SUPPLEMENT TO MODELING AND FORECASTING LARGE REALIZED COVARIANCE MATRICES AND PORTFOLIO CHOICE

Laurent A.F. Callot

VU University Amsterdam, CREATES,

and the Tinbergen Institute

E-mail: l.callot@vu.nl

Anders B. Kock[1]

Aarhus University and CREATES

E-mail: akock@creates.au.dk

Marcelo C. Medeiros

Pontifical Catholic University of Rio de Janeiro

E-mail: mcm@econ.puc-rio.br

Abstract: We consider modeling and forecasting large realized covariance matrices by penalized vector autoregressive models. We consider Lasso-type estimators to reduce the dimensionality and provide strong theoretical guarantees on the forecast capability of our procedure. We show that we can forecast realized covariance matrices almost as precisely as if we had known the true driving dynamics of these in advance. We next investigate the sources of these driving dynamics as well as the performance of the proposed models for forecasting the realized covariance matrices of the 30 Dow Jones stocks. We find that the dynamics are not stable as the data are aggregated from the daily to lower frequencies. Furthermore, we are able beat our benchmark by a wide margin. Finally, we investigate the economic value of our forecasts in a portfolio selection exercise and find that in certain cases an investor is willing to pay a considerable amount in order get access to our forecasts. **This supplement contains additional empirical results as well as the proofs for all theoretical results in the paper.**

Keywords: Realized covariance; shrinkage; Lasso; forecasting; portfolio allocation.

JEL: C22.

Acknowledgements: We would like to thank two anonymous referees and the editor Jonathan Wright for constructive comments which have greatly improved the paper. Parts of the research for this paper were done while the first and second authors were visiting the Pontifical Catholic University of Rio de Janeiro. Its hospitality is gratefully appreciated. MCM's research is partially supported by the CNPq/Brazil. Financial support by the Danish National Research Foundation (grant DNR78, CREATES) is gratefully acknowledged.

[1]Corresponding author: Department of Economics and Business Economics - CREATES, Fuglesangs Allé 4, building 2628, 313, 8210 Aarhus V, Denmark, Phone: +4587165498.

1. Introduction

This document contains two sections. The first one presents supplementary empirical results and the second one contains the mathematical proofs.

2. Supplementary results

This section contains supplementary empirical results mentioned in the main paper. The following tables and figures are included.

- Table sp-1: Number of stocks per industry category.
- Table sp-2: Variable selection results for the VAR(1) including the S&P 500, which should be interpreted as Table 1 in the main text.
- Table sp-3: Variable selection results for weekly aggregated data, which should be interpreted as Table 1 in the main text.
- Table sp-4: Forecast statistics for weekly aggregated data, which should be interpreted as Table 2 in the main text.
- Table sp-5: Variable selection results for monthly aggregated data, which should be interpreted as Table 1 in the main text.
- Table sp-6: Forecast statistics for monthly aggregated data, which should be interpreted as Table 2 in the main text.
- Figure sp-1: Selection frequency of lagged variances.
- Figure sp-2: Selection of S&P 500 variables.

The data used in the paper consists of 30 stocks from the Dow Jones index from 2006 to 2012 with a total of 1474 daily observations[2]. The daily realized covariances are constructed from 5 minutes returns by the method of Lunde et al. (2015). The stocks can be classified in 8 broad categories as shown in Table sp-1.

2.1. Additional results for daily data.

Figure sp-1 investigates the dynamics of the thirty variance equations in more detail at the daily level of aggregation. To be precise, it indicates in what fraction of the 30 equations the lagged variance of the the stock indicated on the y-axis is chosen as an explanatory variable at each point in time on the x-axis. The first thing that can be noticed is the relative stability of the selected variables over the 455 forecasts: the selection frequency of each variable varies very little over time. From Figure sp-1 it is also seen that the variance of IBM is the one which is most often chosen by the Lasso for explaining the dynamics of the realized variances. Actually, IBM is chosen in almost 80 percent of the equations on average. On the other hand, UTX (United Technologies

[2]We are grateful to Asger Lunde for kindly providing us with the data.

Corporation) is the stock whose variance is least important in explaining the realized variances. More precisely, lagged variances of IBM are chosen between seven and eight times as often as those of UTX.

2.2. **Variable selection with the S&P 500.** Table sp-2 reports variable selection results for a VAR(1) estimated by the Lasso with the 30 stocks of the Dow Jones and augmented with the variance of the S&P 500 as a common factor. The results from the model augmented with the S&P 500 (Table sp-2 are very similar to those without that variable (Table 1 in the paper), indicating that the variance of the S&P500 does not behave as a common factor in our models in the sense that it does not crowd out the other variables or is particularly frequently selected.

Figure sp-2 focuses on the significance of the variance of the S&P500 in explaining the dynamics of variances of the Dow Jones stocks. The left panel of Figure sp-2 shows in which of the 30 variance equations of the Dow Jones stocks the lagged variance of the S&P500 is selected. It appears that in most equations the lagged variance is never (or infrequently) selected while being (almost) always retained in 12 of them. The right panel of Figure sp-2 shows which lagged variance of the 30 Dow Jones stocks is selected when modeling the variance of the S&P500. This equation is found to be very sparse (in variance terms at least) and the stocks frequently retained appear not to correspond to those in the left panel.

2.3. **Variable selection and forecasts with weekly aggregates.** Weekly models are estimated using a rolling window of 263 observations resulting in 27 forecasts. The first weekly forecast is made for the first week of 2011.

Table sp-3 reports the average (across estimation windows) fraction of variables from a given category (in rows) selected in equations for stocks belonging to a given category (in columns) at the weekly frequency. As an example, 75% of the lagged variances from the Basic Materials sector are included when modeling variances of the Basic Materials sector while only 12% of the lagged variances of stocks belonging to the Consumer, Non-cyclical sector are included when modeling variances of the Basic Materials sector. The numbers are broken down into variance and covariance terms in order to see whether diagonal and off-diagonal terms have different driving dynamics. There are 30 variance equations and 435 covariance equations. The variance of the S&P 500 was not included.

As in the daily case, the selected models in Table sp-3 for the variance terms are quite sparse, and lags of variance terms are much more important than lags of covariance terms.

Table sp-4 contains the forecast errors at the weekly level of aggregation. It confirms the finding for the daily forecast that the variances are harder to forecast than the covariances. Furthermore, this seems to be the case whether we consider median or maximal absolute forecasts errors.

Focusing on the VAR(1) models, the plain Lasso seems to yield the most precise forecasts. It has a slightly lower median forecast error for the variances than the no-change forecasts and is much more precise when forecasting the covariances 26 periods ahead. This is also seen from the ℓ_2-forecast errors which are around 25 percent lower for the Lasso than for the no-change forecasts.

In contrast to the daily level of aggregation, there no longer seems to be any benefit from adding lags. Neither the forecasts of the variances nor the forecasts of the covariances improve for any of the estimation procedures when including five lags instead of one.

2.4. **Variable selection and forecasts with monthly aggregates.** Monthly models are estimated using a rolling window of 60 observations resulting in seven forecasts of which the first is made January 2011.

Table sp-5 reports the selection frequencies at the monthly level of aggregation. The diagonal pattern for the explanatory variables of the variances disappears almost entirely. The dynamics of the variance equations are now governed mainly by the financial sector and the energy sector, which we interpret as implying that these sectors are driving the long run volatility of the Dow Jones index.

Finally, it can be observed from the bottom panel of Table sp-5 that the basic materials sector is completely dominant in explaining the dynamics of the covariance terms. This is rather surprising in light of the fact that the basic materials sector did not play any particular role for the variance terms.

Table sp-6 shows forecasting results for the monthly level of aggregation. The results confirm that the variances are harder to forecast than the covariances. Among the VAR(1) models the adaLasso and the Lasso with log-matrix transformed data appears to deliver the most precise forecasts, though at these level of aggregation it seems quite difficult to beat the no-change forecasts with 6-month horizon being an exception.

3. Proofs

This section contains the proofs of all theorems and corollaries in the main paper. Section 3.2 contains the proof of Theorem 1.

To prove Theorem 2 (in Section 3.5) we first establish an oracle inequality for the estimation error of the Lasso which leaves the intercept unpenalized. The proof of this result is contained in section 3.4 and relies on several lemmas in Section 3.3 which contains most of the hard work. Section 3.5 contains the proofs of the remaining results in the main paper. We first introduce some notation that we shall encounter throughout the appendix.

3.1. **Notation.** We shall use the following notation throughout the appendix. Let $J_i = \{j : \beta^*_{i,j} \neq 0\} \subseteq \{1, ..., kp\}$ denote the set of non-zero β^*_i and $s_i = |J_i|$ its cardinality. $\bar{s} = \max\{s_1, ..., s_k\}$. We set $\tilde{J}_i = \{1\} \cup (J_i + 1) \subseteq \{1, ..., kp + 1\}$ where the addition is understood elementwise. $\tilde{J}_i$ will be used to index elements in the $(kp + 1) \times 1$ vector $\gamma^*_i = (\omega^*_i, \beta^{*\prime}_i)'$.

For any vector δ in $\mathbb{R}^n$ and a subset $R \subseteq \{1, ..., n\}$ we shall let δ_R denote the vector consisting of only of those elements of δ indexed by R. ι_n denotes the $n \times 1$ vector of ones.

Let $m_i = E(y_{t,i})$ be the mean of the ith variable and set $m_T = \max_{1 \leq i \leq k} |m_i|$. Furthermore, $X_t = (y'_{t-1}, ..., y'_{t-p})'$ corresponding to all elements of Z_t defined in the main paper except for the 1. $X = (X_T, ..., X_1)'$ is $T \times kp$ and $\Gamma = E\left(X_t - m\right)\left(X_t - m\right)'$ denotes its covariance matrix. F denotes the companion matrix corresponding to (1) in the main paper. Finally, $K_T = \frac{\eta_T^2(1+4\exp(9m_T^2))}{\eta_T^2 \wedge \frac{1}{6}}$ which is bounded when η_T^2 and m_T are bounded.

3.2. **Proof of Theorem 1.**

Proof of Theorem 1 in main paper. By the definition of σ^2_t and $\hat{\sigma}^2_t$ one has

$$\left|\hat{\sigma}^2_t - \sigma^2_t\right| = \left|w'(\hat{\Sigma}_t - \Sigma_t)w\right| \leq \left\|(\hat{\Sigma}_t - \Sigma_t)w\right\|_{\ell_\infty} \|w\|_{\ell_1}$$
$$\leq \left\|\hat{\Sigma}_t - \Sigma_t\right\|_\infty \|w\|^2_{\ell_1} \leq \left\|\hat{\Sigma}_t - \Sigma_t\right\|_\infty (1 + c)^2$$

□

3.3. **Auxilliary lemmas.** We start by stating a couple of preparatory lemmas. The first lemma bounds the probability of the maximum of all possible cross terms between explanatory variables and error terms becoming large. This bound will be used in the proof of Lemma 3 below. Throughout this appendix we shall impose

Assumption 1. *The data $\{y_t,\ t = 1 - p, ..., T\}$ is generated by (1) in the main paper where ϵ_t is assumed to be a sequence of i.i.d. error terms with an $N_k(0, \Omega)$ distribution. Furthermore, all roots of $|I_k - \sum_{j=1}^p \Phi_j z^j|$ are assumed to lie outside the unit disc.*

Lemma 1. *Let Assumption 1 be satisfied. Then, for any $L_T > 0$,*

$$P\left(\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}|y_{t-l,i}\epsilon_{t,j}|\geq L_T\right)\leq 2\exp\left(\frac{-L_T}{A\ln(1+T)\ln(1+k)^2\ln(1+p)K_T^2}\right)$$

for some positive constant A.

In order to prove Lemma 1 Orlicz norms turn out to be useful since random variables with bounded Orlicz norms obey useful maximal inequalities. Let ψ be a non-decreasing convex function with $\psi(0)=0$. Then, the Orlicz norm of a random variable X is given by

$$\|X\|_\psi=\inf\left\{C>0: E\psi\left(|X|/C\right)\leq 1\right\}$$

where, as usual, $\inf\emptyset=\infty$. By choosing $\psi(x)=x^p$ the Orlicz norm reduces to the usual L^p-norm since for $X\in L^p$, C equals $E(|X|^p)^{1/p}$. However, for our purpose $\psi(x)=e^x-1$. One has the following maximal inequality:

Lemma 2 (Lemma 2.2.2 from van Der Vaart and Wellner (1996))**.** *Let $\psi(x)$ be a convex, non-decreasing, non-zero function with $\psi(0)=0$ and $\limsup_{x,y\to\infty}\psi(x)\psi(y)/\psi(cxy)<\infty$ for some constant c. Then for any random variables, $X_1,...,X_m$,*

$$\left\|\max_{1\le i\le m}X_i\right\|_\psi\leq K\psi^{-1}(m)\max_{1\le i\le m}\|X_i\|_\psi$$

for a constant K depending only on ψ.

Notice that this result is particularly useful if $\psi^{-1}(x)$ only increases slowly which is the case when $\psi(x)$ increases very fast as in our case.

Proof of Lemma 1. Let $\psi(x)=e^x-1$. First we show that $\left\|\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}y_{t-l,i}\epsilon_{t,j}\right\|_\psi<\infty$. Repeated application of Lemma 2 yields

$$\begin{aligned}&\left\|\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}y_{t-l,i}\epsilon_{t,j}\right\|_\psi\\ &\leq K^4\ln(1+T)\ln(1+k)^2\ln(1+p)\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}\left\|y_{t-l,i}\epsilon_{t,j}\right\|_\psi \qquad (1)\end{aligned}$$

Next, we turn to bounding $\left\|y_{t-l,i}\epsilon_{t,j}\right\|_\psi$ uniformly in $1\leq i,j\leq k$, $1\leq l\leq p$ and $1\leq t\leq T$. First, note that as $y_{t-l,i}\sim N(m_i,\eta_{i,y}^2)$ one has by a Chernoff bound that for any $\lambda>0$

$$P\left(y_{t-l,i}>x\right)=P\left(\lambda y_{t-l,i}>\lambda x\right)\leq\exp\left(m_i\lambda+\frac{1}{2}\eta_{i,y}^2\lambda^2-\lambda x\right)$$

Assume first that $x \geq 3m_i$. Minimizing the right hans side with respect to λ yields $\lambda = \frac{x-m_i}{\eta_{i,y}^2} > 0$ and

$$P\left(y_{t-l,i} > x\right) \leq \exp\left(-\frac{1}{2}\frac{(x-m_i)^2}{\eta_{i,y}^2}\right) = \exp(-m_i^2/2\eta_{i,y}^2)\exp\left(-\frac{1}{2\eta_{i,y}^2}(x^2 - 2m_i x)\right) \leq \exp\left(-\frac{x^2}{6\eta_T^2}\right)$$
$$\leq \exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x^2\right)$$

where we also used $\exp(-m_i^2/2\eta_{i,y}^2) \leq 1$, $\eta_{i,y}^2 \leq \eta_T^2$ and the last estimate is merely to get simpler expressions later. If, on the other hand, $x < 3m_i$ we utilize the trivial bound

$$P\left(y_{t-l,i} > x\right) \leq 1 \leq \exp\left(9m_i^2\left[1 \wedge \frac{1}{6\eta_T^2}\right]\right)\exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x^2\right) \leq \exp\left(9m_T^2\right)\exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x^2\right).$$

Combining the above two displays using worst possible constants yields

$$P\left(y_{t-l,i} > x\right) \leq \exp\left(9m_T^2\right)\exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x^2\right)$$

for all $x > 0$. By symmetry, a similar argument holds for $-y_{t-l,i}$ and so

$$P\left(|y_{t-l,i}| > x\right) \leq 2\exp\left(9m_T^2\right)\exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x^2\right) \tag{2}$$

Furthermore, $\epsilon_{t,j}$ is gaussian with mean 0 and variance $\eta_{j,\epsilon}^2$. Thus, it follows by a standard estimate on gaussian tails (see e.g. Billingsley (1999), page 263) that for any $x > 0$

$$P\left(|y_{t-l,i}\epsilon_{t,j}| > x\right) \leq P\left(|y_{t-l,i}| > \sqrt{x}\right) + P\left(|\epsilon_{t,i}| > \sqrt{x}\right) \leq 2\exp\left(9m_T^2\right)\exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x\right) + 2e^{-x/2\eta_{j,\epsilon}^2}$$
$$\leq 4\exp\left(9m_T^2\right)\exp\left(-\left[1 \wedge \frac{1}{6\eta_T^2}\right]x\right).$$

Hence, $\left\{y_{t-l,i}\epsilon_{t,j}\right\}$ has subexponential tails[3] and it follows from Lemma 2.2.1 in van Der Vaart and Wellner (1996) that $\|y_{t-l,i}\epsilon_{t,j}\|_\psi \leq \frac{1+4\exp(9m_T^2)}{1\wedge\frac{1}{6\eta_T^2}} = \frac{\eta_T^2(1+4\exp(9m_T^2))}{\eta_T^2 \wedge \frac{1}{6}} =: K_T$ Note that K_T is bounded as long as m_T and η_T^2 are bounded from above. Using this in (1) yields

$$\left\|\max_{1\leq t\leq T}\max_{1\leq i\leq k}\max_{1\leq l\leq p}\max_{1\leq j\leq k} y_{t-l,i}\epsilon_{t,j}\right\|_\psi \leq K^4\ln(1+T)\ln(1+k)^2\ln(1+p)K_T$$
$$= A\ln(1+T)\ln(1+k)^2\ln(1+p)K_T := f(T)$$

where $A = K^4$. Finally, by Markov's inequality, the definition of the Orlicz norm, and the fact that $1 \wedge \psi(x)^{-1} = 1 \wedge (e^x - 1)^{-1} \leq 2e^{-x}$,

[3] A random variable X is said to have subexponential tails if there exists constants K and C such that for every $x > 0$, $P(|X| > x) \leq Ke^{-Cx}$.

$$
\begin{aligned}
&P\left(\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}|y_{t-l,i}\epsilon_{t,j}|\ge L_T\right)\\
&=P\left(\psi\left(\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}|y_{t-l,i}\epsilon_{t,j}|/f(T)\right)\ge\psi\left(L_T/f(T)\right)\right)\\
&\le 1\wedge\frac{E\psi\left(\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}|y_{t-l,i}\epsilon_{t,j}|/f(T)\right)}{\psi\left[L_T/f(T)\right]}\\
&\le 1\wedge\frac{1}{\psi\left[L_T/f(T)\right]}\\
&\le 2\exp\left(-L_T/f(T)\right)\\
&=2\exp(-L_T/[A\ln(1+T)\ln(1+k)^2\ln(1+p)K_T])
\end{aligned}
$$

□

Lemma 3. *Let Assumption 1 be satisfied and define*

$$
\mathcal{B}_T=\left\{\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}\left|\frac{1}{T}\sum_{t=1}^{T}y_{t-l,i}\epsilon_{t,j}\right|\vee\max_{1\le j\le k}\left|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{t,j}\right|<\frac{\lambda_T}{2}\right\} \tag{3}
$$

Then,

$$
P(\mathcal{B}_T)\ge 1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-A} \tag{4}
$$

for $\lambda_T=\sqrt{8\ln(1+T)^5\ln(1+k)^4\ln(1+p)^2\ln(k^2p)K_T^2/T}$ *and* A *a positive constant.*

In order to prove Lemma 3 we need the following result, the so-called *useful rule*, on conditional expectations adapted from Hoffmann-Jørgensen (1994) for our purpose.

Lemma 4 ((6.8.14) in Hoffmann-Jørgensen (1994))**.** *Let* $f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ *be measurable such that* $|f(U,V)|$ *is integrable and* $f(U,v)$ *is integrable for* P_V *almost all* $v\in\mathbb{R}$ *(here* P_V *denotes the distribution of* V*), and let* $\phi(v)=E(f(U,v))$*. If, for a sigma field* $\mathcal{G}$*,* V *is measurable with respect to* $\mathcal{G}$ *and* U *is independent of* $\mathcal{G}$*, then we have*

$$
E(f(U,V)|\mathcal{G})=\phi(V)\ P\text{-almost surely}
$$

Proof of Lemma 3. By subadditivity of the probability measure it follows that for any $L_T>0$,

$$
\begin{aligned}
&P\left(\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}\left|\frac{1}{T}\sum_{t=1}^{T} y_{t-l,i}\epsilon_{t,j}\right| \ge \frac{\lambda_T}{2}\right)\\
&= P\left(\bigcup_{i=1}^{k}\bigcup_{l=1}^{p}\bigcup_{j=1}^{k}\left\{\left|\frac{1}{T}\sum_{t=1}^{T} y_{t-l,i}\epsilon_{t,j}\right| \ge \frac{\lambda_T}{2}\right\}\right)\\
&\le P\left(\bigcup_{i=1}^{k}\bigcup_{l=1}^{p}\bigcup_{j=1}^{k}\left\{\left|\frac{1}{T}\sum_{t=1}^{T} y_{t-l,i}\epsilon_{t,j}\right| \ge \frac{\lambda_T}{2}\right\} \cap \bigcap_{t=1}^{T}\bigcap_{i=1}^{k}\bigcap_{l=1}^{p}\bigcap_{j=1}^{k}\left\{|y_{t-l,i}\epsilon_{t,j}| < L_T\right\}\right)\\
&+ P\left(\left\{\bigcap_{t=1}^{T}\bigcap_{i=1}^{k}\bigcap_{l=1}^{p}\bigcap_{j=1}^{k}\left\{|y_{t-l,i}\epsilon_{t,j}| < L_T\right\}\right\}^c\right)\\
&\le \sum_{i=1}^{k}\sum_{l=1}^{p}\sum_{j=1}^{k} P\left(\left|\frac{1}{T}\sum_{t=1}^{T} y_{t-l,i}\epsilon_{t,j}\right| \ge \frac{\lambda_T}{2}, \bigcap_{t=1}^{T}\left\{|y_{t-l,i}\epsilon_{t,j}| < L_T\right\}\right)\\
&+ P\left(\max_{1\le t\le T}\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}|y_{t-l,i}\epsilon_{t,j}| \ge L_T\right)
\end{aligned}
$$

Observe that for $\mathcal{F}_t = \sigma\left(\{\epsilon_s,\ s = 1, ..., t;\ y_s,\ s = 1-l, ..., t\}\right)$ one has that $\left\{y_{t-l,i}\epsilon_{t,j} I_{\{|y_{t-l,i}\epsilon_{t,j}|<L_T\}}, \mathcal{F}_t\right\}_{t=1}^{\infty}$ defines a martingale difference sequence for every $1 \le i, j \le k$ and $1 \le l \le p$ since

$$
E(y_{t-l,i}\epsilon_{t,j} I_{\{|y_{t-l,i}\epsilon_{t,j}|<L_T\}}|\mathcal{F}_{t-1}) = y_{t-l,i}E(\epsilon_{t,j} I_{\{|y_{t-l,i}\epsilon_{t,j}|<L_T\}}|\mathcal{F}_{t-1}) = 0
$$

where the second equality follows from Lemma 4 with $f(\epsilon_{t,j}, y_{t-l,i}) = \epsilon_{t,j} I_{\{|y_{t-l,i}\epsilon_{t,j}|\le L_T\}}$ such that for all $v \in \mathbb{R}$ [4]

$$
\phi(v) = E(f(\epsilon_{t,j}, v)) = E(\epsilon_{t,j} I_{\{|v\epsilon_{t,j}|<L_T\}}) = E(\epsilon_{t,j} I_{\{|\epsilon_{t,j}|<L_T/|v|\}}) = 0
$$

where the last equality follows from the gaussianity of the mean zero $\epsilon_{t,j}$.[5] Hence, $E(\epsilon_{t,j} I_{\{|y_{t-l,i}\epsilon_{t,j}|\le L_T\}}|\mathcal{F}_{t-1}) = \phi(y_{t-l,i}) = 0$ P-almost surely. Next, using the Azuma-Hoeffding inequality[6] on the first term and Lemma 1 on the second term with $L_T = \ln(1+T)^2\ln(1+k)^2\ln(1+p)K_T$ yields,

[4] The argument handles the case $v \neq 0$. The case $v = 0$ follows from omitting the second to last equality and using $E(\epsilon_{t,j}) = 0$. Note that the $y_s,\ s < t$ can be written as functions of $\epsilon_s,\ s < t$ implying that $\mathcal{F}_{t-1}$ is generated by $\epsilon_s, s < t$ only. By the independence of the $\epsilon_t,\ t \in \mathbb{Z}$ it follows that $\epsilon_{t,j}$ and $\mathcal{F}_{t-1}$ are independent.

[5] More precisely, symmetrically truncating a symmetric mean zero variable yields a new variable with mean zero.

[6] The Azuma-Hoeffding inequality is now applicable since we apply it on the set where the summands are bounded by L_T.

$$
\begin{aligned}
&P\left(\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}\left|\frac{1}{T}\sum_{t=1}^{T}y_{t-l,i}\epsilon_{t,j}\right|\ge\frac{\lambda_T}{2}\right)\\
&\le k^2p\cdot 2\exp\left(-\frac{T\lambda_T^2}{8L_T^2}\right)+2\exp\left(-\frac{\ln(1+T)}{A}\right)\\
&=2k^2p\cdot\exp\Big(-\ln(1+T)\ln(k^2p)\Big)+2(1+T)^{-1/A}\\
&=2(k^2p)^{1-\ln(1+T)}+2(1+T)^{-1/A}
\end{aligned}
$$

Furthermore, $\frac{1}{T}\sum_{t=1}^{T}\epsilon_{t,j}\sim N\left(0,\frac{\eta_{j,\epsilon}^2}{T}\right)$ for $j=1,...,k$ such that by standard tail probabilities for gaussian variables (see e.g. Billingsley (1999), page 263) one has

$$
\Big(\Big|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{t,j}\Big|\ge x\Big)\le 2\exp\left(-\frac{Tx^2}{2\eta_{j,\epsilon}^2}\right)\le 2\exp\left(-\frac{Tx^2}{2K_T}\right)
$$

where we used $\eta_{j,\epsilon}^2\le\eta_T^2\le K_T$ and $K_T=\frac{\eta_T^2(1+4\exp(9m_T^2))}{\eta_T^2\wedge\frac{1}{6}}$ as in Lemma 1. Therefore,

$$
\begin{aligned}
P\left(\max_{1\le j\le k}\Big|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{t,j}\Big|\ge\frac{\lambda_T}{2}\right)&\le 2k\exp\Big(-\ln(1+T)^5\ln(1+k)^4\ln(1+p)^2\ln(k^2p)K_T\Big)\\
&\le 2(k^2p)\exp\Big(-\ln(1+T)\ln(k^2p)\Big)=2(k^2p)^{1-\ln(1+T)}
\end{aligned}
$$

where we used $K_T\ge 1$ and $\ln(1+T)^4\ln(1+k)^4\ln(1+p)^2\ge 1$. Therefore, in total we have,

$$
P\left(\max_{1\le i\le k}\max_{1\le l\le p}\max_{1\le j\le k}\left|\frac{1}{T}\sum_{t=1}^{T}y_{t-l,i}\epsilon_{t,j}\right|\vee\max_{1\le j\le k}\Big|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{t,j}\Big|\ge\frac{\lambda_T}{2}\right)\le 4(k^2p)^{1-\ln(1+T)}+2(1+T)^{-1/A}.
$$

$\square$

Lemma 5. *Let* $\lambda_T=\sqrt{8\ln(1+T)^5\ln(1+k)^4\ln(1+p)^2\ln(k^2p)K_T^2/T}$. *Then, on a set with probability at least* $1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-1/A}$ *the following inequalities hold for all* $i=1,...,k$ *for some positive constant* A.

$$
\frac{1}{T}\left\|Z\hat{\gamma}_i-Z\gamma_i^*\right\|^2+\lambda_T\left\|\hat{\gamma}_i-\gamma_i^*\right\|_{\ell_1}\le 4\lambda_T\left\|\hat{\gamma}_{i,\tilde{J}_i}-\gamma_{i,\tilde{J}_i}^*\right\|_{\ell_1}\tag{5}
$$

$$
\left\|\hat{\gamma}_{i,\tilde{J}_i^c}-\gamma_{i,\tilde{J}_i^c}^*\right\|_{\ell_1}\le 3\left\|\hat{\gamma}_{i,\tilde{J}_i}-\gamma_{i,\tilde{J}_i}^*\right\|_{\ell_1}.\tag{6}
$$

Proof of Lemma 5. As the results are equation by equation we shall focus on equation i here but omit the subscript i for brevity. By the minimizing property of $\hat{\gamma}=(\hat{\omega},\hat{\beta}')'$ it follows that

$$
\frac{1}{T}\left\|y-Z\hat{\gamma}\right\|^2+2\lambda_T\left\|\hat{\beta}\right\|_{\ell_1}\le\frac{1}{T}\left\|y-Z\gamma^*\right\|^2+2\lambda_T\left\|\beta^*\right\|_{\ell_1}
$$

which using that $y = Z\gamma^* + \epsilon$ yields

$$\frac{1}{T}\|\epsilon\|^2 + \frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2 - \frac{2}{T}\epsilon' Z(\hat{\gamma}-\gamma^*) + 2\lambda_T\left\|\hat{\beta}\right\|_{\ell_1} \leq \frac{1}{T}\|\epsilon\|^2 + 2\lambda_T\left\|\beta^*\right\|_{\ell_1}$$

Or, equivalently

$$\frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2 \leq \frac{2}{T}\epsilon' Z(\hat{\gamma}-\gamma^*) + 2\lambda_T\left(\left\|\beta^*\right\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1}\right). \tag{7}$$

To bound $\frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2$ one must bound $\frac{2}{T}\epsilon' Z(\hat{\gamma}-\gamma^*)$. Note that on the set $\mathcal{B}_T$ defined in (3) one has

$$\frac{2}{T}\epsilon' Z(\hat{\gamma}-\gamma^*) \leq 2\left\|\frac{1}{T}\epsilon' Z\right\|_{\ell_\infty}\|\hat{\gamma}-\gamma^*\|_{\ell_1} \leq \lambda_T\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1}.$$

Putting things together, on $\mathcal{B}_T$,

$$\frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2 \leq \lambda_T\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1} + 2\lambda_T\left(\left\|\beta^*\right\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1}\right)$$

Adding $\lambda_T\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1}$ yields

$$\begin{aligned}\frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2 + \lambda_T\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1} &\leq 2\lambda_T\left(\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1} + \left\|\beta^*\right\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1}\right)\\ &= 2\lambda_T\left(|\hat{\omega}-\omega^*| + \left\|\hat{\beta}-\beta^*\right\|_{\ell_1} + \left\|\beta^*\right\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1}\right)\end{aligned} \tag{8}$$

Furthermore,

$$\left\|\hat{\beta}-\beta^*\right\|_{\ell_1} + \left\|\beta^*\right\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1} = \left\|\hat{\beta}_J-\beta^*_J\right\|_{\ell_1} + \left\|\beta^*_J\right\|_{\ell_1} - \left\|\hat{\beta}_J\right\|_{\ell_1} \leq 2\left\|\hat{\beta}_J-\beta^*_J\right\|_{\ell_1}$$

Using this in (8) gives

$$\frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2 + \lambda_T\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1} \leq 2\lambda_T\left(|\hat{\omega}-\omega^*| + 2\left\|\hat{\beta}_J-\beta^*_J\right\|_{\ell_1}\right) \leq 4\lambda_T\left(\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\|_{\ell_1}\right)$$

where $\tilde{J} = (1, 1+J)$ and set addition is understood elementwise. This is (5). Next,

$$\lambda_T\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1} \leq 4\lambda_T\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\|_{\ell_1}$$

which is equivalent to

$$\left\|\hat{\gamma}_{\tilde{J}^c}-\gamma^*_{\tilde{J}^c}\right\|_{\ell_1} \leq 3\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\|_{\ell_1}$$

and establishes (6). The lower bound on the probability with which (5)-(6) hold follows from the fact that $P(\mathcal{B}_T) \geq 1 - 4(k^2p)^{1-\ln(1+T)} - 2(1+T)^{-1/A}$ by Lemma 3. □

Next, we introduce the so-called restricted eigenvalue condition. For any $n\times n$ matrix A define

$$\kappa^2=\kappa_A^2(r)=\min\left\{\frac{\delta'\Gamma\delta}{\|\delta_R\|^2}:|R|\leq r,\ \delta\in\mathbb{R}^n\setminus\{0\},\|\delta_{R^c}\|_{\ell_1}\leq 3\|\delta_R\|_{\ell_1}\right\} \tag{9}$$

We note in particular that the restricted eigenvalue of A may be positive even when A is singular. In the sequel we shall be interested in $\kappa^2_{\Psi_T}(s_i)$. We will see that this is positive as long as $\kappa^2_{\Psi}(s_i)$ is positive and Ψ_T is sufficiently close to Ψ. Observe that $\kappa^2_{\Psi}(s_i)$ is positive in particular when Ψ has full rank, a rather innocent assumption. For more details on the restricted eigenvalue condition we refer to Kock and Callot (2015).

The following lemma shows that in order to verify the restricted eigenvalue condition for a matrix it suffices that this matrix is close (in terms of maximum entrywise distance) to a matrix which does satisfy the restricted eigenvalue condition.

Lemma 6. *Let A and B be two positive semi-definite $n\times n$ matrices and assume that A satisfies the restricted eigenvalue condition RE(r) for some κ_A. Then, for $\delta=\max_{1\leq i,j\leq n}|A_{i,j}-B_{i,j}|$, one also has $\kappa_B^2\geq\kappa_A^2-16r\delta$.*

Proof. The proof is similar to Lemma 10.1 in van De Geer et al. (2009). For any (non-zero) $n\times 1$ vector v and $R\subseteq\{1,...,n\},|R|=r$ such that $\|v_{R^c}\|_{\ell_1}\leq 3\|v_R\|_{\ell_1}$ one has

$$\begin{aligned}v'Av-v'Bv&\leq|v'Av-v'Bv|=|v'(A-B)v|\leq\|v\|_{\ell_1}\|(A-B)v\|_{\ell_\infty}\leq\delta\|v\|_{\ell_1}^2\\&\leq\delta 16\|v_r\|_{\ell_1}^2\leq\delta 16r\|v_R\|^2\end{aligned}$$

Hence, rearranging the above, yields

$$v'Bv\geq v'Av-16r\delta\|v_R\|^2$$

or equivalently,

$$\frac{v'Bv}{v_R'v_R}\geq\frac{v'Av}{v_R'v_R}-16r\delta\geq\kappa_A^2-16r\delta$$

Minimizing the left hand side over $\{v\in\mathbb{R}^n\setminus\{0\}:\|v_{R^c}\|_{\ell_1}\leq 3\|v_R\|_{\ell_1}\}$ yields the claim. □

Lemma 7. *Let V be an $n\times 1$ vector with $V\sim N(0,Q)$. Then, for any $\epsilon,M>0$, $P\left(\left|\|V\|^2-E\|V\|^2\right|>\epsilon\right)\leq 2\exp\left(\frac{-\epsilon^2}{8n\|Q\|^2_{\ell_\infty}M^2}\right)+2n\exp\left(-M^2/2\right)$.*

Proof. The statement of the lemma only depends on the distribution of V and so we may equivalently consider $\sqrt{Q}\tilde{V}$ with $\tilde{V}\sim N(0,I)$ where $\sqrt{Q}$ is the matrix square root of Q. Hence,

$$\begin{aligned}&P\left(\left|\|\sqrt{Q}\tilde{V}\|^2-E\|\sqrt{Q}\tilde{V}\|^2\right|>\epsilon\right)\\&\leq P\left(\left|\|\sqrt{Q}\tilde{V}\|^2-E\|\sqrt{Q}\tilde{V}\|^2\right|>\epsilon,\ \|\tilde{V}\|_{\ell_\infty}\leq M\right)+P(\|\tilde{V}\|_{\ell_\infty}>M)\end{aligned} \tag{10}$$

To get an estimate on the first probability we show that on the set $\{\|x\|_{\ell_\infty} \leq M\}$ the function $f(x) = \|\sqrt{Q}x\|^2 = \|\sqrt{Q}(-M \vee x \wedge M)\|^2$ (the minimum and maximum are understood entrywise in the vector x) is Lipschitz continuous. Note that with $g(x) = (-M \vee x \wedge M)$ we can write $f(x) = f(g(x))$ on $\{\|x\|_{\ell_\infty} \leq M\}$. To obtain a bound on the Lipschitz constant note that by the mean value theorem, on $\{\|x\|_{\ell_\infty} \leq M\}$,

$$\begin{aligned}\left|f(x) - f(y)\right| &= \left|f(g(x)) - f(g(y))\right| = \left|f'(c)'(g(x) - g(y))\right| \\ &\leq \left\|f'(c)\right\|_{\ell_\infty} \left\|g(x) - g(y)\right\|_{\ell_1} \leq \left\|f'(c)\right\|_{\ell_\infty} \|x - y\|_{\ell_1} \leq \left\|f'(c)\right\|_{\ell_\infty} \sqrt{n}\|x - y\|\end{aligned}$$

for a point c on the line segment joining $g(x)$ and $g(y)$. Since $c = \mu g(x) + (1-\mu)g(y)$ for some $0 < \mu < 1$ one has $\|c\|_{\ell_\infty} \leq \mu\left\|g(x)\right\|_{\ell_\infty} + (1-\mu)\left\|g(y)\right\|_{\ell_\infty} \leq M$ and so

$$\left\|f'(c)\right\|_{\ell_\infty} \sqrt{n} = \|2Qc\|_{\ell_\infty} \sqrt{n} \leq 2\sqrt{n}\|Q\|_{\ell_\infty} \|c\|_{\ell_\infty} \leq 2\sqrt{n}\|Q\|_{\ell_\infty} M$$

Hence, $f(x)$ is Lipschitz with Lipschitz constant bounded by $2\sqrt{n}\|Q\|_{\ell_\infty} M$. The Borell-Cirelson-Sudakov inequality (see e.g. Massart (2007), Theorem 3.4) then yields that the first probability in (10) can be be bounded by $2\exp\left(\frac{-\epsilon^2}{2(2\sqrt{n}\|Q\|_{\ell_\infty} M)^2}\right)$. Regarding the second probability in (10) note that by the union bound and standard tail probabilities for gaussian variables (see e.g. Billingsley (1999), page 263) one has $P(\|\tilde{V}\|_{\ell_\infty} > M) \leq 2ne^{-M^2/2}$. This yields the lemma. $\square$

Lemma 8. *Let Assumption 1 be satisfied. Then, for any $t, M > 0$, one has*

$$\begin{aligned}&P\left(\max_{1\leq i,j\leq kp+1} |\Psi_{T,i,j} - \Psi_{i,j}| > t\right) \\ &\leq 2(kp+1)^2 \left[2\exp\left(\frac{-t^2T}{8\|Q\|_{\ell_\infty}^2 M^2}\right) + 2T\exp\left(-M^2/2\right) + \exp\left(-\frac{-t^2T}{64\,\|Q\|_{\ell_\infty}(1+4m_T^2)}\right)\right]\end{aligned}$$

where $\|Q\|_{\ell_\infty} \leq 2\,\|\Gamma\|\sum_{i=0}^{T}\|F^i\|$.

Proof. For any $t > 0$, it follows from a union bound

$$P\left(\max_{1\leq i,j\leq kp+1} |\Psi_{T,i,j} - \Psi_{i,j}| > t\right) \leq (kp+1)^2 \max_{1\leq i,j\leq kp+1} P\left(|\Psi_{T,i,j} - \Psi_{i,j}| > t\right) \tag{11}$$

Hence, it suffices to bound $P\left(|\Psi_{T,i,j} - \Psi_{i,j}| > t\right)$ appropriately for all $1 \leq i,j \leq kp+1$. To this end note that by the stationarity of y_t

$$P\left(|\Psi_{T,i,j} - \Psi_{i,j}| > t\right) = P\left(|(Z'Z)_{i,j} - E(Z'Z)_{i,j}| > tT\right)$$

This further implies that it is enough to bound $P\left(|v'Z'Zv - E(v'Z'Zv)| > tT\right)$ for any $t > 0$ and $(kp+1)\times 1$ vector $v = (v_1, v_2')'$ with v_1 a scalar and $\|v\| = 1$ [7]. But for $U = Zv$ this probability equals

[7]Letting v run over the standard basis vectors of $\mathbb{R}^{kp+1}$ yields the result for all diagonal elements. Choosing v to contain only zeros except for $1/\sqrt{2}$ in the ith and jth position and thereafter only zeros except for $1/\sqrt{2}$ in

$P\left(\left|\|U\|^2 - E\|U\|^2\right| > tT\right)$. U is a linear transformation of a multivariate gaussian and hence itself gaussian. Now define $\mu = E(U) = E(Z)v = (\iota_T \; M)v$ where all T rows of M equal the $1 \times kp$ vector $(\iota_p' \otimes m')$ (recall $m = Ey_t$). Therefore, defining $\tilde{U} = U - \mu$ we get

$$\|U\|^2 - E\|U\|^2 = \|\tilde{U} + \mu\|^2 - E\|\tilde{U} + \mu\|^2 = \tilde{U}'\tilde{U} - E\tilde{U}'\tilde{U} + 2\mu'\tilde{U} \leq \left|\|\tilde{U}\| - E\|\tilde{U}\|\right| + 2\left|\mu'\tilde{U}\right| \tag{12}$$

Thus, it suffices to bound $P\left(\left|\|\tilde{U}\| - E\|\tilde{U}\|\right| \geq \frac{tT}{2}\right)$ and $P\left(2\left|\mu'\tilde{U}\right| \geq \frac{tT}{2}\right)$ with high probability. First, note that as $\tilde{U}$ is a mean zero gaussian with covariance matrix

$$Q = E(\tilde{U}\tilde{U}') = E\left(Z - (\iota_T, M)\right)v\left[\left(Z - (\iota_T, \; M)\right)v\right]' = E(X - M)v_2\left[(X - M)v_2\right]' \tag{13}$$

where we used that the first column of Z is ι_T. Hence, it follows from Lemma 7 that for any $M > 0$

$$P\left(\left|\|\tilde{U}\| - E\|\tilde{U}\|\right| \geq \frac{tT}{2}\right) \leq 2\left(2\exp\left(\frac{-t^2T}{32\|Q\|_{\ell_\infty}^2 M^2}\right) + 2T\exp\left(-M^2/2\right)\right) \tag{14}$$

It remains to upper bound $\|Q\|_{\ell_\infty} = \max_{1\leq t\leq T}\sum_{s=1}^T |Q_{t,s}|$. Recall that $X_t' = (y_{t-1}', ..., y_{t-p}')$ is the tth row of X. For any pair of $1 \leq s, t \leq T$ letting $\Gamma_{t-s} = E\left(X_t - m\right)\left(X_s - m\right)'$ (clearly $\Gamma_0 = \Gamma$) and writing y_t in companion form (as an VAR(1)) with companion matrix F in deviations from the mean we get using (13):

$$\begin{aligned}
|Q_{t,s}| &= \left|E\left[\left(X_t - m\right)' v_2 v_2'\left(X_s - m\right)\right]\right| = \left|E\left[v_2'\left(X_s - m\right)\left(X_t - m\right)' v_2\right]\right| \\
&= \left|v'\Gamma_{s-t}v\right| = \begin{cases} |v_2'F^{s-t}\Gamma v_2| & \text{for } s \geq t \\ |v_2'\Gamma(F')^{t-s}v_2| = v_2'\left((F')^{t-s}\right)'\Gamma v_2 = |v_2'F^{t-s}\Gamma v_2| & \text{for } s < t \end{cases}
\end{aligned}$$

where the first inequality in the second case used that a scalar may be transposed without changing its value and Γ is symmetric. Hence, $|Q_{t,s}| = |v'F^{|t-s|}\Gamma v|$ for any pair of $1 \leq s, t \leq T$. By the Cauchy-Schwarz inequality

$$|v'F^{|t-s|}\Gamma v| \leq \|v'F^{|t-s|}\| \, \|\Gamma v\| \leq \|F^{|t-s|}\| \, \|\Gamma\|$$

Putting things together yields (uniformly over $\{v : \|v\| = 1\}$)

$$\|Q\|_{\ell_\infty} = \max_{1\leq t\leq T}\sum_{s=1}^T \|F^{|t-s|}\| \, \|\Gamma\| \leq 2\|\Gamma\|\sum_{i=0}^T \|F^i\|.$$

---

the ith position and $-1/\sqrt{2}$ in the jth position some elementary calculations (available upon request) show that $P\left(|(Z'Z)_{i,j} - E(Z'Z)_{i,j}| > tT\right)$ for $i \neq j$ is bounded by two times $P\left(|(Z'Z)_{i,i} - E(Z'Z)_{i,i}| > tT\right)$. The maximum on the right hand side in (11) is obtained for the off-diagonal elements. This explains the first 2 on the right hand side in (14).

Next, consider the second term, $2\left|\mu'\tilde{U}\right|$, on the right hand side in (12). Note that $2\mu'\tilde{U} \sim N(0, 4\mu' Q\mu)$ such that

$$P\left(\left|2\mu'\tilde{U}\right| \geq \frac{tT}{2}\right) \leq 2\exp\left(-\frac{t^2T^2}{32\mu' Q\mu}\right).$$

It remains to bound $\mu' Q\mu$ from above. To this end, using the definition of μ, observe that

$$\mu' Q\mu \leq \phi_{\max}\left(Q\right)\mu'\mu = \phi_{\max}\left(Q\right)\left(\iota_T v_1 + M v_2\right)'\left(\iota_T v_1 + M v_2\right) = \phi_{\max}\left(Q\right)\left(T v_1^2 + v_2' M' M v_2 + 2 v_1 \iota_T' M v_2\right)$$

Recall from footnote 7 that it suffices to consider $v = (v_1, v_2')' \in \mathbb{R}^{1+kp}$ with two nonzero entries, $\|v\| = 1$ and and v_1 a scalar. Thus, by the Cauchy-Schwarz inequality,

$$v_2' M' M v_2 \leq \left\|M v_2\right\|^2 \leq 4T \max_{1\leq i\leq k} m_i^2$$

where the last inequality used that all T entries of Mv_2 are bounded from above by $2\max_{1\leq i\leq k}|m_i|$ implying $\left\|Mv_2\right\| \leq 2\sqrt{T}\max_{1\leq i\leq k}|m_i|$. By similar arguments

$$2v_1\iota_T' M v_2 \leq 2\left\|v_1\iota_T\right\|\left\|Mv_2\right\| \leq 2\sqrt{T}|v_1| \cdot 2\sqrt{T}\max_{1\leq i\leq k}|m_i| \leq T v_1^2 + 4T\max_{1\leq i\leq k} m_i^2$$

using $2xy \leq x^2 + y^2$ for all $x, y \in \mathbb{R}$. In total, one has

$$\mu' Q\mu \leq \phi_{\max}(Q)\left(T v_1^2 + 4T\max_{1\leq i\leq k} m_i^2 + T v_1^2 + 4T\max_{1\leq i\leq k} m_i^2\right) \leq 2T\left\|Q\right\|_{\ell_\infty}\left(1 + 4\max_{1\leq i\leq k} m_i^2\right)$$

where we used $|v_1| \leq 1$ and $\phi_{\max}(Q) \leq \|Q\|_{\ell_\infty}$ and the latter has been bounded above. Thus,

$$P\left(2\left|\mu'\tilde{U}\right| \geq \frac{tT}{2}\right) \leq 2\exp\left(-\frac{-t^2T}{64\left\|Q\right\|_{\ell_\infty}\left(1 + 4\max_{1\leq i\leq k} m_i^2\right)}\right).$$

In total,

$$\begin{aligned}
&P\left(\max_{1\leq i,j\leq kp+1}\left|\Psi_{T,i,j} - \Psi_{i,j}\right| > t\right) \\
\leq& 2(kp+1)^2\left[2\exp\left(\frac{-t^2T}{8\|Q\|_{\ell_\infty}^2 M^2}\right) + 2T\exp\left(-M^2/2\right) + \exp\left(-\frac{-t^2T}{64\left\|Q\right\|_{\ell_\infty}\left(1 + 4m_T^2\right)}\right)\right].
\end{aligned}$$

as claimed since we have defined $m_T = \max_{1\leq i\leq k}|m_i|$. □

Lemma 9. *Let Assumption 1 be satisfied. Then, on*

$$\mathcal{C}_T = \left\{\max_{1\leq i,j\leq kp+1}\left|\Psi_{T,i,j} - \Psi_{i,j}\right| \leq \frac{(1-q)\kappa_\Psi^2(s+1)}{16(s+1)}\right\} \tag{15}$$

one has for any $0<q<1$ and $s\in\{1,...,kp\}$ that $\kappa^2_{\Psi_T}(s+1)\geq q\kappa^2_{\Psi}(s+1)$. Furthermore,

$$P(\mathcal{C}_T)\geq 1-4(kp+1)^2\exp\left(\frac{-T}{16^3(s+1)^2\|Q\|^2_{\ell_\infty}\log(T)\left[\log\left((kp+1)^2\right)+1\right]}\right)-4(kp+1)^{2(1-\log(T))}$$
$$-2(kp+1)^2\exp\left(-\frac{-(1-q)^2\kappa^4_{\Psi}T}{4\cdot 16^4(s+1)^2\left\|Q\right\|_{\ell_\infty}\left(1+4m_T^2\right)}\right)=1-\pi_q(s)$$

and $\|Q\|_{\ell_\infty}\leq\|\Gamma\|\sum_{i=0}^T\|F^i\|$.

Proof. Define $\kappa_\Psi(s+1)=\kappa_\Psi$ and $\kappa^2_{\Psi_T}(s+1)=\kappa^2_{\Psi_T}$. By Lemma 6 one has that $\kappa^2_{\Psi_T}\geq q\kappa^2_\Psi$ if $\max_{1\leq i,j\leq T}|\Psi_{T,i,j}-\Psi_{i,j}|\leq\frac{(1-q)\kappa^2_\Psi}{16(s+1)}$ (use $R=\tilde{J}$ and $r=s$ in that lemma). It remains to lower bound the measure of $\mathcal{C}_T$. Using $M^2=2\log\left((kp+1)^2\right)\log(T)+2\log(T)$ in Lemma 8 yields

$$P\left(\max_{1\leq i,j\leq kp+1}|\Psi_{T,i,j}-\Psi_{i,j}|>\frac{(1-q)\kappa^2_\Psi}{16(s+1)}\right)$$
$$\leq 4(kp+1)^2\exp\left(\frac{-(1-q)^2\kappa^4_\Psi T}{16^2(s+1)^2 8\|Q\|^2_{\ell_\infty}\left[2\log\left((kp+1)^2\right)\log(T)+2\log(T)\right]}\right)+4(kp+1)^{2(1-\log(T))}$$
$$+2(kp+1)^2\exp\left(-\frac{-(1-q)^2\kappa^4_\Psi T}{16^2(s+1)^2 64\left\|Q\right\|_{\ell_\infty}\left(1+4m_T^2\right)}\right)$$
$$=4(kp+1)^2\exp\left(\frac{-T}{16^3(s+1)^2\|Q\|^2_{\ell_\infty}\log(T)\left[\log\left((kp+1)^2\right)+1\right]}\right)+4(kp+1)^{2(1-\log(T))}$$

(16)
$$+2(kp+1)^2\exp\left(-\frac{-(1-q)^2\kappa^4_\Psi T}{4\cdot 16^4(s+1)^2\left\|Q\right\|_{\ell_\infty}\left(1+4m_T^2\right)}\right)=:\pi_q(s)$$

□

3.4. **Oracle inequality.** We are now ready to state the oracle inequalities which will be used as inputs to establishing an upper bound on the forecast error of the portfolio variance.

Theorem 1. *Set $\lambda_T=\sqrt{8\ln(1+T)^5\ln(1+k)^4\ln(1+p)^2\ln(k^2p)K_T^2/T}$ with $K_T=\frac{\eta_T^2(1+4\exp(9m_T^2))}{\eta_T^2\wedge\frac{1}{6}}$ and $0<q<1$. Then, with probability at least $1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-1/A}-\pi_q(s_i)$ [8], the following inequalities hold for all $i=1,...,k$ for some positive constant A.*

$$\frac{1}{T}\left\|Z\hat{\gamma}_i-Z\gamma_i^*\right\|^2\leq\frac{16}{q\kappa_i^2}(s_i+1)\lambda_T^2 \tag{17}$$

$$\left\|\hat{\gamma}_i-\gamma_i^*\right\|_{\ell_1}\leq\frac{16}{q\kappa_i^2}(s_i+1)\lambda_T. \tag{18}$$

The above inequalities hold on one and the same set which has probability at least $1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-1/A}-\pi_q(\bar{s})$.

[8]Recall that $\pi(s)=4(kp+1)^2\exp\left(\frac{-T}{16^3(s+1)^2\|Q\|^2_{\ell_\infty}\log(T)\left[\log\left((kp+1)^2\right)+1\right]}\right)+4(kp+1)^{2(1-\log(T))}$ $+2(kp+1)^2\exp\left(-\frac{-(1-q)^2\kappa^4_\Psi T}{4\cdot 16^4(s+1)^2\|Q\|_{\ell_\infty}\left(1+4m_T^2\right)}\right)$ which tends to zero fast under general conditions.

Proof. As the results are equation by equation we shall focus on equation i here but omit the subscript i for brevity when no confusion arises. We will work on $\mathcal{B}_T \cap \mathcal{C}_T$ as defined in (3) and (15). By (5), Jensen's inequality and the restricted eigenvalue condition (which is applicable due to (6))

$$\frac{1}{T}\left\|Z(\hat{\gamma}-\gamma^*)\right\|^2 \leq 4\lambda_T\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\|_{\ell_1} \leq 4\lambda_T\sqrt{s+1}\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\| \leq 4\lambda_T\sqrt{s+1}\frac{\left\|Z(\hat{\gamma}-\gamma^*)\right\|}{\kappa_{\Psi_T}\sqrt{T}}$$

Rearranging and using $\kappa^2_{\Psi_T} \geq q\kappa^2$, yields (17). To establish (18) use (5), Jensen's inequality, $\kappa^2_{\Psi_T} \geq q\kappa^2$, and (17):

$$\left\|\hat{\gamma}-\gamma^*\right\|_{\ell_1} \leq 4\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\|_{\ell_1} \leq 4\sqrt{s+1}\left\|\hat{\gamma}_{\tilde{J}}-\gamma^*_{\tilde{J}}\right\| \leq 4\sqrt{s+1}\frac{\left\|Z(\hat{\gamma}-\gamma^*)\right\|}{\kappa_{\Psi_T}\sqrt{T}} \leq \frac{16}{q\kappa^2}(s+1)\lambda_T$$

Combining Lemmas 3 and 9, $\mathcal{B}_T \cap \mathcal{C}_T$ is seen to have at least the stated probability. Regarding the last assertion,

(19)
$$\mathcal{C}_T(\bar{s}) := \left\{\max_{1\leq i,j\leq kp+1}\left|\Psi_{T,i,j}-\Psi_{i,j}\right| \leq \frac{(1-q)\kappa^2_{\Psi}(\bar{s}+1)}{16(\bar{s}+1)}\right\} \subseteq \left\{\max_{1\leq i,j\leq kp+1}\left|\Psi_{T,i,j}-\Psi_{i,j}\right| \leq \frac{(1-q)\kappa^2_{\Psi}(s_i+1)}{16(s_i+1)}\right\}$$

for all $i = 1, ..., k$. On $\mathcal{C}_T(\bar{s})$ it follows from Lemma 9 that $\kappa^2_{\Psi_T}(s_i) \geq q\kappa^2_{\Psi}(s_i)$ for all $i = 1, ..., k$ which is exactly what we used in the arguments above. Hence, (17) and (18) are a valid for all $i = 1, ..., k$ on $\mathcal{B}_T \cap \mathcal{C}_T(\bar{s})$ which has probability at least $1 - 4(k^2p)^{1-\ln(1+T)} - 2(1+T)^{-1/A} - \pi_q(\bar{s})$ by Lemmas 3 and 9. $\square$

3.5. **Remaining proofs of results in main paper.**

Proof of Theorem 2 in the main paper. Since

$$\left\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\right\|_\infty = \left\|\operatorname{vech}\hat{\Sigma}_{T+1}-\operatorname{vech}\Sigma_{T+1}\right\|_\infty = \left\|\hat{y}_{T+1}-y_{T+1}\right\|_{\ell_\infty}$$

we shall bound each entry of $\hat{y}_{T+1}-y_{T+1}$. By assumption

$$y_{T+1,i} = Z'_{T+1}\gamma^*_i + \epsilon_{T+1,i}$$

while

$$\hat{y}_{T+1,i} = Z'_{T+1}\hat{\gamma}_i$$

such that

$$\left|y_{T+1,i}-\hat{y}_{T+1,i}\right| = \left|Z'_{T+1}(\gamma^*_i-\hat{\gamma}_i)+\epsilon_{T+1,i}\right| \leq \left\|Z_{T+1}\right\|_{\ell_\infty}\left\|\hat{\gamma}_i-\gamma^*_i\right\|_{\ell_1} + \left|\epsilon_{T+1,i}\right|$$

Using Theorem 1 above this yields that

$$\left|y_{T+1,i}-\hat{y}_{T+1,i}\right| \le \|Z_{T+1}\|_{\ell_\infty} \frac{16}{q\kappa_i^2} s_i\lambda_T + \left|\epsilon_{T+1,i}\right| \text{ for all } i=1,...,k$$

with probability at least $1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-1/A}-\pi_q(\bar{s})$. Next, by (2)

$$P\left(|y_{t-l,i}|>x\right) \le 2\exp\left(9m_T^2\right)\exp\left(-\left[1\wedge\frac{1}{6\eta_T^2}\right]x^2\right)$$

for all $1\le i\le k$ and $1\le l\le p$ and $P(|\epsilon_{T+1,i}|\ge x)\le 2e^{-x^2/2\eta_T^2}$ for all $1\le i\le k$. This implies[9]

$$\begin{aligned}P\left(\|Z_{T+1}\|_{\ell_\infty}\vee\max_{1\le i\le k}|\epsilon_{T+1,i}|\ge L\right) &\le 2(kp+1)\exp\left(9m_T^2\right)\exp\left(-\left[1\wedge\frac{1}{6\eta_T^2}\right]L^2\right)+2ke^{-L^2/2\eta_T^2}\\ &= 4(kp+1)\exp\left(9m_T^2\right)\exp\left(-\left[1\wedge\frac{1}{6\eta_T^2}\right]L^2\right)\end{aligned}$$

for any $L\ge 1$[10] Choosing $L^2=\tilde{K}_T=\frac{9m_T^2+\ln(kp+1)\ln(T)}{1\wedge\frac{1}{6\eta_T^2}}=\frac{\eta_T^2(9m_T^2+\ln(kp+1)\ln(T))}{\eta_T^2\wedge\frac{1}{6}}$ yields

$$P\left(\|Z_{T+1}\|_{\ell_\infty}\vee\max_{1\le i\le k}|\epsilon_{T+1,i}|\ge L\right)\le 4[kp+1]^{1-\ln(T)} \tag{20}$$

and so

$$\left|y_{T+1,i}-\hat{y}_{T+1,i}\right| \le \sqrt{\tilde{K}_T}\left(\frac{16}{q\kappa_i^2}s_i\lambda_T+1\right)\le\sqrt{\tilde{K}_T}\left(\frac{16}{q\kappa^2}\bar{s}\lambda_T+1\right) \text{ for all } i=1,\dots,k$$

with probability at least $1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-1/A}-\pi_q(\bar{s})-4[kp+1]^{1-\ln(T)}$. □

Proof of Lemma 1 in the main paper. When the Φ_i^* $i=1,...,p$ are known $\left\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\right\|_\infty=\|\epsilon_{T+1}\|_{\ell_\infty}$. Next, for all $1\le i\le k$ one has[11] $P\left(|\epsilon_{T+1,i}|>L\right)\le e^{-L^2/\eta_{T,\epsilon}^2}$. Hence,

$$P\left(\max_{1\le i\le k}|\epsilon_{T+1,i}|>L\right)\le ke^{-L^2/\eta_{T,\epsilon}^2}$$

such that choosing $L^2=\eta_{T,\epsilon}^2\log(k)\ln(T)$ yields

$$P\left(\max_{1\le i\le k}|\epsilon_{T+1,i}|\ge L\right)\le k^{1-\ln(T)}$$

Therefore, with probability at least $1-k^{1-\ln(T)}$ one has $\left\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\right\|_\infty\le\eta_{T,\epsilon}\sqrt{\ln(k)\ln(T)}$. □

Proof of Corollary 1 in the main paper. Combine Theorems 1 and 2. □

Proof of Corollary 2 in the main paper. Combine Theorems 1 and 2 noting that the former is valid uniformly over $\{w\in\mathbb{R}^n:\|w\|_{\ell_1}\le 1+c\}$. □

[9]The stationarity of y_t and ϵ_t implies that η_T^2 is also an upper bound on the variance of $\epsilon_{T+1,i}$ for all $i=1,...,k$.
[10]Here we choose $L\ge 1$ for convenience in order to also bound the constant 1 in Z_t with probability 1.
[11]The stationarity of ϵ_t implies that $\eta_{T,\epsilon}^2$ can also be used as an upper bound on the variance of all entries of ϵ_{T+1}.

Proof of Theorem 3 in main paper. By Corollary 6.2.32 in Horn and Johnson (1991) it follows that

$$\begin{aligned}\left\|e^{\hat{\Sigma}_{T+1}}-e^{\Sigma_{T+1}}\right\|_\infty &\le \left\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\right\|_\infty e^{n\|\hat{\Sigma}_{T+1}\|_\infty}e^{k\|\Sigma_{T+1}\|_\infty}\\ &\le \left\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\right\|_\infty e^{n(2\|\Sigma_{T+1}\|_\infty+\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\|_\infty)}\end{aligned} \tag{21}$$

where we used that $n\left\|\cdot\right\|_\infty$ is a submultiplicative norm on $n\times n$ matrices. Recall that we are assuming the same VAR structure as in Theorem 2 but for the log-matrix transformed data. Therefore we can use the same probabilistic bounds. In particular, with probability at least $1-4[kp+1]^{1-\ln(T)}$

$$\|\Sigma_{T+1}\|_\infty=\max_{1\le i\le k}|y_{T+1,i}|\le\|Z_T\|_{\ell_\infty}\max_{1\le i\le k}\|\gamma_i^*\|_{\ell_1}+\max_{1\le i\le k}|\epsilon_i|\le\sqrt{\tilde{K}_T}\left(\max_{1\le i\le k}\|\gamma_i^*\|_{\ell_1}+1\right) \tag{22}$$

where we used (20) and $\tilde{K}_T=\frac{\eta_T^2(9m_T^2+\ln(kp+1)\ln(T))}{\eta_T^2\wedge\frac{1}{6}}$. In the case where the γ_i^* are unknown and estimated by the Lasso we furthermore utilize the upper bound from Theorem 2 to conclude $\|\hat{\Sigma}_{T+1}-\Sigma_{T+1}\|_\infty$ to conclude

$$\begin{aligned}\left\|e^{\hat{\Sigma}_{T+1}}-e^{\Sigma_{T+1}}\right\|_\infty &\le \sqrt{\tilde{K}_T}\left(\frac{16}{q\kappa^2}\bar{s}\lambda_T+1\right)e^{n\left(2\sqrt{\tilde{K}_T}\left[\max_{1\le i\le k}\|\gamma_i^*\|_{\ell_1}+1\right]+\sqrt{\tilde{K}_T}\left[\frac{16}{q\kappa^2}\bar{s}\lambda_T+1\right]\right)}\\ &=\sqrt{\tilde{K}_T}\left(\frac{16}{q\kappa^2}\bar{s}\lambda_T+1\right)e^{n\sqrt{\tilde{K}_T}\left(2\left[\max_{1\le i\le k}\|\gamma_i^*\|_{\ell_1}+1\right]+\left[\frac{16}{q\kappa^2}\bar{s}\lambda_T+1\right]\right)}\end{aligned}$$

with probability at least $1-4(k^2p)^{1-\ln(1+T)}-2(1+T)^{-1/A}-\pi_q(\bar{s})-4[kp+1]^{1-\ln(T)}$.

If, on the other hand, the γ_i^* were known, Lemma 1 in the main paper in combination with (21) and (22) yields

$$\left\|e^{\hat{\Sigma}_{T+1}}-e^{\Sigma_{T+1}}\right\|_\infty\le\sqrt{\eta_{T,\epsilon}^2\log(k)\ln(T)}e^{n\left(2\sqrt{\tilde{K}_T}\left[\max_{1\le i\le k}\|\gamma_i^*\|_{\ell_1}+1\right]+\sqrt{\tilde{K}_T}\right)}$$

with probability at least $1-4(k^2p)^{1-\ln(1+T)}-k^{1-\ln(T)}$. □

References

Billingsley, P. (1999). *Convergence of probability measures.* John Wiley & Sons.

Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics. Vol. 1.* Chapman & Hall.

Horn, R. A. and C. R. Johnson (1991). *Topics in matrix analysis.* Cambridge University Press.

Kock, A. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics 186*, 325–344.

Lunde, A., N. Shephard, and K. Sheppard (2015). Econometric analysis of vast covariance matrices using composite realized kernels. *Journal of Business and Economic Statistics*. Forthcoming.

Massart, P. (2007). *Concentration inequalities and model selection*, Volume 10. Springer.

van De Geer, S. A., P. Bühlmann, et al. (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics 3*, 1360–1392.

van Der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer.

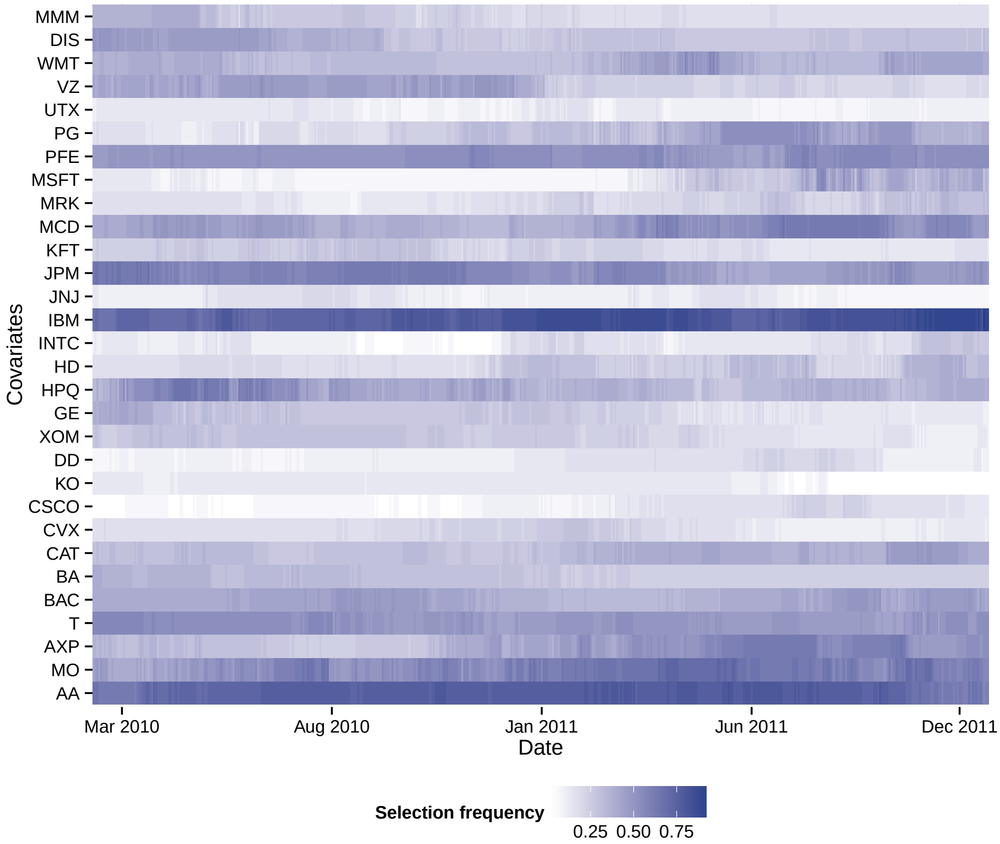

FIGURE SP-1. Selection frequency of the stocks in the columns. At each point in time on the x-axis we calculate the selection frequence across the 30 variance equations of each of the stocks on the y-axis. The results are based on a VAR(1) model estimated by the Lasso at the daily level of aggregation.

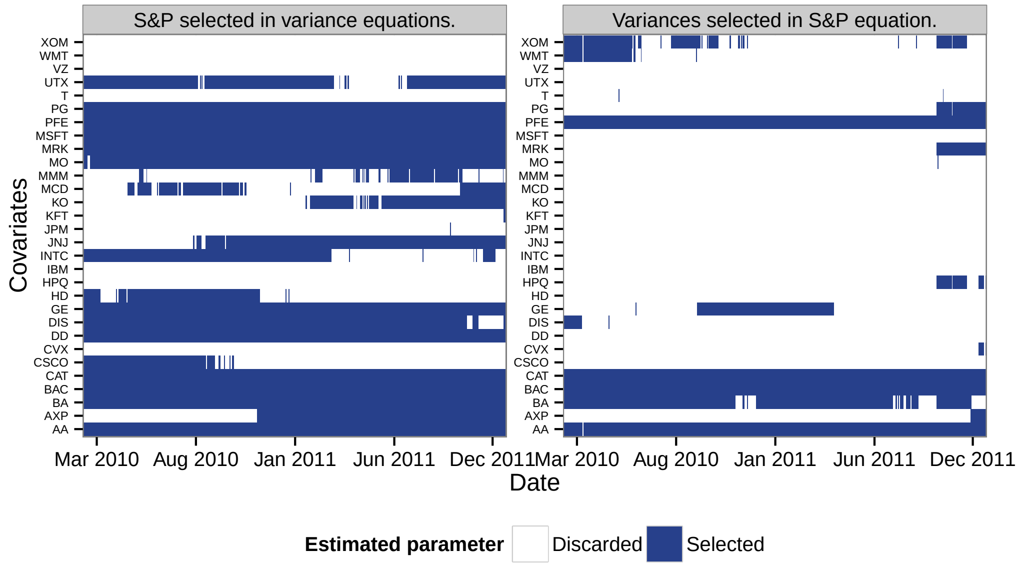

FIGURE SP-2. Lagged variance of the S&P selected in the variance equations of the Dow Jones stocks (left panel) and lagged variances of the Dow Jones stocks selected in the equation of the variance of the S&P 500 (right panel). The results are based on a VAR(1) model estimated by the Lasso at the daily level of aggregation.

Basic Materials 2	Technology 4	Consumer Cyclical 3	Consumer Non-cyclical 7
Energy 2	Financial 3	Industrial 5	Communication 4

Table sp-1. Number of stocks per category. 30 Dow Jones stocks.

		Basic Materials	Consumer, Non-cyclical	Financial	Communications	Industrial	Energy	Technology	Consumer, Cyclical	S&P 500
		Variance								
Variance	Basic Materials	0.75	0.19	0.03	0.27	0.04	0.08	0.44	0.23	0.34
	Consumer, Non-cyclical	0.00	0.20	0.72	0.07	0.17	0.05	0.03	0.10	0.04
	Financial	0.13	0.28	0.18	0.11	0.25	0.00	0.16	0.56	0.01
	Communications	0.06	0.40	0.19	0.19	0.12	0.02	0.09	0.11	0.00
	Industrial	0.29	0.26	0.37	0.17	0.00	0.55	0.18	0.30	0.34
	Energy	0.12	0.10	0.06	0.07	0.00	0.00	0.44	0.02	0.17
	Technology	0.00	0.05	0.21	0.09	0.78	0.00	0.01	0.03	0.17
	Consumer, Cyclical	0.09	0.11	0.02	0.28	0.00	0.02	0.15	0.05	0.13
	S&P500	0.68	0.05	0.01	0.35	0.00	0.12	0.50	0.25	1.00
Covariance	Basic Materials	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Financial	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Communications	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Industrial	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.00
	Energy	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.01
	Technology	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	Consumer, Cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S&P500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Covariance								
Variance	Basic Materials	0.28	0.06	0.01	0.04	0.05	0.24	0.04	0.06	0.12
	Consumer, Non-cyclical	0.24	0.31	0.21	0.08	0.11	0.12	0.12	0.13	0.01
	Financial	0.01	0.04	0.25	0.03	0.01	0.07	0.03	0.04	0.03
	Communications	0.33	0.27	0.27	0.35	0.24	0.41	0.32	0.28	0.22
	Industrial	0.01	0.01	0.02	0.02	0.06	0.01	0.01	0.00	0.00
	Energy	0.06	0.01	0.05	0.01	0.01	0.28	0.04	0.04	0.03
	Technology	0.28	0.17	0.13	0.20	0.16	0.13	0.34	0.15	0.08
	Consumer, Cyclical	0.03	0.15	0.21	0.10	0.17	0.23	0.07	0.40	0.03
	S&P500	0.66	0.23	0.48	0.06	0.03	0.33	0.41	0.21	0.51
Covariance	Basic Materials	0.05	0.07	0.07	0.07	0.06	0.07	0.06	0.06	0.04
	Consumer, Non-cyclical	0.02	0.03	0.03	0.02	0.01	0.03	0.02	0.02	0.01
	Financial	0.05	0.07	0.10	0.06	0.03	0.05	0.05	0.05	0.03
	Communications	0.04	0.05	0.05	0.06	0.02	0.05	0.04	0.04	0.04
	Industrial	0.10	0.09	0.09	0.09	0.14	0.07	0.09	0.08	0.04
	Energy	0.08	0.08	0.09	0.08	0.07	0.13	0.09	0.09	0.06
	Technology	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.03	0.02
	Consumer, Cyclical	0.02	0.03	0.03	0.03	0.02	0.05	0.03	0.03	0.02
	S&P500	0.06	0.04	0.04	0.07	0.03	0.08	0.06	0.04	0.08

TABLE SP-2. Fraction of variables selected from each sector (in row) when modeling variables from the sector in the columns. Model: VAR(1), Lasso, daily, Dow-Jones augmented with S&P 500.

		Basic Materials	Consumer, Non-cyclical	Financial	Communications	Industrial	Energy	Technology	Consumer, Cyclical
		Variance Equations							
Lagged variance	Basic Materials	0.75	0.13	0.08	0.35	0.08	0.00	0.41	0.19
	Consumer, Non-cyclical	0.12	0.43	0.08	0.15	0.04	0.00	0.02	0.08
	Financial	0.17	0.54	0.78	0.33	0.66	0.19	0.18	0.51
	Communications	0.26	0.08	0.04	0.27	0.04	0.00	0.08	0.04
	Industrial	0.00	0.00	0.00	0.00	0.69	0.00	0.01	0.00
	Energy	0.47	0.20	0.19	0.20	0.00	0.81	0.31	0.32
	Technology	0.22	0.17	0.01	0.07	0.00	0.02	0.55	0.00
	Consumer, Cyclical	0.19	0.32	0.06	0.20	0.25	0.06	0.23	0.64
Lagged covariance	Basic Materials	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	Financial	0.00	0.01	0.02	0.01	0.00	0.00	0.01	0.00
	Communications	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Industrial	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	Energy	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01
	Technology	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
	Consumer, Cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Covariance Equations							
Lagged variance	Basic Materials	0.43	0.12	0.09	0.11	0.11	0.27	0.13	0.07
	Consumer, Non-cyclical	0.07	0.26	0.07	0.10	0.03	0.17	0.08	0.08
	Financial	0.12	0.33	0.79	0.20	0.06	0.46	0.16	0.23
	Communications	0.23	0.17	0.10	0.28	0.17	0.14	0.24	0.18
	Industrial	0.04	0.01	0.00	0.02	0.02	0.00	0.03	0.00
	Energy	0.25	0.14	0.31	0.04	0.02	0.38	0.18	0.17
	Technology	0.29	0.14	0.06	0.12	0.10	0.07	0.36	0.16
	Consumer, Cyclical	0.24	0.26	0.18	0.15	0.37	0.29	0.18	0.51
Lagged covariance	Basic Materials	0.06	0.05	0.03	0.07	0.05	0.06	0.05	0.04
	Consumer, Non-cyclical	0.03	0.05	0.04	0.04	0.04	0.04	0.04	0.04
	Financial	0.06	0.07	0.10	0.07	0.04	0.05	0.05	0.06
	Communications	0.04	0.04	0.05	0.06	0.04	0.05	0.05	0.04
	Industrial	0.06	0.05	0.05	0.07	0.14	0.04	0.05	0.05
	Energy	0.04	0.05	0.03	0.05	0.04	0.10	0.07	0.05
	Technology	0.03	0.02	0.03	0.03	0.02	0.03	0.05	0.02
	Consumer, Cyclical	0.03	0.03	0.03	0.03	0.02	0.03	0.04	0.04

TABLE SP-3. Fraction of variables selected from each sector (in row) when modeling variables from the sector in the columns. Model: VAR(1), Lasso, weekly.

		AMedAFE			AMaxAFE			ℓ_2		
Model	h	A	D	O	A	D	O	A	D	O
No-Change	1	0.11	0.29	0.11	1.62	1.62	0.53	4.49	2.97	3.2
Benchmark	10	0.59	1	0.58	4.54	4.54	1.86	16.83	8.35	14.36
	26	0.96	1.46	0.94	7.26	7.26	3.26	27.44	13.05	24
DCC(1,1)	1	5.09	4.62	5.00	3.10	3.10	3.60	3.96	3.47	4.40
EWMA ($\lambda = 0.96$)	1	5.36	4.93	5.18	3.30	3.30	3.81	4.15	3.70	4.58
VAR(1), Lasso	1	1.45	1.07	1.36	0.98	0.98	1.53	1.22	1.01	1.40
	10	1.05	0.96	1.03	0.99	0.95	1.27	1.04	0.95	1.07
	26	0.72	0.79	0.72	0.89	0.88	0.86	0.78	0.83	0.76
VAR(1), Lasso	1	1.36	0.90	1.27	0.90	0.87	1.91	1.17	0.91	1.38
Post Lasso OLS	10	1.17	1.17	1.17	1.13	1.07	1.51	1.25	1.11	1.28
	26	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf
VAR(1), adaptive Lasso	1	1.36	1.10	1.36	1.02	1.02	1.47	1.23	1.05	1.39
Initial estimator: Lasso	10	1.10	0.96	1.10	1.02	0.95	1.39	1.10	0.96	1.15
	26	0.73	0.78	0.73	0.89	0.87	0.89	0.79	0.83	0.77
VAR(1), Lasso	1	1.18	1.14	1.09	0.94	0.94	1.15	1.06	1.01	1.13
Log-matrix transform	10	0.97	1.01	0.97	0.98	0.98	0.98	0.98	0.99	0.97
	26	0.83	0.79	0.84	0.90	0.90	0.91	0.87	0.85	0.88
VAR(1), Lasso	1	1.45	1.07	1.36	0.98	0.98	1.55	1.23	1.02	1.41
Not eigenvalue cleaned	10	1.05	0.96	1.03	0.99	0.95	1.27	1.04	0.96	1.07
	26	0.72	0.79	0.72	0.93	0.90	0.89	0.79	0.85	0.76
VAR(1), Lasso	1	1.45	1.07	1.45	0.98	0.98	1.57	1.27	1.01	1.43
Including S&P 500	10	1.03	0.96	1.03	1.00	0.96	1.29	1.10	0.96	1.09
	26	0.72	0.79	0.72	0.89	0.88	0.86	0.82	0.83	0.76
VAR(5), Lasso	1	1.36	1.03	1.27	0.94	0.94	1.55	1.17	0.99	1.33
	10	1.05	0.95	1.05	0.96	0.96	1.10	1.02	0.95	1.06
	26	0.73	0.79	0.72	0.90	0.89	0.83	0.79	0.84	0.76
VAR(5), Lasso	1	1.18	0.90	1.09	0.89	0.86	1.85	1.11	0.89	1.30
Post Lasso OLS	10	1.00	0.91	1.00	1.00	0.97	1.34	1.02	0.94	1.05
	26	0.81	0.87	0.81	1.02	1.00	1.01	0.89	0.95	0.87
VAR(5), adaptive Lasso	1	1.36	1.10	1.27	1.02	1.02	1.57	1.20	1.05	1.34
Initial estimator: Lasso	10	1.08	0.96	1.09	0.97	0.93	1.25	1.07	0.94	1.11
	26	0.73	0.79	0.73	0.90	0.88	0.86	0.79	0.84	0.77
VAR(5), Lasso	1	1.09	1.10	1.09	0.91	0.91	1.15	1.04	0.97	1.11
Log-matrix transform	10	0.95	0.99	0.97	0.95	0.95	0.97	0.96	0.96	0.96
	26	0.83	0.79	0.84	0.90	0.90	0.91	0.87	0.85	0.88
VAR(5), Lasso	1	1.36	1.03	1.36	0.94	0.94	1.47	1.21	0.99	1.33
Including S&P 500	10	1.05	0.96	1.05	0.99	0.95	1.20	1.08	0.95	1.06
	26	0.74	0.79	0.73	0.91	0.90	0.84	0.84	0.85	0.77

TABLE SP-4. Forecasts precision measures with weekly observations. Absolute measures (gray background) and relative to benchmark (white background). A: All equations, D: Diagonal equations, O: Off-diagonal equations.

		Basic Materials	Consumer, Non-cyclical	Financial	Communications	Industrial	Energy	Technology	Consumer, Cyclical
		Variance Equations							
Lagged variance	Basic Materials	0.10	0.19	0.21	0.30	0.48	0.06	0.27	0.34
	Consumer, Non-cyclical	0.00	0.10	0.01	0.01	0.06	0.00	0.00	0.02
	Financial	0.00	0.52	0.65	0.40	0.68	0.10	0.10	0.35
	Communications	0.14	0.14	0.09	0.19	0.19	0.08	0.17	0.15
	Industrial	0.00	0.03	0.01	0.01	0.00	0.04	0.16	0.05
	Energy	0.57	0.35	0.28	0.31	0.10	0.61	0.41	0.37
	Technology	0.12	0.06	0.00	0.04	0.03	0.01	0.23	0.01
	Consumer, Cyclical	0.01	0.06	0.09	0.13	0.08	0.02	0.11	0.09
Lagged covariance	Basic Materials	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Financial	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	Communications	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.01
	Industrial	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Energy	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	Technology	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.01
	Consumer, Cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Covariance Equations							
Lagged variance	Basic Materials	0.69	0.89	0.94	0.78	0.75	0.90	0.75	0.77
	Consumer, Non-cyclical	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.00
	Financial	0.18	0.35	0.49	0.26	0.09	0.10	0.14	0.24
	Communications	0.29	0.40	0.42	0.34	0.29	0.41	0.35	0.39
	Industrial	0.39	0.13	0.11	0.05	0.05	0.38	0.26	0.15
	Energy	0.62	0.63	0.61	0.52	0.17	0.70	0.63	0.57
	Technology	0.05	0.01	0.02	0.01	0.02	0.02	0.05	0.00
	Consumer, Cyclical	0.42	0.45	0.50	0.42	0.38	0.44	0.40	0.37
Lagged covariance	Basic Materials	0.01	0.01	0.02	0.01	0.00	0.03	0.01	0.01
	Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Financial	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00
	Communications	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05
	Industrial	0.02	0.03	0.05	0.03	0.01	0.07	0.03	0.03
	Energy	0.02	0.02	0.02	0.02	0.03	0.05	0.02	0.02
	Technology	0.04	0.06	0.06	0.06	0.03	0.05	0.04	0.04
	Consumer, Cyclical	0.02	0.03	0.04	0.02	0.02	0.04	0.02	0.02

TABLE SP-5. Fraction of variables selected from each sector (in row) when modeling variables from the sector in the columns. Model: VAR(1), Lasso, monthly.

		AMedAFE			AMAaxFE			ℓ_2		
Model	h	A	D	O	A	D	O	A	D	O
No-Change	1	0.11	0.22	0.1	1.28	1.28	0.67	4.34	2.56	3.38
Benchmark	2	0.44	0.71	0.43	2.99	2.99	1.33	12.16	5.8	10.6
	6	0.87	1.29	0.86	6.15	6.15	2.84	24.49	11.34	21.64
EWMA ($\lambda = 0.96$)	1	6.27	8.32	6.70	4.43	4.43	3.34	4.97	4.92	5.07
VAR(1), Lasso	1	1.64	1.05	1.80	0.95	0.77	1.57	1.27	0.84	1.49
	2	1.48	1.10	1.51	1.17	1.09	1.77	1.43	1.06	1.53
	6	0.93	0.87	0.93	0.96	0.91	1.03	0.98	0.88	0.98
VAR(1), Lasso	1	2.64	1.50	2.90	1.42	1.06	2.60	2.08	1.17	2.50
Post Lasso OLS	2	1.59	1.23	1.60	1.29	1.21	1.95	1.59	1.19	1.69
	6	1.31	1.68	1.31	1.32	1.23	2.31	1.60	1.49	1.59
VAR(1), adaptive Lasso	1	1.36	1.05	1.50	0.88	0.83	1.33	1.15	0.86	1.32
Initial estimator: Lasso	2	1.20	1.11	1.21	1.15	1.09	1.41	1.20	1.06	1.25
	6	0.85	0.85	0.85	0.91	0.89	0.89	0.88	0.87	0.87
VAR(1), Lasso	1	0.91	0.91	1.00	0.89	0.89	0.91	0.93	0.93	0.93
Log-matrix transform	2	1.07	1.11	1.07	1.17	1.17	1.12	1.08	1.11	1.08
	6	0.92	0.84	0.92	0.98	0.98	0.99	0.94	0.92	0.95
VAR(1), Lasso	1	2.64	1.32	2.90	1.80	1.20	3.39	2.21	1.16	2.67
Not eigenvalue cleaned	2	1.95	0.99	2.02	1.62	1.09	2.63	1.95	1.01	2.13
	6	2.66	1.05	2.74	2.12	1.05	3.99	2.69	1.03	2.94
VAR(1), Lasso	1	1.64	1.09	1.80	1.02	0.84	1.75	1.32	0.88	1.50
Including S&P 500	2	1.50	1.08	1.51	1.16	1.11	1.66	1.49	1.06	1.53
	6	0.90	0.85	0.91	0.91	0.88	1.00	0.98	0.84	0.94
VAR(5), Lasso	1	1.27	1.14	1.40	0.82	0.77	1.33	1.09	0.83	1.25
	2	1.00	1.03	1.00	1.09	1.09	1.13	1.02	1.01	1.03
	6	0.84	0.87	0.83	0.99	0.99	0.94	0.87	0.91	0.85
VAR(5), Lasso	1	1.36	1.18	1.50	1.04	0.94	1.75	1.35	0.97	1.56
Post Lasso OLS	2	1.18	1.03	1.19	1.03	0.97	1.40	1.19	0.99	1.25
	6	1.24	1.01	1.24	0.98	0.96	1.43	1.21	0.97	1.26
VAR(5), adaptive Lasso	1	1.18	1.05	1.30	0.90	0.81	1.42	1.09	0.82	1.24
Initial estimator: Lasso	2	1.02	1.03	1.02	1.11	1.11	1.17	1.04	1.02	1.05
	6	0.77	0.84	0.77	0.96	0.96	0.82	0.81	0.89	0.78
VAR(5), Lasso	1	0.91	0.95	1.00	0.97	0.97	0.97	0.97	0.97	0.97
Log-matrix transform	2	1.02	1.14	1.02	1.13	1.13	1.08	1.05	1.09	1.04
	6	0.95	0.99	0.94	1.03	1.03	1.03	0.98	1.00	0.97
VAR(5), Lasso	1	1.36	1.14	1.50	0.86	0.86	1.28	1.15	0.83	1.30
Including S&P 500	2	1.00	1.01	1.00	1.05	1.05	1.17	1.08	0.98	1.06
	6	0.91	0.88	0.91	0.96	0.96	1.03	0.96	0.88	0.92

TABLE SP-6. Forecasts precision measures with monthly observations. Absolute measures (gray background) and relative to benchmark (white background). A: All equations, D: Diagonal equations, O: Off-diagonal equations.