NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Nikolay Arefiev

IDENTIFICATION OF MONETARY POLICY SHOCKS WITHIN A SVAR USING RESTRICTIONS CONSISTENT WITH A DSGE MODEL

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 125/EC/2016

IDENTIFICATION OF MONETARY POLICY SHOCKS WITHIN A SVAR USING RESTRICTIONS CONSISTENT WITH A DSGE MODEL

NIKOLAY AREFIEV

Abstract. I identify and estimate the monetary policy rule and the monetary policy shocks within a structural vector autoregression model for the US economy. I make two contributions to the literature. First, for identification I propose to use restrictions consistent with the literature on dynamic stochastic general equilibrium (DSGE) models. Typical DSGE model produces more restrictions than is required for the identification, so overidentifying restrictions can be tested against the data. The second contribution is a new method of testing the overidentifying restrictions. This method divides the set of identifying restrictions into subsets, and tests each subset independently of the others. This method does not reject most restrictions produced by the DSGE model. The only rejections provide evidence that the Federal Reserve uses delayed information about the inflation in policy making. The proposed approach to identification helps explain and solve the price puzzle problem reported in the previous literature.

Keywords: graphical identification; sparse SVAR; price puzzle.

JEL codes: C30, E52.

1. Introduction

The purpose of this paper is to identify and estimate the monetary policy rule and the response of the economy to monetary policy shocks within a Structural Vector Autoregression (SVAR) model for the US economy. The literature on SVARs usually identifies monetary policy shocks using restrictions imposed on the matrix of contemporaneous effects and on the impulse response functions (Sims, 1980, 1986, 1992;

*February 17, 2016.

The National Research University Higher School of Economics. E-mail: n.arefiev@gmail.com. Address: office 708A, Shabolovka 31, 119049, Moscow, Russia. Cell phone: +7 926 825 7147.

**This research was inspired by online course at Coursera "Probabilistic Graphical Models" lectured by Daphne Koller. I thank Alina Arefeva, Antoine d'Autume, Svetlana Bryzgalova, Jean-Bernard Chatelain, Boris Demeshev, Jean-Pierre Drugeon, Jean-Marie Dufour, Bulat Gafarov, James Hamilton, Oleg Itskhoki, Maarten Janssen, Grigory Kantorovich, Ramis Khabibulin, Sergey Kusnetzov, Jessie Li, Judea Pearl, Anatoly Peresetsky, Christopher Sims, Alain Trognon, Ilya Voskoboynikov, and Mark Watson for helpful comments and suggestions. I also thank participants of groups 7inR and Macroteam, and other colleagues from University Paris-1 Panthéon-Sorbonne and from The Higher School of Economics for fruitful discussions. This study (research grant No 14-01-0088) was supported by The National Research University Higher School of Economics' Academic Fund Program in 2014/2015.

Blanchard and Quah, 1993; Christiano et al., 1999; Zha, 1999; Hanson, 2004; Uhlig, 2005; Sims and Zha, 2006). However, restrictions on the matrix of contemporaneous effects are not consistent with the literature on Dynamic Stochastic General Equilibrium (DSGE) models (Smets and Wouters, 2003, 2007; Christiano et al., 2005). These models predict that the matrix of contemporaneous effects is dense. Instead, DSGE models imply restrictions on the matrices of lagged effects, and I propose to use those restrictions for identification.

The DSGE-based restrictions overidentify the SVAR, so many of them can be tested against the data. I introduce the Transformed Concentration Network (TCNW), and use it to test the overidentifying restrictions. There are alternative tests for these restrictions, such as the J-test (Sargan (1958); Hansen (1982)) and the likelihood ratio test. However, these tests do not specify which identifying assumptions are tested, and which are not. In the case where the overidentifying restrictions are not rejected, there may exist a continuum of other observationally equivalent models, which pass these tests equally well. If the identifying assumptions are rejected, these tests are not informative on which assumptions are violated. The method proposed in this paper has the following two advantages. First, for a large class of models it guarantees that if the identification restrictions are not correct, they are asymptotically rejected in almost all parameter points. Therefore, if the identifying assumptions are accepted, the parameters of the other observationally equivalent models passing these tests equally well lay in a parameter space of measure zero. For other models, including the model considered in this paper, the method can distinguish between the identification assumptions that can be empirically verified and the assumptions that are justified only theoretically. So the method points at the identifying assumptions that cannot be asymptotically rejected even if they are incorrect. The second advantage is that if the identifying restrictions are rejected, the TCNW indicates the subset of the assumptions that do not hold.

The method proposed in this paper stems from the literature on causality and probabilistic graphical models. The theory of probabilistic graphical models has a lot of successful applications in computer science (see Koller, 2009; Pearl, 2009; Chen and Pearl, 2014), and it become popular in econometrics[1]. However, the approach developed in this literature has a firm theoretical background only for recursive models. The method proposed in this paper is designed to work not only with recursive but also with cyclical models.

The estimated TCNW is consistent with most identifying assumptions predicted by the DSGE model. Many identifying restrictions for the monetary policy rule equation are testable, and most of them are not rejected by the TCNW. The only restriction that was rejected is that the Federal Reserve does not use

[1]See Ahelegbey et al. (2014); Kwon and Bessler (2011); Bryant and Bessler (2015); Demiralp et al. (2014); Hoover (2005); Oxley et al. (2009); Phiromswad (2014); Reale and Wilson (2001); Wilson and Reale (2008); Fragetta and Melina (2013).

delayed information about the GDP deflator inflation and the commodity price inflation in policy decisions. This discovery motivates me to relax the respective restrictions when I estimate the SVAR model. This identification produces impulse response functions consistent with the macroeconomic theory without any anomalies such as the price puzzle.

I use the TCNW to diagnose the reasons of for the *price puzzle* problem reported in the previous literature. The price puzzle is a prediction of an estimated SVAR model that right after the restrictive monetary policy shock the inflation initially significantly rises and only then start to decline (Sims (1992); Christiano et al. (1999); Zha (1999); Hanson (2004); Keating et al. (2014)). This result is at odds with the macroeconomic theory, which predicts that the restrictive policy shock leads to a decline in inflation.

I argue that the price puzzle problem arises because of misspecification, where a high dimensional monetary policy rule is estimated within a low dimensional model. In low dimensional models the estimated monetary policy shocks are contaminated with the residuals that arise from the dimensionality reduction. The dimensionality reduction can produce many different biases, including the price puzzle bias. This explanation is consistent with those of Bernanke and Federal Reserve Board (2005). In addition to the discussion provided by Bernanke et al., I show exactly how the dimensionality reduction produces the price puzzle bias in low dimensional models.

The remaining part of the paper is organized as follows. In Section 2 I specify the SVAR model. In Section 3 I show how the structure of a typical DSGE model helps to choose restrictions for identification of the SVAR. In Section 4 I introduce the TCNW, and use it to verify the identifying restrictions. In Section 5 I diagnose the reasons for he price puzzle reported in the previous literature. Section 6 reports the estimated impulse response functions, Section 7 concludes.

2. Model

2.1. **SVAR.** The Structural Vector Autoregression (SVAR) model is:

$$\mathbf{A}_0 Y_t = \sum_{l=1}^{p} \mathbf{A}_l Y_{t-l} + \varepsilon_t, \tag{1}$$

where Y is $n \times 1$ vector of endogenous variables, $\mathbf{A}_i$ are matrices of parameters, p is the number of lags, and ε is the vector of structural shocks. Vector Y has been centralized, so the constant term in (1) is zero. The structural shocks are assumed to be orthogonal, and the standard deviation is normalized to one, so the covariance matrix is the identity matrix, $\mathbb{E}\left(\varepsilon_t \varepsilon_t^T\right) = \mathbf{I}$. The main diagonal of $\mathbf{A}_0$ is normalized to be positive. I combine matrices $\mathbf{A}_i$ into matrix $\mathbf{A} = [\mathbf{A}_0 \ \ -\mathbf{A}_1, \ldots, -\mathbf{A}_p]$, and vectors Y_{t-i} into vector

$X_t = \left[Y_t^T \ Y_{t-1}^T \ Y_{t-2}^T \ \ldots \ Y_{t-p}^T\right]^T$, so model (1) can be rewritten as:

$$\mathbf{A}X_t = \varepsilon_t \tag{2}$$

Let Z_t be the predetermined part of X_t: $Z_t = \left[Y_{t-1}^T \ Y_{t-2}^T \ \ldots \ Y_{t-p}^T\right]^T$, and $\mathbf{A}_L$ the respective part of $\mathbf{A}$: $\mathbf{A}_L = -\left[\mathbf{A}_1, \ \mathbf{A}_2 \ldots, \mathbf{A}_p\right]$. Matrix $\mathbf{A}_0$ is referred to as the matrix of contemporaneous effects, and $\mathbf{A}_L$ is matrix of lagged effects. I assume that the unconditional distribution of Z_t can be modeled as:

$$\mathbf{S}Z_t = \epsilon_t, \tag{3}$$

where $\mathbf{S}$ is nonsingular square matrix of size $pn \times pn$. The covariance of ϵ is normalized to the identity matrix, $\mathbb{E}\left(\epsilon_t \epsilon_t^T\right) = \mathbf{I}$. Residuals ϵ_t and ε_t are independent. Define $\mathcal{E}_t = \left[\varepsilon_t^T \ \epsilon_t^T\right]^t$. By construction, $\mathbb{E}(\mathcal{E}\mathcal{E}^T) = \mathbf{I}$. The whole model can be written as:

$$\mathbf{P}X_t = \mathcal{E}_t, \tag{4}$$

where

$$\mathbf{P} = \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_L \\ \mathbf{0}_{n\times n} & \mathbf{S} \end{pmatrix}$$

2.2. **Data and specification.** Vector Y includes the federal interest rate r, the inflation rate π measured as the growth rate of the GDP deflator, the commodity price inflation rate π^c, the GDP growth rate g, the capacity utilization rate c, and the unemployment rate u. This choice of variables is explained in Section 5. The data on π, π^c, g, c, and u come seasonally adjusted, and the data on r, π, π^c and g are annualized.

I use the US quarterly data from 1967:Q1 to 2007:Q4. I exclude the period of unconventional monetary policy, which began in 2008. I include 2 lags for the following reasons. Model selection criteria, such as BIC, AIC and others, suggest chosing between 1 and 4 lags. If I use only one lag, I cannot reliably identify the monetary policy rule, because there is no evidence that the contemporaneous or the first lag inflation affects the federal interest rate. Models with 2, 3, and 4 lags produce similar results, so a particular decision within these three models does not affect the conclusions. I choose 2 lags to simplify the identification procedure.

3. Identification

I use the structure of a typical DSGE model to formulate restrictions for identification of the SVAR model. DSGE models usually include a lot of frictions. A typical set of included frictions consists of Calvo pricing (Calvo, 1983; Clarida et al., 1999; Bils and Klenow, 2004), consumption habit formation (Abel, 1990;

Campbell and Cochrane, 1999; Fuhrer, 2000), capital or investment adjustment costs, and search-and-match at the labor market (Pissarides, 2000; Merz, 1995; Gertler et al., 2008). Each friction produces an equation, where the dependent variables is expressed as a function of other contemporary variables, its own expected value, and its own lagged values. The New Keynesian Phillips curve is a good example:

$$\pi_t = \alpha_1 \pi^{\mathbb{E}}_{t+1} + \alpha_2 \mathrm{MC_t} + \alpha_3 \pi_{t-1} + \alpha_4 \pi_{t-2} + \varepsilon^{\pi}_t \tag{5}$$

where $\pi^{\mathbb{E}}_{t+1}$ is the expected inflation, MC is the marginal cost function, ε^{π}_t is the inflation structural shock, and $\alpha_1, \ldots, \alpha_4$ are parameters. The marginal cost function depends on contemporary variables, such as the capacity utilization or the GDP growth rate, but it does not depend on any lagged variables. The state in the DSGE model is represented by Y_t, so the expected inflation is a function of all variables entering into vector Y_t, but as the marginal cost function, it also does not depend on any lagged variable. Therefore, all coefficients in the row for π in matrix $\mathbf{A}_0$ may be not zero, but only the coefficients before the lagged values of π are not constrained in the matrices of lagged effects.

The other equations associated with frictions have a similar structure. The dependent variable in each such equation is a function of its own expected value. Since the expected value can be represented as a function of Y_t, all coefficients in the respective row of $\mathbf{A}_0$ must be unconstrained. It also depends on its own lags, but not on the lagged values of the other variables. For this reason I do not impose any restrictions on the matrix of contemporaneous effects, and I make the following three identifying assumptions for the matrices of lagged effects.

First, I assume each variable may depend on its own first or second lags. Therefore, the main diagonals of $\mathbf{A}_1$ and $\mathbf{A}_2$ are not constrained. The assumption that the diagonals of $\mathbf{A}_1$ and $\mathbf{A}_2$ are unconstrained is similar to the Minnesota prior introduced by Litterman (1979); however, the motivation for this assumption is different. The motivation for the Minnesota prior is that the logarithms of most variables follow a process close to the random walk. However, I consider most variables in growth rates, where the random walk is converted into the white noise. Instead, I motivate this assumption by frictions.

Second, in Section 4.4 I provide evidence that the Federal Reserve uses delayed information about the GDP deflator inflation and the commodity price inflation when it sets the interest rate. Because of informational delays, the interest rate may depend on the first and second lags of the inflation and commodity price inflation, so I remove the constraints on coefficients before π and π^c in the equation for r in matrices $\mathbf{A}_1$ and $\mathbf{A}_2$.

TABLE 1. Summary of identification restrictions

	$\mathbf{A}_0$						$\mathbf{A}_1$						$\mathbf{A}_2$					
	r	u	c	g	π	π^c	$\mathbb{L}r$	$\mathbb{L}u$	$\mathbb{L}c$	$\mathbb{L}g$	$\mathbb{L}\pi$	$\mathbb{L}\pi^c$	$\mathbb{L}^2r$	$\mathbb{L}^2u$	$\mathbb{L}^2c$	$\mathbb{L}^2g$	$\mathbb{L}^2\pi$	$\mathbb{L}^2\pi^c$
r	*	*	*	*	*	*	*	*	*	0	*	*	*	0	0	0	*	*
u	*	*	*	*	*	*	0	*	*	0	0	0	0	*	0	0	0	0
c	*	*	*	*	*	*	0	*	*	0	0	0	0	0	*	0	0	0
g	*	*	*	*	*	*	0	*	*	*	0	0	0	0	0	*	0	0
π	*	*	*	*	*	*	0	*	*	0	*	0	0	0	0	0	*	0
π^c	*	*	*	*	*	*	0	*	*	0	0	*	0	0	0	0	0	*

"0" for the coefficients constrained to zero, "*" for the unconstrained coefficients.
Each row represents the corresponding equation in (1), columns correspond to variables included ("*") or excluded ("0") from this equation.

The third identification assumption for the matrices of lagged effects is the following. Observe that the model described only by the previous inclusions may be misspecified for the following reason. Consider the following production function:

$$Q = A\,(cK)^{\alpha}\,(L\,(1-u))^{1-\alpha}\,, \tag{6}$$

where Q is the output, A is the stochastic productivity parameter, K is the capital, L is the total labour force, and α is parameter. Product cK gives the amount of capital used in production, and $L(1-u)$ gives the number of employed workers. Assuming that K and L are approximately constant in the short run, the first-difference log-linearized version of (6) is:

$$g = \Delta \ln A + \alpha\Delta c - (1-\alpha)\Delta u \tag{7}$$

where Δ is the first difference operator. If I do not relax the constraints before the lagged values of u and c in the equation for g, I would estimate the model where g depends on c and u, but not on the first differences of c and u, so the model would be misspecified. To fix this problem, in each equation that includs the contemporaneous values of u or c I add the lagged values of the respective variables. That is, I copy all inclusions from the second and third columns of $\mathbf{A}_0$ into the respective columns of $\mathbf{A}_1$.

4. CHECKING IDENTIFICATION USING TRANSFORMED CONCENTRATION NETWORK

In this section I introduce the transformed concentration matrix, and then use it to define the Transformed Concentration Network (TCNW). Then I show how the TCNW can be used for testing the identifying restrictions, and apply it to test the identifying restrictions summarized in Table 1.

4.1. **Concentration and transformed concentration matrices.** The concentration matrix, also known as the precision matrix, is defined as the inverse covariance matrix:

$$\mathbf{C} = \mathbb{E}(XX^T)^{-1} \tag{8}$$

From (4) I obtain:

$$\mathbf{I} = \mathbb{E}\left(\mathcal{E}\mathcal{E}^T\right) = \mathbb{E}\left(\mathbf{P}XX^T\mathbf{P}^T\right) = \mathbf{P}\mathbf{C}^{-1}\mathbf{P}^T \tag{9}$$

Matrix manipulations gives:

$$\begin{aligned} \mathbf{C} &= \mathbf{P}^T\mathbf{P} \\ &= \begin{pmatrix} \mathbf{A}_0^T\mathbf{A}_0 & \mathbf{A}_0^T\mathbf{A}_L \\ \mathbf{A}_L^T\mathbf{A}_0 & \mathbf{A}_L^T\mathbf{A}_L + \mathbf{S}^T\mathbf{S} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} \end{aligned} \tag{10}$$

The transformed concentration matrix is defined as:

$$\tilde{\mathbf{C}} = \begin{pmatrix} \mathbf{A}_0^T\mathbf{A}_0 & \mathbf{A}_0^T\mathbf{A}_L \\ \mathbf{A}_L^T\mathbf{A}_0 & \mathbf{A}_L^T\mathbf{A}_L \end{pmatrix} \equiv \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \tilde{\mathbf{C}}_{22} \end{pmatrix} \tag{11}$$

The concentration matrix can be estimated using its definition (8). The transformed concentration can be estimated using the following lemma:

Lemma 1. $\tilde{\mathbf{C}}_{22} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$

Proof. By definition (11), I have: $\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} = \mathbf{A}_L^T\mathbf{A}_0\left(\mathbf{A}_0^T\mathbf{A}_0\right)^{-1}\mathbf{A}_0^T\mathbf{A}_L = \mathbf{A}_L^T\mathbf{A}_L = \tilde{\mathbf{C}}_{22}$ □

4.2. **Transformed concentration network.** I use the the transformed concentration matrix introduced in the previous section to define the transformed concentration network as:

Definition 1 (TCNW)**.** The Transformed Concentration Network (TCNW) is an undirected graph, which nodes are the random variables of the structural model, and where nodes x_i and x_j are adjacent if and only if the respective coefficient of the transformed concentration matrix is not zero: $\tilde{c}_{ij} \neq 0$.

The transformed concentration network is useful for analysis of identification assumptions because of the following property:

Proposition 1 (TCNW and the model's structure)**.** *Consider any two random variables of the structural model, say x_i and x_j, $i \neq j$.*

- *If x_i is excluded from each structural equation where x_j is included then $\tilde{c}_{ij} = 0$, and edge $x_i x_j$ is absent in the TCNW.*
- *If there exists at least one equation where x_i and x_j are both included, then $\tilde{c}_{ij} \neq 0$ in almost all parameter points, and edge $x_i x_j$ is present in the TCNW in almost all parameter points.*

Proof. Consider block $\mathbf{C}_{11} = \mathbf{A}_0^T \mathbf{A}_0$ of the transformed concentration matrix. Let a_{ij}^0 and c_{ij} be elements of $\mathbf{A}_0$ and $\mathbf{C}_{11}$. By the rule of matrix multiplication, I have: $c_{ij} = \sum_{k=1}^{n} a_{ki}^0 a_{kj}^0$. Variable x_i is included into the k^{th} equation if and only if $a_{ki}^0 \neq 0$. If x_j is excluded from each equation where x_i is included, then for each k such that $a_{ki}^0 \neq 0$ I have $a_{kj}^0 = 0$, so each summand is zero, and $c_{ij} = 0$. If there exists equation k where both variables are included then $a_{ki}^0 a_{kj}^0 \neq 0$, and the sum is zero only if this summand is exactly offset by the other non-zero summands, which does not happen in almost all parameter points.

The proof for blocks $\mathbf{C}_{12}$, $\mathbf{C}_{21}$ and $\tilde{\mathbf{C}}_{22}$ is the same as for block $\mathbf{C}_{11}$. □

Let *marker* for a particular equation be a variable, which is included only into this structural equation. For example, $\mathbb{L}r$ is marker for the first structural equation in Table 1, associated with the Taylor rule. I use the following two properties of the markers to check the identifying assumptions, which directly follow from Proposition 1:

Corollary 1.1 (Markers)**.** *In almost all parameter points, marker associated with structural equation i is adjacent in the TCNW only to the variables included into equation i. Markers associated with different structural equations are not adjacent.*

Corollary 1.2 (Sufficient condition for full testable identification)**.** *Assume that there exists at least one marker associated with each structural equation. Then each inclusion and exclusion restriction used for the identification of the structural model can be tested in almost all parameter points.*

4.3. **Estimation of the TCNW.** There are two steps of the estimation of the TCNW. At the first step I use block bootstrap to test individual hypotheses associated with particular edges in the TCNW. At the second step I make a correction for the multiple hypothesis testing problem.

Consider the first step. For each i and j, $i \neq j$, the individual null hypothesis is that the respective entry of the transformed concentration matrix is zero, $\mathcal{H}_0 : \tilde{c}_{ij} = 0$, and the alternative is $\mathcal{H}_1 : \tilde{c}_{ij} \neq 0$.

At each iteration of the bootstrap procedure I repeat the following steps:

(1) Construct a resample X_1 by shuffling overlapping blocks of X of length[2] 4.

[2]There is no consensus in the literature how to choose the length of one block for the block bootstrap procedure. I use the recommendation to include approximately $\frac{3}{4}\sqrt[3]{T}$ observations into each block, where T is the total number of observations.

(2) Centralize X_1, that is, subtract the mean.

(3) Calculate the empirical concentration matrix, $\mathbf{C}_1 = \left(\frac{X_1^T X_1}{T-1}\right)^{-1}$, where T is the number of observations.

(4) Calculate the transformed concentration matrix using Lemma 1.

Let freq_{ij} be the number of iteration of the bootstrap procedure where $\tilde{c}_{ij}$ is positive minus the number of iterations where $\tilde{c}_{ij}$ is negative divided by the total number of iterations where $(X^T X)$ is not singular. The p-value associated with the null hypothesis is calculated using:

$$\text{pvalue}_{ij} = 1 - \left|\text{freq}_{ij}\right| \tag{12}$$

The p-value is not calculated for $i = j$.

At the second step I use the procedure of Benjamini and Hochberg (1995) to account for the multiple hypothesis testing problem. Let $m = n(p+1)$ be the number of variables in vector X, where n is the number of variables in Y_t, and p is the number of lags. There are $N = m\,(m-1)\,/2$ individual hull hypotheses, so a multiple hypothesis testing procedure is required. Let *total discoveries* be the number of rejected null hypotheses, and *false discoveries* be the number of wrongly rejected null hypotheses. The false discovery rate is defined as the ratio of the false discoveries to the total discoveries if the number of total discoveries is positive, and defined to be zero if the number of total discoveries is zero. Benjamini and Hochberg (1995) prove that the following procedure controls the expected false discovery rate below or at level q^*: estimate individual p-values $p_1, p_2, \ldots, p_N$, sort p-values in increasing order $p_1 \le p_2, \le \ldots, \le p_N$, find the largest k for which $p_k < \frac{k}{N} q^*$, reject the null hypotheses associated with $p_1, \ldots, p_k$, and accept the null hypotheses associated with $p_{k+1}, \ldots, p_N$. The q-value associated with the i^{th} null hypothesis is defined as the minimal value of q^* for which the i^{th} null hypothesis is rejected.

Benjamini and Hochberg (1995) assume independence of individual null hypotheses, but this assumption may be violated when the TCNW is estimated. That is, when two null hypothesis are true and one of them is rejected, this may be more likely that the other hypothesis is also rejected. To see why, assume the model has been generated by T independent realizations of $\mathcal{E}$ written into $n \times T$ matrix $\hat{\mathcal{E}}$, and let $\hat{\mathbf{E}}$ be defined by $\hat{\mathbf{E}} = \left(\hat{\mathcal{E}}\hat{\mathcal{E}}^T\right)^{-1}$. By construction, the expected value of $\hat{\mathbf{E}}$ is the identity matrix. Each element of the empirical transformed concentration matrix can be expressed as $\hat{\tilde{c}}_{ij} = \sum_{k,l} p_{ik} \cdot p_{jl} \cdot \hat{e}_{kl}$, where $\hat{e}_{kl}$ is the respective entry of $\hat{\mathbf{E}}$. If the null hypothesis is correct for some $\tilde{c}_{i_1 j_1}$ and $\tilde{c}_{i_2 j_2}$, then $\mathbb{E}\left(\tilde{c}_{i_1 j_1}\right) = \mathbb{E}\left(\tilde{c}_{i_2 j_2}\right) = 0$. However, if for a particular realization of the residuals the linear combination of $\hat{e}_{kl}$ defining $\hat{\tilde{c}}_{i_1 j_1}$ is outside the confidence interval, this may be more likely that another linear combination of $\hat{e}_{kl}$, defining $\hat{\tilde{c}}_{i_2 j_2}$, is also

Table 2. Markers

Equation index, structural shock	Equation name	Markers
1. ε^r	Taylor rule	$\mathbb{L}r$, $\mathbb{L}^2r$
2. ε^u	Unemployment equation	$\mathbb{L}^2u$
3. ε^c	Capacity utilization equation	$\mathbb{L}^2c$
4. ε^g	GDP growth rate equation	$\mathbb{L}g$, $\mathbb{L}^2g$
5. ε^π	Phillips curve for GDP deflator inflation	no markers
6. $\varepsilon^{\pi c}$	Phillips curve for the commodity price inflation	no markers

outside of the confidence interval. Therefore, the test statistics may have positive dependency. Benjamini and Yekutieli (2001) prove that the procedure described above correctly controls the expected false discovery rate at the level below or equal to q^* in the case of positive dependency.

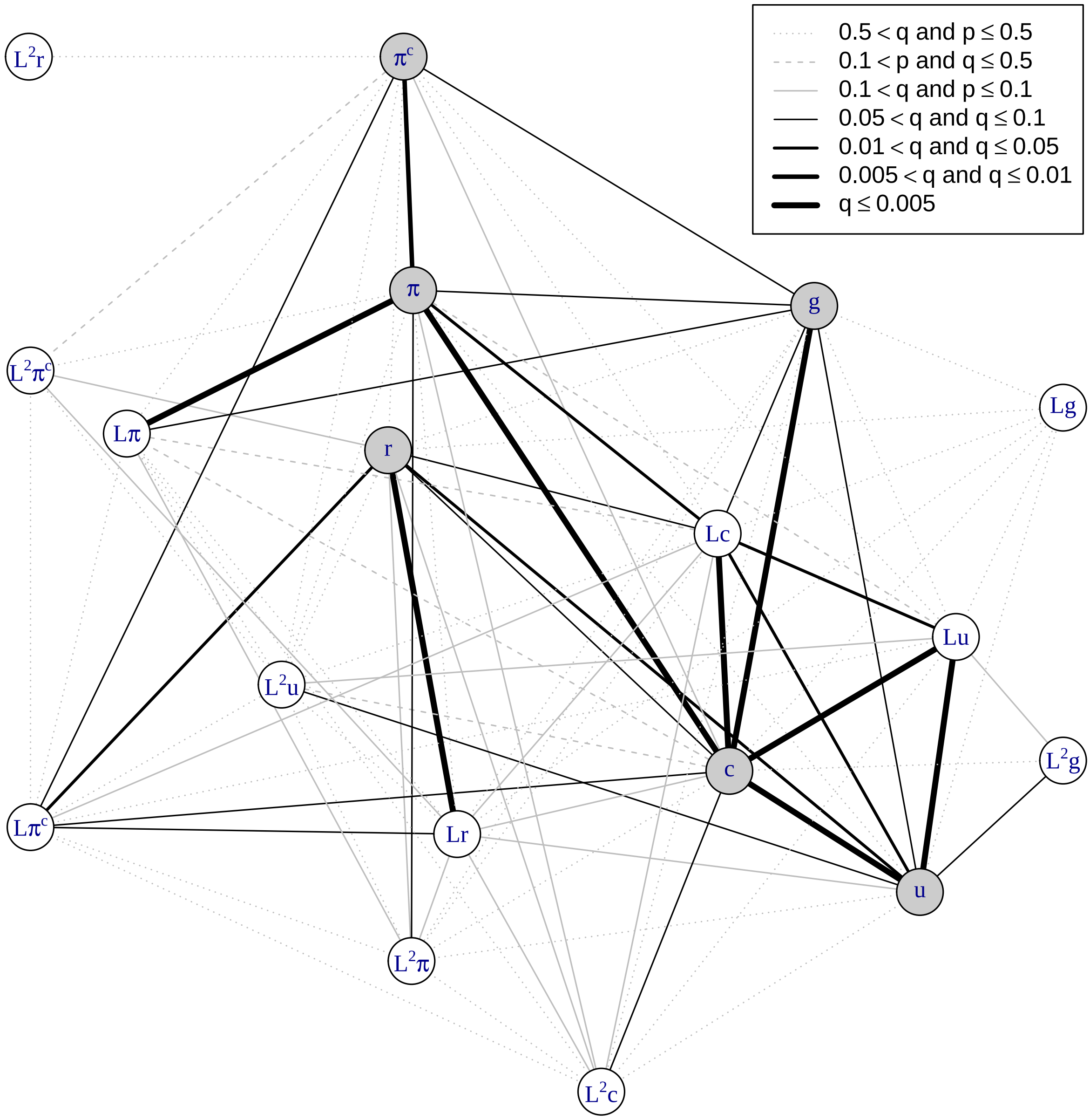
4.4. **Verifying identification restrictions.** Now I use the TCNW to test the identification restrictions summarized in Table 1. Based on these restrictions, the markers available for each structural equation are listed in Table 2. Each of the first four equations has at least one marker, so the theoretical TCNW suffices to distinguish between these equations. The Phillips curves for the GDP deflator inflation and for the commodity price inflation, however, do not have markers. Moreover, in Table 1 I see that each variable entering into one of the Phillips curve equation also enters into the Taylor rule. Therefore, identification restrictions summarized in Table 1 produce a TCNW that can separate the Taylor rule from the 2nd, 3rd, and 4th equation, but the TCNW cannot separate the Taylor rule from the Phillips curve equations. That is, the identification of the Taylor rule is only partially testable.

The estimated concentration network is depicted in Figure 1. Black edges in this figure are significant at the 10% q-value level, and the thicker is the edge the lower is the associated q-value. Gray solid edges are significant at the 10% p-value level, but they are not significant when I make the correction for the multiple hypothesis testing. Gray dashed and dotted edges are not significant even individually.

Observe that $\mathbb{L}^2r$ and $\mathbb{L}g$ are not adjacent significantly to any other node. Using the terminology of instrumental variables, these instruments are either weak or irrelevant, so they cannot be used as markers. However, even without these instruments each of the first four structural equations has a marker listed in Table 2.

To check the identifying assumptions summarized in Table 1, verify that the properties of the markers predicted by Corollary 1.1 are not rejected. Namely, make the following two checks. First, verify that the markers associated with different equations are not adjacent. Second, verify that the marker associated with each equation is adjacent only to the variables included into this equation. In Figure 1 I see that some edges

Figure 1. Transformed concentration network.
Notation p is used for the p-values associated with individual hypotheses, and q is the upper bound for the expected false discovery rate.



predicted by the identifying assumptions summarized in Table 1 are not significant; however, the imposed exclusion restrictions are not rejected by the TCNW.

How does the interest rate respond to the inflation shocks? The TCNW in Figure 1 provides no evidence that there is at least one structural equation that includes the contemporary values of r and π. For example, if the contemporary value of inflation where included into the monetary policy rule equation, π would be adjacent to r and $\mathbb{L}r$. Similarly, if r where included into the Phillips curve equation, it would be adjacent to π and $\mathbb{L}\pi$. There is no these edges in the estimated TCNW, so, the direct effects between r and π are either weak or absent. However, there is a weak evidence that r is affected by the second lag of π: edge $r - \mathbb{L}^2\pi$ is significant at 10% p-value level, although it is not significant when I make the correction for the multiple hypothesis testing procedure. Theory, however, predicts that r responds to shocks in π, so r must be adjacent to at least one of the lags of π. I interpret this observation as a weak evidence that the federal reserve uses the delayed information about π in its policy decisions.

A similar reason applies to the commodity price inflation. The commodity price inflation is not adjacent to the contemporaneous values of r or $\mathbb{L}^2r$. However, edge $r - \mathbb{L}\pi^c$ is significant at 1% q-value level. I interpret this observation as a strong evidence that the federal reserve uses delayed information about π^c in policy making.

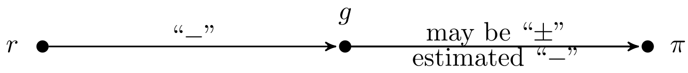
5. Price puzzle diagnostics

As Bernanke and Federal Reserve Board (2005) argue, the monetary policy rule cannot be identified within low dimensional models. In Figure 1, the lagged interest rate, which is marker for the Taylor rule equation, is adjacent to $\mathbb{L}\pi^c$, and there is weak evidence that it is adjacent to u, c, and $\mathbb{L}^2\pi$. Therefore, there is weak evidence that the Taylor responds in different way to four different signals, which implies that it includes at least 5 variables. If I estimate a model with less than 5 variables, then this four-dimensional signal is compressed to fit a lower dimensional space before it reaches the interest rate. The residuals of this compression are added with some weights to the residuals of the estimated monetary policy rule equation, so the residuals of the estimated monetary policy equation are contaminated with other shocks. The estimated IRFs for the monetary policy shock in this case are biased.

A particular bias produced by the signal dimensionality reduction depends on particular restrictions used for identification. In this Sections I explain how a set of restrictions consistent with the TCNW produces the price puzzle bias in low dimensional models. That is, small models predict that in response to a restrictive monetary policy shock the inflation rate does not decline monotonically, as the theory predicts, but it initially goes up, and only then starts to decrease. The price puzzle has been reported in many low dimensional SVAR models, which motivates me to diagnose the price puzzle problem using the TCNW introduced in this paper.

5.1. **Price puzzle bias in a three-variable model.** Consider the model that includes the federal interest rate r, the GDP growth rate g, and the inflation rate π. Observe that in TCNW depicted in Figure 1 there

FIGURE 2. Transmission channel of monetary policy and signal loss in the three-variable model.



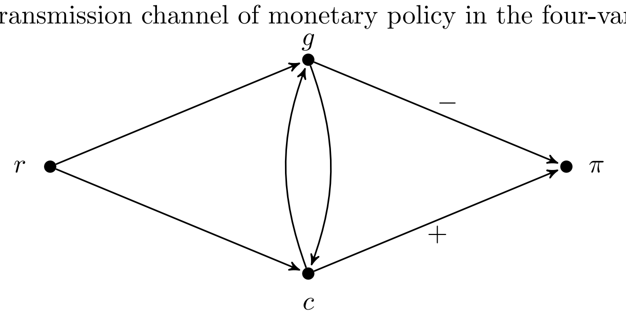
is no evidence that r directly affects the contemporary value of π, so this effect is either weak or absent in the true data-generated model. The same property holds for the TCNW estimated for the three-variable model, and in all specifications that I have considered. To be precise, assume that the direct effect of r on π is absent in the true data-generating model, that is, $r \not\to \pi$.

Consider the transmission mechanism from the monetary policy to the inflation rate depicted in Figure 2. By the identification assumption discussed above, there is no direct effect of r on π, so this effect is fully mediated by g. In this case, the contemporaneous effect of r on π is the product of the effect of r on g and of the effect of g on π. I interpret an exogenous increase in the federal interest rate as a restrictive monetary policy shock, which decreases the output. Therefore, the effect of r on g is negative. The theory predicts that the effect of g on π is ambiguous. If the GDP growth rate has increased because of an aggregate demand shock, the inflation will go up, but if it has increased because of an aggregate supply shock, the inflation will go down. In the estimated model, the effect of r on g and the effect of g on π are both negative.

In spite of the fact that the estimated effects of r on g and of g on π are consistent with the theory, the contemporaneous effect of the monetary policy shock on the inflation rate, which equals to the product of these two effects, is not consistent with the theory. An increase in the interest rate decreases the output, but this effect is misclassified in the estimated model as a negative aggregate supply shock, so the model predicts that the inflation will go up. This is how too low dimensionality of the estimated model can produce the price puzzle.

5.2. **A solution to the price puzzle problem.** Consider the model with r, c, g and π, where I use the same identification assumption, $r \not\to \pi$. In this model, the influence of r on π is mediated by two variables, c and g, and since the mediator is two-dimensional, it can can distinguish between the aggregate demand and aggregate supply shocks. The transmission mechanism from the monetary policy to the inflation rate for this model is depicted in Figure 3. A shock that increases the GDP growth rate without affecting the capacity utilization is interpreted as an aggregate supply shock, so the expected inflation decreases. If the capacity utilization rate has increased but the GDP growth rate has not changed, this is interpreted as a simultaneous positive aggregate demand shock and a negative aggregate supply shock, so the inflation will go up. Since the mediator can distinguish between the aggregate demand and aggregate supply shocks, it

Figure 3. Transmission channel of monetary policy in the four-variable model.



can correctly classify an exogenous increase in r as an aggregate demand shock, and the price puzzle effect diminishes.

The IRFs for the four-variable model are closer to the prediction of the macroeconomic theory than the IRFs for the three-variable model. In particular, the price puzzle effect is not significant in four-variable model. Nevertheless, these impulse response functions are also not entirely consistent with the theory. From the theoretical perspective I expect that in response to a restrictive monetary policy shock the capacity utilization decreases. In the estimated model, however, it first jumps up, and only then smoothly decreases.

A possible explanation for this result is the confounding effect. For example, the unemployment rate and the commodity price inflation may affect the contemporaneous values of r and c. If this is the reason why I observe the jump in the capacity utilization rate, I need to include the confounders into the estimated model to avoid the bias. When I include the unemployment rate and the commodity price inflation rate, the positive jump in the capacity utilization rate disappears. The estimated model, therefore, includes r, u, c, g, π, and π^c.

6. Estimation of the impulse response functions

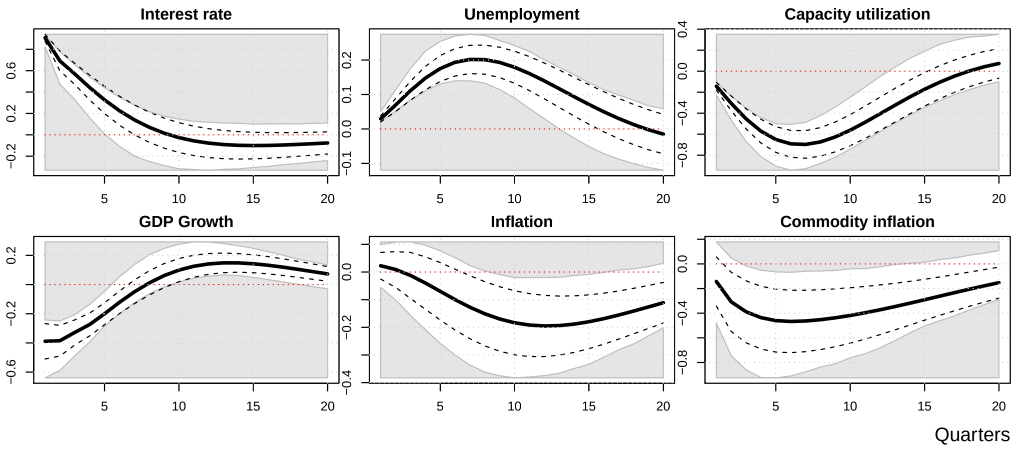
The parameters of the structural model are estimated using the maximum likelihood estimator. As the starting point for the likelihood maximization I use parameters estimated by two-stage least squares procedure. The confidence intervals are estimated by bootstrapping the residuals.

The estimated IRFs for the monetary policy shock are reported in Figure 4a.The black thick line in each panel of Figure 4a represents the mean, the dashed lines lie with one standard deviation from the mean, and the gray zone indicates the outside 90% bootstrap confidence intervals.

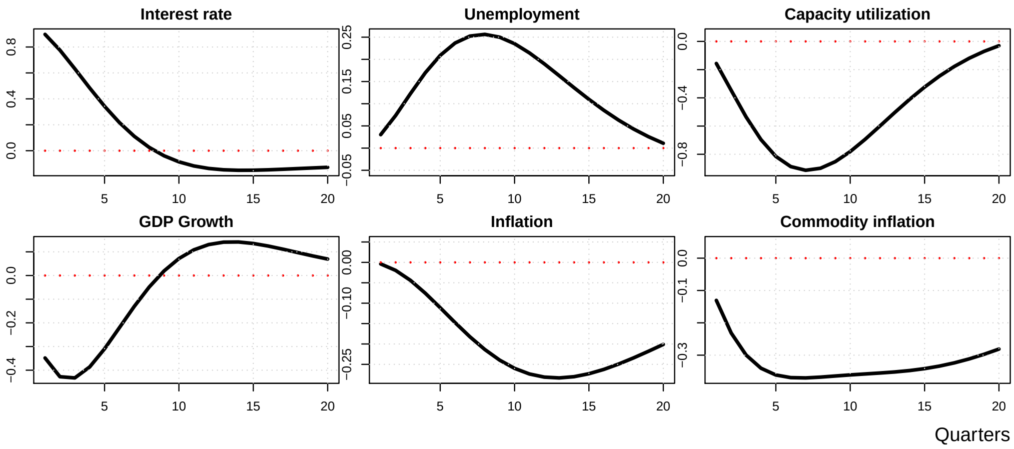
The estimated IRFs are consistent with the macroeconomic theory. As the theory predicts, a restrictive monetary policy shock increases the unemployment and decreases the capacity utilization. The GDP growth rate is below its long-run level just after the shock, but it rises above the long-run level two years later, so the GDP level eventually recovers. The inflation rate immediately goes down. The commodity price inflation

Figure 4. Estimated impulse response functions for the restrictive monetary policy shock. The thick line in each panel is the expected impulse response function, the dashed lines lay within one standard deviation from the mean, and the gray zone indicate the outside 90% confidence intervals.

a) Theoretical identification



b) Robustness check: pure data-based identification



also decreases immediately, but faster than the deflator inflation, which is consistent with the believe that the commodity price inflation is a leading indicator for the deflator inflation (Garner (1989)).

To check the robustness of the identifying assumptions, I make the following two experiments. First, I verify that relaxing a few random exclusions and adding a few random TCNW-consistent exclusions only slightly modifies the estimated IRFs. Second, I consider the pure data-based identification, where I do not use any theoretical identifying assumptions, and rely only on assumptions implied by the estimated TCNW. Namely, I restrict to zero all coefficients associated with the edges insignificant at 10% q-value level. For example, since the edge between the contemporaneous values of π and r is not significant in the TCNW on Figure 1, parameters a^0_{51} and a^0_{15} associated with the direct effects $r \to \pi$ and $\pi \to r$ are restricted to zero. The threshold at the 10% q-value level is arbitrary, however, the estimated IRFs do not change too much when I vary the threshold between 10% and 50%. I do not report the confidence intervals for the pure data-based identification, because estimation of them would involve the sequential hypothesis testing problem.

The IRFs for the pure data-based identification are shown in Figure4b. The IRFs for the data-based and theoretical identification are slightly different in the magnitudes. Nevertheless, in the considered application, the pure data-based identification produces a reasonable approximation for the theoretical identification.

7. Conclusions

I use restrictions produced by a typical DSGE model for identification of the SVAR. The literature on SVARs usually imposes restrictions on the impulse response functions and on the matrix of contemporaneous effects, making it sparse. The DSGE literature, however, predicts that the matrix of contemporaneous effects is dense, but the matrix of lagged effects is sparse. I impose restrictions on the matrix of lagged rather than contemporaneous effects, which is consistent with the DSGE literature.

A typical DSGE model produces more restrictions than is required for identification, so many identifying restrictions are testable. I use the TCNW for testing the overidentified restrictions. In the considered application, theoretically justified TCNW can separate the Taylor rule equation from three other equations of the structural model.

In the estimated TCNW, there is no evidence that the Federal Reserve takes into account the contemporary values of inflation and commodity price inflation in policy making. However, there is weak evidence that it uses delayed information about the inflation, and strong evidence that it uses delayed information about the commodity price inflation.

The method reveals the misidentification problems producing the price puzzle in the previous literature. The misidentification arises as a result of the loss of dimensionality of the structural model. Using theoretical identification restrictions, which are not rejected by the data, I achieve a robust identification of the monetary policy rule. The estimated IRFs are consistent with the theory, producing no anomalies such as the price puzzle.

8. Computational details

The SVAR model was estimated in `R` (R Core Team, 2012) using packages `igraph` (Csardi and Nepusz, 2006) and `NLopt` (Johnson, 2014).

References

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching Up with the Joneses," *American Economic Review*, 80, 38 – 42, papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association.

Ahelegbey, D. F., M. Billio, and R. Casarin. (2014): "Bayesian Graphical Models for Structural Vector Autoregressive Processes," *Journal of applied Econometrics*, forthcoming.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.

Benjamini, Y. and D. Yekutieli (2001): "The Control of the False Discovery Rate in Multiple Testing under Dependency," *The Annals of Statistics*, 29, 1165–1188.

Bernanke, B. S. and Federal Reserve Board (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 422.

Bils, M. and P. J. Klenow (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.

Blanchard, O. J. and D. Quah (1993): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 83, 655–673.

Bryant, H. L. and D. A. Bessler (2015): "Conditions Sufficient to Infer Causal Relationships Using Instrumental Variables and Observational Data," *Computational Economics*, forthcoming.

Calvo, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12, 383 – 398.

Campbell, J. Y. and J. H. Cochrane (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107, 205 – 251.

Chen, B. and J. Pearl (2014): "Graphical Tools for Linear Structural Equation Modeling," *Psychometrica*, forthcoming.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamics Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1 – 45.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What Extend?" in *Handbook of Macroeconomics, vol. 1A*, North Holland: Amsterdam.

Clarida, R., J. Galí, and M. Gertler (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, 1661 – 1707.

Csardi, G. and T. Nepusz (2006): "The igraph software package for complex network research," *InterJournal*, Complex Systems, 1695.

Demiralp, S., K. Hoover, and S. Perez (2014): "Still puzzling: Evaluating the price puzzle in an empirically identified structural vector autoregression," *Empirical Economics*, 46, 701–731, cited By 0.

Fragetta, M. and G. Melina (2013): "Identification of monetary policy in SVAR models: a data-oriented perspective," *Empirical Economics*, 45, 831–844.

Fuhrer, J. C. (2000): "Habit formation in consumption and its implications for monetary-policy models," *The American Economic Review*, 90, 367 – 390.

Garner, C. A. (1989): "Commodity Prices: Policy Target or Information Variable?: Note," *Journal of Money, Credit and Banking*, 21, 508–514.

Gertler, M., L. Sala, and A. Trigari (2008): "An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining," *Journal of Money, Credit and Banking*, 40, 1713–1764.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hanson, M. S. (2004): "The "price puzzle" reconsidered," *Journal of Monetary Economics*, 51, 1385–1413.

Hoover, K. D. (2005): "Automatic Inference of the Contemporaneous Causal Order of a System of Equations," *Econometric Theory*, 69–77.

Johnson, S. G. (2014): "The NLopt nonlinear-optimization package," http://ab-initio.mit.edu/nlopt.

Keating, J. W., L. J. Kelly, and V. J. Valcarcel (2014): "Solving the price puzzle with an alternative indicator of monetary policy," *Economics Letters*, 124, 188 – 194.

Koller, D. (2009): *Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)*, The MIT Press.

Kwon, D.-H. and D. Bessler (2011): "Graphical Methods, Inductive Causal Inference, and Econometrics: A Literature Review," *Computational Economics*, 38, 85–106.

Litterman, R. (1979): "Techniques of forecasting using vector autoregressions," *Federal Reserve of Minneapolis Working Paper*.

Merz, M. (1995): "Search in the labor market and the real business cycle," *Journal of Monetary Economics*, 36, 269 – 300.

Oxley, L., M. Reale, and G. Wilson (2009): "Constructing Structural VAR models with conditional independence graphs," *Mathematics and Computers in Simulation*, 79, 2910–2916.

Pearl, J. (2009): *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 2nd ed.

Phiromswad, P. (2014): "Measuring monetary policy with empirically grounded identifying restrictions," *Empirical Economics*, 46, 681–699, cited By 0.

Pissarides, C. A. (2000): *Equilibrium Unemployment Theory*, MIT Press, Cambridge, Mass., USA., 2nd ed.

R Core Team (2012): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Reale, M. and G. Wilson (2001): "Identification of vector AR models with recursive structural errors using conditional independence graphs," *Statistical Methods and Applications*, 10, 49–65.

Sargan, J. D. (1958): "The Estimation of Economic Relationships using Instrumental Variables," *Econometrica*, 26, 393–415.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

——— (1986): "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2 – 16.

——— (1992): "Interpreting the macroeconomic time series facts: the effects of monetary policy," *European Economic Review*, 36, 2–16.

Sims, C. A. and T. Zha (2006): "Were There Regime Switches in US Monetary Policy?" *American Economic Review*, 96, 54–81.

Smets, A. and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 97, 586 – 606.

Smets, F. and F. Wouters (2003): "An estimated stochastic general equilibrium model of the Euro Area," *Journal of the European Economic Association*, 1, 1123 – 1175.

Uhlig, H. (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Wilson, G. T. and M. Reale (2008): "The sampling properties of conditional independence graphs for I(1) structural VAR models," *Journal of Time Series Analysis*, 29, 802–810.

Zha, T. (1999): "Block Recursion and Structural Vector Autoregressions," *Journal of Econometrics*, 90, 291–316.

NIKOLAY AREFIEV
The National Research University Higher School of Economics. E-mail: n.are_ev@gmail.com. Address: o_ce 708A, Shabolovka
31, 119049, Moscow, Russia.

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.