SVAR Identification with High-Frequency Macroeconomic Data

Fulvio Corsi[1] Luigi Longo[2] Francesco Cordoni[1]

[1]Dipartimento di Economia e Management, Università di Pisa

[2]IMT School for Advanced Studies, Lucca

Workshop on *"Model Evaluation and Causal Search: Empirical and Experimental Approaches"* (University of Pisa)

Contents

Introduction and previous studies

Problem of Identification in macroeconomics: the structural and causal relations between macroeconomic series can be achieved by *imposing restrictions* from economic theory

- Bernanke (1986) imposes short-run restrictions, Blanchard and Quah (1989) long-run restrictions, while Uhlig (2005) imposes restrictions on the sign of IRFs;

or by *exploiting the statistical properties of the data*

- Sentana and Fiorentini (2001), Rigobon (2003), Lütkepohl and Netsunajev (2017) exploit heteroskeasticity of the data;
- Moneta et al. (2013), Lanne et al. (2017), Gourieroux et al. (2017), Cordoni and Corsi (2019), Lanne and Luoto (2021) achieve identification through non-Gaussianity of the residuals;
- Foroni and Marcellino (2014) and (2016) use a mix-frequency approach containing both high-frequency and low-frequency data.

Our contribution

- **Motivation**:
  - We start from the theoretical observation that the identification problem of SVAR models is fundamentally a problem of variable aggregation (Foroni and Marcellino 2014 and 2016);
  - In fact, low-frequency shocks generated by time-aggregation are a complex non-linear mixture of high-frequency variables,
    $\Rightarrow$ impossible to recover the correct structural relations using only linear transformation of low-frequency information.
- **Methodology**:
  - We propose a new identification method of SVAR models based on the *nowcasting* of macroeconomic data;
  - The idea is that contemporaneous relations between macroeconomic quantities tend to vanish as the observation frequency increases;
  - We present both theoretical arguments and empirical evidences that by increasing the observation frequency of the macroeconomic variables is possible to fully identify all the structural shocks with no further information.

The identification problem in macroeconomics

Considering a VAR of order 1:

$$\boldsymbol{y}_t = A\boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \quad t = 1, 2, \dots, T \tag{1}$$

We can write it in its structural form:

$$B_0\boldsymbol{y}_t = B_0 A\boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = 1, 2, \dots, T \tag{2}$$

where

$$\text{cov}(\boldsymbol{e}_t) = I, \quad \boldsymbol{u}_t = Z\boldsymbol{e}_t, \quad Z = B_0^{-1}. \tag{3}$$

We can estimate the empirical covariance matrix of the VAR residuals:

$$\text{cov}(\boldsymbol{u}_t) \equiv \Sigma_u = B_0^{-1} B_0^{-1\prime} = ZZ' \tag{4}$$

- **Identification problem:** Σ_u is symmetric with $K(K+1)/2$ independent elements while B_0 has K^2 independent elements;
- $\Rightarrow$ Need to impose restrictions on B_0.

High-frequency and low-frequency VAR models

We denote the high-frequency VAR (HF-VAR) of order 1 as:

$$\boldsymbol{y}_t^H = A\boldsymbol{y}_{t-1}^H + \boldsymbol{u}_t^H, \quad t = 1, 2, \ldots, T \tag{5}$$

The related low-frequency VAR (LF-VAR) is:

$$\boldsymbol{y}_\tau^L = \Phi\boldsymbol{y}_{\tau-1}^L + \boldsymbol{u}_\tau^L, \quad \tau = m, 2m, \ldots, T \tag{6}$$

With point-in-time sampling aggregation, the LF-VAR can be written as:

$$\boldsymbol{y}_\tau^L = A^m\boldsymbol{y}_{\tau-1}^L + \boldsymbol{u}_\tau^L,\ \Phi = A^m \tag{7}$$

With variance-covariance matrix of the residuals:

$$\text{cov}\left(\boldsymbol{u}_\tau^L\right) = \Omega^L = \Omega^H + A\Omega^H A' + A^2\Omega^H A^{2'} + \cdots + A^{m-1}\Omega^H A^{m-1'} \tag{8}$$

Notice that lead-lag dependence in the HF-VAR (matrix A) appears to be contemporaneous in the LF-VAR.

Structural HF-VAR

The HF-VAR can be written in its structural form:

$$B_0 \boldsymbol{y}_{t^*}^H = B_0 A \boldsymbol{y}_{t^*-1}^H + \boldsymbol{e}_t^*, \quad t^* = 1, 2, \ldots, T^* \tag{9}$$

where t^* is the temporal index related to the sample interval Δ^*, i.e. the structural frequency.

The relation between the structural shocks and the reduced form residuals of the HF-VAR is

$$\boldsymbol{u}_{t^*}^H = Z \boldsymbol{e}_{t^*}$$

so that

$$\mathrm{cov}(\boldsymbol{u}_{t^*}^H) = Z Z'$$

.

Structural HF-VAR

- Standard short-run restrictions on B_0 are traditionally based on timing assumptions, i.e. that a given macroeconomic variable does not react within the considered period (e.g. one quarter) to a change in another variable;

- In order to formalize this idea, let us denote with Δ the time interval (e.g., daily, monthly, quarterly) at which data are sampled;

- and let $R \in \mathbb{R}^{N \times N}$ be a *matrix of reaction times* where the generic element r_{ij} represents the minimum time required for variable i to react to a change in variable j;

- Hence, we will have a short-run restriction represented by a zero entry in the corresponding element of the matrix B_0 whenever $r_{ij} > \Delta$, i.e. when the reaction time of i on j will be larger than the sampling interval Δ.

Identification of structural HF-VAR

- **Assumption 1.** *The high-frequency structural shocks are zero mean process with covariance matrix equal to the identity.*
- **Assumption 2.** *The HF-VAR model is non singular, i.e., the number of (high-frequency) structural shocks is equal to the number of (high-frequency) variables.*

  We discuss in the paper how Assumption 2 could be relaxed.

- **Theorem 2.3.** *Under Assumptions 1, 2, there exists a sampling frequency* Δ^* *such that the HF-VAR and HF-SVAR representations coincide, i.e.,* $\boldsymbol{u}_{t^*}^H$ *are effectively the (high-frequency) structural shocks* $\boldsymbol{e}_{t^*}$ *up to scaling.*

Identification of structural HF-VAR

- Theoretically we only need Assumption 1, and 2 to prove that B_0 is diagonal;
- Empirically, B_0 might be diagonal by chance if exist an orthogonal and non-diagonal matrix such that: $\boldsymbol{u}_{t^*}^H = C\boldsymbol{e}_{t^*}^H$;
- We need another assumption to use ICA approach by Gouriéroux et al. (2017) and test whether C is in the class of the identity:

- **Assumption 3.** *The high-frequency structural shocks are i.i.d., where their components are mutually independent with at most one Gaussian distribution.*

- Thanks to Assumption 3 we may estimate the orthogonal matrix C and test whether this matrix is in the equivalence class of the identity or not through the Wald-test of Gourieroux et al. (2017).

High-frequency heterogenous VAR (HF-HVAR)

- Considering a daily frequency as the highest possible frequency, we define an HF-HVAR inspired by Corsi (2009):

$$B_0\boldsymbol{y}_t = B_d\boldsymbol{y}^{(d)} + B_w\boldsymbol{y}^{(w)} + B_m\boldsymbol{y}^{(m)} + B_q\boldsymbol{y}^{(q)} + \boldsymbol{e}_t \qquad (10)$$

  where

$$\boldsymbol{y}^{(d)} = \boldsymbol{y}_{t-1}, \boldsymbol{y}^{(w)} = \sum_{i=2}^{5}\boldsymbol{y}_{t-i}, \boldsymbol{y}^{(m)} = \sum_{i=6}^{22}\boldsymbol{y}_{t-i}, \boldsymbol{y}^{(q)} = \sum_{i=23}^{66}\boldsymbol{y}_{t-i}$$

- It is a restricted least square regression;

- The motivation behind the HF-HVAR is to solve the curse of dimensionality problem arising when using high frequency data with macroeconomic quantities.

Data

The high-frequency macroeconomic data used in this empirical analysis are kindly supplied by Square Macro LTD (Beber et al. (2015)). They provide US nowcasted daily data for the sample from 1996-01-02 to 2020-03-10. We used the following variables:

- Inflation: π_t (*Square Macro*);
- Output-sentiment: O_t (*Square Macro*);
- Federal Funds effective rate: i_t (*FRED*);

In a second analysis we include a fourth variable:

- S&P500 index: r_t (FRED).

Monte Carlo analysis

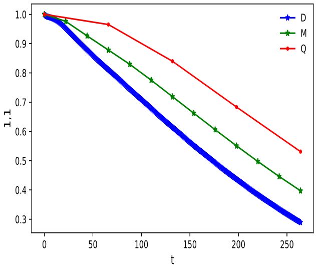
We simulate an HF-HVAR(1) using matrix of coefficients estimated with the empirical data (3-variables VAR):

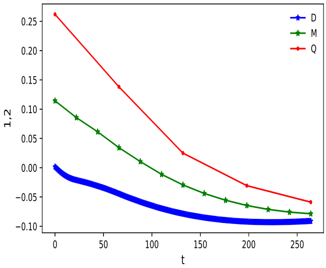
$$\boldsymbol{y}_t^d = A_d \boldsymbol{y}^{(d)} + A_w \boldsymbol{y}^{(w)} + A_m \boldsymbol{y}^{(m)} + A_q \boldsymbol{y}^{(q)} + \boldsymbol{u}_t^d \tag{11}$$

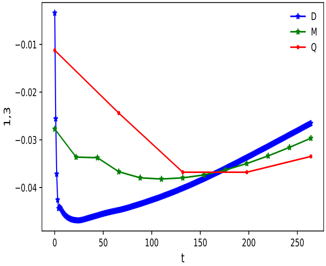
where $\boldsymbol{u}_t^d \sim N\left(0, I_N\right)$.

We generate $\boldsymbol{y}_t^m$ and $\boldsymbol{y}_t^q$ by aggregating the original process over 22 and 66 periods. We estimate their covariance matrix and we identify the IRFs with Cholesky decomposition.

Monte Carlo analysis: results







Monte Carlo analysis: results

	Response of variable:		
Shock of 1 ⇒	1	2	3
Daily	6.068e-5	7.768e-5	0.019
Monthly	0.209	0.195	1.379
Shock of 2 ⇒	1	2	3
Daily	0.005	0.006	1.02e-5
Monthly	0.129	0.106	0.828
Shock of 3 ⇒	1	2	3
Daily	3.101e-5	2.095e-4	2.675e-4
Monthly	0.037	0.217	0.226

Table 1: MSE of the estimated IRF and the theoretical IRF. In the table results are shown as the ratio between the reported frequency (first column) and the quarterly frequency.

Empirical analysis

- We estimate a daily (heterogenous estimation), weekly, monthly and quarterly (standard LS estimation);
- Number of lags is selected via AIC on the aggregated (quarterly) data ($p = 1$); this is consistent at all the frequencies;
- We compute the Frobenius distance from the identity of the correlation matrix of VAR residuals;
- We estimate $\hat{C}$ as in Gouriéroux et. al (2017) and we test whether is in the class of the identity $\mathscr{P}(I)$ at each frequency.

Empirical analysis: contemporaneous correlation

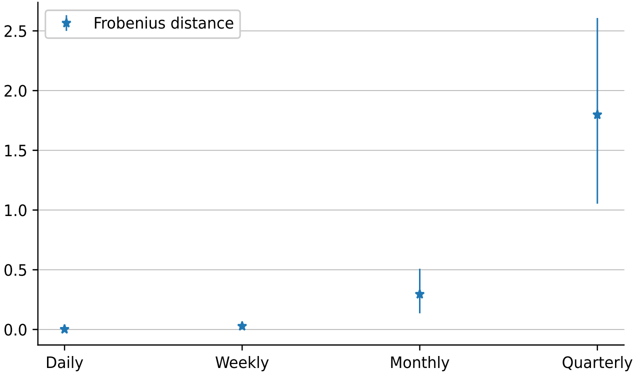


Figure: Frobenius distance of the correlation matrix of reduced-form VAR residuals estimated at each frequency from the identity matrix. HF-HVAR estimation is used for the daily frequency.

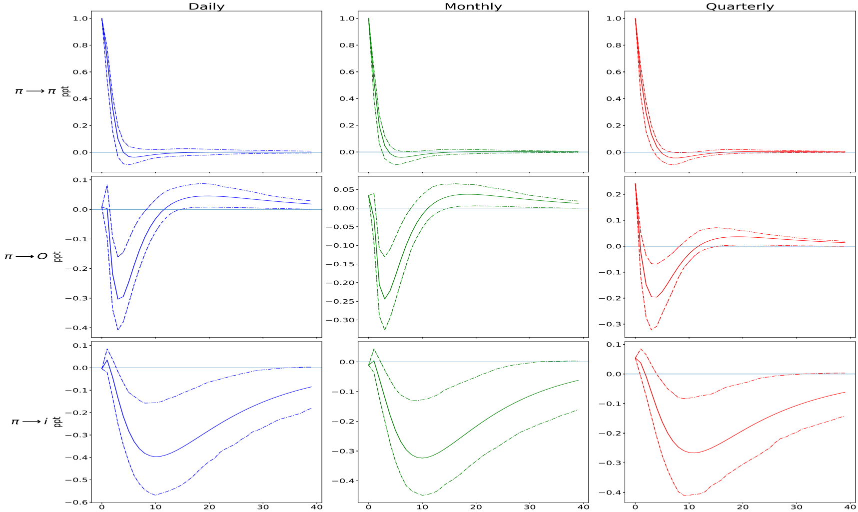
Empirical analysis: Wald test results

	Order 1			Order 2			Order 3		
Freq.	Dist.	Stat.	p-value	Dist.	Stat.	p-value	Dist.	Stat.	p-value
D	0.001	0.231	0.972	0.001	0.719	0.869	0.001	0.732	0.866
W	0.039	27.573	<1e-05	0.021	16.276	<1e-03	0.021	16.286	<1e-03
M	0.187	51.856	<1e-10	0.209	57.483	<1e-11	0.223	60.426	<1e-12
Q	0.756	64.373	<1e-13	1.525	112.406	<1e-16	0.803	65.34	<1e-13

Table 3: Wald-test results following Gouriéroux et al. (2017). Under the null hypothesis, the estimated ICA orthogonal matrix $\widehat{C}$ should be in $\mathscr{P}(I)$. The columns show the test results when the variables ordering changes. Order 1 = $[\pi_t, O_t, i_t]$, Order 2 = $[\pi_t, i_t, O_t]$, Order 3 = $[i_t, \pi_t, O_t]$, From the first column of each table: data frequency, i.e., daily (D), weekly (W), monthly (M) and quarterly (Q), distance of $\widehat{C}$ to $\mathscr{P}(I)$, the related Wald-test statistic and p-value. The pseudo-maximum likelihood are set to four distinct Laplace distributions with scale equal to $15, 20, 10$, respectively.

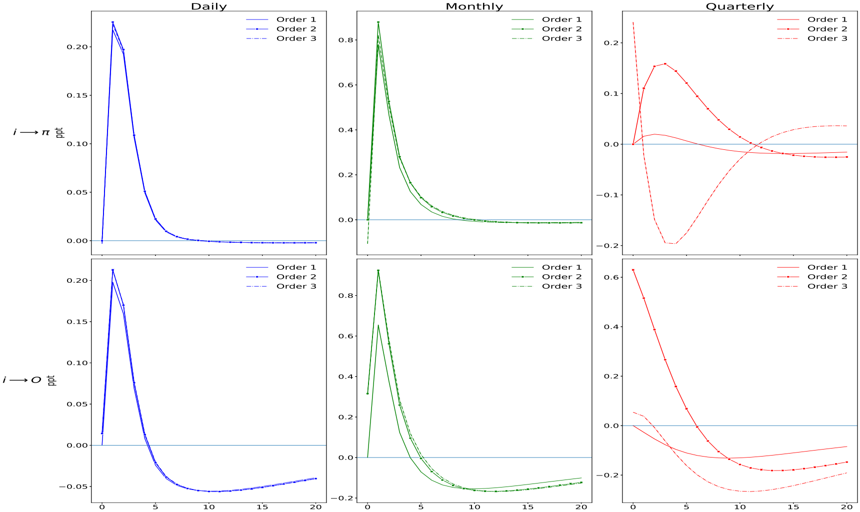
Empirical analysis: contractionary supply shock

Order 1: $[\pi_t, O_t, i_t]$.



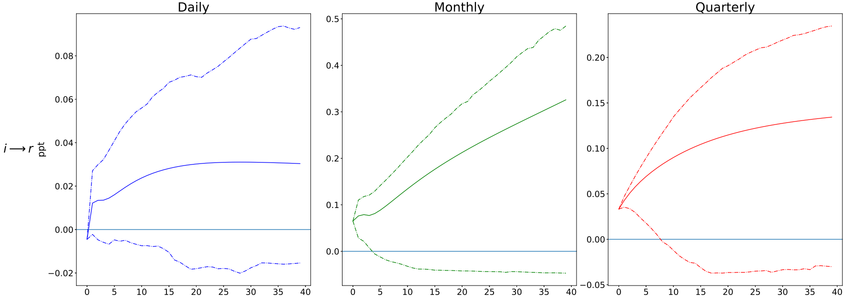
Empirical analysis: monetary policy (robustness)

Order 1: $[\pi_t, O_t, i_t]$; Order 2: $[\pi_t, i_t, O_t]$; Order 3: $[i_t, \pi_t, O_t]$.



Empirical analysis: monetary policy and financial interdependence

Order: $[\pi_t, O_t, i_t, r_t]$. The response of O_t and π_t to a monetary policy shock is the same of the 3-variables VAR.



Positive contemporaneous reaction of the financial markets at low-frequency (odd for economic theory).

Conclusion

- We provided a well-founded theoretical argument to show that by increasing the observation frequency of the variables is possible to fully identify all the structural shocks with no further information;
- We show that our theoretical argument holds in the empirical application;
- Our methodology is robust to changes in the variables ordering of the VAR;
- Interest rate hikes might convey positive Central Bank's optimistic view of economic outlook (the response of the output in the short-term is positive);
- The results of the methodology are invariant to the inclusion of a stock market index in the model.

Thank You For The Attention!

luigi.longo@imtlucca.it

The paper is available on SSRN