Maastricht University

Granger Causality Testing in High-Dimensional VARs: a Post-Double-Selection Procedure

Citation for published version (APA):

Hecq, A., Margaritella, L., & Smeekes, S. (2019). Granger Causality Testing in High-Dimensional VARs: a Post-Double-Selection Procedure. (arXiv e-prints; No. 1902.10991). arXiv.org at Cornell University Library.

Document status and date:
Published: 28/02/2019

Document Version:
Early version, also known as pre-print

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 23 Sep. 2019

arXiv:1902.10991v1 [econ.EM] 28 Feb 2019

Granger Causality Testing in High-Dimensional VARs: a Post-Double-Selection Procedure

Hecq, A., Margaritella, L., Smeekes, S.

Department of Quantitative Economics
School of Business and Economics
Maastricht University

February, 2019

Abstract

In this paper we develop an LM test for Granger causality in high-dimensional VAR models based on penalized least squares estimations. To obtain a test which retains the appropriate size after the variable selection done by the lasso, we propose a post-double-selection procedure to partial out the effects of the variables not of interest. We conduct an extensive set of Monte-Carlo simulations to compare different ways to set up the test procedure and choose the tuning parameter. The test performs well under different data generating processes, even when the underlying model is not very sparse. Additionally, we investigate two empirical applications: the money-income causality relation using a large macroeconomic dataset and networks of realized volatilities of a set of 49 stocks. In both applications we find evidences that the causal relationship becomes much clearer if a high-dimensional VAR is considered compared to a standard low-dimensional one.

Keywords: Granger causality, Post-double-selection, vector autoregressive models, high-dimensional inference
JEL: C55, C12, C32

1. Introduction

With the increase of data availability, high-dimensional (HD) econometric and statistical models have gained a lot of interest over the last twenty years. Economics, statistics and finance have seen a rapid increase of applications involving time series in high-dimensional systems. Central to many of these applications is the vector autoregressive (VAR) model that allows for a flexible modelling of dynamic interactions between multiple time series. In this paper we develop a simple method to test for Granger causality in high-dimensional VARs (HD-VARs) with potentially many variables.

The second and third author thank NWO for financial support. Previous versions of this paper have been presented at NESG 2018, CFE-CM Statistics 2018 and EC^2 2018. We gratefully acknowledge the comments by participants at these conferences. In addition, we thank Marcelo Medeiros, Anders B. Kock, Etienne Wijler and Benoit Duvocelle for their valuable feedback. All remaining errors are our own. Address correspondence to: Luca Margaritella, Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: l.margaritella@maastrichtuniversity.nl

The notion of Granger causality captures predictability given a particular information set (Granger, 1969). By increasing the latter, and thus considering high(er)-dimensional VAR models, one can reduce the effect of potential omitted variables, and hence obtain a clearer picture on the relation between the variables of interest. However, the number of parameters in a VAR increases quadratically with the number of time series included; an unrestricted VAR(p) has K^2p coefficients to be estimated, where K is the number of series and p is the lag-length. As the time series dimension T is typically fairly small for many economic applications, the data do not contain sufficient information to estimate the parameters and consequently standard least squares and maximum likelihood methods become unreliable, resulting in estimators with high variance that overfit the data.

While the econometric literature has often been focused on allowing for high dimensionality in VARs through the use of factor models (see e.g. Bernanke et al., 2005, Chudik and Pesaran, 2016) or Bayesian methods (Bańbura et al., 2010), recent years have seen an increase in *regularized*, or penalized, estimation of *sparse* VARs based on popular methods from statistics such as the lasso (Tibshirani, 1996) and elastic net algorithms (Zou and Hastie, 2005). These methods impose sparsity by setting a (data-driven) selection of the coefficients to zero. Compared to factor models, such sparsity-seeking methods have often an advantage of interpretability, as in many economic and financial aplications, it appears natural to believe that the most important dynamic interactions among a large set of variables can be adequately captured by a relatively small – but unknown – number of 'key' variables. As such, the use of these methods for estimating HD-VAR models has also increased significantly in recent years.

Recent theoretical advances in regularized VAR estimation include Kock and Callot (2015), who developed non-asymptotic oracle inequalities for the lasso in stationary VAR models, establishing upper bounds for its prediction and estimation error. Large VAR models have also been studied in Song and Bickel (2011) where the time series set up is exploited by taking into account different strengths of dependence in the model. Davis et al. (2016) proposed a two-step procedure to fit sparse VARs by means of employing ℓ_1-penalized likelihood estimators using information criteria and spectral coherence. The consistency of these ℓ_1-penalized estimators has been established by Basu and Michailidis (2015) who derive non-asymptotic bounds on the estimation errors for many stable non-Gaussian processes.

While regularized estimation theory for high-dimensional time series is now well established, performing inference on HD-VARs, such as testing for Granger causality, still remains a non-trivial matter. As is well known, performing inference after model selection (post-selection inference) is complicated as the selection step invalidates 'standard' inference where the uncertainty regarding the selection is ignored (see Leeb and Pötscher, 2005). While recent years have seen enormous advances in the literature on valid post-selection inference, the statistical literature has mainly focused on independent data.

Therefore, the complexities introduced by the temporal and cross-sectional dependencies in the VAR mean that standard post-selection inference methods are not automatically applicable.

Existing literature on Granger causality testing in HD-VARs therefore has so far not considered post-selection inferential procedures. Wilms et al. (2016) propose a bootstrap Granger causality test in HD-VARs, but they do not account for post-selection issues. Similarly, Skripnikov and Michailidis (2018) investigate the problem of jointly estimating multiple network Granger causal models in VARs with sparse transition matrices using lasso-type methods, but focus mostly on estimation rather than testing. Song and Taamouti (2017) focus on statistical procedures for testing indirect/spurious causality in highly dimensional scenarios, but consider factor models rather than regularized regression techniques.

In this paper we build on the post-double-selection approach proposed by Belloni et al. (2014b), to develop a valid post-selection test of Granger causality in HD-VARs. The finite-sample performance depends heavily on the exact implementation of the method. In particular, the tuning parameter selection in the penalized estimation is crucial. We therefore perform an extensive simulation study to investigate the finite-sample performance of the different ways to set up the test in order to be able to give some practical recommendations. In addition, we revisit the long-standing economic discussion on the money-income causality relation (Sims, 1972) using the high-dimensional FRED-QD dataset, and investigate the construction of networks of realized volatilities using a sample of 49 financial stocks modeled as an heterogeneous VAR (Corsi, 2009). In both empirical exercises we are able to demonstrate how our approach allows for obtaining much sharper conclusions than standard low-dimensional VAR techniques.

The remainder of the paper is as follows: Section 2 introduces the high-dimensional VAR model, the regularization techniques employed, and the issue of choosing the tuning parameter. Section 3 discusses the post-selection inferential problem and consequently proposes our post-double-selection procedure for Granger causality testing in high-dimensional VAR models. Section 4 reports the results of the Monte-Carlo simulations. Section 5.1 investigates a first empirical application by testing for Granger causality between money and income. Section 5.2 investigates a second empirical application on networks in realized volatilities. Section 6 is left for conclusions and some final remarks.

2. High-dimensional VAR models and regularization

2.1. The model

Let $y_1, \dots, y_T$ be a K-dimensional multiple time series process, where $y_t = (y_{1,t}, \dots, y_{K,t})'$ is generated by a VAR(p) process

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \epsilon_t, \qquad t = p+1, \dots, T\ , \tag{1}$$

where for notation simplicity we omit the vector of intercepts. $A_1, \ldots, A_p$ are $K \times K$ parameter matrices and ϵ_t is a sequence of martingale difference sequence error terms with a non-singular $(K \times K)$ covariance matrix Σ_ϵ. We further assume that all the roots of $|I_K - \sum_{j=1}^{p} A_j z^j|$ lie outside the unit disc, such that the lag polynomial process is invertible.

Given the VAR model stated in (1), we are interested in testing whether variable k Granger causes variable i.[1]It is convenient to rewrite each line of (1) into the stacked representation

$$y_i = X\beta_i + \epsilon_i \quad i = 1, \ldots, K \ , \tag{2}$$

where $y_i = (y_{i,p+1}, \ldots, y_{i,T})'$ is the $T-p \times 1$ vector of observations on the i-th variable for $i = 1, \ldots, K$, $X = (Z_{p+1}, \ldots, Z_T)'$ being the $T - p \times Kp$ matrix of covariates containing for each t the $Kp \times 1$ vector of explanatory variables $Z_t = (y'_{t-1}, \ldots, y'_{t-p})'$. Finally β_i is the Kp vector of coefficients and $\epsilon_i = (\epsilon_{i,p+1}, \ldots, \epsilon_{i,T})$ is the $T-p \times 1$ vector of error terms. Call X_{GC} the columns of X containing the Granger causing variables; that is, the p lags of $y_{k,t}$, and let $\beta_{i,GC}$ be the corresponding coefficients. Testing for Granger causality then implies testing

$$H_0 : \beta_{i,GC} = 0 \quad \text{against} \quad H_1 : \beta_{i,GC} \neq 0.$$

As (2) is a high-dimensional regression, we will apply regularization techniques to select the relevant variables from X. Next to the selection of variables, we also need to select the number of lags p.

As far as p is concerned, there exists an important advantage for considering a high-dimensional VAR(p) framework over a small scale setting. There are indeed theoretical reasons to favor a small p, say from 1 to 2 in large VARs. This neglected feature of high-dimensional systems is justified by having a look at what is called the final equation representation (FER, see Zellner and Palm, 1974, 1975, 2004) of a finite dimensional VAR(p).

In order to investigate the mechanisms underlying the marginalization features (i.e. ARMA or VARMA) generated from a VAR(p), let us rewrite equation (1) in its polynomial form such as

$$(I - A_1 L - \ldots - A_p L^p)\, y_t \equiv A(L) y_t = \epsilon_t. \tag{3}$$

Premultiplying both sides of (3) by the adjoint of the matrix polynomial $A(L)$, i.e. by

[1]Although we focus mainly on testing Granger casuality for individual variables, the procedure can easily be extended to testing blocks of variables; see Remark 2 for details.

$A^*(L) = \det[A(L)]A(L)^{-1}$, leads to

$$\det[A(L)]y_t = A^*(L)\epsilon_t. \tag{4}$$

It is observed in (4) that each element of y_t follows an ARMA$(Kp, (K-1)p)$. This means that the univariate ARMA models derived from a VAR(p) with for instance $p = 2$ lags and $K = 100$ series are of orders ARMA(200, 198). Obviously these are maximal orders and smaller numbers are obtained in case of existing cancellations between the AR and the MA part. For instance the presence of a reduced rank structure and/or block diagonal matrices lead to lower implied dynamic models (Cubadda et al., 2009, Hecq et al., 2016). Without any restriction the VAR(p) can even yield long memory processes for $K \to \infty$ (see Chevillon et al., 2018). From the same tools, we can also derive that partial systems are VARMA and not VARs anymore. As an example, a bivariate system derived from the same VAR(p) with $p = 2$ and $K = 100$ will be a VARMA(100,98), a model that will be typically approximate by a VAR with potentially many lags. Hence, given the lag lengths usually estimated in empirical macro, namely $p = 4$ or 8, on quarterly data with a small set of series say $K = 4$ or 5, it is plausible to assume that the data generating process of the high-dimensional VAR has a small p.

As such, we can fix p to a small number and do not have to consider the joint estimation of lag length and variables, such as done in Nicholson et al. (2018) for example. We will evaluate the robustness of the choice of p both in Monte Carlo simulations and in empirical illustrations.

2.2. The lasso estimator

As β_i is high-dimensional when Kp is large relative to T, least squares estimation is not appropriate, and a structure must be imposed on β_i to be able to estimate it consistently. In particular, we assume *sparsity* of β_i; that is, we assume that β_i can accurately be approximated by a coefficient vector with a (significant) portion of the coefficients equal to zero.[2]

The sparsity assumption validates the use of variable selection to obtain the subset of relevant covariates in explaining y_i, thereby reducing the dimensionality of the system without having to sacrifice predictability. We use the lasso to simultaneously perform the variable selection and the estimation of the parameters by solving

$$\hat{\beta}_{i,lasso} = \arg\min_{\beta_i} \left(\frac{1}{T} ||y_i - X\beta_i||_2^2 + \lambda ||w_i'\beta_i||_1 \right), \tag{5}$$

[2]Formally we can make the distinction between exact sparsity, which implies that at most s elements of β_i are non-zero with $s << T$, and approximate sparsity, which allows all regressors to potentially have a non-zero effect on the dependent variable, but no more than s are needed for accurately approximating it (see e.g. Belloni et al., 2011a). As we are not interested in variable selection as such, approximate sparsity suffices for our purposes.

where for any n-dimensional vector x, $||x||_q = \left(\sum_{j=1}^{n} |x_j|^q\right)^{1/q}$ is the standard ℓ_q-norm, λ is a non-negative tuning parameter determining the strength of the penalty, and w_i is a vector of weights corresponding to the parameters in β_i. For the standard lasso all weights are either equal to one, or equal to zero (if this parameter should not be penalized).

The lasso estimator combines shrinking parameter estimates towards zero (proportional to increasing λ) and a variable selection as the penalty function is non-differentiable at zero. However, the convexity of the ℓ_1-penalty ensures that fast algorithms can be used to compute the solution efficiently (Friedman et al., 2010). In general the lasso does not provide consistent variable selection as it selects too many variables, and does not therefore have the "oracle property" of being able to select the right set of variables with probability 1 asymptotically. Zou (2006) proposes the *adaptive lasso* with parameter-specific weights w_i – different from 0 or 1 – in (5) to obtain the oracle properties, provided a proper choice of the tuning parameter λ. However, note that for our purpose, oracle properties are not very relevant; we wish to eliminate the effects of the other "nuisance" variables on the relation between the variables tested for Granger causality, but we do not need to identify which of these nuisance variables matter.

The adaptive lasso in the context of VARs have been studied by Kock and Callot (2015). They show that it is still an oracle procedure in the time series setting. Lower bounds on the finite sample probabilities of selecting the right model are derived and employed to establish with high probability the true sparsity pattern. Conditions (sufficient) for sign consistency comprise a "beta-min" condition on the minimal magnitude of the non-zero coefficients to avoid them being set too close to zero to detect. Furthermore, asymptotic equivalence of the estimates of the non-zero parameters with the oracle-assisted least squares is assessed. Basu and Michailidis (2015) contributed further to the sparse VAR(p) estimation literature. Conversely to previous works (e.g. Loh and Wainwright, 2011) which assume tight dependence conditions on the model parameters or on the transition matrix, they verify that appropriate restricted eigenvalues and deviation conditions hold with high probability. Furthermore, these conditions are sufficient for consistency of the VAR(p) models and especially for every stable VAR under ℓ_1-penalization. Medeiros and Mendes (2016) extended the above conditions and especially the restricted eigenvalue condition with conditionally heteroskedastic errors. Under this more general framework, they manage to show how the adaptive lasso is again able to retrieve the oracle property.

2.3. Tuning parameter selection

A crucial problem in ℓ_1-regularization techniques is the choice of the tuning parameter λ. The task is to find a proper balance between the fit and the model complexity in the variance-bias trade off. In the lasso, $\lambda >> 0$ implies a strong variable selection, hence a larger bias. At the opposite, $\lambda \approx 0$ lets the lasso converge in the limit to the standard

OLS estimator, thus paying the price of not performing any variable selection.

Minimizing an information criterion (IC) in order to determine an appropriate data-driven λ is one way to deal with dependent data as in the time series setting. Let $S \subseteq \{1, \ldots, Kp\}$ denote a subset of variables in X with X_S the columns of X containing only the variables in S. Furthermore, $\beta_{i,S}$ is the subvector of β_i and $\hat{\beta}_{i,S}$ the corresponding subset of estimated parameters. Then λ is selected by minimizing

$$\arg\min_S \left(\frac{1}{T} ln ||y_i - X_S \hat{\beta}_{i,S}||^2 + \frac{1}{T} C_T df \right), \tag{6}$$

where df is the degrees of freedom after the penalization procedure is applied, i.e. the number of non-zero coefficients selected by (adaptive) lasso; C_T is the penalty specific to each criterion. We consider three of them: the *Akaike information criterion* (AIC) Akaike (1974) with $C_T = 2$, the *Bayesian information criterion* (BIC) by Schwarz (1978) with $C_T = \ln(T)$, and the *Extended Bayesian information criterion* (EBIC) by Chen and Chen (2008) with $C_T = \ln(T) + 2\gamma \ln(Kp)$. We consider EBIC as Chen and Chen (2012) argue that BIC fails to correctly address the variable-selection phenomenon in scenarios where the parameter space is substantially higher than the sample size, leading to select a model with spurious covariates in high-dimensional settings. Throughout this paper we calculate EBIC with $\gamma = 0.5$.

ICs provide a simple and fast method to select the tuning parameter. An alternative approach is to plug in estimates of theoretically optimal values (see e.g Bickel et al., 2009, Belloni and Chernozhukov, 2013, Belloni et al., 2011b). The tuning parameter λ is derived as the upper bound on the gradient of the criterion function (i.e. the score). Requiring with high probability $\lambda \geq c||X'\epsilon_i||_\infty/T$, where X is the matrix of covariates and c is an absolute constant, comes as natural in the penalized regression framework where bias towards zero is introduced to help reducing the variance which drives the estimator away from the true value. Since ϵ_i is not known in practice, one can rely on a Gaussian approximation $\lambda = \frac{c\sigma}{\sqrt{T}} \Phi^{-1}\left(1 - \frac{\alpha}{2Kp}\right)$. Otherwise, penalty loadings (ω_i) are used e.g. in Belloni and Chernozhukov (2013), Belloni et al. (2011a), Chernozhukov et al. (2016) to introduce self-normalization of the first-order condition of the lasso problem. This allows to apply moderate deviation theory results (see Jing et al., 2003) to bound deviations of the maximal element of the score vector. More specifically, given the lasso in (5), the penalty loadings are set to $\omega_i = \sqrt{\mathbb{E}X^2}$ and λ is chosen as $\lambda = 2c\hat{\sigma}T^{-1/2}\Phi^{-1}(1 - \alpha/(2Kp))$ where c is a constant set to $= .5$, $\alpha \in (0, 1)$, $\hat{\sigma}$ is the initial estimate of the standard deviation of ϵ_i and $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the standard Gaussian distribution.

Perhaps the most popular way to choose the tuning parameter is cross-validation (CV), although CV is not always appropriate in the time series setup without modifications

(Bergmeir et al., 2015). To estimate the tuning parameter with CV in a time series setup requires to employ a rolling K-fold cross validation (TSCV) in order to gradually train the series avoiding to lose their dependence. When compared to the theoretical approach, TSCV looks appealing since it does not require an estimate of σ. However, as observed in Chetverikov et al. (2016), when applied to the lasso, TSCV typically yields small values of λ thus still gaining fast convergence rate but at the price of less variable selection.

3. Post-Double-Selection Granger causality test

In this section we propose our high-dimensional Granger causality test. Before going into the details of the procedure, we first motivate why we need a post-selection inferential procedure.

One might be tempted to simply perform the (adaptive) lasso as in (5), setting $w_{i,GC} = 0$, and then testing whether $\beta_{i,GC} = 0$, potentially after re-estimating the VAR equation by OLS on only the selected variables. However, this ignores the fact that the final, selected, model is actually random and a function of the data. The randomness contained in the selection step means the post-selection estimators do not converge uniformly to a normal distribution, as the potential omitted variable bias from omitting (weakly) relevant variables in the selection step is too large to maintain uniformly valid inference.

In a sequence of papers (see e.g Leeb and Pötscher, 2005), Leeb and Pötscher addresses these issues, proving impossibility results for estimating the distribution of post model-selection estimators, hence showing how the distributional properties of such ℓ_1-penalized estimators are more complex than might appear at first glance. Consistent or conservative model selection postulated by oracle properties, in fact only holds for large-enough parameters, breaking down for small parameters. Technically, this means that distributions of post-selection estimators only converge point-wise in the parameter space to normal distributions, and as such these asymptotics fail to deliver a proper approximation of finite-sample behavior, unless one is willing to rule out small parameters by imposing assumptions on the minimal magnitude of non-zero coefficients (also known as *beta-min* conditions, see e.g. van de Geer et al. 2011). As such, conditions for perfect selection are way too restrictive, requiring a sharp separation of non-zero coefficients from zero in order for the post-model selection estimator to converge at the usual rate.

To address these issues, several approaches to valid post-selection inference (sometimes referred to as "honest inference") have been developed in recent years based on various philosophies, such as simultaneous inference across models (Berk et al., 2013), inference conditional on selected models (Lee et al., 2016), or debiasing (desparsifying) the lasso estimates (Van de Geer et al., 2014, Zhang and Zhang, 2014). In this paper, we focus on the double selection approach developed by Belloni, Chernozhukov and co-authors; see e.g. (Belloni et al., 2014a) for an overview. This approach is tailored for the lasso, easy to implement, and can be extended to dependent data.

Particularly relevant for our setting is Belloni et al. (2014b), who develop a *post-double-selection* approach to construct uniform inference for treatment effects in partially linear models with high-dimensional controls using the lasso. Two lasso estimations of both the outcome and the treatment variable on all the controls are performed, and a final post-selection least squares estimation is conducted of the outcome variable on the treatment variable and all the controls selected in one of the two steps. The striking difference with the usual post (single) model-selection is the double variable-selection step, which ensures to substantially diminish the omitted variable bias and ensuring the errors of the final model are (close enough to) orthogonal with respect to the treatment. The authors proved uniform validity of the procedure under a wide range of DGPs, including heteroskedastic and non-Gaussian errors.

Chernozhukov et al. (2018) extend the analysis of estimation and inference for highly-dimensional systems in regressions, allowing for (weak) temporal and cross-sectional dependency, both in the errors and the covariates. Regularization techniques for dimensionality reduction are applied iteratively in the system and the overall penalty is jointly chosen by a block multiplier bootstrap procedure. Oracle properties and bootstrap consistency of the test procedure are derived. Furthermore, simultaneous valid inference is obtained via algorithms employing least square or least absolute deviation after (double) lasso selection step(s).

In this paper we consider a direct extension of the approach of Belloni et al. (2014b) to the HD-VAR. Although our approach is closely related to that of Chernozhukov et al. (2018) our method is simpler and faster to implement. We now provide the details of our method. In addition to the already defined matrix X_{GC} containing the lags of the Granger causing variable, let X_{-GC} denote the submatrix of X containing the remaining columns. Further, let $X_{GC}^{(j)}$ denote the column of X_{GC} containing the j-th lag of the Granger causing variable.

Post-double-selection Granger causality LM (PDS-LM) test

---

[1] Estimate $y_i = X_{-GC}\beta_i^{(0)} + \epsilon_i$ by means of (adaptive) lasso as in (5), obtaining estimates $\hat{\beta}_i^{(0)}$, and define the set of selected variables $\hat{S}_0 = \{m : |\hat{\beta}_{m,i}^{(0)}| > 0,\ m = 1, \dots, (K-1)p\}$.

[2] For $j = 1, \dots, p$ perform further steps of selection by regressing the lags of the potential Granger causing variable, $X_{GC}^{(j)}$, onto X_{-GC}

$$\hat{\beta}_i^{(j)} = \arg\min_{\beta_i^{(j)}} \left(||X_{GC}^{(j)} - X_{-GC}\beta_i^{(j)}||_2^2 + \lambda ||w_i^{(j)\prime}\beta_i^{(j)}||_1 \right), \quad j = 1, \dots, p, \tag{7}$$

and let $\hat{S}_j = \{m : |\hat{\beta}_{m,i}^{(j)}| > 0,\ m = 1, \dots, (K-1)p\}$.

[3] Collect all variables kept by the lasso in Steps 1 and 2 by letting $S = \cup_{j=0}^{p} \hat{S}_j$ and define X_S the $T - p \times S^*$ submatrix of X corresponding to the selected coefficients S, where S^* denotes the cardinality of S. Then estimate by OLS

$$y_i = X_S \beta_{i,S} + \xi_i, \tag{8}$$

and store the residuals $\hat{\xi}_i = y_i - X_S \hat{\beta}_{i,S}$.

[4] Regress $\hat{\xi}_i$ onto the variables retained by the previous regularization steps plus the Granger causality variables $(X_{S \,\cup\, GC})$:

$$\hat{\xi}_i = X_{S \,\cup\, GC} \mathcal{B}_i + \nu_i, \tag{9}$$

and calculate the R^2 of (9).

[5a] Reject H_0 if $TR^2 > q_{\chi^2_p}(1-\alpha)$, where $q_{\chi^2_p}(1-\alpha)$ is the $1-\alpha$ quantile of the χ^2 distribution with p degrees of freedom.

[5b] Reject H_0 if $\left(\frac{T-S^*-p}{p}\right)\left(\frac{R^2}{1-R^2}\right) > q_{\mathcal{F}_{p,T-S^*-p}}(1-\alpha)$, where $q_{\mathcal{F}_{p,T-S^*-p}}(1-\alpha)$ is the $1-\alpha$ quantile of the $\mathcal{F}$ distribution with p and $T - S^* - p$ degrees of freedom .

The asymptotic χ^2 distribution of our test under the null hypothesis, and consequently the validity of our procedure, can be established from the results in Belloni et al. (2014a), by extending their arguments to allow for martingale difference error terms. Validity of our method follows in a similar way. Lasso estimation in time series, and particularly VAR models, is theoretically investigated in Basu and Michailidis (2015), Kock and Callot (2015) and Medeiros and Mendes (2016).[3] As discussed in these papers, conditions on sparsity, dependence and restricted eigenvalues, need to be satisfied for the lasso to consistently estimate the parameters. Our finite-order, exponentially decaying, stable VAR model satisfies the dependence conditions. Furthermore, the restricted eigenvalue condition can be shown to hold as well for "well-behaved" VAR processes that are not too high-dimensional and have innovations with sufficient moments existing; we refer to Basu and Michailidis (2015), Kock and Callot (2015) and Medeiros and Mendes (2016) for further details. For the macroeconomic application we consider in Section 5.1, these assumptions are likely to be satisfied after performing the necessary transformations. The key assumption remains sparsity of course, such that the lasso can reliably be applied. We investigate the sensitivity of our method to the degree of sparsity in our simulation study in Section 4.

[3] Chernozhukov et al. (2018) also prove asymptotic normality of parameter estimators from similar post-selection methods in models allowing for various kinds of temporal dependence. However, their inferential procedure is different from ours.

Remark 1. Given that we have $p+1$ steps of selection, it would be more appropriate to refer to our method as "post-$(p+1)$-selection" approach. For expositional simplicity however we stick to the post-double-selection name, as this is the common name for such a procedure, and conveys the essence of our method equally well.

Remark 2. Although we focus on the case where we have a single variable Granger causing and a single one being Granger caused, the method can also be extended to testing Granger causality of blocks of multiple variables. The step to multiple Granger causing variables is straightforward: one still just needs to estimate a single equation for variable i, but the composition of X_{GC} and accordingly $\beta_{i,GC}$ changes as X_{GC} will now contain the lags of multiple variables. The full algorithm goes through with those changes. The modifications for having multiple Granger causing variables is more involved, as in this case one should estimate the equations corresponding to the variables of interest jointly as a system. In particular, this involves re-parameterizing the system as a single equation (see e.g. Basu and Michailidis, 2015, p. 1556); once this is done and the matrix with Granger causing variables is properly defined accordingly, the algorithm goes through as described above. Alternatively one could perform the Granger causality test equation-by-equation, and then use a multiple testing procedure to correct the size of the overall test.

Remark 3. In Step 2 we propose not to consider the GC variables in the first regularization and insert them back at Step 4. Alternatively, the GC variable(s) can be left in the regression, such that, we regress on the full X matrix. In this case there are then two further possibilities by either penalizing these variables or not. Simulations for these two alternatives have been carried out and in practice we do not find significant differences among the three in terms of size and power. The approach proposed in Step 1 delivers the best results in terms of size.

Remark 4. When $T \approx Kp$, i.e. when the time series length approaches the number of covariates, information criteria and time series cross-validation tend to break down and select too many covariates in order to perform a post-selection by OLS. To overcome this issue we propose to place a lower bound on the penalty to ensure that in each relevant equation of the VAR at most $c\ T$ variables are selected, for some $0 < c < 1$. In our simulation and empirical studies we set $c = 0.5$. Although the lower bound is easy to implement for both ICs and TSCV, the computational time is thoroughly different, with TSCV taking more than twice the running time of a simple IC. Note that, as we have $p+1$ selection steps, the possibility remains that different variables are selected in each steps, making the number of variables in the union S still too large to perform the post-selection OLS, although this problem is likely to occur far less often. This can be addressed by ensuring that fewer than $T/(p+1)$ variables are selected in each selection step. We do not impose this stricter bound in general, as it will often be much too strict. Instead, we recommend to only address this issue if it arises in practice by an ad-hoc increase of the lower bound on the penalty.[4]

Remark 5. Although our Granger causality test has a χ^2 distribution under the null hypothesis asymptotically, in smaller samples the test might still suffer from the usual small-sample approximation error. As such we propose a finite-sample correction to the test in Step 5b, which in our simulation studies improved the size of our test.

Remark 6. Instead of the lasso or the adaptive lasso, one can use any algorithm that performs variable selection. In particular, the Elastic Net of Zou and Hastie (2005) that adds an ℓ_2-penalty in addition to the ℓ_1-penalty of the lasso, might be of interest. The additional penalty ensures that the Elastic Net is strictly convex, and as a consequence tends to select highly correlated variables as a group together, whereas the lasso would tend to select only one of these variables (Zou and Hastie, 2005). Given the typically strong correlations between many economic variables, this appears particularly useful for our context. However, we used the Elastic Net for both the simulations and the empirical application, and in both cases we found that the results are widely comparable to those of lasso. Therefore we chose to omit them from the paper; they are available upon request.

Remark 7. One can also perform a standard Wald test of Granger causality instead of the LM test, by regressing the variable of interest on the GC variables plus the variables selected in Step 2 ($X_{S \cup GC}$) and testing for the significance of the coefficients of X_{GC}. While asymptotically the LM and Wald tests behave equally, differences might arise in small samples. We investigated the Wald version of the test in simulations as well, with results reported in Appendix A, Table A.3. In general, differences between the two methods are negligible. However, for the Wald test, occasionally we run into the problem described in Remark 4, where even with the imposed lower bound on the penalty, too many variables are selected for performing a post-selection OLS. For this reason we prefer the LM version.

4. Monte-Carlo Simulations

4.1. Setup

We now evaluate the finite-sample performance of our proposed Granger causality test in a Monte-Carlo simulation study. We consider three Data Generating Processes (DGPs)

[4]Although it happens less often, the theoretical plug-in method for the tuning parameter occasionally also selects too many variables to make the post-OLS estimation infeasible. However, for this method no easy solution is available for bounding the penalty. One could increase the constant in the plug-in expression, thus strengthening the penalty, but this would be a rather ad-hoc adjustment. In particular, imposing the lower bound for the other methods only limits the allowed range of the tuning parameter, forcing the minimization to choose another (local) minimum that can still be far away from the boundary and justified graphically. For the plug-in method it is however difficult to justify the right amount of the increase, as the tuning parameter will be fixed to that value, and thus the chosen increase is rather arbitrary.

inspired by Kock and Callot (2015):

$$\text{DGP1:} \qquad y_t = \begin{bmatrix} 0.5 & 0 & \dots & 0 \\ 0 & 0.5 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0.5 \end{bmatrix} y_{t-1} + \epsilon_t, \tag{10}$$

$$\text{DGP2:} \qquad y_t = \begin{bmatrix} (-1)^{|i-j|}a^{|i-j|+1} & \dots & (-1)^{|i-j|}a^{|i-j|+1} \\ (-1)^{|i-j|}a^{|i-j|+1} & \dots & (-1)^{|i-j|}a^{|i-j|+1} \\ \vdots & \ddots & \vdots \\ (-1)^{|i-j|}a^{|i-j|+1} & \dots & (-1)^{|i-j|}a^{|i-j|+1} \end{bmatrix} y_{t-1} + \epsilon_t, \tag{11}$$

with $a = 0.4$.

$$\text{DGP3:} \qquad y_t = \begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A \end{bmatrix} y_{t-1} + \epsilon_t, \quad \text{with} \underbrace{A}_{5\times 5} = \begin{bmatrix} 0.15 & \cdots & 0.15 \\ \vdots & \ddots & \vdots \\ 0.15 & \cdots & 0.15 \end{bmatrix}. \tag{12}$$

DGP1 respects the sparsity assumption while in DGP2 the entries are set to decrease exponentially fast in the distance from the main diagonal and hence the sparsity assumption is not met. Finally DGP3 is a block-diagonal system. Note that as written above, DGP1 satisfies the null of no Granger causality from unit 2 to 1, while DGP2 and DGP3 do not. Therefore, we adapt DGP 1 for the power analysis by setting the coefficient in position $(2, 1)$ equal to 0.2. Conversely, we set the same coefficient equal to zero for DGP2 and DGP3 for the size analysis.

Following Section 2, we pick our time series of interest y_i and y_k with $i = 2$, $k = 1$. Here we consider for simplicity $p = 1$ lag, namely the same lag-length as in the DGPs, so $j = 1$. Then we have:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{K,t} \end{bmatrix} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,K} \\ \beta_{2,GC} & \beta_{2,2} & \cdots & \beta_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{K,1} & \beta_{K,2} & \cdots & \beta_{K,K} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{K,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{K,t} \end{bmatrix}.$$

Hence, for each DGP we test the hypothesis that $y_{1,t}$ does not Granger cause $y_{2,t}$:

$$H_0 : \beta_{2,GC} = 0 \quad \text{against} \quad H_1 : \beta_{2,GC} \neq 0$$

using our proposed PDS-LM test.

Table 1 reports the size and power of the test for 1000 replications by using different combinations of time series length $T = (50, 100, 200, 500)$ and number of variables in the system $K = (10, 20, 50, 100)$ and a fixed lag-length $p = 1$. All the rejection frequencies are reported using a burn-in period of fifty observations.

For each scenario, AIC, BIC and EBIC are compared with the theoretical choice of the tuning parameter λ^{th} and time series cross validation λ^{TSCV} as described in Subsection 2.3. To obtain $\hat{\sigma}$, the initial (conservative) estimate of the standard deviation of ϵ_i, the least squares estimator is run between y_i and the five most correlated regressors implying conservative starting values for λ and its loadings. This estimate is updated iteratively (see Belloni et al., 2012, for details).

Simulations are also reported for different types of covariance matrices of the error terms. We employ a Toepliz-version for calculating the covariance matrix as $\Sigma_{i,j} = \rho^{|i-j|}$ by using two scenarios of correlation: $\rho = (0, 0.7)$. The first no-correlation is equivalent to set $\Sigma_{i,j} = I_{i,j}$, where I is the identity matrix.

In the Appendix we provide some additional simulation results. First, we investigate the Wald version of our test in Table A.3. Second, in Table A.4 we investigate the effects of miss-specification of the lag length by estimating the over-specified VAR($p+1$) instead of the true-order VAR(p). Third, in Table A.5 we report the results for the size of a bivariate Granger causality test for a non-sparse DGP when using a standard Wald (F) test. This test is obviously sensitive to omitted variable bias, and our goal is to demonstrate its effect. Finally, although all results reported here use the finite sample correction in Step 5b of the algorithm, we also investigated the differences with Step 5a. We comment on these results in the next subsection. All results not reported in this paper are available from the authors upon request.

Table 1: Simulation results for the PDS-LM Granger causality test

DGP	Size/Power	ρ	T	50					100					200					500				
			K	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}
			10	6.7	6.6	7.6	6.8	6.7	6.7	6.1	6.0	7.0	6.5	5.5	4.2	4.5	5.2	5.2	4.3	4.1	4.3	4.2	4.3
1	Size	0	20	7.4	6.2	8.3	6.4	6.1	6.0	4.9	5.3	6.1	5.5	4.0	4.7	5.4	4.0	4.3	4.3	4.1	4.3	4.1	4.7
			50	7.1	6.0	8.8	6.6	5.5	7.4	7.0	7.3	7.3	6.3	6.9	5.9	6.0	7.4	7.7	6.8	6.5	6.7	6.5	6.5
			100	6.8	6.9	10.8	NA	4.5	7.1	5.4	7.3	6.1	4.3	6.5	4.7	4.9	5.7	5.8	5.8	3.9	4.0	5.2	5.7
			10	30.1	30.7	33.5	30.7	30.3	58.3	58.9	60.8	58.6	56.8	89.1	89.3	89.6	88.8	89.4	99.9	99.9	99.9	99.9	99.9
1	Power	0	20	22.9	28.8	33.5	24.8	22.4	52.8	55.1	57.7	54.0	51.6	85.6	88.0	89.0	85.3	84.1	99.9	100	100	99.9	99.8
			50	13.5	25.8	36.9	17.2	11.0	38.7	53.8	59.0	44.7	35.7	78.2	85.8	87.0	80.2	75.1	99.9	99.9	99.9	99.9	99.8
			100	9.1	21.3	35.7	NA	7.8	20.4	52.4	59.0	29.7	17.4	61.6	85.4	86.7	70.0	54.8	99.5	100	100	99.8	98.9
			10	6.2	6.4	8.2	5.7	6.0	5.0	5.0	6.3	5.2	5.1	4.7	4.6	4.9	4.7	5.1	3.9	3.9	3.9	3.8	3.9
2	Size	0	20	7.7	6.3	8.9	6.9	7.0	5.4	5.4	6.7	5.5	5.4	5.8	5.9	5.4	6.1	5.9	4.7	3.8	4.3	4.4	4.8
			50	7.0	6.5	9.4	6.5	6.7	8.0	6.7	10.3	8.2	5.7	7.3	6.2	6.7	7.9	8.6	6.3	6.6	6.4	6.7	6.5
			100	7.5	7.9	10.8	NA	4.8	6.5	6.3	12.4	6.7	4.3	5.4	5.8	7.2	4.6	4.7	6.1	4.3	5.0	6.4	4.7
			10	18.1	20.1	24.5	18.5	16.7	37.9	39.9	42.2	38.0	37.1	64.7	64.7	66.9	64.8	64.4	97.4	97.4	97.5	97.4	97.4
2	Power	0	20	16.3	20.6	31.5	17.7	14.1	35.3	40.5	47.0	35.8	33.6	64.4	67.5	69.8	64.7	63.6	97.2	97.5	97.5	97.5	97.0
			50	9.5	17.2	31.3	11.7	6.7	24.8	37.5	49.3	29.0	19.4	57.0	66.5	72.1	60.2	54.1	95.0	96.2	96.8	95.2	94.5
			100	8.7	15.8	38.6	NA	4.9	15.5	36.8	53.8	20.3	11.0	44.8	65.2	75.3	52.0	38.0	94.7	97.3	97.7	95.0	92.7
			10	5.3	5.0	7.3	5.3	4.8	5.7	6.0	8.9	5.6	5.4	3.9	4.8	8.6	4.3	3.8	4.1	4.0	3.9	3.9	4.1
3	Size	0	20	4.2	5.7	7.4	5.5	4.7	4.6	5.7	11.0	4.3	4.2	5.2	6.6	13.3	4.9	4.8	4.7	4.4	4.2	4.8	4.6
			50	7.5	7.3	10.7	6.4	6.4	6.6	8.1	13.7	6.0	7.3	6.9	8.7	18.2	6.5	6.5	4.9	5.3	7.8	5.2	5.0
			100	6.9	7.2	10.6	NA	5.7	6.3	8.7	13.9	5.7	5.3	4.7	8.3	18.9	3.9	4.7	4.4	5.5	8.8	4.5	5.4
			10	15.2	21.3	28.7	16.2	14.9	31.6	38.5	49.2	32.2	30.4	58.4	61.7	65.3	58.2	58.0	95.2	95.6	95.7	95.5	95.2
3	Power	0	20	13.3	21.4	32.4	13.8	11.2	29.5	40.5	56.8	29.8	26.7	56.5	63.5	73.8	56.5	55.6	94.0	94.6	94.8	94.2	94.1
			50	11.8	23.3	35.2	12.1	9.7	24.3	43.7	59.4	26.1	20.0	50.5	62.1	77.6	49.8	48.1	91.1	92.8	93.6	91.8	90.5
			100	9.3	21.3	36.1	NA	8.1	14.0	47.2	63.8	17.9	11.9	41.4	64.2	79.6	44.5	32.3	90.0	94.0	95.0	90.2	87.1
			10	5.8	5.1	5.4	5.5	5.8	5.8	5.7	5.2	5.7	5.2	5.1	4.6	4.6	4.9	4.6	4.3	4.0	4.2	4.4	4.3
1	Size	0.7	20	7.3	6.3	5.8	5.8	6.4	6.0	6.1	6.2	5.9	5.4	5.1	5.5	5.5	5.5	4.9	5.0	4.9	4.9	4.5	5.3
			50	7.3	5.6	7.8	7.5	5.7	6.4	7.3	5.7	6.8	5.8	6.5	4.9	4.2	5.8	6.5	5.7	4.8	4.8	5.2	5.7
			100	7.2	7.8	10.1	NA	5.3	8.0	6.4	6.3	7.8	6.3	6.9	5.8	4.4	4.8	6.3	5.6	3.6	3.6	5.3	6.5
			10	19.0	18.2	18.9	19.6	17.6	33.7	34.2	34.3	34.3	33.2	56.6	57.1	56.8	57.4	54.9	93.7	94.2	94.2	94.3	93.8
1	Power	0.7	20	15.3	18.3	19.5	19.2	14.1	29.2	30.7	31.6	30.0	27.3	54.4	56.8	56.6	56.1	52.8	92.8	94.1	94.1	93.5	92.6
			50	10.4	14.1	21.0	14.2	9.3	24.0	30.4	32.7	30.5	20.0	50.1	57.1	58.8	55.8	46.2	90.7	92.0	92.5	91.8	89.7
			100	9.1	14.2	23.5	NA	8.1	13.2	29.2	33.6	22.8	10.8	34.1	53.6	55.7	47.3	27.2	88.1	94.0	94.5	91.0	62.0

Notes: Size and Power for the different DGPs described in Section 4.1 are reported for 1000 replications. $T = (50, 100, 200, 500)$ is the time series length, $K = (10, 20, 50, 100)$ the number of variables in the system, the lag-length is fixed to $p = 1$. ρ indicates the correlation employed to simulate the time series with the Toeplitz covariance matrix.

4.2. Results

Our proposed approach shows a good performance in terms of size and (unadjusted) power for all DGPs considered. Both for the setting of no correlation and high correlation of errors, sizes are in the vicinity of 5% and power is increasing with the sample size T.

Only moderate size distortion is visible in large systems for small samples (e.g. $K \geq 50$, $T = 50$). As expected, the test procedure works remarkably well for the sparse DGP1 in high dimensions. However, size properties under the non-sparse DGP2 do not deviate much from its sparse counterpart, although for both DGP2 and DGP3 we do observe a slight deterioration of size when the dimension of the system increases.

Interestingly, the three different information criteria show substantially different behavior. EBIC, due to its very stringent nature, tends to perform well only in very large systems, while it is comparatively worse than BIC and even AIC in small samples. We have to add though that the good performance of AIC in particular is somewhat inflated by the imposed lower bound on the penalty; unreported simulations show that without the lower bound AIC performs significantly worse, often selecting too many variables rendering the post-OLS estimation infeasible. The one advantage of using EBIC as information criterion to tune λ in the $K >> T$ settings when T is small (e.g. $T = 50, 100$) is the possibility to avoid the lower bound on the penalty. However, since this comes at a price of more size distortion, we recommend the use of BIC instead, along with the $0.5T$ lower bound on the penalty.

Comparing our test to the bivariate VAR in Table A.5, it is clear that our proposed PDS-LM is very robust to omitted variable bias, unlike the bivariate test, whose size distortions increase with both the sample size and the number of variables, with sizes of 45% observed for the sample sizes we consider in our application in 5.2. We will further elaborate on this difference in our empirical applications in Section 5. The results of robustness to misspecification of the lag length order: $p = 2$ instead of $p = 1$, are reported in Table A.3 in Appendix A. As the size distortions across the range of considered DGPs are only marginally higher for large K and T comparatively small, the test appears to be quite robust to this misspecification. Again, BIC seems to be the best choice for tuning the penalty for all DGPs. Unreported simulations (available upon request) further show that the finite sample adjustment for the test performed in Step 5b of the algorithm is able to substantially reduce size distortions in small samples compared to the asymptotic version of Step 5a.

5. Empirical Applications

5.1. Money-Income Causality

In this section, we investigate Granger causality between *Real Money* (M1) and *Real GDP* (GDP). Whether real output is driven by changes in real money supply is a common controversial debate in empirical macroeconomics and macroeconometric literature. Sims (1972) argues that a causal relationship occurs between money and GDP. Nevertheless, empirical findings show that GDP does not feed back in forecasting money. We now investigate this relation in a high-dimensional VAR, allowing for interactions with many variables to be able to isolate the effect of money on GDP and vice versa. Our approach avoids that omitted variable bias distorts the findings. We use the FRED-QD, a quarterly databases for Macroeconomic Research by the Federal Reserve Bank of St. Louis.[5] The time span of the data we have used goes from third quarter of 1959 until the third quarter of 2015 for a total of 257 variables. After some necessary cleaning of the dataset we are left with $T = 222$ quarters and $K = 204$ macroeconomic indicators. In the HD-VAR, we use $p = 2$ lags of each variable.

We compare our testing procedure to a simple bivariate VAR (*2-VAR*) model

$$\begin{bmatrix} \Delta \log(M1)_t \\ \Delta \log(GDP)_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} \end{bmatrix} \begin{bmatrix} \Delta \log(M1)_{t-j} \\ \Delta \log(GDP)_{t-j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \tag{13}$$

by testing the lower triangularity of the coefficient matrices. We also compare it to a four-variate VAR (*4-VAR*) model

$$\begin{bmatrix} \Delta \log(M1)_t \\ \Delta \log(GDP)_t \\ \Delta(TB3MS)_t \\ \Delta^2 \log(CPI)_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} & \phi_{13}^{(j)} & \phi_{14}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} & \phi_{23}^{(j)} & \phi_{24}^{(j)} \\ \phi_{31}^{(j)} & \phi_{32}^{(j)} & \phi_{33}^{(j)} & \phi_{34}^{(j)} \\ \phi_{41}^{(j)} & \phi_{42}^{(j)} & \phi_{43}^{(j)} & \phi_{44}^{(j)} \end{bmatrix} \begin{bmatrix} \Delta \log(M1)_{t-j} \\ \Delta \log(GDP)_{t-j} \\ \Delta(TB3MS)_{t-j} \\ \Delta^2 \log(CPI)_{t-j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{bmatrix}, \tag{14}$$

where $\Delta(TB3MS)$ is the variation of nominal interest rates (3-months treasury bill) and $\Delta^2 \log(CPI)$ the variation of inflation. Finally, we compare our method with the factor-augmented VAR (FAVAR) as introduced by Bernanke et al. (2005). Following McCracken and Ng (2016), we estimate the static factors with PCA from the original (standardized) dataset, excluding the series of GDP and M1. The significant factors are selected by means of the $PC_p = \frac{\log(\min(K,T))}{\min(K,T)}$ criterion of Bai and Ng (2002). We obtain a total of eight factors $(F_1, \dots, F_8)$; the top three series most correlated with each factor are reported

[5] A detailed description of the variables in the dataset and the relative transformations performed to achieve stationarity of each series is available on the FRED-QD website (https://research.stlouisfed.org/econ/mccracken/fred-databases/). R codes are available on the GitHub page of the corresponding author (https://github.com/Marga8).

Table 2: Money-Income Causality

$\overrightarrow{Cause}$	λ	PDS-LM	S^*	2-VAR	p	4-VAR	p	FAVAR	p
M1→GDP	AIC	**.6808**	217	**.0460**	2	**.2355**	5	**.0827**	2
	BIC	**.0283**	16	**.0701**	1	**.0661**	1	**.7620**	1
	EBIC	**.0258**	14						
	TH	**.0455**	88						
	TSCV	**.6172**	126						
GDP→M1	AIC	**.4206**	137	**.0780**	2	**.8145**	5	**.2131**	2
	BIC	**.1744**	35	**.0690**	1	**.9206**	1	**.0029**	1
	EBIC	**.7030**	21						
	TH	**.5401**	69						
	TSCV	**.6094**	112						

Notes: The λ column indicates the method used to set the tuning parameter. The PDS-LM column reports the p-values of the Granger causality tests when using the PDS-LM method. The column S^* shows the cardinality of the set S, that is, the number of coefficients selected by the lasso. The 2-VAR, 4-VAR and FAVAR columns report the p-values of the test applied to the bivariate, four-variate and factor-augmented VARs respectively, while the columns denoted by p report the number of lags selected with the relevant information criteria for these VARs.

in the Appendix B. Hence, we test for Granger causality using the FAVAR model

$$\begin{bmatrix} \Delta\log(GDP)_t \\ \Delta\log(M1)_t \\ F_{1,t} \\ F_{2,t} \\ \vdots \\ F_{8,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ \vdots \\ c_{10} \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \phi_{1,1}^{(j)} & \phi_{1,2}^{(j)} & \cdots & \phi_{1,10}^{(j)} \\ \phi_{2,1}^{(j)} & \phi_{2,2}^{(j)} & \cdots & \phi_{2,10}^{(j)} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{10,1}^{(j)} & \phi_{10,2}^{(j)} & \cdots & \phi_{10,10}^{(j)} \end{bmatrix} \begin{bmatrix} \Delta\log(GDP)_{t-j} \\ \Delta\log(M1)_{t-j} \\ F_{1,t-j} \\ F_{2,t-j} \\ \vdots \\ F_{8,t-j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \vdots \\ \epsilon_{10,t} \end{bmatrix}. \quad (15)$$

For 2-VAR, 4-VAR and FAVAR models we select the lag-length via AIC or BIC. We exclude EBIC since BIC is already parsimonious enough for selecting lags in low-dimensional VARs.

Table 2 reports the results from the Granger causality tests. We find that our PDS-LM method provides strong evidence of Granger causality for real money to real output when the lasso is appropriately tuned with (E)BIC or the theoretical plug-in method. When AIC or time series cross-validation is used, too many variables are selected, resulting in a counter-intuitive high p-value.[6] Granger causality in the opposite direction is instead always clearly rejected.

The two small VAR models provide conflicting evidence on Granger causality from money to GDP, as well as the other direction. The 2-VAR shows a weak significance in the $M1 \rightarrow GDP$ causal relation, yet the 4-VAR is very far from rejecting the null

[6] In order to obtain the p-value for the PDS-LM test with AIC we had to adjust the penalty lower bound to $0.5(T) - 4$.

hypothesis, although the difference between the two VAR models might here be attributed to differences in lag length selection. In the opposite direction, the differences between the two small VAR models are even much larger, with the 2-VAR having p-values below 0.1 and the 4-VAR above 0.8, regardless of lag length selection.

The factor-augmented VAR appears to be very sensitive to the selected lag length. Adding a single lag from $p = 1$ to $p = 2$ decreases the p-value for Granger causality from M1 to GDP by nearly 70%, while in the other direction an increase is observed from highly significant to decisively not significant. It appears that, even though BIC selects one lag, this is insufficient to capture enough dynamics. We also observed that when manually increasing the lag length to four the p-values appear to remain stable, and are qualitatively similar as those of the appropriately tuned PDS-LM test. However, we have not investigated the estimation of the number of factors, which is notoriously difficult, instead going with the established choice of McCracken and Ng (2016). It is likely that uncertainty about the number of factors will further increase the variability of the outcome of the test using the FAVAR.

To wrap up, our results show empirically that by increasing the information set by considering a high-dimensional VAR model, and thus allowing for the potential interaction of many other indicators, one is able to reduce the effect of the omitted variables and thereby retrieve a clearer picture of the causal relations between money and outcome.

5.2. Networks in Realized Volatilities

For our second application we investigate Granger causality between volatilities of stock returns using the dataset used in Hecq et al. (2016). Stock prices are obtained from TickData and consists of daily transaction prices for 49 large capitalization stocks from the NYSE, AMEX and NASDAQ. The period we investigate is from January 1, 1999 to December 31, 2007 (2239 trading days). Table B.7 provides a list of ticker symbols and company names. Prices have been sampled at 5-minute frequency using the interpolation method of Barndorff-Nielsen et al. (2009). For each series we consider the Median Truncated Realized Variance ($MedRV$), as an estimator of the realized variance that is robust to jumps:

$$MedRV_t \equiv \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M-2} \right) \sum_{j=2}^{M-1} \text{med} \left(|r_{t,j}| |r_{t,j-1}| |r_{t,j+1}| \right)^2, \tag{16}$$

where $r_{t,j}$ are the high frequency intra-day returns, observed for M intra-day 5-min periods considered each day.

Given the time series of realized volatilities as defined in (16), we employ a multivariate version of the heterogeneous autoregressive model (VHAR) of Corsi (2009) to model their joint behavior. To formally define the VHAR model, we log-transform the series and we stack the logarithmic MedRV into a vector $y_{i,t}$. The VHAR specification is given by the following model:

$$\begin{bmatrix} y_{1,t}^{(day)} \\ y_{2,t}^{(day)} \\ \vdots \\ y_{K,t}^{(day)} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{bmatrix} + \begin{bmatrix} \beta_{1,1}^{(1)} & \beta_{1,2}^{(1)} & \cdots & \beta_{1,K}^{(1)} \\ \beta_{2,1}^{(1)} & \beta_{2,2}^{(1)} & \cdots & \beta_{2,K}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{K,1}^{(1)} & \beta_{K,2}^{(1)} & \cdots & \beta_{K,K}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1,t-1}^{(day)} \\ y_{2,t-1}^{(day)} \\ \vdots \\ y_{K,t-1}^{(day)} \end{bmatrix} + \begin{bmatrix} \beta_{1,1}^{(2)} & \beta_{1,2}^{(2)} & \cdots & \beta_{1,K}^{(2)} \\ \beta_{2,1}^{(2)} & \beta_{2,2}^{(2)} & \cdots & \beta_{2,K}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{K,1}^{(2)} & \beta_{K,2}^{(2)} & \cdots & \beta_{K,K}^{(2)} \end{bmatrix} \begin{bmatrix} y_{1,t-1}^{(week)} \\ y_{2,t-1}^{(week)} \\ \vdots \\ y_{K,t-1}^{(week)} \end{bmatrix} +$$
$$+ \begin{bmatrix} \beta_{1,1}^{(3)} & \beta_{1,2}^{(3)} & \cdots & \beta_{1,K}^{(3)} \\ \beta_{2,1}^{(3)} & \beta_{2,2}^{(3)} & \cdots & \beta_{2,K}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{K,1}^{(3)} & \beta_{K,2}^{(3)} & \cdots & \beta_{K,K}^{(3)} \end{bmatrix} \begin{bmatrix} y_{1,t-1}^{(month)} \\ y_{2,t-1}^{(month)} \\ \vdots \\ y_{K,t-1}^{(month)} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{K,t} \end{bmatrix}$$

where

$$y_{i,t}^{(week)} := \frac{1}{5}\sum_{j=0}^{4} y_{i,t-j}^{(day)} \qquad \text{and} \qquad y_{i,t}^{(month)} := \frac{1}{22}\sum_{j=0}^{21} y_{i,t-j}^{(day)}$$

are the vectors containing the average volatility over the last 5 (week) and 22 (month) days. Granger causality in this context represents contagion, or spillover, of volatility from one asset to another. To test for the null hypothesis of no Granger causality / no volatility spillovers from $y_{k,t}$ to $y_{i,t}$ against the alternative of spillovers, we test

$$H_0 : \beta_{i,k}^{(1)} = \beta_{i,k}^{(2)} = \beta_{i,k}^{(3)} = 0 \qquad \text{vs.} \qquad H_1 : \beta_{i,k}^{(1)}, \beta_{i,k}^{(2)}, \beta_{i,k}^{(3)} \neq 0.$$

We perform this test for every (i, k)-pair to obtain the full 48×48 network of spillover effects.

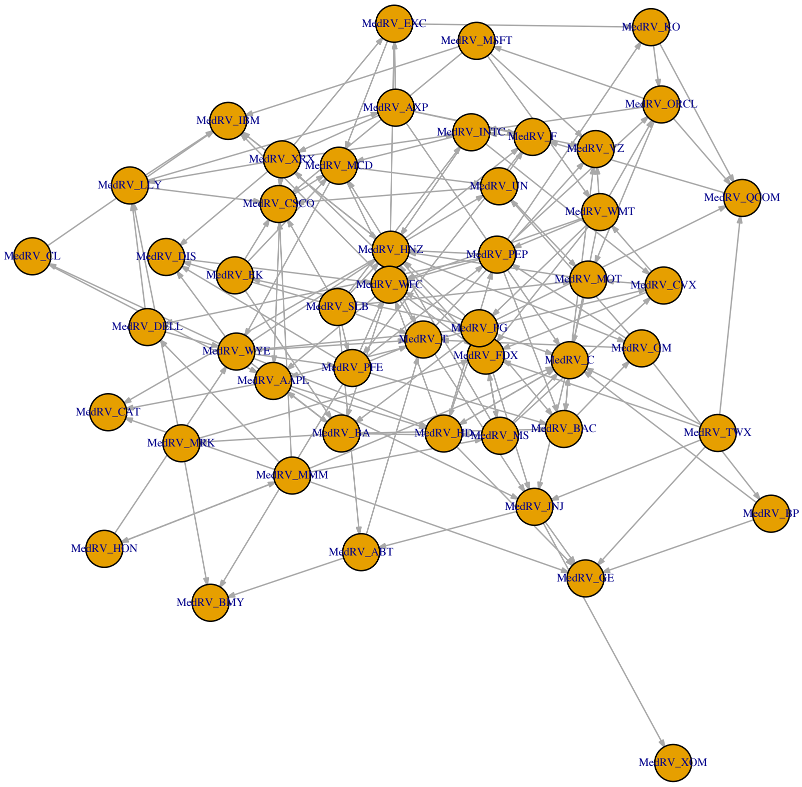
As heteroskedasticity is likely present in these data, we robustify the PDS-LM procedure by implementing the heteroskedasticity-robust LM test such as for example described in Wooldridge (2015, Ch. 8). The full algorithm for the heteroskedasticity-robust PDS-LM test is given in Appendix B.[7]

We now report the results of our spillover tests for the volatility network. We use BIC to select the tuning parameter of the lasso, and perform the Granger causality tests with a 1% significance level.[8] Figure 1 reports the contagion network of volatilities estimated with the high-dimensional VHAR. Maintaining the VHAR modeling structure of realized volatilities, we compare our post-double selection method with (heteroskedasticity-robust)

[7] In the presence of heteroskedasticity, one might prefer the Wald version of the test, as this can be corrected in the standard way by using heteroskedasticty-robust standard errors. Empirically we found hardly any differences between the LM and Wald versions.

[8] We do not perform a correction for multiple testing, as this would only qualitatively affect our results. Moreover, our goal is not to identify exactly the set of spillovers, but to get a feeling of the relations between two variables at a time. As such, we believe a multiple testing correction is not needed, though it can be easily implemented.

Figure 1: Volatility Contagion Network



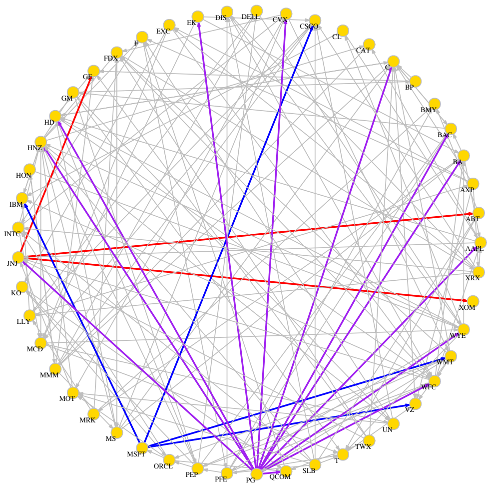


Figure 2: PDS-LM VHAR spillover network

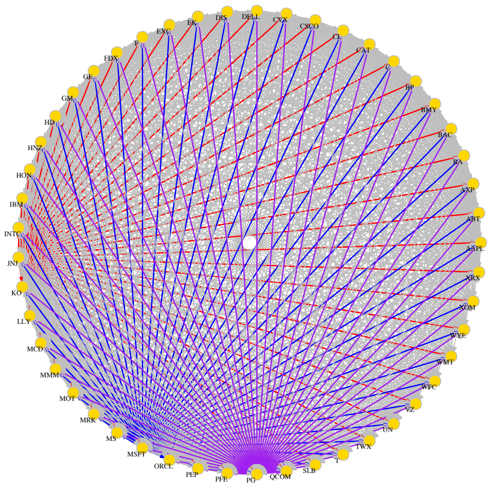


Figure 3: Bivariate VHAR spillover network

bivariate Granger causality tests for each pair of stocks.

Our PDS-LM method identifies a volatility contagion network which consists of 164 connections, while the bivariate tests detect a network consisting of 2082 connections. In Figures 2 and 3 we highlight the contagion directions of Johnson & Johnson, Microsoft and Procter & Gamble. It is evident from these figures how the use of our PDS-LM VHAR instead of a bivariate VHAR is able to clear the picture of the spillover network. The highlighted stocks have respectively 47, 48 and 48 connections out of 48 when using the bivariate VHAR. However, when using our PDS-LM VHAR, 3, 4, 11 are respectively detected. These results are in line with our simulation results, confirming that bivariate Granger causality testing in VAR models is seriously affected by omitted variable bias in high-dimensional systems.

As a next step, we use our identified networks to find clusters of closely connected stocks, or communities as they are called in graph theory. Communities are groups of densely connected nodes with fewer connections across groups. In order to represent volatility spillover communities in the graph we make use the Newman and Girvan (2004) algorithm based on edge-betweenness. The edge betweenness for edge e is defined as

$$\sum_{s,t \neq e} \frac{\sigma_{st}(e)}{\sigma_{st}} \tag{17}$$

where σ_{st} is total number of shortest paths from node s to node t and $\sigma_{st}(e)$ is the number of shortest paths passing through e. The edge with the highest betweenness is sequentially removed and the betweenness is recalculated at each step until the best partitioning of the

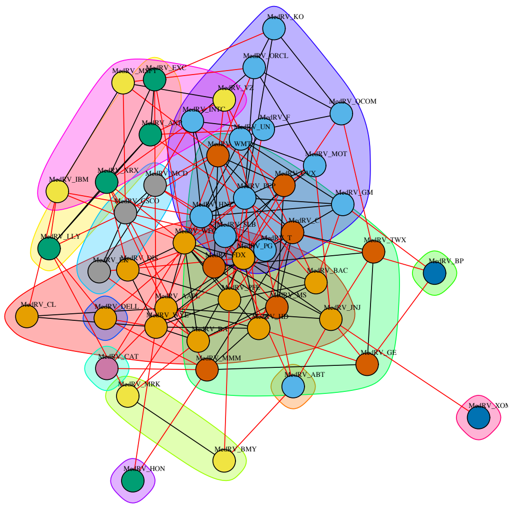


Figure 4: PDS-LM VHAR volatility clusters

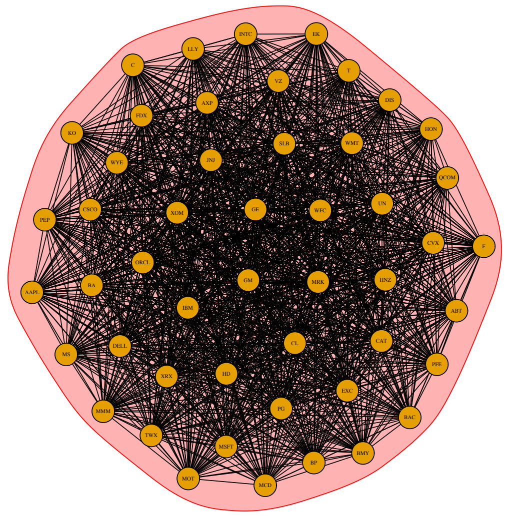


Figure 5: Bivariate VHAR volatility clusters

network is found. Figures 4 and 5 report the graphs of the clustered network respectively for our proposed PDS-LM tests and the standard bivariate tests.

The result for the PDS-LM method highlights three main clusters of volatility spillovers, respectively containing 12,11 and 8 nodes and other 10 smaller clusters containing between 1 and 4 nodes. We can recognize several patterns in the different clusters in terms of the sectors to which the relevant stocks belong. The largest cluster contains telecommunication/informatics stocks like Qualcomm, Oracle Corp., Motorola, Intel Corp and AT&T Corp. but also companies producing capital goods as food and beverages like Coca Cola Co., Pepsico Inc., Unilever N.V., H J Heinz Co. and capital goods as motors like Ford Motor Co. and General Motors. The second largest cluster comprises many health care providers as well as consumer services of cosmetic types like Colgate Palmolive Co., Pfizer Inc, Johnson & Johnson and Procter & Gamble along with a few financial and industry stocks like Bank of America, Wells Fargo & Co., Boeing Co. Finally, the third largest cluster contains financial as well as energy stocks like Morgan Stanley, Citigroup, General Elec. and Chevron Corp.

Instead, when using the bivariate VHAR method only one cluster is identified. Although this result was expected from our previous analysis of the spillover network, this picture reinforces our claim that using bivariate Granger causality testing to identify spillover networks, as well as volatility clusters, is not informative in high-dimensional systems.

6. Conclusion

We propose an LM test in order to test for Granger causality in high-dimensional VAR models. We employ a post-double selection procedure using the lasso to select the set of relevant covariates in the system. The double selection step allows to substantially reduce the omitted variable bias and thereby allowing for valid post-selection inference on the parameters.

We provide an extensive simulation study to evaluate the performance of our method in finite samples, paying particular attention to the tuning of the penalty parameter. We compare different information criteria, time series cross-validation and a plug-in method based on theoretical arguments, and find that generally BIC and the theoretically tuned penalty perform best. However, to use information criteria in systems with a significantly larger number of variables than observations, a lower bound on the penalty parameter is needed to prevent too many variables being selected.

The simulations also show that, when properly tuned, our proposed PDS-LM test attains good results both for size and power under different DGPs. Especially, it is shown to be robust both to non-sparse settings as well as to lag-length overspecification.

We also empirically investigate usefulness of our method in two applications. First, we investigate the causal, dynamic relation between real money (M1) and output (GDP) using the FRED-QD macroeconomic dataset in comparison to a standard bivariate VAR, a four-variate VAR with also nominal interest rate and inflation, and a factor-augmented VAR (FAVAR). We find that the low-dimensional methods often yield conflicting results, whereas the results of the high-dimensional methods are more in line with economic theory. In addition, the PDS-LM method appears less sensitive to specification issues than the FAVAR.

In a second empirical study we apply our PDS-LM method to a high-dimensional VHAR process in order to construct a contagion network of volatility spillovers for 49 large capital stocks. We compare our method with standard bivariate Granger-causality tests and derive clusters of volatility contagion via the edge betweenness algorithm. Again we find that by increasing the information set through considering a high-dimensional VAR model in the estimation, we are able to obtain more realistic effects than in the low-dimensional models.

Finally, we mention some limitations of the present method, and discuss room for extensions. First, although theoretical results for all the "components" of our method are available, and the finite-sample performance of our method gives little reason to doubt its validity, we do not provide a formal proof of uniform validity of our approach. Given the focus of this paper on finite-sample properties, such a proof is outside the scope of this paper, though clearly of great interest. Second, note that unlike Belloni et al. (2014a), we do not give a "truly" causal interpretation to the established Granger causalities. In how far Granger causality is a useful concept to study true causality is (and has long been) open to debate, see for example (Eichler, 2013) and the references therein. Moreover, though it appears desirable and in line with Granger's 1969 original intentions to make the information set as large as possible, it is well known in the literature on graphical models (see Eichler, 2013) for causality that considering only the full model is not sufficient for establishing true causal relations from Granger causal ones. However, the analysis of the full model is a necessary element for a study of causality in a graphical framework. It would therefore be an interesting avenue for further research to study how the method proposed here could fit into such a graphical framework.

References

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics*, 43(4):1535–1567.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Belloni, A., Chernozhukov, V., and Hansen, C. (2011a). Inference for high-dimensional sparse econometric models. arXiv e-print 1201.0220.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni, A., Chernozhukov, V., and Kato, K. (2014b). Uniform post-selection inference for least absolute deviation regression and other z-estimation problems. *Biometrika*, 102(1):77–94.

Belloni, A., Chernozhukov, V., and Wang, L. (2011b). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806.

Bergmeir, C., Hyndman, R. J., Koo, B., et al. (2015). A note on the validity of cross-validation for evaluating time series prediction. Monash University Department of Econometrics and Business Statistics Working Paper 10.

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013). Valid post-selection inference. *Annals of Statistics*, 41(2):802–837.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120(1):387–422.

Bickel, P. J., Ritov, Y., Tsybakov, A. B., et al. (2009). Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37(4):1705–1732.

Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771.

Chen, J. and Chen, Z. (2012). Extended bic for small-n-large-p sparse glm. *Statistica Sinica*, pages 555–574.

Chernozhukov, V., Hansen, C., and Spindler, M. (2016). High-dimensional metrics in R. arXiv e-print 1603.01700.

Chernozhukov, V., Härdle, W. K., Huang, C., and Wang, W. (2018). Lasso-driven inference in time and space. arXiv e-print 1806.05081.

Chetverikov, D., Liao, Z., and Chernozhukov, V. (2016). On cross-validated lasso. arXiv e-print 1605.02214.

Chevillon, G., Hecq, A., and Laurent, S. (2018). Generating univariate fractional integration within a large var (1). *Journal of Econometrics*, 204(1):54–65.

Chudik, A. and Pesaran, M. H. (2016). Theory and practice of gvar modelling. *Journal of Economic Surveys*, 30(1):165–197.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Cubadda, G., Hecq, A., and Palm, F. C. (2009). Studying co-movements in large multivariate data prior to multivariate modelling. *Journal of Econometrics*, 148(1):25–35.

Davis, R. A., Zang, P., and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, 25(4):1077–1096.

Eichler, M. (2013). Causal inference with multiple time series: principles and problems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371:20110613.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, pages 424–438.

Hecq, A., Laurent, S., and Palm, F. C. (2016). On the univariate representation of bekk models with common factors. *Journal of Time Series Econometrics*, 8(2):91–113.

Jing, B.-Y., Shao, Q.-M., Wang, Q., et al. (2003). Self-normalized cramér-type large deviations for independent random variables. *Annals of Probability*, 31(4):2167–2215.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186(2):325–344.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *Annals of Statistics*, 44(3):907–927.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21:21–59.

Loh, P.-L. and Wainwright, M. J. (2011). High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. In *Advances in Neural Information Processing Systems*, pages 2726–2734.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Medeiros, M. C. and Mendes, E. F. (2016). ℓ_1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *Journal of Econometrics*, 191(1):255–271.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Nicholson, W., Wilms, I., Bien, J., and Matteson, D. (2018). High dimensional forecasting via interpretable vector autoregression. arXiv e-print 1412.5250.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review*, pages 540–552.

Skripnikov, A. and Michailidis, G. (2018). Joint estimation of multiple network granger causal models. *Econometrics and Statistics*.

Song, S. and Bickel, P. J. (2011). Large vector auto regressions. arXiv e-print 1106.3915.

Song, X. and Taamouti, A. (2017). A better understanding of granger causality analysis: A big data environment.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society - Series B*, pages 267–288.

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics*, 42(3):1166–1202.

van de Geer, S., Bühlmann, P., and Zhou, S. (2011). The adaptive and the thresholded lasso for potentially misspecified models (and a lower bound for the lasso). *Electronic Journal of Statistics*, 5:688–749.

Wilms, I., Gelper, S., and Croux, C. (2016). The predictive power of the business and bank sentiment of firms: A high-dimensional Granger causality approach. *European Journal of Operational Research*, 254(1):138–147.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.

Zellner, A. and Palm, F. C. (1974). Time series analysis and simultaneous equation econometric models. *Journal of Econometrics*, 2(1):17–54.

Zellner, A. and Palm, F. C. (1975). Time series and structural analysis of monetary models of the US economy. *Sankhya: The Indian Journal of Statistics, Series C*, 37:12–56.

Zellner, A. and Palm, F. C. (2004). *The Structural Econometric Time Series Analysis Approach*. Cambridge University Press.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.

Appendix A. Additional Simulation Results

Table A.3: Simulation results for the PDS-WALD Granger causality test

DGP	Size/Power	ρ	T	50					100					200					500				
			K	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}
			10	7.2	7.0	6.3	7.1	7.1	7.0	6.3	5.8	7.2	6.6	5.5	4.3	4.6	5.3	5.3	4.3	4.1	4.3	4.2	4.3
1	Size	0	20	8.2	6.5	6.8	6.8	6.7	6.2	5.1	5.3	6.6	5.9	4.2	4.8	5.5	4.0	4.3	4.3	4.1	4.3	4.2	4.7
			50	8.2	6.6	8.7	7.2	6.1	7.6	6.8	7.3	7.3	6.6	7.1	5.9	5.8	7.4	7.9	6.8	6.5	6.7	6.5	6.6
			100	7.8	7.8	9.2	NA	5.5	7.5	5.2	4.3	6.5	4.6	6.6	4.7	5.0	5.8	5.9	5.8	3.9	4.0	5.2	5.8
			10	31.3	31.3	32.8	31.6	31.1	58.6	59.5	61.7	58.9	57.5	89.1	89.3	89.6	88.8	89.5	99.9	99.9	99.9	99.9	99.9
1	Power	0	20	23.8	28.0	33.3	26.3	23.4	53.4	56.0	59.9	54.5	52.2	85.6	88.2	89.2	84.4	87.1	99.9	100	100	99.9	99.8
			50	15.7	24.4	38.4	18.1	12.6	39.9	54.5	61.2	45.2	36.2	78.8	86.1	87.4	80.3	75.5	99.9	99.9	99.9	99.9	99.8
			100	12.2	21.0	38.5	NA	9.8	21.2	52.4	59.0	29.7	18.5	61.7	85.8	87.2	70.1	56.1	99.5	100	100	99.8	98.9
			10	6.8	6.7	7.6	6.2	6.4	5.2	4.7	5.7	5.5	5.1	5.1	4.8	5.1	4.9	5.1	3.9	3.9	3.9	3.8	3.9
2	Size	0	20	8.4	5.5	6.1	7.1	7.9	5.5	5.4	5.2	5.5	5.4	5.9	5.9	5.6	6.1	5.9	4.7	3.8	4.3	4.4	4.8
			50	8.2	7.2	10.0	6.8	7.6	8.5	6.2	6.4	8.4	6.1	7.6	6.3	6.2	8.1	8.6	6.4	6.6	6.4	6.7	6.7
			100	7.9	9.3	8.6	NA	5.7	7.0	5.0	8.1	6.7	4.3	5.4	5.6	5.8	4.6	4.8	6.1	4.3	5.1	6.4	4.7
			10	18.7	21.0	21.2	18.9	17.8	38.3	40.0	41.8	38.3	37.4	65.3	65.2	67.3	65.2	64.7	97.4	97.4	97.6	97.4	97.4
2	Power	0	20	16.7	20.3	25.7	18.9	15.0	35.7	39.4	42.1	36.4	34.5	64.8	67.5	69.7	64.8	63.8	97.2	97.5	97.5	97.5	97.2
			50	11.0	14.9	26.7	12.6	7.8	25.2	37.0	44.0	29.7	20.3	57.2	67.0	71.9	60.5	54.4	95.0	96.2	96.8	95.2	94.5
			100	11.3	14.7	29.5	NA	5.9	16.3	34.4	47.6	21.0	11.3	45.1	65.2	74.9	52.3	38.3	94.7	97.3	97.7	95.2	92.7
			10	5.6	5.4	7.3	5.6	5.3	6.0	6.0	8.9	5.6	5.8	3.9	4.9	8.7	4.4	3.8	4.1	4.1	4.0	4.0	4.2
3	Size	0	20	4.6	5.8	7.5	6.0	5.4	5.0	5.7	11.0	4.4	4.4	5.4	6.6	13.3	5.0	4.9	4.7	4.4	4.2	4.8	4.6
			50	8.1	7.5	10.7	7.4	7.0	6.9	8.2	13.8	6.3	7.6	7.0	8.7	18.3	6.7	6.5	5.0	5.3	7.8	5.2	5.0
			100	7.7	7.6	10.6	NA	6.7	6.2	8.9	14.0	6.0	5.4	4.7	8.5	18.9	4.1	5.0	4.4	5.5	8.8	4.5	5.4
			10	16.2	21.9	29.2	16.7	15.6	32.2	38.9	49.4	32.8	30.9	58.6	61.8	65.5	58.4	58.2	95.2	95.6	95.7	95.5	95.2
3	Power	0	20	14.2	21.8	32.6	14.4	12.1	29.9	40.7	57.2	30.2	27.3	56.7	63.7	73.9	56.8	55.6	94.1	94.6	94.8	94.2	94.1
			50	12.3	23.5	35.2	12.5	10.8	24.9	43.9	59.5	26.2	21.3	50.8	62.2	77.7	50.2	48.3	91.2	92.9	93.6	91.8	90.5
			100	9.8	21.9	36.1	NA	9.5	14.9	47.4	63.8	18.5	13.1	41.7	64.4	79.8	44.8	32.7	90.0	94.2	95.2	90.2	87.3
			10	6.2	5.5	5.8	6.0	6.4	5.9	5.9	5.4	6.0	5.3	5.1	4.7	4.7	5.1	4.8	4.5	4.1	4.2	4.6	4.3
1	Size	0.7	20	7.8	6.3	5.9	6.3	6.6	6.5	6.2	6.2	6.2	5.6	5.1	5.5	5.5	5.6	5.1	5.1	4.9	4.9	4.5	5.3
			50	9.0	6.3	8.6	7.9	6.3	6.9	7.5	6.0	7.1	6.0	6.6	5.1	4.3	5.8	6.5	5.7	4.8	4.9	5.2	5.8
			100	8.0	8.6	10.4	8.8	5.7	8.6	6.5	6.5	8.1	6.9	7.0	5.8	4.6	4.9	6.4	5.6	3.6	3.7	5.3	6.5
			10	19.7	18.9	19.9	20.5	19.1	34.7	35.2	35.2	34.7	34.1	56.7	57.3	57.1	57.8	55.0	93.7	94.2	94.2	94.3	93.9
1	Power	0.7	20	16.2	19.3	20.4	19.7	14.8	29.7	31.0	32.2	30.4	28.0	54.7	57.1	56.8	56.5	53.3	92.8	94.2	94.2	93.5	92.6
			50	11.3	15.0	21.8	14.7	10.4	24.9	30.8	33.7	31.2	20.6	50.6	57.3	59.0	56.1	46.4	90.7	92.0	92.5	91.8	89.8
			100	9.6	15.1	24.6	10.5	10.1	15.4	29.6	34.0	23.2	11.3	34.5	53.8	56.0	47.7	27.5	88.2	94.1	94.6	91.0	84.8

Notes: Size and Power for the different DGPs described in Section 4.1 are reported for 1000 replications. $T = (50, 100, 200, 500)$ is the time series length, $K = (10, 20, 50, 100)$ the number of variables in the system, the lag-length is fixed to $p = 1$. ρ indicates the correlation employed to simulate the time series with the Toeplitz covariance matrix. NAs are placed whenever the post-OLS estimation was not feasible due to $S^* > T$.

Table A.4: Simulation results for the PDS-LM Granger causality test (Overspecified lag-length)

DGP	Size/Power	ρ	T	50					100					200					500				
			K	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}	AIC	BIC	EBIC	λ^{th}	λ^{TSCV}
1	Size	0	10	7.2	6.7	7.9	6.8	6.5	6.0	6.9	6.7	6.9	5.0	5.7	5.3	5.4	5.4	5.5	5.4	5.6	5.5	5.2	5.1
			20	8.6	9.0	10.3	8.8	6.7	6.5	5.5	6.0	5.1	5.4	5.6	5.5	5.4	5.6	5.7	5.0	3.8	3.8	4.6	4.6
			50	NA	7.5	8.9	NA	NA	8.9	5.4	6.8	6.4	7.0	7.1	5.1	4.8	6.1	5.8	5.7	4.8	4.9	4.7	5.8
			100	NA	NA	10.8	NA	NA	NA	5.9	8.0	NA	NA	7.1	4.7	4.6	6.0	5.4	6.0	5.2	5.5	5.7	5.5
1	Power	0	10	20.7	24.6	28.8	22.0	18.8	45.6	49.0	50.4	46.9	44.1	82.4	82.9	84.1	82.9	81.6	99.7	99.8	99.8	99.7	99.8
			20	13.7	19.0	28.8	17.1	10.8	33.2	45.3	48.6	36.6	30.0	73.3	79.2	79.8	74.2	70.8	99.6	99.6	99.6	99.6	99.6
			50	NA	NA	32.3	NA	NA	16.7	43.8	50.4	25.7	17.3	56.1	77.6	80.1	64.3	51.3	98.1	99.7	99.7	98.4	97.3
			100	NA	NA	33.3	NA	NA	NA	38.4	48.8	NA	NA	20.3	76.7	79.8	98.4	23.1	96.0	99.8	99.8	98.0	94.2
2	Size	0	10	5.5	6.2	8.1	5.6	5.0	6.5	6.6	7.6	6.7	6.1	5.8	5.4	6.1	5.5	5.7	5.1	5.0	4.9	5.1	5.1
			20	7.3	7.1	10.7	7.6	5.8	5.4	5.0	7.5	5.0	4.6	5.0	4.9	6.5	5.6	4.6	4.4	4.4	4.7	4.2	4.2
			50	NA	5.2	8.4	NA	NA	6.0	6.9	10.8	4.6	5.5	8.3	5.7	8.0	5.3	6.8	6.3	5.1	4.7	4.2	5.0
			100	NA	NA	10.9	NA	NA	NA	6.9	12.6	NA	4.6	7.6	5.3	9.7	6.7	5.4	6.1	5.2	5.2	5.5	5.4
2	Power	0	10	13.0	16.2	21.7	14.2	12.5	29.0	29.6	34.3	29.5	27.1	52.6	52.5	54.9	52.6	51.9	94.1	94.3	94.3	94.6	94.2
			20	11.1	16.8	28.8	15.1	10.4	23.6	29.1	40.0	25.5	20.3	52.5	55.2	58.7	53.7	48.9	92.6	92.7	92.8	93.1	91.6
			50	NA	NA	28.1	NA	NA	9.4	29.0	47.0	18.5	9.1	38.3	53.6	62.7	43.6	32.5	88.7	92.1	93.6	89.2	86.9
			100	NA	NA	34.3	NA	NA	NA	30.1	51.8	NA	5.7	13.6	54.5	65.2	33.0	13.6	83.9	93.6	95.2	88.0	78.2
3	Size	0	10	5.5	6.3	8.8	6.3	4.7	6.1	7.2	12.8	6.0	5.6	4.8	5.8	11.4	5.0	4.6	5.1	5.2	4.9	5.0	4.9
			20	6.2	6.0	10.0	6.4	6.5	5.3	6.9	14.2	4.1	5.2	5.8	5.6	14.3	5.0	5.0	4.2	3.5	3.8	3.4	3.9
			50	NA	7.1	11.1	NA	NA	8.0	7.7	18.4	5.9	6.0	6.2	7.1	20.9	5.5	5.4	4.8	4.5	5.7	4.6	4.5
			100	NA	NA	11.5	NA	NA	NA	9.3	17.8	NA	6.3	7.4	7.5	24.9	5.7	5.4	3.7	5.1	6.1	4.3	3.2
3	Power	0	10	10.1	14.7	26.0	9.8	8.6	22.2	27.7	43.8	22.0	21.8	43.2	44.3	53.5	43.8	43.0	86.7	87.4	87.7	87.3	86.7
			20	9.0	16.2	30.6	9.8	8.0	15.6	27.8	50.8	18.5	13.6	39.7	44.7	61.6	41.1	37.5	85.3	86.7	87.1	86.3	84.5
			50	NA	15.9	31.3	NA	NA	10.5	33.1	58.1	18.4	11.5	32.1	47.5	69.7	38.2	25.9	78.3	81.5	82.6	79.9	75.8
			100	NA	NA	32.0	NA	NA	NA	34.3	59.3	NA	NA	14.2	49.4	76.9	27.4	6.8	14.1	83.9	85.6	83.0	61.1
1	Size	0.7	10	7.6	6.9	5.8	6.5	6.6	5.5	5.5	5.7	4.9	5.6	4.7	4.6	5.5	4.9	5.2	4.8	4.8	4.8	4.8	4.8
			20	6.9	7.0	7.3	6.9	6.9	6.3	7.1	5.8	6.0	5.6	5.1	4.7	4.9	5.5	5.3	5.4	5.6	5.3	5.7	5.5
			50	NA	NA	9.3	NA	NA	8.6	6.1	5.4	8.0	7.1	7.8	4.6	4.4	5.9	7.5	6.5	5.3	4.9	5.8	6.3
			100	NA	NA	11.2	NA	NA	NA	6.1	6.0	NA	5.2	6.3	5.1	4.5	5.8	5.3	5.9	4.3	4.6	5.2	6.3
1	Power	0.7	10	13.0	13.6	14.4	14.2	13.0	22.7	24.0	25.3	23.6	22.4	42.1	43.4	43.9	43.2	41.7	86.0	87.5	87.5	86.9	86.1
			20	10.2	11.5	13.9	11.5	9.9	19.6	21.1	20.9	21.4	16.7	39.4	42.3	43.2	41.7	36.5	84.8	86.8	86.7	86.0	83.4
			50	NA	NA	18.6	NA	NA	12.6	22.3	23.6	19.8	11.8	30.6	43.6	46.4	39.3	26.8	80.7	85.8	86.0	83.1	77.2
			100	NA	NA	18.3	NA	NA	NA	19.9	23.1	NA	NA	14.3	41.3	44.2	30.3	14.5	72.3	85.2	86.1	80.5	61.5

Notes: Size and Power for the different DGPs described in Section 4.1 are reported for 1000 replications. $T = (50, 100, 200, 500)$ is the time series length, $K = (10, 20, 50, 100)$ the number of variables in the system, the lag-length is fixed to $p = 1$. ρ indicates the correlation employed to simulate the time series with the Toeplitz covariance matrix. NAs are placed whenever the post-OLS estimation was not feasible due to $S^* > T$.

Table A.5: Simulation results for the bivariate Granger causality test

DGP	Size/Power	ρ	$K\backslash T$	50	100	200	500
			10	5.9	6.6	7.8	11.8
2	Size	0	20	5.6	5.9	7.8	11.8
			50	4.3	7.0	9.7	14.5
			100	5.5	6.7	8.9	13.9

Notes: Size is reported for DGP 2, as described in Section 4.1, for 1000 replications. $T = (50, 100, 200, 500)$ is the time series length, $K = (10, 20, 50, 100)$ the number of variables in the system, the lag length is fixed to $p = 1$. ρ indicates the correlation employed to simulate the time series with the Toeplitz covariance matrix.

Appendix B. Additional Material for the Empirical Applications

Heteroskedasticity-robust PDS-LM Granger causality test

[1]-[3] As in the original PDS-LM algorithm.

[4] Regress $X_{GC}^{(j)}$ onto X_S and store the residuals $\hat{u}_i^{(j)}$

[5] Compute the element-wise products $\pi_i^{(j)} := \hat{\xi}_i \hat{u}_i^{(j)}$ and regress a vector of ones on $\pi_1^1, \dots, \pi_i^{(p)}$ (without constant) and store the sum of squared residuals (SSR) from this regression.

[6] Reject H_0 if $T - SSR > q_{\chi_p^2}(1-\alpha)$, where $q_{\chi_p^2}(1-\alpha)$ is the $1-\alpha$ quantile of the χ^2 distribution with p degrees of freedom.

Table B.6: Correlations

Factors	Series	ρ
F_1	USPRIV	-0.925
	USGOOD	-0.923
	PAYEMS	-0.911
F_2	AAAFFM	-0.708
	T5YFFM	-0.692
	HOUST	-0.671
F_3	DGDSRG3Q086SBEA	-0.857
	CUSR0000SAC	-0.856
	PCECTPI	-0.840
F_4	CES9093000001	-0.540
	CES9092000001	-0.513
	USFIRE	-0.479
F_5	OPHPBS	0.493
	AWHMAN	-0.481
	NWPIx	-0.467
F_6	CONSPI	0.540
	S&P 500	-0.446
	S&P: indust	-0.435
F_7	GS1	0.486
	TB6MS	0.481
	TB3MS	0.457
F_8	PCEPILFE	0.414
	DSERRG3Q086SBEA	0.383
	DHCERG3Q086SBEA	0.382

Notes: the table reports the top-three macroeconomic series from the FRED-QD dataset most correlated (Pearson ρ) with each estimated factor ($F_1, \dots, F_8$).

Table B.7: Stocks used in Section 5.2

N.	Symbol	Issue name	N.	Symbol	Issue name
1	AAPL	APPLE INC	26	JNJ	JOHNSON &JOHNSON
2	ABT	ABBOTT LABORATORIES	27	KO	COCA COLA CO
3	AXP	AMERICAN EXPRESS CO	28	LLY	ELI LILLY & CO
4	BA	BOEING CO	29	MCD	MCDONALDS CORP
5	BAC	BANK OF AMERICA	30	MMM	3M COMPANY
6	BMY	BRISTOL MYERS SQ	31	MOT	MOTOROLA
7	BP	BP plc	32	MRK	MERCK & CO
8	C	CITIGROUP	33	MS	MORGAN STANLEY
9	CAT	CATERPILLAR	34	MSFT	MICROSOFT CP
10	CL	COLGATE-PALMOLIVE CO	35	ORCL	ORACLE CORP
11	CSCO	CISCO SYSTEMS	36	PEP	PEPSICO INC
12	CVX	CHEVRON CORP	37	PFE	PFIZER INC
13	DELL	DELL INC	38	PG	PROCTER & GAMBLE
14	DIS	WALT DISNEY CO	39	QCOM	QUALCOMM
15	EK	EASTMAN KODAK	40	SLB	SCHLUMBERGER N.V.
16	EXC	EXELON CORP	41	T	AT&T CORP
17	F	FORD MOTOR CO	42	TWX	TIME WARNER
18	FDX	FEDEX CORP	43	UN	UNILEVER N.V.
19	GE	GENERAL ELEC	44	VZ	VERIZON COMMS
20	GM	GENERAL MOTORS	45	WFC	WELLS FARGO & CO
21	HD	HOME DEPOT INC	46	WMT	WAL-MART STORES
22	HNZ	H J HEINZ CO	47	WYE	WEYERHAEUSER CO
23	HON	HONEYWELL INTL	48	XOM	EXXON MOBIL
24	IBM	INTL BUS MACHINE	49	XRX	XERON CORP
25	INTC	INTEL CORP			