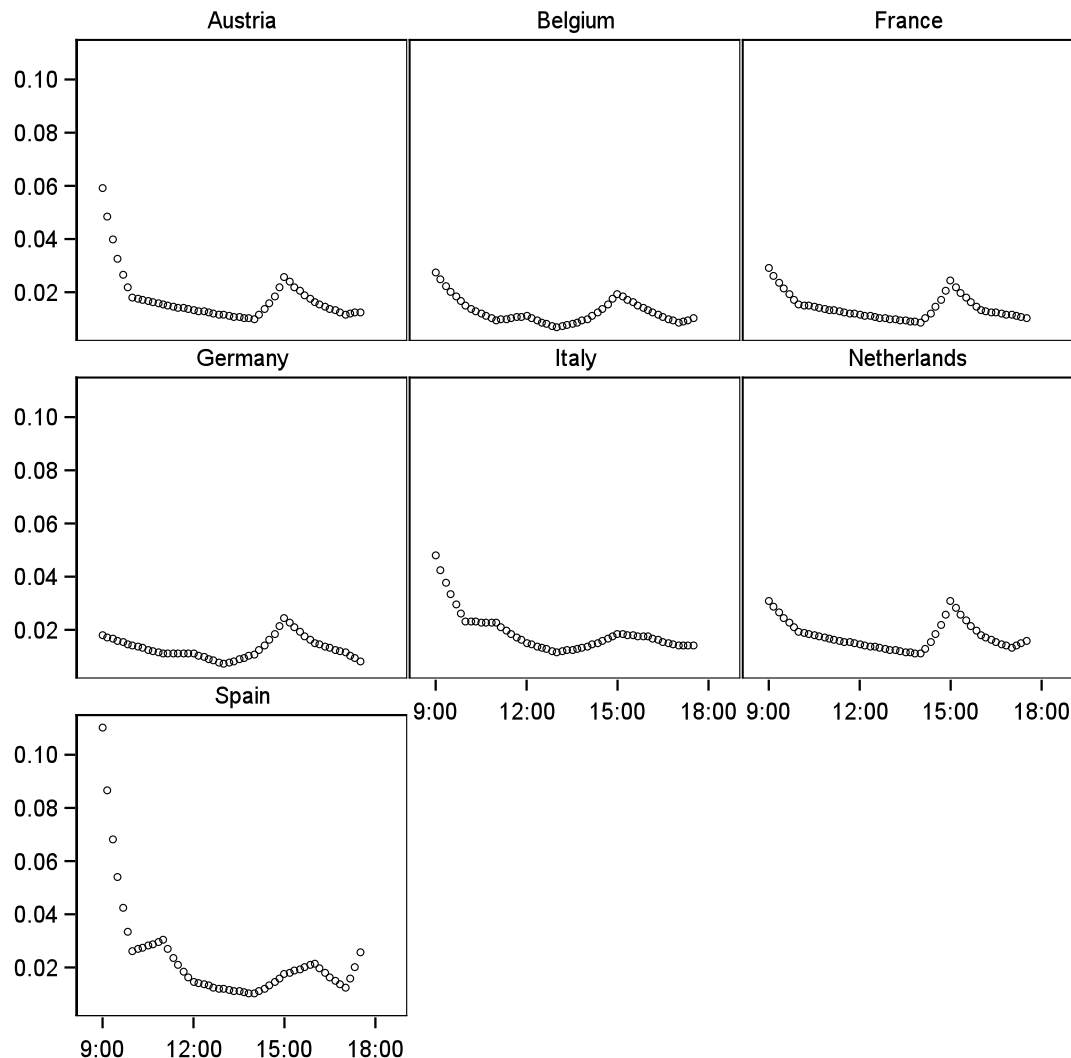
The other group contains Italy and Spain. The diurnal volatility pattern for these countries shows the largest volatility at the open of the trading day and not at the open of the US market. The volatility of

**Table 8**

Summary statistics of intraday 10-minute returns. The 10-minute returns are derived from the clean series. Moreover, the returns from 8:15 to 8:50 are removed from the final sample.

Country	N	Mean	St.D.	Skew.	Kurt. (excess)
Austria	22956	0.0003	0.046	1.236	65.278
Belgium	22961	0.0004	0.042	0.402	28.379
France	22968	0.0004	0.042	0.226	13.366
Germany	22957	0.0002	0.046	0.111	9.490
Italy	22979	0.0008	0.089	−2.366	108.35
Netherlands	22943	0.0002	0.048	−0.100	14.118
Spain	22861	0.0007	0.105	−0.330	75.015

Italian and Spanish government bonds seems mainly driven by domestic and European news. At 10:00, volatility has significantly decreased although it later bounces back slightly and then it trends down to the lowest daily level at around 13:00. Again, we notice the effect of the opening of the American markets which increases volatility until 16:00. In the final half-hour, the volatility of Spanish bonds increases further. Overall, one common point that the seven countries share is that the volatility opens at a high level. This could be due to market makers competing less aggressively at the opening or to a greater



**Fig. 7.** Diurnal components of the 7 European countries. The diurnal component is specified as  $\delta_0 \cdot \exp\left(\sum_{j=1}^m \delta_j^* (\Delta_n - k_{j-1})_+\right)$  where  $\Delta_n - k_{j-1} > 0$  when  $\Delta_n > k_{j-1}$  and  $\Delta_n - k_{j-1} = 0$  otherwise,  $\Delta_n = \frac{n}{N}$ ,  $n = 0, 1, \dots, N$ . There are 8 knots set for each trading hour and an extra knot set for the final half-hour for bonds of Belgium, Germany, Italy, and Spain. Three knots at 11:00, 12:00, and 13:00 are omitted for estimation reason for bonds of Austria, France and the Netherlands.



**Table 9**

Estimation of the intraday volatility model. The intraday unit GARCH is specified as  $q_{t,n} = 1 - \alpha - \beta + \alpha \left( \frac{r_{t,n-1}^2}{s_{n-1}h_t} \right) + \beta q_{t,n-1}$ . It is jointly estimated with the diurnal pattern, which is specified as an exponential piecewise linear function of time interval:  $s_n = \delta_0 * \exp \left( \sum_{j=1}^m \delta_j * (\Delta_n - k_{j-1})_+ \right)$  where  $\Delta_n - k_{j-1} > 0$  when  $\Delta_n > k_{j-1}$  and  $\Delta_n - k_{j-1} = 0$  otherwise,  $\Delta_n = n/N$ ,  $n = 0, \dots, N$ . T-statistics are shown in parentheses.

Country	$\alpha$	$\beta$	$\delta_0$	$\delta_1$	$\delta_2$	$\delta_3$	$\delta_4$	$\delta_5$	$\delta_6$	$\delta_7$	$\delta_8$	$\delta_9$
Austria	0.0651 (17.2530)	0.8917 (129.8230)	0.0717 (17.0070)	-10.302 (-21.7838)	8.9880 (17.1380)				9.6131 (21.6807)	-12.195 (-16.1481)	0.9862 (1.1660)	4.1739 (2.5207)
Belgium	0.0866 (17.5083)	0.8399 (96.3974)	0.0311 (16.6534)	-5.5979 (-9.7082)	1.8813 (2.0185)	4.8921 (6.2388)	-5.2441 (-6.6089)	7.3146 (8.5084)	2.5396 (3.0232)	-8.9567 (-11.3734)	-0.6500 (-0.7396)	6.5569 (3.8864)
France	0.0805 (16.8397)	0.8661 (98.0778)	0.0322 (16.6469)	-5.4253 (-10.6934)	4.1413 (7.3420)				10.2893 (22.7116)	-14.333 (-18.5009)	4.1029 (4.5570)	-0.5934 (-0.3536)
Germany	0.0532 (12.4611)	0.8994 (100.5398)	0.0218 (12.6202)	-4.1331 (-4.5372)	3.8132 (2.3187)	-0.5480 (-0.4385)	-2.6307 (-3.1607)	6.8180 (8.0356)	3.7221 (3.9193)	-11.131 (-12.1327)	1.9272 (1.8381)	0.3657 (0.1536)
Italy	0.1053 (17.7867)	0.8484 (93.7866)	0.054 (15.1586)	-6.2639 (-11.0628)	5.9838 (6.4409)	-3.1780 (-3.9693)	1.1216 (1.5069)	3.8395 (4.6574)	1.0415 (1.1765)	-3.0627 (-3.4437)	-1.3971 (-1.5149)	2.4838 (1.5213)
Netherlands	0.0753 (14.5030)	0.8204 (51.8162)	0.029 (17.3643)	-3.3011 (-6.5688)	2.0024 (3.6026)				10.4359 (22.5383)	-13.924 (-17.8215)	2.7475 (3.0339)	1.2561 (0.7248)
Spain	0.0598 (21.0195)	0.9039 (202.3823)	0.1400 (15.9266)	-12.429 (-21.4210)	13.6642 (14.0371)	-7.6326 (-8.5132)	4.5944 (5.5998)	0.5584 (0.7805)	5.8541 (7.8211)	-2.7180 (-3.2709)	-6.6631 (-7.7462)	17.2191 (10.7740)

uncertainty about the bond prices right after the overnight period.

#### 4.2.2. Intraday volatility

Interestingly, we model daily and intraday volatility in the same manner with a GARCH(1,1) process. The two GARCH(1,1) models enable us to compare the behavior of daily and intraday volatility. From Table 9, we can see that most of the spline parameters  $\delta_1 - \delta_9$  as well as the GARCH parameters are significant. Notably, the relative magnitudes of  $\alpha$  and  $\beta$  change dramatically across countries, with Spanish bonds showing the highest persistence of volatility, probably due to the general success of capturing the periodicity of intraday volatility, whereas the  $\beta$  of the Netherlands is the lowest among the 7 countries. The volatility of the 10-year bonds of Austria, Belgium, France, Germany and Italy maintains the characteristics of the daily GARCH volatility. Italy still has a relatively low  $\beta$  and the highest  $\alpha$ . The overall scale of volatility is partially reflected in parameter  $\delta_0$ , which is the constant in the spline equation. Still Spain has the highest  $\delta_0$ , with Italy and Austria following it. None of the other countries has a constant exceeding 0.05. It seems that the estimation of intraday volatility of Spain and the Netherlands provides a different picture from daily volatility. However, the dynamics of intraday volatility still vary significantly across countries.

### 5. Forecast evaluation

In view of the general success of GARCH(1,1) model in forecasting daily volatility of bond markets (see Hansen & Lunde, 2005), we want to compare the forecast performance of our model against the GARCH(1,1) model estimated for daily returns. The out-of-sample period covers the first two months of 2014. We filter the intraday observations using the best filter derived from the in-sample estimation (see Section 3.4). Since the bonds of Italy and Germany require adaptive filtering, which utilizes future information, we restrict the bonds to have a percentage spread less than 50 basis point. In addition, if there is a new issue during the out-of-sample period, we switch to the new bond according to the rule described in the Appendix.

Four criteria are considered to evaluate the forecast performance, namely mean square error (MSE), quasi-likelihood based error<sup>14</sup> (QLIKE), mean absolute error (MAE), and correlation between volatility forecast and benchmark volatility, which is approximated by the realized volatility of raw 2-hour returns. The validity of using raw 2-hour return to compute realized volatility is proven in the robustness check

of our filtering MSE result (see Section 3.4). As it is shown by Patton (2011), the “MSE” and “QLIKE” loss functions, which lead to unbiased predictors, give a consistent ranking of volatility forecasts when the benchmark is a noisy volatility proxy. The “MAE” loss function, although it may not have the nice properties of the “MSE” and “QLIKE” criteria, is robust to outliers. The “CORR” function generally measures the closeness between the patterns of volatility forecasts and the volatility proxies. The one-day-ahead forecast of the daily GARCH(1,1) model for day  $t$  is denoted as  $h_{1,t}^f$  while the forecast from the intraday model is labeled as  $h_{2,t}^f$ .

$$MSE(h_{i,t}^f) = \frac{1}{T} \sum_{t=1}^T (h_{i,t}^f - \hat{V}_t)^2$$

$$MAE(h_{i,t}^f) = \frac{1}{T} \sum_{t=1}^T |h_{i,t}^f - \hat{V}_t|$$

$$QLIKE(h_{i,t}^f) = \frac{1}{T} \sum_{t=1}^T (\log(h_{i,t}^f) + \hat{V}_t/h_{i,t}^f)$$

$$CORR(h_{i,t}^f) = \frac{1}{T} \sum_{t=1}^T (h_{i,t}^f - \bar{h}_{i,t}^f)(\hat{V}_t - \bar{\hat{V}}_t)$$

where  $i = 1$  or  $2$ .

The forecasting schemes for the two models are now laid out for the purpose of fair comparison, i.e. using all the information which can be processed by each model before day  $t$ . In Section 4.1 and Table 7, we have already seen the predictive power of the daily GARCH(1,1) model for Italy and Spain. The parameters involved in forecasting are derived from a fixed-sample and all daily forecasts are generated from these parameters. In order to use new information to improve the daily model's forecast, we estimate the GARCH(1,1) model whenever a new day can be included in the fitting sample and produce the forecast for the next day. The forecasts generated by the dynamic sample approach, can be substantially different from those generated by fixed sample approach especially for Italy and Spain (see Table 6 for the volatile period of Italian and Spanish bonds during 2012). For the intraday model, the one-day-ahead forecast  $h_{2,t}^f$  is equal to  $(h_{1,t}^f \sum_{n=1}^N s_n q_{t,n}^f)$  where  $q_{t,n}^f$  is a  $n$ -step-ahead forecast generated from the intraday GARCH(1,1) model. For the first interval every day, the  $q_{t,n}^f$  is initialized by  $1/N \sum_{n=1}^N r_{t-1,n}^2 / (h_{t-1,n}^f s_n)$ . Obviously, both methods exclude the information that becomes available during the forecasting day and the forecast from the GARCH model estimated on daily data is nested in  $h_{2,t}^f$ . It is also evident that the extra predictive power as compared to the daily model stems from the diurnal and intraday GARCH components. The accuracy of  $h_{2,t}^f$  relies on the success of estimating the fixed diurnal

<sup>14</sup> This is a likelihood based loss function that asymmetrically penalizes over- and under-prediction.

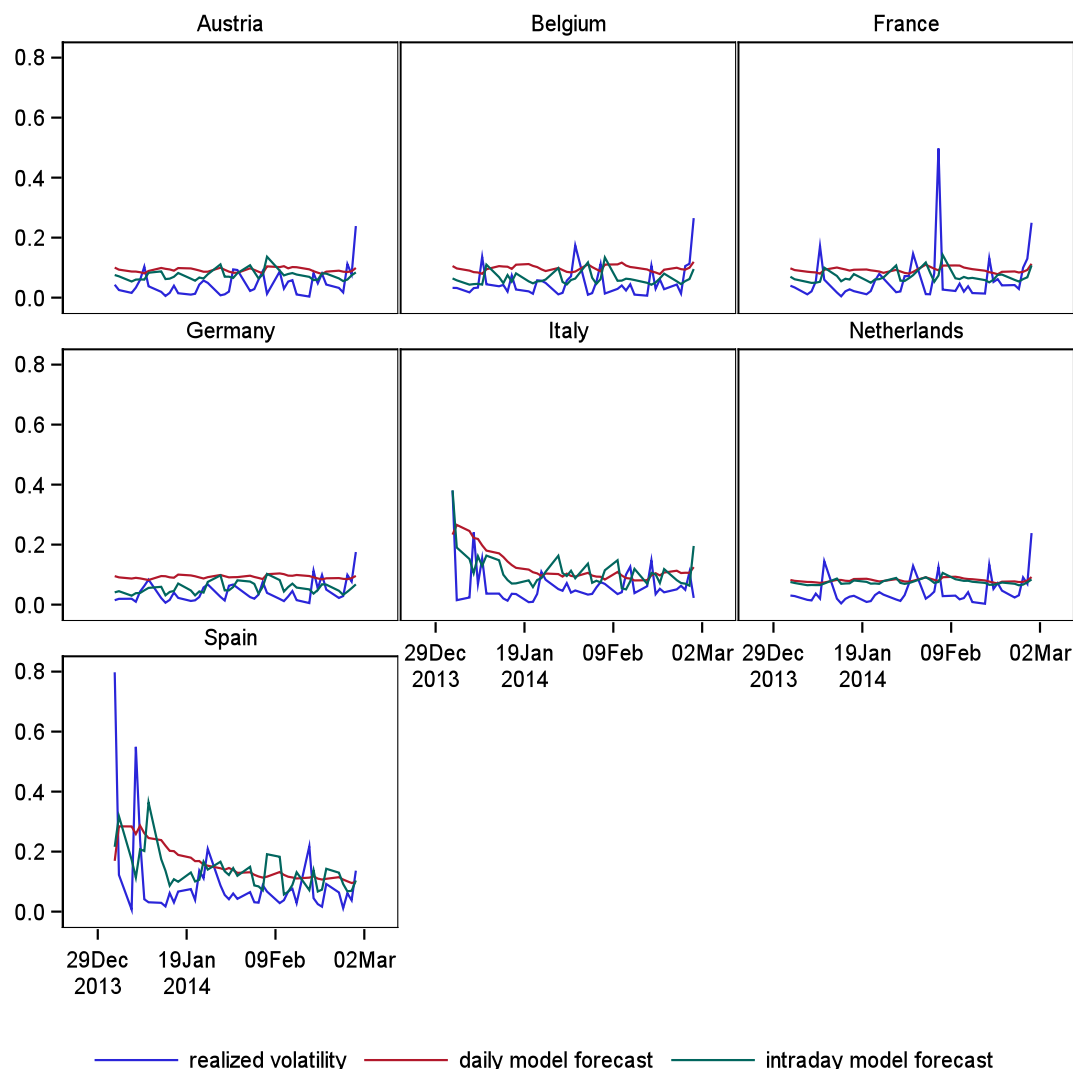
**Table 10**

The Diebold and Mariano (1995) test for comparing the predictive ability of the daily GARCH(1,1) model and the intraday multiplicative component GARCH model. Negative values show the preference to the intraday model. \*\*\*, \*\*, \* denote 1%, 5%, 10% significance respectively

Country	MSE	QLIKE	MAE	CORR
Austria	−0.0012**	−0.1046***	−0.0131***	−0.0001
Belgium	−0.0017***	−0.1370***	−0.0209***	−0.0000
France	−0.0008	−0.0743	−0.0130***	−0.0002
Germany	−0.0023***	−0.2260***	−0.0245***	−0.0002***
Italy	−0.0026	−0.0582	−0.0117	−0.0008
Netherlands	−0.0004**	−0.0353***	−0.0043***	−0.0000
Spain	−0.0014	−0.0521	−0.0139	0.0001

that  $\sum_{n=1}^N s_n = 1$ . Since  $E(q_{t,n}) = 1$ , the intraday model will provide very little information if  $q_{t,n}^f$  stays close to its unconditional expectation and if the summation of  $s_n$  is 1. In other words, if this is indeed the case, then  $h_{1,t}^f$  and  $h_{2,t}^f$  would be identical. The normalization of the intraday volatility pattern is a common practice in fitting and forecasting intraday volatility. Taylor and Xu (1997), for example, standardize the sum of their variance seasonal pattern when studying foreign exchange volatility. Table 10 presents the Diebold and Mariano (1995) test<sup>15</sup> for forecast performance comparison between the two models. A negative value indicates that the component GARCH model which uses information from the intraday model produces better volatility forecasts than the daily model.

The forecast daily volatility is presented in Fig. 8 for four major European countries. The correlation between the daily GARCH(1,1)



**Fig. 8.** Forecast plots for different countries. The blue line represents the realized volatility computed from 2-hour returns. The red line is the daily volatility forecast from the daily GARCH(1,1) model. The green line is the forecast given by the intraday GARCH model.

component and an adequate specification of the GARCH component. The intraday periodicity  $s_n$  is assumed to be unchanged during the out-of-sample period. In fact, we can view the  $\sum_{n=1}^N s_n q_{t,n}^f$  as a factor that modifies the daily GARCH(1,1) forecast according to a larger information set. If the intraday information is indeed relevant, it will improve the daily GARCH(1,1) forecast. To measure the extra information content we propose to re-estimate the intraday model with a daily dynamic-sample forecast and normalize the diurnal component so

forecast and the intraday component GARCH forecast is around 0.4 for Austria, Belgium, France and Germany while it increases to roughly 0.6 for Italy and Spain and reaches 0.8 for the Netherlands. However, the low correlation does not necessarily indicate a better forecast ability, as

<sup>15</sup> We do not use Giacomini and White (2006)'s test in the forecasting evaluation as it is more computationally demanding and it requires iterative estimations of the models.

it is seen below that the intraday component GARCH model is more suitable for forecasting the volatility of the Dutch bonds. From Fig. 8 we can see that the two forecasts tend to diverge when there is little variation of returns from the previous trading day. This can be explained by the nature of  $q_{t,n}^f$  – the multi-step-ahead forecast which is a component of  $h_{2,t}^f$ . The half-life of  $q_{t,n}^f$  is roughly 15 (or even lower for some countries) intervals, which corresponds to two-and-a-half hours whereas the half-life of  $h_{1,t}^f$  is around 20 days! Therefore, when there is a shock followed by a few quiet trading days, the daily GARCH(1,1) model will generally over-predict the daily volatility but the intraday model is capable of quickly giving a low volatility forecast.

It turns out that the intraday model provides a superior forecast for most of the less volatile bonds, whereas there is no “winner model” for Italian and Spanish bonds. The “MSE” and “QLIKE” measures both confirm the better forecast accuracy of the intraday model and Fig. 8 suggests that the daily GARCH(1,1) model generally produces too high a volatility forecast for safer government bonds. For “CORR”, which measures the synchronicity of volatility forecasts and the volatility proxy, neither of the two models seems to be better than the other. An insignificantly different forecast performance is expected for Italy and Spain, as the two models are both fitted to a high-volatility environment but the volatility is very low during the out-of-sample period. On the other hand, since the volatility of the other five bond series is always low, the intraday model does provide extra information to the daily GARCH(1,1) forecast. One exception is French bonds. Only the “MAE” loss function gives a significant result, which may be explained by the sudden spike in the middle of the forecasting period. The other measures are easily influenced by this outlier. Overall, we do see that the intraday data can be employed to improve the daily volatility forecast. In the robustness check, we investigate the possibility that the over-prediction generated by the daily GARCH(1,1) model is due to the omission of the overnight movements in realized volatility computation. We redo the Diebold and Mariano (1995) test, adding the square of the overnight returns<sup>16</sup> to the realized volatility. The test result does not change significantly.

## 6. Conclusion

In this paper, we study the daily and intraday volatility of the long-term government bonds of seven European countries during the sovereign debt crisis. A new specification of intraday periodicity, along with a unit GARCH(1,1) model, is formulated under the framework of Engle and Sokalska (2012). We utilize a long time series to focus on the volatility of 10-year government bonds. Several filters are presented and tested against the benchmark inspired by Bandi and Russell (2008) using the data of the MTS interdealer market. It appears that the filters based on the spread are most suitable for our data. The necessity of filtering suggests that only part of the high-frequency information is relevant for estimating longer horizon volatility. The risk of Italian and Spanish bonds is emphasized in both daily and intraday estimations. Daily volatility of Italian and Spanish bonds exhibits much higher sensitivity to shocks and lower persistence than volatility of bonds of other European countries. The daily GARCH estimations show that these unique features of Italian and Spanish volatility are even exacerbated during the sovereign bond crisis despite the ECB direct intervention in the market with bond purchases. Although the ECB purchases were somewhat successful in controlling yields, they may have contributed to further reducing liquidity and increasing volatility. At the intraday level, the existence of diurnal periodicity is confirmed and captured successfully. The volatility transmission from US to European markets is evident in all countries. The evaluation of the forecasting ability of the daily GARCH(1,1) model and the intraday multiplicative

component GARCH model demonstrates that the intraday information is able to improve the volatility forecast accuracy especially for less volatile bonds. As final points, we encourage risk managers to adopt proper modeling tools for managing and predicting the risk of government bond portfolios. Also, we suggest that policy makers and regulators pay particular attention to the effects that yield-targeting interventions may have on liquidity conditions of government bonds thus potentially causing additional transitory volatility. This issue was only marginally addressed in this paper and deserves further investigation.

## Appendix A. Appendix

We describe the details of variable construction, including the criteria for determining eligible bonds for our analysis, and the procedures we use for constructing series of on-the-run bond returns. In particular, the main reason for switching from off-the-run bonds to on-the-run bonds is to maintain constant the time to maturity and insure that bonds are sufficiently liquid so that the mid-quote price is a good proxy for the underlying bond price. We consider bonds with remaining time to maturity ranging from 8.5 years to 11.5 years. The lower bound for the selection is in accordance with the usual minimum remaining time to maturity for a bond to be qualified in a long-term bond futures contract (see the Eurex Exchange Long Term Bond Futures Contract). The upper bound is determined to have the same distance to 10 year as the lower bound. We select only one on-the run bond for each period and each country. Beber, Brandt, and Kavajecz (2009) use a tighter maturity band for 10-year bonds (9.5–10.5 years) for their study on the relationship between credit default swaps (CDS) and sovereign yield spreads during crisis periods. The CDS contracts are explicitly written on the same range of bonds. While Dunne, Moore, and Portes (2007) define long-term bonds with maturity of 6.6–13.5 years, which is broad enough to examine the benchmark status. We adjust the range of maturities according to the specific issuing frequencies of European countries during the sovereign bond crisis. For example, Germany auctioned in total 10 bonds while Austria did not issue any new 10-year bonds from 2006 to 2011. Nevertheless, we find bonds for all countries falling into our 10-year maturity band.

With the passage of time and new issues, the current benchmark bond loses its status. In order to have an accurate view of the crisis period and preserve the quality of the data used for our study, we have to change our benchmark bond whenever the existing benchmark bond does not comply with our maturity standard or there is a new auction. The rolling-over approach is a common solution for the periodic issues and changes in seasonality of benchmark bonds (see Fleming & Lopez, 1999 for GOVPX data and Bollerslev et al., 2000 for US long-term bond futures data). On each switching date, the return is computed from the prices of the old bond and the returns are always computed using data from the same bond. We choose different policies to deal with switching bonds for liquidity or maturity reasons. If the maturity of the current benchmark bond falls below 8.5 years, the switching is triggered immediately.<sup>17</sup> However, if there is a new auction, we choose to delay the introduction of the new bond and the exclusion of the old bond by one month. According to Pasquariello and Vega (2009), there is a significant liquidity and price heterogeneity of newly issued benchmark bonds and the just off-the-run bonds across maturities in US market. They demonstrate that for 10-year US bonds, the liquidity condition of the on-the-run bonds is improved after 10 days since the auction. Diaz, Merrick Jr., and Navarro (2006) also find that the liquidity measured by relative traded volume is different between off/on-the-run 10-year Spanish government bonds. The authors illustrate that an on-the-run bond does not instantly gain benchmark status. We therefore do not replace old bonds with new bonds immediately.

<sup>16</sup> The overnight return is the log of the mid-quote price at 9:00 minus the log of the mid-quote price at 17:30 the previous day.

<sup>17</sup> If the designated switching date is a market holiday, the switching will be postponed to the nearest market open date. If two bonds both qualify for this category we choose the most recent one.

## References

- Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3, 5–39.
- Andersen, T. G., & Bollerslev, T. (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*, 53, 219–265.
- Apergis, N., Barunfk, J., & Lau, M. C. K. (2017). Good volatility, bad volatility: What drives the asymmetric connectedness of Australian electricity markets? *Energy Economics*, 66, 108–115.
- Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3–30.
- Bandi, P. M., & Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75, 339–369.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, 12, C1–C32.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64, 253–280.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17, 457–477.
- Beber, A., Brandt, M. W., & Kavajecz, K. A. (2009). Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market. *Review of Financial Studies*, 22, 925–957.
- Blume, M. E., Keim, D. B., & Patel, S. A. (1991). Returns and volatility of low-grade bonds 1977–1989. *Journal of Finance*, 46, 49–74.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
- Bollerslev, T., Cai, J., & Song, F. M. (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US treasury bond market. *Journal of Empirical Finance*, 7, 37–55.
- 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., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies*, 31, 2729–2773.
- Brownlees, C. T., & Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, 51, 2232–2245.
- Cheung, Y. C., de Jong, F., & Rindi, B. (2005). Trading European sovereign bonds: The microstructure of the MTS trading platforms. *Working Paper 432*. The European Central Bank.
- Chou, R. Y. (1988). Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics*, 3, 279–294.
- Christiansen, C. (2007). Volatility-spillover effects in European bond markets. *European Financial Management*, 13, 923–948.
- Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28, 777–795.
- Darbha, M., & Dufour, A. (2013). Microstructure of euro-area government bond market. In H. K. Baker, & H. Kiyomaz (Eds.), *Market microstructure in emerging and developed markets*. Hoboken, NJ: John Wiley.
- De Bruyckere, V., Gerhardt, M., Schepens, G., & Vander Vennet, R. (2013). Bank/sovereign risk spillovers in the European debt crisis. *Journal of Banking & Finance*, 37, 4793–4809.
- De Goeij, P., & Marquering, W. (2004). Modeling the conditional covariance between stock and bond returns: A multivariate GARCH approach. *Journal of Financial Econometrics*, 2, 531–564.
- Deo, R., Hurvich, C., & Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics*, 131, 29–58.
- Diaz, A., Merrick, J. J., Jr., & Navarro, E. (2006). Spanish treasury bond market liquidity and volatility pre- and post-European monetary union. *Journal of Banking & Finance*, 30, 1309–1332.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13, 253–263.
- Dimitriou, D., Kenourgios, D., & Simos, T. (2013). Global financial crisis and emerging stock market contagion: A multivariate FIAPARCH-DCC approach. *International Review of Financial Analysis*, 30, 46–56.
- Dunne, P. G., Moore, M. J., & Portes, R. (2007). Benchmark status in fixed-income asset markets. *Journal of Business Finance & Accounting*, 34, 1615–1634.
- Ehrmann, M., & Fratzscher, M. (2017). Euro area government bonds - Fragmentation and contagion during the sovereign debt crisis. *Journal of International Money and Finance*, 70, 26–44.
- Engle, R. (2002a). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20, 339–350.
- Engle, R. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics*, 17, 425–446.
- Engle, R., & Engle, R. (2007). Execution risk. *Journal of Portfolio Management*, 33(2), 34–44.
- Engle, R. F., & Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, 1, 237–245.
- Engle, R. F., & Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21, 1187–1222.
- Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66, 1127–1162.
- Engle, R. F., & Sokalska, M. E. (2012). Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. *Journal of Financial Econometrics*, 10, 54–83.
- Fleming, M. (2001). Measuring treasury market liquidity. *Staff Report 133*. FRB of New York.
- Fleming, M., & Lopez, J. (1999). Heat waves, meteor showers, and trading volume: An analysis of volatility spillovers in the US treasury market. *Staff Report 82*. FRB of New York.
- Gençay, R., Dacorogna, M., Müller, U. A., Pictet, O., & Olsen, R. (2001). *An introduction to high-frequency finance*. Cambridge: Academic Press.
- Ghysels, E., Idier, J., Manganelli, S., & Vergote, O. (2017). A high-frequency assessment of the ECB securities markets programme. *Journal of the European Economic Association*, 15, 218–243.
- Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20, 873–889.
- Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24, 127–161.
- Harris, L. (2002). *Trading and exchanges: Market microstructure for practitioners*. Oxford: Oxford University Press.
- Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46, 179–207.
- Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies*, 6, 191–212.
- Hasbrouck, J. (2018). High-frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis*, 53, 613–641.
- Jones, C. M., Lamont, O., & Lumsdaine, R. L. (1998). Macroeconomic news and bond market volatility. *Journal of Financial Economics*, 47, 315–337.
- Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *Journal of Finance*, 72, 967–998.
- Longstaff, F. A., & Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two factor general equilibrium model. *Journal of Finance*, 47, 1259–1282.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347–370.
- Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle, & D. L. McFadden (Vol. Eds.), *Handbook of econometrics*. Vol.4. North-Holland: Elsevier.
- Pasquariello, P., & Vega, C. (2009). The on-the-run liquidity phenomenon. *Journal of Financial Economics*, 92, 1–24.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160, 246–256.
- Poon, S.-H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41, 478–539.
- Taylor, S. J., & Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, 4, 317–340.
- Wu, W., Lau, M. C. K., & Vigne, S. A. (2017). Modelling asymmetric conditional dependence between Shanghai and Hong Kong stock markets. *Research in International Business and Finance*, 42, 1137–1149.