A Large t-Vector AutoRegressive model for Volatility Spillover Analysis

Luca Barbaglia*, Christophe Croux and Ines Wilms

Faculty of Economics and Business, KU Leuven, Belgium

January 26, 2017

Abstract. The Vector AutoRegressive (VAR) model is often used to analyze volatility spillovers, that is interdependences among volatilities of financial assets. These models typically contain log transformed volatilities as time series and assume normally distributed errors. However, recent advances in the volatility literature show that log transformed volatilities might exhibit deviations from the normal distribution, in particular due to fat-tails. Therefore, we propose to analyze volatility spillovers using a VAR model with errors following a *multivariate Student t* distribution with unknown degrees of freedom. In this way, we account for the possible fat-tailed distribution of the VAR model errors in a data-driven manner. Moreover, we study volatility spillovers among a *large* number of commodities. To this end, we use *penalized* estimation of the VAR model with t-distributed errors. Using network analysis, we reveal bidirectional volatility spillovers between energy and biofuel, and between energy and agriculture commodities. Finally, our results indicate that the distribution of the residuals of the VAR model is fat-tailed.

Keywords: Biofuel, EM algorithm, Lasso, Multivariate t distribution, Vector AutoRegressive model, Volatility spillover.

JEL classification: C58, C32, Q02

*Corresponding author: Faculty of Economics and Business, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium. *Email address*: luca.barbaglia@kuleuven.be

1 Introduction

In recent years, financial markets have shown high interconnectedness both at the levels of returns and of return volatilities. In this paper, we focus on the interdependences among return volatilities, called volatility spillover in the remainder. As volatility is a measure of risk, analysts and investors are interested in better understanding volatility and its spillovers.

Among others, Diebold and Yilmaz (2015) analyze volatility spillovers based on Vector AutoRegressive (VAR) models in a two step approach. First, they obtain a measure of volatility by computing realized variances (e.g. McAleer and Medeiros, 2008) or realized ranges (e.g. Martens and van Dick, 2007). Second, they consider a VAR model containing the logarithmic transformed volatilities as time series to estimate the spillovers. An advantage of this procedure is its simplicity, making it easily adaptable to different empirical applications. Nevertheless, two main limitations remain: (i) the errors of the standard VAR model do not follow a distribution with heavy tails, thereby not accounting for the frequent occurrence of extreme observations which are typical of volatilities (Callot et al., 2016) and, (ii) the number of time series in the VAR model is limited since the number of parameters to be estimated increases quadratically with the number of time series included (Diebold and Yilmaz, 2015, p183).

The two above limitations have already been acknowledged in the literature. On the second one there is general consensus on the need to further investigate high-dimensional VAR models, even though only few works on this topic are available (e.g. Barigozzi and Hallin, 2017). The first limitation has not been addressed yet. The use of a VAR model with normally distributed errors is mainly motivated by the fact that log volatilities should be approximately normal (Diebold and Yilmaz, 2015, p18), which is supported by empirical applications (e.g. Andersen et al., 2001; Thomakos and Wang, 2003) and asymptotic theory (Barndorff-Nielsen and Shephard, 2002). Nevertheless, subsequent studies emphasize that the type of volatility measure might have an impact on its distribution. Considering realized variances, Corsi et al. (2008) and Hassler et al. (2016) highlight that the log volatility of the S&P 500 index based on 5min intra-day returns exhibits deviations from the Gaussian distribution. Considering realized ranges, Christensen et al. (2009) prove that the log volatility follows a mixed multivariate normal, and Caporin and Velo (2015) show that the log volatility has heavier tails than the normal. Hence, the normality assumption on the log volatilities does not always hold and thus it seems appropriate to specify a more general distribution for the errors of the VAR model containing the log volatilities.

We analyze volatility spillovers taking into account the above described limitations. To this end, we rely on the work of Diebold and Yilmaz (2015), but we (i) use a VAR model with *t-distributed errors*. We *estimate the degrees of freedom* of the t distribution, in contrast to prior studies who take them to be fixed (e.g. Franses and Lucas, 1998; Finegold and Drton, 2011). As such, we determine the degree of fat-tailedness

of the VAR residuals in a data-driven manner. (ii) We use a *penalized* estimator of the VAR model such that no limitations need to be imposed on the number of included time series.

We use the proposed large t-VAR approach to investigate the existence of volatility spillovers among $J = 10$ energy, biofuel and agricultural commodities (see Serra and Zilberman, 2013, and references therein). The reference period is 2012-2016, over which we employ a rolling window to study the evolution of the volatility spillovers. Our results highlight fat-tailed VAR residuals throughout all time windows. Besides, we observe volatility spillovers between biofuels and those agricultural commodities that can be used as inputs in biofuel production. Moreover, we find volatility spillovers between energy and agriculture, regardless of the fact whether those crops can be used for biofuel production or not.

The remainder of this article is structured as follows. Section 2 reviews the VAR model with Gaussian errors and introduces the VAR model with Student t errors. The corresponding penalized estimators are subsequently discussed. Section 3 outlines the algorithm. Simulation studies in Section 4 show the good performance of the proposed estimator. Section 5 presents the data, the definition of volatility spillovers and the network analysis tools to visualize them. Section 6 presents the results of the volatility spillover analysis. Finally, Section 7 concludes.

2 Models and estimators

Let $\boldsymbol{y}_t$ be a J-dimensional vector of log volatilities for $1 \leq t \leq T$, with T the time series length. We take the logarithmic transformation of the volatilities, which ensures the positivity of the volatility forecasts (e.g. Callot et al., 2016). We consider a stationary VAR of order P for the log volatilities

$$\boldsymbol{y}_t = \sum_{p=1}^{P} \boldsymbol{B}_p \boldsymbol{y}_{t-p} + \boldsymbol{e}_t, \tag{1}$$

where $\boldsymbol{B}_p$ is the $J \times J$ matrix of autoregressive coefficients for lags $1 \leq p \leq P$, and $\boldsymbol{e}_t$ is a J-dimensional vector of error terms with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}$. Without loss of generality, we assume that all log volatility time series are mean centered such that no intercept is included.

For ease of notation, we rewrite model (1) in matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E},$$

where the $N \times J$ matrix $\boldsymbol{Y} = [\boldsymbol{y}'_{P+1}, \ldots, \boldsymbol{y}'_T]'$, with $N = T - P$. Let $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_P]$ be the $N \times JP$ matrix of lagged time series, where $\boldsymbol{X}_p = [\boldsymbol{y}'_{P+1-p}, \ldots, \boldsymbol{y}'_{T-p}]'$ is a $N \times J$ matrix for $1 \leq p \leq P$. Finally, the $JP \times J$ matrix of autoregressive coefficients is $\boldsymbol{B} = [\boldsymbol{B}'_1, \ldots, \boldsymbol{B}'_P]'$ and $\boldsymbol{E}$ is the $N \times J$ error matrix.

In the next subsections we first review the penalized estimator for the Gaussian VAR model, and then we introduce the penalized estimator for the VAR model with Student t errors. To this end, we build on the t-lasso estimator for sparse graphical models proposed by Finegold and Drton (2011).

2.1 Gaussian model

Assume that the error terms $\boldsymbol{e}_t$ follow a multivariate normal distribution $N_J(\mathbf{0}, \boldsymbol{\Sigma})$. The penalized maximum likelihood (PML) estimator of $\boldsymbol{B}$ with Gaussian errors is obtained by minimizing the negative log likelihood:

$$\widehat{\boldsymbol{B}} = \underset{\boldsymbol{B}}{\operatorname{argmin}} \ \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] + \lambda \sum_{i,j=1}^{J} \sum_{p=1}^{P} |B_{p,ij}|, \tag{2}$$

where $\operatorname{tr}(\cdot)$ is the trace operator, $\lambda > 0$ is a regularization parameter associated to the l_1-penalty (Tibshirani, 1996) on the ij^{th} entry of $\boldsymbol{B}_p$ denoted by $[\boldsymbol{B}_p]_{ij} = B_{p,ij}$. This penalty ensures that the estimation $\boldsymbol{B}$ is feasible even if the number of parameters exceeds the time series length. Moreover, it sets some elements of $\widehat{\boldsymbol{B}}$ exactly equal to zero. The larger the regularization parameter λ, the sparser $\widehat{\boldsymbol{B}}$. Other type of penalties on $\boldsymbol{B}$ could be used, like the group lasso of Yuan and Lin (2006).

In line with Rothman et al. (2010), we account for correlated errors by including the estimation of the inverse error covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. We extend (2) and jointly estimate $\boldsymbol{B}$ and $\boldsymbol{\Omega}$ by minimizing the negative log likelihood:

$$(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Omega}}) = \underset{\boldsymbol{B},\boldsymbol{\Omega}}{\operatorname{argmin}} \ \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\boldsymbol{\Omega}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] - \frac{1}{2}\log|\boldsymbol{\Omega}| + \lambda \sum_{i,j=1}^{J} \sum_{p=1}^{P} |B_{p,ij}| + \gamma \sum_{i \neq j}^{J} |\omega_{ij}|, \tag{3}$$

where ω_{ij} is the ij^{th} element of $\boldsymbol{\Omega}$, and $\gamma > 0$ is the regularization parameter associated to the l_1-penalty on the off-diagonal elements of the inverse error covariance matrix (Friedman et al., 2008). This l_1-penalty ensures that the estimation of $\boldsymbol{\Omega}$ is feasible even if the number of parameters exceeds the time series length. Furthermore, it sets some elements of $\widehat{\boldsymbol{\Omega}}$ to zero. The larger the regularization parameter γ, the sparser $\widehat{\boldsymbol{\Omega}}$. We refer to the estimator in (3) as the *"Gaussian Lasso"*.

2.2 Student t model

We depart from the normality assumption and assume that the error terms are distributed according to a multivariate t distribution $t_{J,\nu}(\mathbf{0}, \boldsymbol{\Psi})$, where ν are the degrees of freedom and $\boldsymbol{\Psi}$ is the scale matrix, with associated variance-covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\nu/(\nu - 2)$ if $\nu > 2$. The associated density function is given by

$$\frac{\Gamma((\nu + J)/2)|\boldsymbol{\Psi}|^{-1/2}}{(\pi\nu)^{J/2}\Gamma(\nu/2)[1 + \delta_t(\boldsymbol{e}_t, \boldsymbol{\Psi})/\nu]^{(\nu+J)/2}},$$

with $\delta_t(\boldsymbol{e}_t, \boldsymbol{\Psi}) = \boldsymbol{e}_t' \ \boldsymbol{\Psi}^{-1} \ \boldsymbol{e}_t$ (Kotz and Nadarajah, 2004, p1).

Let $\boldsymbol{\Theta}_t$ be a J-dimensional random vector distributed as a multivariate normal $N_J(\mathbf{0}, \boldsymbol{\Psi})$ independent of the random variable τ_t distributed as a Gamma $\Gamma(\nu/2, \nu/2)$. Then, the Student t-distributed random vector $\boldsymbol{e}_t$ can be written as a Gaussian scale-mixture of $\boldsymbol{\Theta}_t$ and τ_t (e.g. Kotz and Nadarajah, 2004, p2). That is, $\boldsymbol{e}_t = \boldsymbol{\Theta}_t/\sqrt{\tau_t}$ follows a t distribution $t_{J,\nu}(\mathbf{0}, \boldsymbol{\Psi})$. Hence, $\boldsymbol{e}_t$ has a normal conditional distribution

$$\boldsymbol{e}_t|\tau_t \sim N_J\left(\mathbf{0}, \ \frac{\boldsymbol{\Psi}}{\tau_t}\right)$$

and

$$\tau_t|\boldsymbol{e}_t \sim \Gamma\left(\frac{\nu+J}{2},\ \frac{\nu+\delta_t(\boldsymbol{e}_t,\boldsymbol{\Psi})}{2}\right),$$

where the arguments of the Gamma distribution are respectively the shape and scale parameters (Liu and Rubin, 1995). By the properties of the Gamma distribution

$$E\left[\tau_t|\boldsymbol{e}_t\right] = \frac{\nu+J}{\nu+\delta_t(\boldsymbol{e}_t,\boldsymbol{\Psi})}. \tag{4}$$

To simplify the notation we consider from now on the vector forms $\boldsymbol{\tau} = [\tau_1,\ldots,\tau_N]'$ and $\boldsymbol{\delta}(\boldsymbol{e}_t,\boldsymbol{\Psi}) = [\delta_1(\boldsymbol{e}_t,\boldsymbol{\Psi}),\ldots,\delta_N(\boldsymbol{e}_t,\boldsymbol{\Psi})]'$.

The joint estimator of $\boldsymbol{B}$ and $\boldsymbol{\Omega}$ with t-distributed errors is obtained by minimizing the negative log likelihood:

$$(\widehat{\boldsymbol{B}},\widehat{\boldsymbol{\Omega}}\ |\tilde{\boldsymbol{\tau}}) = \underset{\boldsymbol{B},\boldsymbol{\Omega}}{\operatorname{argmin}}\ \frac{1}{2N}\operatorname{tr}\left[\tilde{\boldsymbol{\tau}}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})\ \boldsymbol{\Omega}\ (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})'\right] - \frac{1}{2}\log|\boldsymbol{\Omega}| + \lambda\sum_{i,j=1}^{J}\sum_{p=1}^{P}|B_{p,ij}| + \gamma\sum_{i\neq j}^{J}|\omega_{ij}|, \tag{5}$$

where we exploit the fact that the errors have a normal conditional distribution, that $\boldsymbol{\Psi}^{-1}$ is proportional to $\boldsymbol{\Omega}$, and with $\tilde{\boldsymbol{\tau}}$ a $N\times N$ diagonal matrix containing on its diagonal the expected realization of $\boldsymbol{\tau}$ as in (4). The estimator in equation (5) is the weighted version of the estimator in equation (3) with weights $\boldsymbol{\tau}$. We refer to the estimator in (5) as the *"t-Lasso"*. A special case of (5) is the t-lasso of Finegold and Drton (2011) for graphical models, where only the interdependences between responses are modeled and no predictors are considered.

3 Algorithm

As common in previous studies (Franses and Lucas, 1998; Finegold and Drton, 2011), we first assume that the degrees of freedom ν are known and discuss the EM (Expectation Maximization) algorithm used to obtain the t-Lasso estimator with known degrees of freedom. However, in practice ν is not known and needs to be estimated. To this end, we use the ECM (Expectation Conditional Maximization) algorithm to obtain the t-Lasso estimator with unknown degrees of freedom (Liu and Rubin, 1995).

3.1 EM algorithm with known ν

For known degrees of freedom ν, the solution of the objective function in (5) is approximated using the EM algorithm (see Finegold and Drton, 2011 for graphical models). We treat $\boldsymbol{\tau}$ as a hidden variable and estimate $\boldsymbol{B}$ and $\boldsymbol{\Omega}$. In the E-step, we compute the conditional expectation of the weights $\boldsymbol{\tau}$ according to (4). In the M-step, we solve the optimization problem in (5) conditional on the value of $\boldsymbol{\tau}$ found in the E-step and obtain the estimates $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\Omega}}$. This M-step corresponds exactly to solving the problem with Gaussian errors for weighted observations $\boldsymbol{Y}^* = \boldsymbol{\tau}^*\boldsymbol{Y}$ and $\boldsymbol{X}^* = \boldsymbol{\tau}^*\boldsymbol{X}$, where $\boldsymbol{\tau}^*$ is the $N\times N$ diagonal matrix, with

Algorithm 1 Expectation Maximization (EM) algorithm

Input $\mathbf{Y}$, $\mathbf{X}$, degrees of freedom ν and desired accuracy ε.

Initialization Set the initial values of $\widehat{\boldsymbol{\Omega}}^{(0)} = \mathbf{I}_J$ and $\widehat{\boldsymbol{B}}_p^{(0)} = \boldsymbol{I}_J$ for $1 \leq p \leq P$.

Iteration Iterate the following steps for $m = 0, 1, 2, \ldots$:

E-step Compute the weights:

$$\widehat{\boldsymbol{\tau}}^{(m+1)} = \frac{\nu + J}{\nu + \widehat{\boldsymbol{\delta}}^{(m)}(\boldsymbol{e}_t, \boldsymbol{\Psi})}.$$

M-step Compute $\widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}}^{(m+1)}$ using Algorithm 3 in Appendix B with inputs $\mathbf{Y} = \widehat{\boldsymbol{\tau}}^{*(m+1)}\boldsymbol{Y}$, and $\mathbf{X} = \widehat{\boldsymbol{\tau}}^{*(m+1)}\boldsymbol{X}$.

Convergence Iterate until the relative change in the value of the objective function in (5) in two successive iterations is smaller than ε.

Output $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^{(m+1)}$.

elements $[\boldsymbol{\tau}^*]_{tt} = ([\tilde{\boldsymbol{\tau}}]_{tt})^{1/2}$. Thus, each observation at time t gets weight $(\tau_t)^{1/2}$. Algorithm 1 gives the EM algorithm.

3.2 ECM algorithm with unknown ν

For unknown degrees of freedom ν, the EM algorithm can be extended to the ECM algorithm to include the estimation of ν (e.g. Liu and Rubin, 1995). The M-step is replaced by two constrained maximizations (CM) steps, where first $(\boldsymbol{B}, \boldsymbol{\Omega})$ and then ν are estimated. An E-step is introduced before each CM-step, such that the weights $\boldsymbol{\tau}$ are estimated twice in each iteration. This results is a multi-cycle version of the EM algorithm.

To estimate ν, rewrite the complete data log likelihood as the sum of the log likelihood of $(\boldsymbol{B}, \boldsymbol{\Omega})$ and the log likelihood of ν (for details, see Liu and Rubin, 1995). The negative log likelihood function of $(\boldsymbol{B}, \boldsymbol{\Omega}, \nu)$ is

$$L(\boldsymbol{B}, \boldsymbol{\Omega}, \nu) = -N \log(\Gamma(\frac{\nu + J}{2})) + N \log(\Gamma(\frac{\nu}{2})) + \frac{N}{2} \log(\det(\boldsymbol{\Omega})) \\ - \frac{N\nu}{2} \log(\nu) + \frac{N(\nu + J)}{2} \log(\nu + \boldsymbol{\delta}(\boldsymbol{e}_t, \boldsymbol{\Psi})). \quad (6)$$

Given $\boldsymbol{B}$ and $\boldsymbol{\Omega}$, we minimize (6) with respect to ν. The estimate of ν is given by the solution to the equation

$$-\phi(\frac{\nu}{2}) + \log(\frac{\nu}{2}) + \frac{1}{N} \sum_{i=1}^{N} (\log(\tau_i) - \tau_i) + 1 + \frac{1}{N} \sum_{i=1}^{N} \left[\phi\left(\frac{\nu + J}{2} \right) - \log\left(\frac{\nu + J}{2} \right) \right] = 0, \quad (7)$$

where $\phi(\cdot)$ is the derivative of the log Gamma function. Algorithm 2 gives the ECM algorithm.

Algorithm 2 Expectation Conditional Maximization (ECM) algorithm

Input $\mathbf{Y}$, $\mathbf{X}$, and desired accuracy ε.

Initialization Set the initial values of $\widehat{\nu}^{(0)} = 1000$, $\widehat{\boldsymbol{\Omega}}^{(0)} = \mathbf{I}_J$ and $\widehat{\boldsymbol{B}}_p^{(0)} = \boldsymbol{I}_J$ for $1 \leq p \leq P$.

Iteration Conditional on $\widehat{\nu}^{(m)}$, iterate the following steps for $m = 0, 1, 2, \ldots$:

E-step 1 Compute the weights:

$$\widehat{\boldsymbol{\tau}}^{(m+\frac{1}{2})} = \frac{\widehat{\nu}^{(m)} + J}{\widehat{\nu}^{(m)} + \widehat{\boldsymbol{\delta}}^{(m)}(\boldsymbol{e}_t, \boldsymbol{\Psi})}.$$

CM-step 1 Compute $\widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}}^{(m+1)}$ using Algorithm 3 with inputs $\mathbf{Y} = \widehat{\boldsymbol{\tau}}^{*(m+\frac{1}{2})}\boldsymbol{Y}$, and $\mathbf{X} = \widehat{\boldsymbol{\tau}}^{*(m+\frac{1}{2})}\boldsymbol{X}$.

E-step 2 Compute the weights:

$$\widehat{\boldsymbol{\tau}}^{(m+1)} = \frac{\widehat{\nu}^{(m)} + J}{\widehat{\nu}^{(m)} + \widehat{\boldsymbol{\delta}}^{(m+1)}(\boldsymbol{e}_t, \boldsymbol{\Psi})}.$$

CM-step 2 Compute $\widehat{\nu}^{(m+1)}$ with a one-dimensional search on (7).

Convergence Iterate until the relative change in the value of the objective function in (5) in two successive iterations is smaller than ε.

Output $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}^{(m+1)}$, $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^{(m+1)}$ and $\widehat{\nu} = \widehat{\nu}^{(m+1)}$.

4 Simulations

We analyze the performance of the *t-Lasso*, i.e. the solution of equation (5). We use both the t-Lasso with fixed degrees of freedom (at the true value) and with estimated degrees of freedom. Their performance is compared to two alternative estimators: the *Least Squares* (LS), and the *Gaussian Lasso*, i.e. the solution of equation (3). The LS is the standard non-penalized benchmark, the Gaussian Lasso is the benchmark for Gaussian models.

Data generating process. We simulate from a VAR of order $P = 2$ with $J = 10$ time series. The dimensions of the VAR are in line with the ones in the application to be discussed in Section 5. The data generating process is:

$$\boldsymbol{y}_t = \boldsymbol{B}_1\boldsymbol{y}_{t-1} + \boldsymbol{B}_2\boldsymbol{y}_{t-2} + \boldsymbol{e}_t,$$

for $P + 1 \leq t \leq T = 100$. The autoregressive coefficient matrices are highly sparse in a structured manner: B_1 and B_2 have the same sparsity structure with non-zero elements (0.4 for B_1 and 0.2 for B_2) on the main diagonal and on the first row. The error terms $\boldsymbol{e}_t$ follow a multivariate Student $t_{J,\nu}(\mathbf{0}, \boldsymbol{\Psi})$, with

Table 1: Mean Absolute Estimation Error for the four estimators and different degrees of freedom ν.

ν	LS	Gaussian	t-Lasso	
		Lasso	ν fixed	ν estimated
1	0.638	0.187	0.088	0.089
2	0.135	0.096	0.088	0.088
3	0.107	0.090	0.088	0.089
5	0.098	0.090	0.088	0.089
10	0.095	0.091	0.089	0.090
∞	0.093	0.091	0.091	0.091

$\nu \in \{1, 2, 3, 5, 10, \infty\}$. In the special case $\nu = 1$, the distribution is a multivariate Cauchy distribution, whereas for $\nu \to \infty$ the distribution is a multivariate normal.[1] The ij^{th} entry of $\boldsymbol{\Psi}$ is $[\boldsymbol{\Psi}]_{i,j} = 0.1^{|i-j|}$, such that inverse error covariance matrix $\boldsymbol{\Omega}$ is a band matrix. We take $S = 1000$ simulations runs.

Performance measures. The different estimators are compared in terms of their estimation accuracy. To evaluate the estimation accuracy, we use the Mean Absolute Estimator Error

$$\text{MAEE}(\boldsymbol{B}, \widehat{\boldsymbol{B}}) = \frac{1}{S}\frac{1}{PJ^2}\sum_{S=1}^{S}\sum_{i,j=1}^{J}\sum_{p=1}^{P}|\widehat{B}_{p,ij}^{(s)} - B_{p,ij}^{(s)}|,$$

where $\widehat{B}_{p,ij}^{(s)}$ is the ij^{th} entry of the estimate $\widehat{\boldsymbol{B}}_p^{(s)}$ in simulation run s.

Simulation results. Table 1 reports the MAEEs for the four estimators and the different values of the degrees of freedom ν. The t-Lasso with ν fixed always achieves the best MAEE. It is closely followed by the t-Lasso with ν estimated. There is no considerable loss in estimation accuracy when estimating ν. Both t-Lasso estimators perform significantly better than the Gaussian Lasso (the difference in estimation accuracy is tested with a paired t-test, all p-values < 0.01). The margin by which the t-Lasso estimators outperform the Gaussian Lasso decreases for larger degrees of freedom. In particular, for $\nu \to \infty$ there is no significant difference between the t-Lasso estimators and the Gaussian Lasso. As $\nu \to \infty$, a multivariate t distribution can be viewed as the approximation of a multivariate normal. Hence, as expected, the Gaussian Lasso and the t-Lasso estimators show more similar performances as the degrees of freedom ν increase. The t-Lasso estimators also significantly outperform the LS for all values of ν: they improve estimation accuracy by, for instance, 17% for $\nu = 3$.

[1] For $\nu \to \infty$, we take the t-Lasso with fixed $\nu = 1000$ since there is practically no difference between a multivariate t with $\nu = 1000$ and a multivariate normal. Likewise, we limit the maximum value of the estimated degrees of freedom to 1000.

Finally, Figure 1 shows the frequencies of the estimated degrees of freedom by the t-Lasso, for the different settings $\nu \in \{1, 2, 3, 5, 10, \infty\}$. For all settings, the estimated degrees of freedom are closely centered around the true value (vertical red line). The variance of the estimated degrees of freedom is the lowest for small values of ν, it increases for larger values of ν. The average (averaged over all simulation runs) estimated degrees of freedom are 0.94 (for $\nu = 1$), 2.10 (for $\nu = 2$), 3.16 (for $\nu = 3$), 5.30 ($\nu = 5$), 11.14 ($\nu = 10$), 713.63 ($\nu \to \infty$). Hence, the degrees of freedom are accurately estimated.

5 Data and volatility spillover analysis

In this section, we first present the data. Then, we define volatility spillovers based on the forecast error variance decomposition. Finally, we present a network analysis tool showing these volatility spillovers.

5.1 Data

We study $J = 10$ agricultural (corn, wheat, soybean, sugar, cotton, coffee), energy (crude oil, gasoline, natural gas) and biofuel (ethanol) commodities. We use daily information about the opening, highest and lowest prices of the nearest to maturity contracts traded in the corresponding future markets. Data are available on Thomson Reuters Eikon. The time span ranges from January 3^{rd} 2012 to October 28^{th} 2016, thus $T = 1070$ daily observations, where we discarded all days where the trade volume of at least one contract is zero.

First, we obtain univariate volatility measures using the realized daily high-low range proposed by Parkinson (1980), see Appendix A for more details. Table 2 reports the kurtosis of the log volatilities. Although the logarithmic transformation should provide a better approximation to the Gaussian distribution (Diebold and Yilmaz, 2015, p18), the log volatilities are leptokurtic (i.e. there is evidence of fat tails) and present strong deviations from the normal distribution. We check for stationarity of the estimated volatilities with univariate unit root tests and the pooled unit root test of Levin et al. (2002) and find strong evidence in favor of stationarity (p-values < 0.01).

Table 2: Kurtosis of log realized ranges.

	CRUO	GASO	NATG	ETHA	CORN	WHEA	SOYB	SUGA	COTT	COFF
Kurtosis	6.85	5.75	6.64	3.63	5.91	6.54	6.86	4.78	3.47	4.18

Notes: CRUO for crude oil, GASO for gasoline, NATG for natural gas, ETHA for ethanol, CORN for corn, WHEA for wheat, SOYB for soybean, SUGA for sugar, COTT for cotton and COFF for coffee.

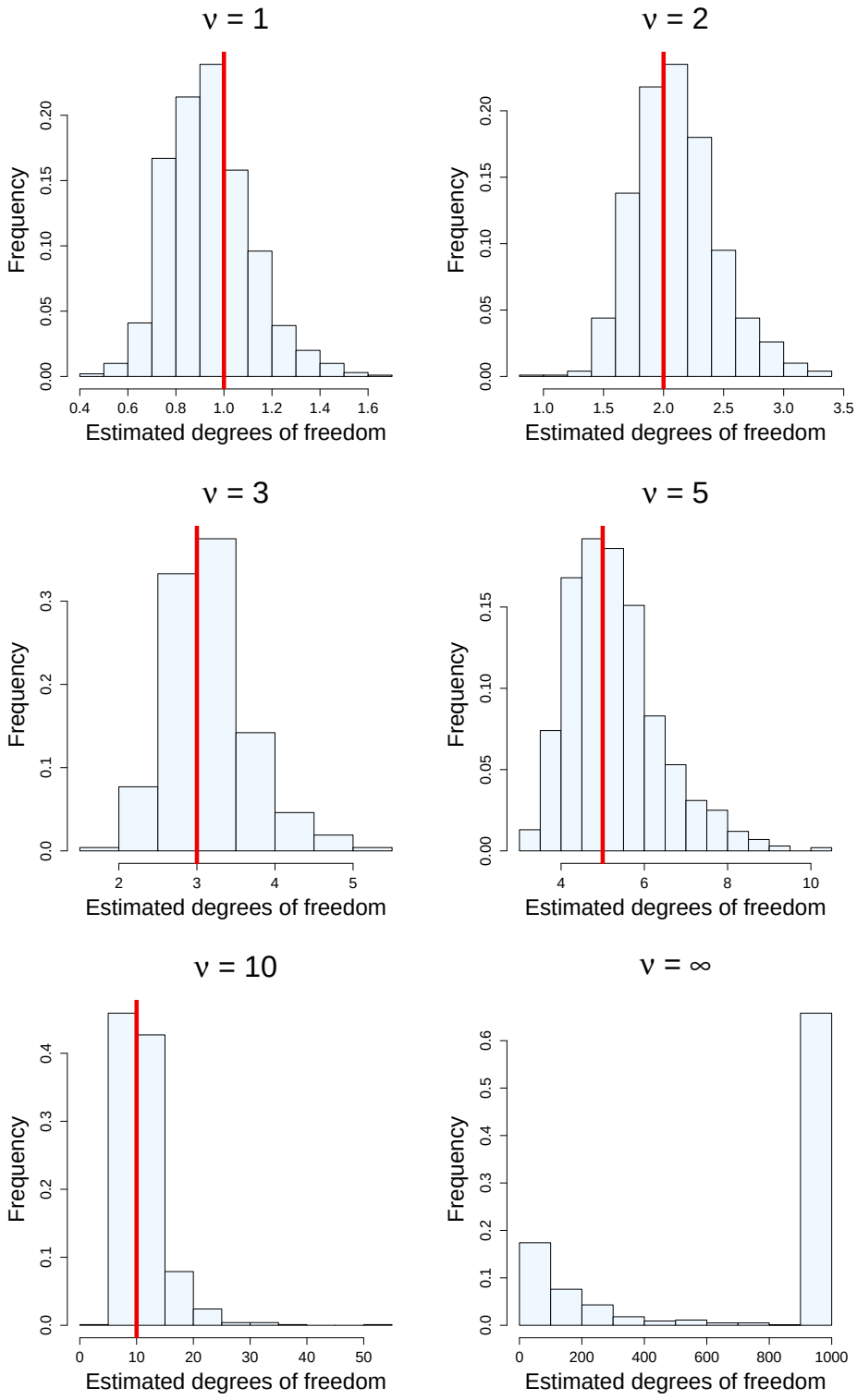

Figure 1: Frequency of the estimated degrees of freedom, for each setting $\nu = 1, 2, 3, 5, 10, \infty$ in the simulation study. The true value of the degrees of freedom is indicated by the vertical red line (dark gray on a gray scale).

Second, we employ a rolling window over the period 2012-2016 with window size $W = 220$ days. We consider a forecast horizon $h = 5$ days. At each time point $t = W, \ldots, T - W$, we jointly model the log volatilities in a VAR(P) model, where P is selected via Bayesian Information Criterion BIC. We assume that the error terms follow a multivariate t distribution and estimate the degrees of freedom using the t-Lasso. Then, we study volatility spillovers based on forecast error variance decomposition, as described in the following subsection.

5.2 Forecast error variance decomposition

We summarize the volatility spillovers by means of the forecast error variance decomposition. Consider the Vector Moving Average (VMA) representation of the VAR(P) in model (1)

$$\boldsymbol{y}_t = \sum_{p=0}^{\infty} \boldsymbol{\theta}_p \boldsymbol{e}_{t-p},$$

where $\boldsymbol{\theta}_p$ is the moving average coefficient matrix at lag p (cfr. Wold's representation theorem, Lütkepohl, 2005, p25). Let $\widehat{\boldsymbol{y}}_{t+h}$ be the h-step ahead forecast for time $t + h$ made at time t with forecast horizon h. Then, the h-step-ahead forecast error for the j^{th} component of $\boldsymbol{y}_t$ is

$$\widehat{y}_{t+h,j} - y_{t+h,j} = \sum_{p=0}^{h-1} \theta_{p,j1} e_{t+h-p,1} + \ldots + \theta_{p,jJ} e_{t+h-p,J},$$

where $e_{t,j}$ is the j^{th} component of $\boldsymbol{e}_t$, and $\theta_{p,jk}$ is the jk^{th} entry of $\boldsymbol{\theta}_p$.

The $h-$step ahead forecast error variance decomposition measures the fraction of the h-step ahead forecast error variance of one time series due to the shocks in the other time series (including its own). To obtain these variance decompositions, the correlations between the error terms $\boldsymbol{e}_t$ need to be taken into account. This is typically done either by Cholesky's decomposition or by the generalized variance decomposition. The former guarantees the orthogonality of the shocks, with the drawback that the associated variance decompositions depend on the ordering of the time series (Lütkepohl, 2005, p66). Therefore, we follow Diebold and Yilmaz (2014) and use the *generalized variance decomposition* (Koop et al., 1996; Pesaran and Shin, 1998), which is invariant to the time series ordering.

Although Diebold and Yilmaz (2014) assume normally distributed errors, the generalized variance decomposition can be extended to other error distributions. Assuming that $\boldsymbol{e}_t$ follows a multivariate t distribution, we rely on the fact that the mean of a conditional multivariate t distribution is the same as the mean of a conditional multivariate normal distribution (Ding, 2016). Then, the generalized variance decomposition in case of t-distributed errors is equal to the one with normally distributed errors. Following Diebold and Yilmaz (2014), the proportion of the h-step ahead forecast error variance of commodity j due to commodity k is

$$w_{h,jk} = \frac{\sigma_{kk}^{-1} \sum_{p=0}^{h-1} (\boldsymbol{n}_j' \boldsymbol{\theta}_p \boldsymbol{\Sigma} \boldsymbol{n}_k)^2}{\sum_{p=0}^{h-1} \boldsymbol{n}_j' \boldsymbol{\theta}_p \boldsymbol{\Sigma} \boldsymbol{\theta}_p' \boldsymbol{n}_j},$$

where σ_{kk} is the kk^{th} entry of $\boldsymbol{\Sigma}$, and $\boldsymbol{n}_k$ is the selection vector of length J with the k^{th} element equal to one and zeros elsewhere. As the error terms are not orthogonal, in general $\sum_{k=1}^{J} w_{h,jk} \neq 1$. We follow Diebold and Yilmaz (2014) and consider the normalized variance decomposition

$$\tilde{w}_{h,jk} = \frac{w_{h,jk}}{\sum_{k=1}^{J} w_{h,jk}},$$

which is in the interval $[0, 1]$ by construction. The above normalized variance decomposition corresponds analytically to the generalized variance decomposition proposed by Lanne and Nyberg (2016).

We then define the *volatility spillover* from commodity k to commodity j (Diebold and Yilmaz, 2015) based on the h-step ahead forecast error variance as

$$s_{h,k\to j} = 100 * \tilde{w}_{h,jk}. \tag{8}$$

Moreover, we define the *volatility spillover index* as

$$s_h = \sum_{\substack{j\neq k \\ j,k=1}}^{J} s_{h,k\to j}, \tag{9}$$

as a single measure of the overall volatility spillovers. Notice that, contrary to the volatility spillover, the volatility spillover index does not express a percentage, but only an overall proxy of the magnitude of the volatility spillovers.

5.3 Network analysis

We visualize the volatility spillovers using a network analysis tool (e.g. Diebold and Yilmaz, 2015). The nodes in the network are the different commodities. An edge from commodity k to commodity j is drawn if $s_{k\to j}$ from equation (8) is non-zero; the edge width represents the magnitude of the volatility spillover. Hence, our network is directed and weighted.

This network is, however, not necessarily sparse. Indeed, the sparsity of $\widehat{\boldsymbol{B}}$ is not necessarily preserved in the estimated VMA coefficient matrices $\widehat{\boldsymbol{\theta}}$. Consequently, the network will picture many non-zero volatility spillovers, although many of them may be quite small. For practical reason, it is therefore desirable to only visualize the most important spillovers. To this end, we give the network where only a certain portion of the largest volatility spillovers is visualized.

6 Results

In this section, we first present the rolling window analysis (cfr. Section 5.1). Second, we picture the networks showing the volatility spillovers (cfr. Section 5.3). Finally, we forecast the volatilities and we show the good performance of the t-Lasso in terms of forecast accuracy.

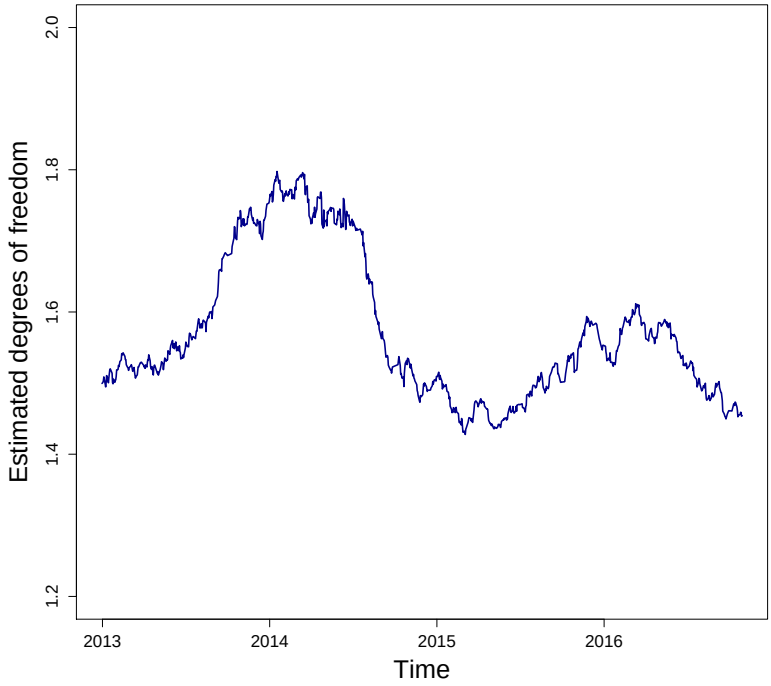

Figure 2: Estimated degrees of freedom of the multivariate t distribution of the VAR residuals by the t-Lasso for a 220-day rolling window. The x-axis represents the ending date of each window.

6.1 Rolling window analysis

Estimated degrees of freedom. Figure 2 reports the estimated degrees of freedom of the multivariate t distribution of the VAR residuals at each time point t, being t the end point of each time window. The average value of $\widehat{\nu}$ is 1.57, with maximum 1.80 and minimum 1.43. Overall, we observe that the estimated degrees of freedom are very low. This confirms the existence of heavy tails in the joint distribution of the VAR residuals and justifies the use of the t-Lasso rather than the Gaussian Lasso.

If we look at the evolution of $\widehat{\nu}$ over time, we detect lower values of the estimated degrees of freedom in the time windows ending by the second half of 2014 and by the first months of 2015. Volatile commodity markets characterized the second half of 2014, mainly driven by the drop of almost 50% in crude oil price (Knittel and Pindyck, 2016). Smaller values of $\widehat{\nu}$ indicate more pronounced extreme realizations in the VAR residuals and reflect the less predictable behavior of commodity markets in those months.

Volatility spillover index. Figure 3 reports the evolution of the volatility spillover index (cfr. equation (9)) for the t-Lasso as a function of time t, being t the end point of each time window. We observe that the volatility spillover index experiences large swings over time. In particular, in the second half of 2014 we

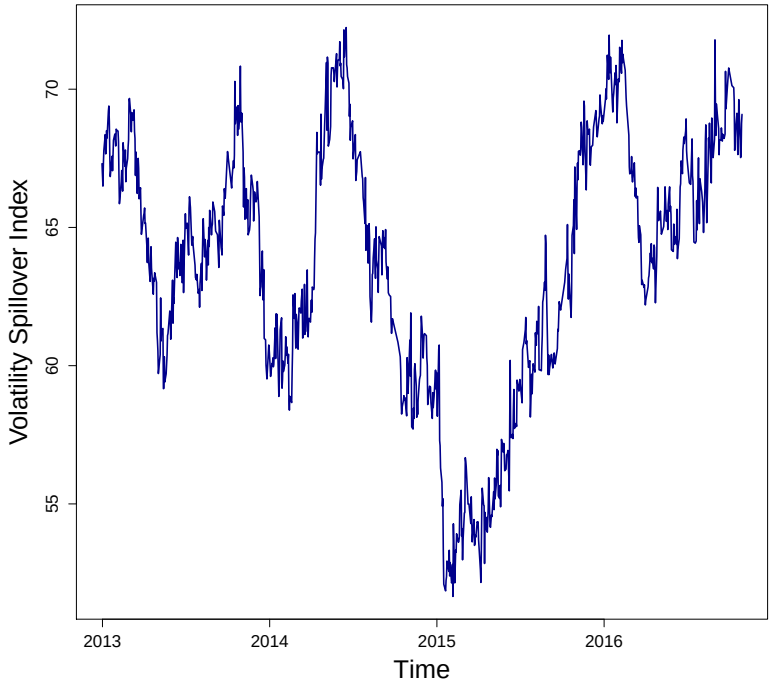

Figure 3: Volatility spillover index for the t-Lasso for a 220-day rolling window. The x-axis represents the ending date of each window.

detect a large drop in the overall level of volatility spillovers. This drop is not permanent and during 2015 the volatility spillover index returns to the level prior to 2014.

The downturn in volatility spillover index can also be linked to the fall of crude oil price that occurred in the second half of 2014. Crude oil price experienced large downturns driven by (i) a larger supply from OPEC and non-OPEC countries and (ii) a weak global demand due to the slowdown of the world economy, notably the Chinese one. The drop in energy prices made biofuel less profitable and implied lesser and weaker volatility spillovers among energy, biofuel and agriculture commodities.

6.2 Network analysis

We visualize the volatility spillovers in a network of connected commodities at a specific ending date of the 220-day rolling window. Figure 4 top-row presents the network for the time window ending on October 28^{th} 2016, the last time point in our data set. We draw the network of the largest 15% volatility spillovers (cfr. Section 5). For instance, a directed edge is drawn from gasoline (GASO) to crude oil (CRUO) since the volatility spillover from gasoline towards crude oil is different from zero and belongs to the largest 15% volatility spillovers. The edge width represents the magnitude of the volatility spillover. Figure 4 also reports

the networks for the time windows ending on February 5^{th} 2015 (bottom left) and on June 17^{th} 2014 (bottom right). For simplicity, from now on we refer to the three networks only by the year of the ending date of their windows. The 2015 and 2014 one correspond to the time windows with the lowest and highest values of volatility spillover index, respectively.

Link energy-biofuel. In all networks we find volatility spillovers among energy and biofuel commodities. In the 2015 and 2014 networks, gasoline is the only energy commodity connected by an edge with ethanol. As ethanol is often blended in gasoline for consumption (Serra and Zilberman, 2013), it is not surprising to observe volatility spillovers between the two commodities in times of high energy prices (i.e. network 2014) and in times of large energy price changes (i.e. network 2015).

Link energy-agriculture. Bidirectional volatility spillovers between energy and agriculture are detected in all networks confirming the findings of Rezitis (2015). In the 2016 network, natural gas shows numerous and large spillovers from and towards agricultural commodities, whereas in the 2015 and 2014 networks, gasoline is the most connected energy commodity with agriculture. Both natural gas and gasoline are key commodities in the agricultural sector: the former is the major production cost for fertilizers, whereas the latter is the most important fuel (together with diesel). If energy prices are high (e.g. network 2014), fuel has a large impact on agricultural commodities, which could explain the numerous volatility spillovers from/to gasoline. Conversely, as energy prices are low and fuel is cheap (e.g. network 2016), variations in fertilizer prices largely affect agricultural commodities volatilities, which could explain the many volatility spillovers from/to natural gas.

In all networks, various agricultural commodities are involved in volatility spillovers to and from energy. In general, volatility spillovers from and to energy involve both agricultural crops that are primary inputs for biofuel production - like wheat, soybean and sugar -, and crops that do not have a direct link with biofuels, - like coffee -, as in Rezitis (2015).

Link biofuel-agriculture. Volatility spillovers between biofuel and agriculture are found only in the 2016 and 2014 networks. These volatility spillovers are bidirectional and involve wheat, sugar and cotton: this is consistent with our expectation since these crops can be used for ethanol production. In the 2015 network, we observe no volatility spillovers between biofuel and agriculture. In that period energy prices substantially dropped (Knittel and Pindyck, 2016): this made biofuel less attractive compare to other standard fuels and could have resulted in weaker volatility spillovers from and to agriculture.

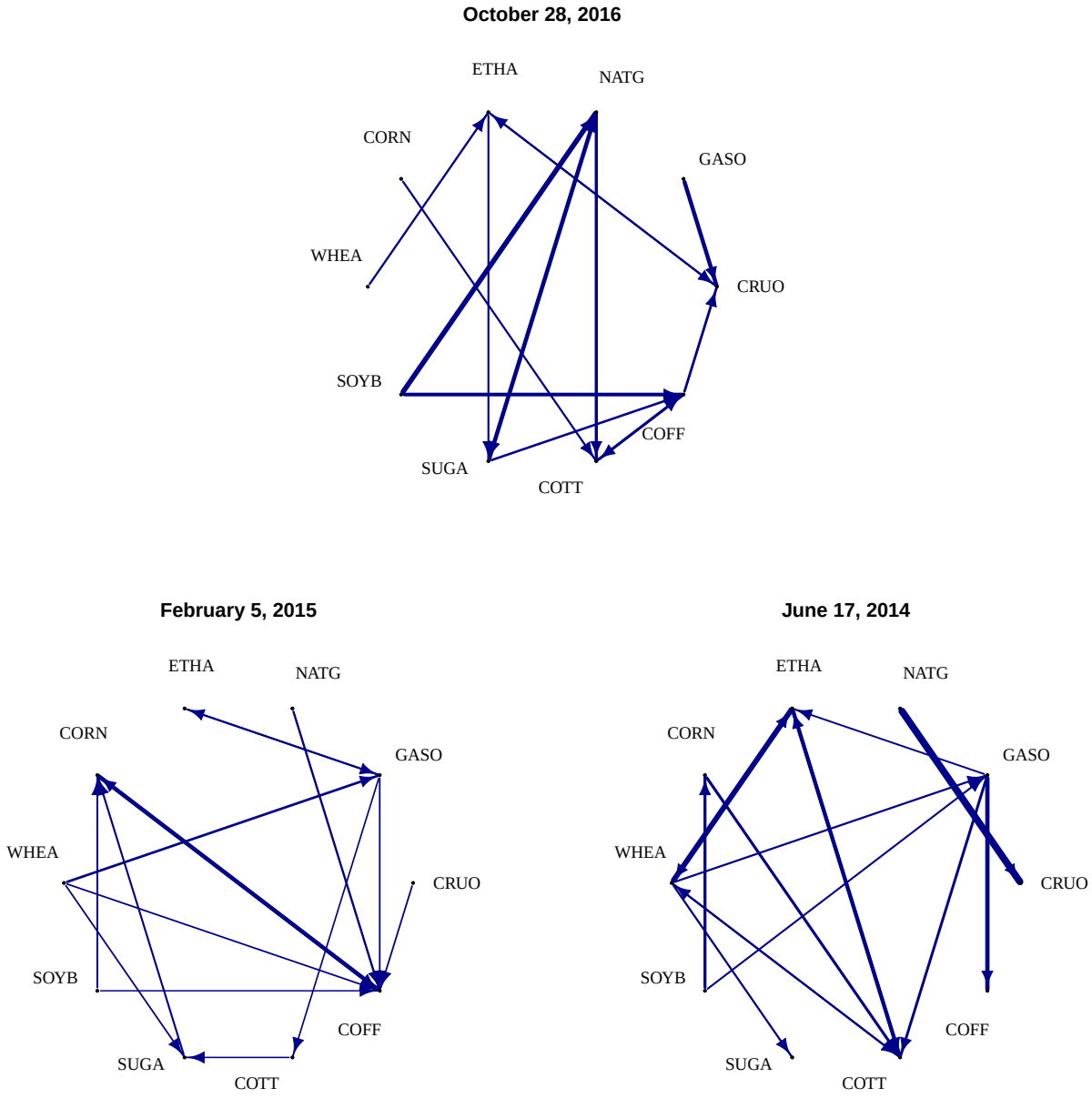

Figure 4: Commodity networks of volatility spillovers for different ending times of the 220-day rolling windows: October 28^{th} 2016 (top), February 5^{th} 2015 (bottom left) and June 17^{th} 2014 (bottom right). A directed edge is drawn from one commodity to another if the associated volatility spillover is estimated non-zero and belongs to the largest 15% volatility spillovers of that particular network. The edge width represents the magnitude of the volatility spillover.

6.3 Forecast accuracy

We now obtain forecasts for the commodity volatilities. In particular, we forecast the vector of log volatilities $\widehat{\boldsymbol{y}}_t^{(t-h)} = [\widehat{y}_{t,1}^{(t-h)}, \ldots, \widehat{y}_{t,J}^{(t-h)}]$ based on the time window $[t-W+1, t]$, where t is the end point of the time window of size W and $\widehat{y}_{t,j}^{(t-h)}$ is the h-step ahead forecast of series j for time t made at time $t-h$. The forecast is obtained using the VAR model in (1), with autoregressive coefficients $\widehat{\boldsymbol{B}}$ obtained with the t-Lasso with ν estimated, the Gaussian Lasso and the LS. We compare these three estimators in terms of forecast accuracy by computing the Mean Absolute Forecast Error

$$\text{MAFE}(\widehat{\boldsymbol{v}}^2, \boldsymbol{v}^2) = \frac{1}{N-h-W+1} \frac{1}{J} \sum_{t=W}^{N-h} \sum_{j=1}^{J} |\widehat{v}_{t,j}^{2,(t-h)} - v_{t,j}^2|,$$

where $v_{t,j}^2$ is the volatility of series j obtained with the daily realized range and $\widehat{v}_{t,j}^{2,(t-h)} = \exp(\widehat{y}_{t,j}^{(t-h)})$ is the volatility forecast for time t made at time $t-h$ for the time window $[t-W+1, t]$. The smaller the value of the MAFE, the more accurate the volatility forecasts.

Results. Table 3 reports the results of the MAFE for the three estimators for different forecast horizons, namely $h = 1$, $h = 5$ and $h = 20$ days. Notice that the MAFE is computed for each time window and then averaged across all of them. The t-Lasso attains the best value of the MAFE for all forecast horizons. The difference in forecast accuracy with the other estimators is always significant based on a Diebold-Mariano test (Diebold and Mariano, 1995). For forecast horizon $h = 1$, the t-Lasso gives an improvement in terms of MAFE of 2% and of 42% over the Gaussian Lasso and the LS, respectively. The higher the forecast horizon, the better the performance of the t-Lasso compared to the Gaussian Lasso: for instance, for $h = 20$ our estimator gives an improvement of 21%.

Figure 5 reports the evolution of the MAFE (for $h = 20$) for the t-Lasso (blue solid line) and the Gaussian Lasso (red dashed line) as a function of time t, being t the end point of each time window. We do not plot the MAFE for the LS since its values are much larger. The t-Lasso attains a lower MAFE than the Gaussian Lasso in 99.9% of the time windows and the difference in forecast accuracy is confirmed to be significant in all time windows by a Diebold-Mariano test.

Overall, we find that the t-Lasso attains a better forecast performance than the Gaussian Lasso and the LS, regardless of the forecast horizon or the presence of volatility spillovers. Moreover, the good performance of our estimator in terms of forecast accuracy suggests that it is preferable to use the t-Lasso for the volatility spillovers analysis which is based on forecast error variance decomposition.

Table 3: Mean Absolute Forecast Error for the t-Lasso, the Gaussian Lasso and the LS for forecast horizon $h = 1, 5, 20$ on the 220-day rolling window.

Horizon	t-Lasso	Gaussian Lasso	LS
$h = 1$	0.473	0.482	0.821
$h = 5$	0.501	0.537	0.855
$h = 20$	0.629	0.797	0.932

7 Discussion

This paper studies volatility spillovers in a Vector AutoRegressive model accounting for the fat-tailedness of commodity volatilities. We extend the work of Diebold and Yilmaz (2015) and propose a penalized estimator of the VAR model with errors following a multivariate t distribution, of which we include the estimation of the degrees of freedom. Our estimator includes lasso type penalties and is computable even for a large number of relatively short time series. Alternatively, a Bayesian VAR approach could have been employed, as proposed in the contemporary work by Lee et al. (2016). Our simulation study shows that the proposed t-Lasso attains a better performance in terms of estimation accuracy than standard Gaussian and the Least Squares. On the one hand, the t distribution for the errors better captures the spikes typical of commodity volatility. On the other hand, the penalized estimator ensure good estimation accuracy even for VAR models that contain a large number of time series relative to the time series length.

We study the dynamics of volatility spillovers between $J = 10$ energy, biofuel and agricultural commodities using a rolling window approach. Our findings highlight that the distributions of the log volatilities and of the residuals of the VAR model are fat-tailed. Furthermore, we visualize volatility spillovers using networks built on forecast error variance decomposition accounting for t-distributed errors. We find evidence of bidirectional volatility spillovers between energy and agricultural commodities, regardless of the fact of being biofuel crops or not (Rezitis, 2015).

Our analysis is built on the volatility spillover definition by Diebold and Yilmaz (2014) and relies on generalized forecast error variance decomposition. Nevertheless, other definitions could have been used. For instance, we redid the analysis using Cholesky's decomposition (Diebold and Yilmaz, 2009) and the spectral decomposition (Hafner and Herwartz, 2006). In both cases, we obtain results comparable to the ones reported, with only minor changes in the magnitudes of the volatility spillovers. The same holds when using other volatility spillover definitions based on impulse response functions. For instance, we redid the analysis defining a volatility spillover as the sum of squared generalized impulse responses and draw comparable conclusions. Detailed results are available from the authors upon request.

Further research might follow several trajectories. On the one hand, it could be interesting to use

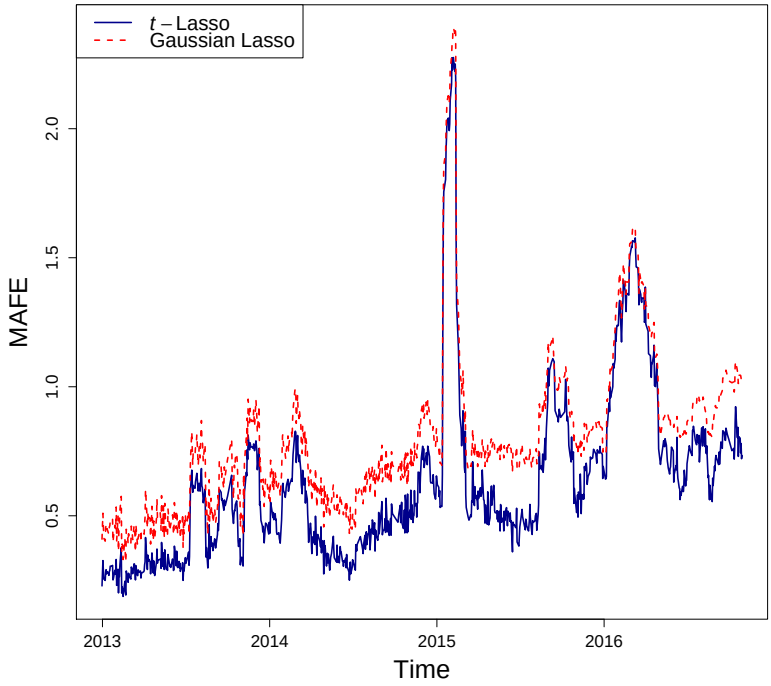

Figure 5: Mean Absolute Forecast Error for a 220-day rolling window and forecast horizon $h = 20$ for the t-Lasso (solid) and the Gaussian Lasso (dashed). The x-axis represents the ending date of each window.

other volatility measures (McAleer and Medeiros, 2008). Data on the opening, lowest and highest daily prices are easier to obtain than intra-day high-frequency returns, motivating our choice of studying daily realized ranges. However, depending on the realized measure used, the distribution of the log volatility could be less or more heavy-tailed and this might have different implications in terms of risk management (for instance, see Brownlees and Gallo (2010) for a Vale-at-Risk comparison between different realized measures of volatility). On the other hand, one might investigate the existence of asymmetries in the VAR errors, for instance by considering a multivariate skewed t distribution (Kotz and Nadarajah, 2004, p98) or by basing the volatility spillover analysis on realized semi-variances as in Baruník et al. (2015). With respect to the commodity application, it could be interesting to further investigate the relation between energy, biofuel and agricultural commodities allowing for structural changes in the volatility. Teterin et al. (2016) use a Fourier Flexible Form to explore the existence of volatility shifts using crude oil and corn future prices: a penalized approach would overcome the limitation in terms of number of time series.

Acknowledgments. We gratefully acknowledge support from the FWO (Research Foundation Flanders, contract number 12M8217N) and from the GOA/12/014 project of the Research Fund KU Leuven.

A Volatility measure

Following Parkinson (1980), we obtain a measure of volatility of a future contract using the high-low daily range estimator. Consider the daily information about opening price $O_{t,j}$, the highest price $H_{t,j}$ and the lowest price $L_{t,j}$ attained on date $1 \leq t \leq T$ for commodity $1 \leq j \leq J$. The high-low range estimator for the daily variance is

$$v_{t,j}^2 = \frac{1}{4\log(2)}(h_{t,j} - l_{t,j})^2,$$

where $h_{t,j} = \log(H_{t,j}) - \log(O_{t,j})$ and $l_{t,j} = \log(L_{t,j}) - \log(O_{t,j})$ are the maximum and minimum daily return, respectively. For a review on realized range measures of volatility see Shu and Zhang (2006) or Martens and van Dick (2007). The time series entering the VAR of equation (1) are the log trasformations of the realized ranges $\widehat{v}_{t,j}$ of commodity j at time t, that is $\mathbf{y}_t = [\log(\widehat{v}_{t,j}), \ldots, \log(\widehat{v}_{t,J})]'$. Notice that the original volatility series can be obtained by applying the exponent transformation.

B Algorithm Gaussian Lasso

The algorithm used to obtain the Gaussian Lasso is given in Algorithm 3. First, we solve for the autoregressive parameter $\boldsymbol{B}$ conditional on the inverse error covariance matrix $\boldsymbol{\Omega}$ using a block coordinate descent algorithm as in Friedman et al. (2007). Second we solve for $\boldsymbol{\Omega}$ conditional on $\boldsymbol{B}$, using the Graphical Lasso algorithm of Friedman et al. (2008). We iteratively repeat the two steps until converge of the objective function is reached.

Selection of regularization parameters. When solving for $\boldsymbol{B}|\boldsymbol{\Omega}$, we use a grid of regularization parameters λ and search for the optimal one minimizing the Bayesian Information Criterion (BIC)

$$\text{BIC}_\lambda = -2\text{log}L_\lambda + \text{df}_\lambda \text{log}(N),$$

where $\text{log}L_\lambda$ is the estimated likelihood, i.e. the first term in (3), using regularization parameter λ, and df_λ is the number of non-zero components of $\widehat{\boldsymbol{B}}_\lambda$. Likewise, when solving for $\boldsymbol{\Omega}|\boldsymbol{B}$ we use a grid of regularization parameters γ and search for the optimal one minimizing the BIC

$$\text{BIC}_\gamma = -2\text{log}L_\gamma + \text{df}_\gamma \text{log}(N),$$

where df_γ is the number of non-zero lower diagonal elements of $\widehat{\boldsymbol{\Omega}}$.

Algorithm 3 Algorithm for PML estimation with Gaussian errors

Input $\mathbf{Y}$, $\mathbf{X}$, and desired accuracy ε.

Initialization Set the initial value of $\widehat{\boldsymbol{\Omega}}^{(0)} = \mathbf{I}_J$.

Iteration Iterate the following steps for $m = 0, 1, 2, \ldots$:

Solving for $\boldsymbol{B}|\boldsymbol{\Omega}$. Compute $\widehat{\boldsymbol{B}}^{(m+1)}$:

$$\widehat{\boldsymbol{B}}^{(m+1)} = \underset{\boldsymbol{B}}{\operatorname{argmin}} \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\widehat{\boldsymbol{\Omega}}^{(m)}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'\right] + \lambda \sum_{i,j=1}^{J} \sum_{p=1}^{P} |B_{p,ij}|.$$

Solving for $\boldsymbol{\Omega}|\boldsymbol{B}$. Compute $\widehat{\boldsymbol{\Omega}}^{(m+1)}$:

$$\widehat{\boldsymbol{\Omega}}^{(m+1)} = \underset{\boldsymbol{\Omega}}{\operatorname{argmin}} \frac{1}{2N} \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(m+1)})\boldsymbol{\Omega}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}^{(m+1)})'\right] - \frac{1}{2}\log|\boldsymbol{\Omega}| + \gamma \sum_{i \neq j}^{J} |\omega_{ij}|.$$

Convergence Iterate until the relative change in the value of the objective function in (3) in two successive iterations is smaller than ε.

Output $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}^{(m+1)}$ and $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}^{(m+1)}$.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., Heiko, E., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61(1), 43–76.

Barigozzi, M., Hallin, M., 2017. A network analysis of the volatility of high dimensional financial series. Journal of the Royal Statistical Society: Series C (Applied Statistics) doi:10.1111/rssc.12177.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(2), 253–280.

Baruník, J., Kočenda, E., Vácha, L., 2015. Volatility spillovers across petroleum markets. The Energy Journal 36(3), 309–330.

Brownlees, C. T., Gallo, G. M., 2010. Comparison of volatility measures: A risk management perspective. Journal of Financial Econometrics 8(1), 29–56.

Callot, L. A. F., Kock, A. B., Medeiros, M. C., 2016. Modeling and forecasting large realized covariance matrices and portfolio choice. Journal of Applied Econometrics 10.1002/jae.2512.

Caporin, M., Velo, G. G., 2015. Realized range volatility forecasting: Dynamic features and predictive variables. International Review of Economics and Finance 40, 98–112.

Christensen, K., Podolskij, M., Vetter, M., 2009. Bias-correcting the realized range-based variance in the presence of market microstructure noise. Finance and Stochastics 13(2), 239–268.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27(1-3), 46–78.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253–263.

Diebold, F. X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Economic Journal 119(534), 158–171.

Diebold, F. X., Yilmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics 182(1), 119–134.

Diebold, F. X., Yilmaz, K., 2015. Financial and macroeconomics connectedness: A network approach to measurement and monitoring. Oxford University Press, New York, US.

Ding, P., 2016. On the conditional distribution of the multivariate t distribution. The American Statistician 70(3), 293–295.

Finegold, M., Drton, M., 2011. Robust graphical modeling of gene networks using classical and alternative t-distributions. The Annals of Applied Statistics 5(2A), 1057–1080.

Franses, P., Lucas, A., 1998. Outlier detection in cointegration analysis. Journal of Business & Economic Statistics 16(4), 459–468.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., 2007. Pathwise coordinate optimization. The Annals of Applied Statistics 1(2), 302–332.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3), 432–441.

Hafner, C. M., Herwartz, H., 2006. Volatility impulse responses for multivariate GARCH models: An exchange rate illustration. Journal of International Money and Finance 25(5), 719–740.

Hassler, U., Rodrigues, P. M. M., Rubia, A., 2016. Quantile regression for long memory testing: A case of realized volatility. Journal of Financial Econometrics 14(4), 693–724.

Knittel, C. R., Pindyck, R. S., 2016. The simple economics of commodity price speculation. American Economic Journal: Macroeconomics 8(2), 85–110.

Koop, G., Pesaran, M. H., Potter, S. M., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74(1), 119–147.

Kotz, S., Nadarajah, S., 2004. Multivariate t distributions and their applications. Cambridge University Press, Cambridge, UK.

Lanne, M., Nyberg, H., 2016. Generalized forecast error variance decomposition for linear and nonlinear multivariate models. Oxford Bulletin of Economics and Statistics 78(4), 595–603.

Lee, N., Choi, H., Kim, S.-H., 2016. Bayes shrinkage estimation for high-dimensional VAR models with scale mixture of normal distributions for noise. Computational Statistics and Data Analysis 101(C), 250–276.

Levin, A., Lin, C.-F., Chu, C.-S. J., 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. Journal of Econometrics 108(1), 1–24.

Liu, C., Rubin, D. B., 1995. ML estimation of the t distribution using EM and its extensions, ECM and ECME. Statistica Sinica 5, 19–39.

Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer, Heidelberg, Germany.

Martens, M., van Dick, D., 2007. Measuring volatility with the realized range. Journal of Econometrics 138(1), 181–207.

McAleer, M., Medeiros, M. C., 2008. Realized volatility: A review. Econometric Reviews 27(1-3), 10–45.

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. The Journal of Business 53(1), 61–65.

Pesaran, H. H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. Economics Letters 58(1), 17–29.

Rezitis, A. N., 2015. The relationship between agricultural commodity prices, crude oil prices and US dollar exchange rates: A panel VAR approach and causality analysis. International Review of Applied Economics 29(3), 403–434.

Rothman, A. J., Levina, E., Zhu, J., 2010. Sparse multivariate regression with covariance estimation. Journal of Computational and Graphical Statistics 19 (4), 947–962.

Serra, T., Zilberman, D., 2013. Biofuel-related price transmission literature: A review. Energy Economics 37, 141–151.

Shu, J., Zhang, J. E., 2006. Testing range estimators of historical volatility. The Journal of Futures Markets 26(3), 297–313.

Teterin, P., Brooks, R., Enders, W., 2016. Smooth volatility shifts and spillovers in U.S. crude oil and corn futures markets. Journal of Empirical Finance 38(A), 22–36.

Thomakos, D. D., Wang, T., 2003. Realized volatility in the futures markets. Journal of Empirical Finance 10(3), 321–353.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 58(1), 267–288.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68(1), 49–67.