



**UNIVERSITY
OF TURKU**

Turku School of
Economics

PUZZLING PRICES:

**Information insufficiency and misidentification of monetary policy shocks in VAR
models**

Department of Economics

Master's thesis

Author:

Juho Pitkäranta

Supervisor:

Prof. Jouko Vilmunen

April 8, 2024

Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Master's thesis

Subject: Economics

Author: Juho Pitkäranta

Title: Puzzling Prices: Information insufficiency and misidentification of monetary policy shocks in VAR models

Supervisor: Professor Jouko Vilmunen

Number of pages: 68 pages + appendices 11 pages

Date: April 8, 2024

Traditional low-dimensional VAR models do not contain enough information to ensure reliable identification of monetary policy shocks and thus tend to produce a price puzzle. The often used commodity price and exchange rate controlling is found to be negligible at best and evidence is shown that the seeming successfulness of traditional models can be attributed to trend components turning the used economic activity variable to a crude output gap estimate. Traditional Cholesky identification is found to be unjustified for most applications and to underestimate the effect sizes of monetary shocks especially in lower dimensional models. Models including a measure for the output gap, data-rich FAVAR models and statistical identification are found to be more successful in identifying monetary shocks. These models indicate, that there is no price puzzle for the euro area. A 25 basis point monetary shock is found to have a 10 to 15 basis point negative effect to the euro area HICP inflation rate, peaking around 12 months and converging to zero around 30 months after impact.

Key words: price puzzle, vector autoregression, monetary policy, statistical identification, information sufficiency.

Pro gradu -tutkielma

Oppiaine: Taloustiede

Tekijä: Juho Pitkäranta

Otsikko: Hintapähkinä: informaatiovaje ja rahapolitiikkashokkien identifiointi VAR-malleissa

Ohjaaja: Professori Jouko Vilmunen

Sivumäärä: 68 + liitteet 11 sivua

Päivämäärä: 1.4.2024

Perinteisten mataladimensioisten VAR-mallien informaatiojoukko ei mahdollista rahapolitiikkashokkien luotettavaa identifiointia, jonka takia ne tuottavat usein niin kutsutun hintapähkinän. Tulokseni näyttävät, että informaatiojoukon laajentaminen tavanomaisilla hyödykehinta- ja valuuttakurssikontrolleilla ei tuota merkittävää parannusta tähän, ja perinteisten mallien onnistuminen selittyy malleissa käytetyillä trendikomponenteilla, jotka muuttavat käytetyn tuotantomuuttujan lineaariseksi tuotantokuiluestimaatiksi. Empiiriset tulokseni hylkäävät rekursiivisen Cholesky-identifikaation lähes jokaisessa mallissa. Lisäksi Cholesky-identifikaatio vaikuttaisi tuottavan todellista pienempiä estimaatteja mataladimensioisissa malleissa. Tuotantokuilun sisältävät mallit, isoista aineistoista estimoidut FAVAR-mallit sekä tilastollinen identifikaatio mahdollistavat rahapolitiikkashokkien identifikaation luotettavammin. Näiden mallien tuottamien tulosten perusteella euroalueella ei ole havaittavissa hintapähkinää. Tulosteni perusteella 25 korkopisteen rahapolitiikkashokilla on 10-15 korkopisteen negatiivinen vaikutus inflaatioasteeseen noin 12 kuukauden viiveellä, jonka jälkeen vaikutus konvergoituu nolnaan noin 30 kuukauden mennessä.

Avainsaat: hintapähkinä, vektoriautoregressio, rahapolitiikka, tilastollinen identifikaatio, informaatiovaje.

TABLE OF CONTENTS

1	Introduction	10
1.1	The Price Puzzle	10
1.2	The aim of this thesis	10
2	Theoretical background	13
2.1	What exactly is a monetary policy shock?	13
2.2	Model setup	13
2.3	Impulse response functions	14
2.4	The confidence interval problem	15
2.5	Fundamentalness	16
2.6	What to expect? Implications from the IS-MP framework	17
3	Previous literature	20
3.1	Model specification	20
3.1.1	Indicators for future inflation	20
3.1.2	The output gap, Phillips curve and inflation	22
3.1.3	Different monetary policy regimes	23
3.1.4	Factor augmentation	25
3.1.5	The cost channel explanation	28
3.2	Identification	32
3.2.1	The long lags of monetary policy	33
3.2.2	Long run restrictions in a VECM	34
3.3	Statistical identification	35
3.3.1	Structural breaks in variance regimes	36
3.3.2	SVAR-GARCH	37
3.3.3	Smooth transition in variance regimes	38
3.3.4	Non Gaussian maximum likelihood	39
3.4	Conclusions from previous literature	41
4	Empirical evaluation	42
4.1	The Benchmark model	42
4.2	A Mainstream model	44
4.3	Including the output gap	46
4.3.1	A note on linear trends	46
4.3.2	A more sophisticated approach	48
4.4	The FAVAR approach	49
4.4.1	One common factor	50
4.4.2	Four common factors	52
4.4.3	D-FAVAR expansion	54
4.4.4	Some notes on the estimation of common factors	57
4.5	Practical concerns regarding integrated variables	57
4.6	Some notes on the identified monetary shocks	60

5	Concluding remarks.	63
	References	65
A	Appendix: Estimated impulse responses	69

Abbreviations

BBE	Bernanke, Boivin and Eliaz
BN	Beveridge-Nelson decomposition
BoF	Bank of Finland
BVAR	Bayesian vector autoregression
CEE	Christiano, Eichenbaum, and Evans
D-FAVAR	Dynamic factor augmented vector autoregression
DFM	Dynamic factor model
DGP	Data generation process
DSGE	Dynamic stochastic general equilibrium
ECB	European Central Bank
FAVAR	Factor augmented vector autoregression
FED	The Federal reserve system
FFR	Fed Funds Rate
GARCH	Generalized autoregressive conditional heteroskedasticity
NAIRU	Non-accelerating inflation rate of unemployment
NGML	Non-Gaussian maximum likelihood
ST	Smooth transition
SVAR	Structural vector autoregression
VAR	Vector autoregression
VECM	Vector error correction model
VMA	Vector moving average

List of Figures

Figure 1	The cost channel in action	29
Figure 2	Simulated example of Gaussian and non-Gaussian shocks, $n=1000$	40
Figure 3	Benchmark impulse response estimates, Cholesky identified with 0.9-percent moving block bootstrap intervals (grey), 1000 simulations.	43
Figure 4	Benchmark impulse response estimates, statistical identification.	43
Figure 5	Exchange rates as inflation expectation proxy. Grey areas 0.9-percent moving block bootstrap intervals, 1000 simulations.	45
Figure 6	Commodity prices and effective Fed Funds rate as exogenous controls, Cholesky identified.	46
Figure 7	Benchmark and mainstream models estimated with linear trends. Moving block 90-percent bootstrap intervals in grey, 1000 simulations.	47
Figure 8	Benchmark model with quarterly data and the real GDP, Cholesky identified. Moving block 90-percent bootstrap intervals in grey, 1000 simulations.	47
Figure 9	Benchmark and mainstream models estimated with linear trends, Cholesky identified. Moving block 90-percent bootstrap intervals in grey, 1000 simulations.	48
Figure 10	FAVAR with one common factor, Cholesky identified. Grey areas 0.9-percent moving block bootstrap intervals, 1000 simulations.	50
Figure 11	FAVAR with one common factor, unrestricted statistical identification.	51
Figure 12	FAVAR with one common factor, restricted statistical identification.	51
Figure 13	FAVAR with four common factors, Cholesky identified. Grey areas 0.9-percent moving block bootstrap intervals, 1000 simulations.	52
Figure 14	FAVAR with four common factors, statistical identification and additional restrictions.	52
Figure 15	Impulse responses for selected variables in X_t estimated from FAVAR with four common factors.	54
Figure 16	D-FAVAR with four common factors. Grey areas 90% moving block bootstrap intervals, 1000 simulations.	55
Figure 17	Comparison of static and dynamic estimates for four common factors, normalized to mean zero and unity standard deviation.	56
Figure 18	Impulse responses for selected variables in X_t estimated from D-FAVAR with four common factors.	56
Figure 19	FAVEC with four common factors.	59
Figure 20	Euro area HICP inflation between 2000 and 2020.	60

Figure 21	Some monetary shocks identified from traditional low-dimensional models.	61
Figure 22	Some monetary shocks identified from better performing models.	62
Figure 23	Benchmark model, Cholesky identified	69
Figure 24	Benchmark model, GARCH identified	70
Figure 25	Benchmark model, NGML identified	70
Figure 26	Exchange rate model, Cholesky identified	71
Figure 27	Exchange rate model, GARCH identified	71
Figure 28	Model with linear estimate for the output gap, Cholesky identified	72
Figure 29	Model with linear estimate for the output gap, GARCH identified	72
Figure 30	VAR(1) with BN-BVAR estimate for the output gap, Cholesky identified	73
Figure 31	VAR(2) with BN-BVAR estimate for the output gap, Cholesky identified	73
Figure 32	One common factor FAVAR, Cholesky identified	74
Figure 33	One common factor FAVAR, ST identified	74
Figure 34	Four common factor FAVAR, NGML identified	75
Figure 35	Four common factor FAVAR, Cholesky identified	76
Figure 36	Four common factor FAVAR, partially restricted ST identified	77
Figure 37	Four dynamic common factor FAVAR, partially restricted NGML identified	78
Figure 38	Four dynamic common factor FAVEC, partially restricted NGML identified	79

List of Tables

Table 1	Information criteria for different controlled model specifications.	45
Table 2	Principal components estimated from a large data set of macroeconomic variables	49
Table 3	Johansen procedure test statistics for the dataset used in the D-FAVAR model.	59

1 Introduction

1.1 The Price Puzzle

Ever since VAR models were introduced to the econometric literature by Sims (1980), they have become the workhorse of marcoeconometricians. However, to this day the question of the impact of monetary policy on inflation remains controversial. This question was first raised by Sims (1992) who pondered the initial rise in prices following a contractionary monetary policy shock. Against all intuition and theoretical models, an exogenous shock seemed to initially have a positive effect on prices which turned into a negative one only after several years after the shock. This counter intuitive finding has since been well documented in other works as well, which has led to the coining of the term “price puzzle”. Although several solutions to this problem have been proposed, to this day no convergence or consensus has been achieved. This is evident from the fact that long after Sims (1992, p. 981) used commodity prices and exchange rates to control for inflation expectations, new solutions are still proposed and evaluated, for example Giordani (2001), Bernanke, Boivin, and Elias (2005), Rusnák, Havranek, and Horváth (2013) and Estrella (2015). To add insult to injury, many of these attempts only seem dampen the price puzzle instead of removing it.

The price puzzle raises some fundamental questions regarding the research on dynamic effects of monetary policy with VARs. First: As discussed by Christiano, Eichenbaum, and Evans (1999), the consensus on the effects of monetary policy shocks is quite broad and non-specific. How are we to reliably evaluate the effects of monetary policy, when there is no consensus even on the requisites for specification? Second: if the puzzle is indeed the result of information insufficiency, what does it mean for other variable responses besides inflation? As noted by Bernanke, Boivin, and Elias (2005), information insufficiency would lead to the whole model being misspecified and thus the estimated responses of all the variables faulty. The identification of these models raise some questions too, as the generally *ad hoc* recursive identification procedures remain dominant to this day.

1.2 The aim of this thesis

As stated by Cochrane (2023), with monetary VARs it is important to be clear with the questions one is asking. Even if one were to uncover the structural relationships between

variables of interest, this doesn't necessarily answer the question of how ECB's next rate hike will effect the economy, as the hike should be endogenous given any coherent monetary rule. So why should we even be interested in these exogenous shocks? To answer this, let us take a parallel example from microeconometrics. Angrist and Lavy (1999, p. 550) find that both reading and math results correlate strongly with class sizes. Using an instrumental variable approach, the causality from class sizes to scores is found to be negative however. The initial positive correlation is attributed to selection biases, ie. endogenous variation. One could say the econometrician does not have the same information as the parents who select their school districts to get their kids to certain schools. Finding the causal relationship essentially entails forcing a modelling framework, where parents put their children into a particular school at random and then taking a particular kid, reassign them at random into a new class with different class size "for the heck of it" and watch what happens to their scores. One could be even tempted to call this random reassignment a "shock" and then, to loosely use the words of Cochrane (2023), say "but *the parents never do this*. Ask them. There are no shocks as defined".

So why should we care about this instrumented causal relationship? Put simply, if one would want to make educated decisions on educational policy, growing class sizes because the strong correlation between them and scores fits reality with its endogenous variation more realistically would obviously be a horrendous idea. In a similar manner, making educated monetary policy decisions obviously requires knowledge of the structural causal relations which are revealed by identified shocks. Knowing the effects of these shocks, be they how theoretical in nature as they be, is paramount: even if class sizes and scores correlate and the selective behaviour of parents may be much more interesting than a theoretical instrumentation using a near millennia old arbitrary rule, the causal link between them has to be acquired, as if there is no causal link, why should the policy maker even bother with restricting class sizes? In the same vain, if there is no sizable, relatively quick structural causal effect from monetary policy to inflation, why should a inflation targeting central bank even bother with an official interest rate as its instrument? Thus, making educated endogenous policy decisions require information of the causal effects, which are uncovered by identifying these shocks.

As VARs can be estimated by minimal assumptions compared to other model types, they are a powerful tool in the toolkit of an econometrician. They are also important in validating and calibrating other model types. As such, finding clear requisites for the specification and identification of monetary VARs is a crucial task.

The aim of this thesis is to review different solutions to the price puzzle. The scope of this review will be focused on model specification from the viewpoint of information suffi-

ciency and proper identification of shocks. In Section 2 I will present the VAR framework upon which this thesis will be built upon. In section 3 I will be reviewing solutions proposed in the existing econometric literature as well as more general potential pitfalls and solutions. In section 4 I will be evaluating empirically the performance of these solutions as well as offering my own. As most of the literature uses quarterly datasets from the US between the 50s and the 00s, the empirical section of this thesis will be conducted on monthly data from the euro area between 2000 and 2019: as more and more studies with mostly overlapping data is conducted, the more likely it is for a minor difference to produce a Type II error and thus the results to vary from chance alone.

The scope of this thesis will for the most part be empirical. As will be evident later, even theoretical models with reasonable empirical calibration provide too varying results to give us conclusive answers, indicating the question at hand to be empirical in nature in the end.

It is worth noting that whereas usually empirical research is conducted by first building reasonable assumptions and then evaluating the results, I will be turning this the other way around. In other words, the aim is to “brute force” the expected results and then evaluate the plausibility of the assumptions needed to reach the predetermined results.

2 Theoretical background

2.1 What exactly is a monetary policy shock?

Let's define monetary policy shocks in line with Christiano, Eichenbaum, and Evans (1999, p. 71) as

$$S_t = f(\Omega_t) + v_t, \quad (1)$$

where S_t is the instrument of the monetary authority (for example the official interest rates of the ECB), Ω_t is the informational set available to the monetary authority and f is a linear function capturing the behaviour and reactions to other variables of the monetary authority when setting the policy instrument (this could be the Taylor's rule for example). The monetary policy shock is the disturbance term v_t . In other words, it is the part not explained by the monetary authority's reaction function and the available information set. This of course implies that the shock is exogenous and orthogonal to other shocks.

But where does this shock come from? The policy instrument is controlled by the monetary authority and thus any change (or lack thereof) is by definition a conscious decision by them. As Cochrane (2023) put it, why would the FED just add 25 basis point to the appropriate rate "at random, just for the heck of it"? However, as outlined by Christiano, Eichenbaum, and Evans (1999, pp. 71–72), there are several reasonable interpretations for these shocks. First, stochastic shifts in the preferences of the monetary authority. For example, the composition of the board of a central bank could stochastically shift due to personal reasons. One could think of Tuomas Vähimäki taking Olli Rehn's seat in the ECB's governing council due to Rehn's presidential campaign (BoF 2023). Second, the monetary authority might give into public pressure and deviate from its general feedback rule. Third, economic variables usually get revised several times afterwards, as the initial values usually contain certain amount of measurement error. Thus when making decisions, the monetary authority is restricted to imprecise data corrected only after the fact and thus leads to skewed decisions.

2.2 Model setup

In line with most textbooks (see for example Kilian and Lütkepohl 2017, p. 23 or Lütkepohl 2005, p. 13) let $Y_t = (y_{1t} \ y_{2t} \ \dots \ y_{kt})'$ be a $k \times 1$ vector of variables at time t . By regressing these variables on the past values $Y_{t-1}, Y_{t-2} \dots$ we get the reduced form VAR(p)

model

$$Y_t = \mathbf{v} + \sum_{i=1}^p A_i Y_{t-i} + u_t, \quad (2)$$

where \mathbf{v} is the vector of constants, A_i is the matrix of coefficients relating the dynamic effects of the variables to each other at lag i and u_t is the error term vector with covariance matrix Σ_u . As outlined in Kilian and Lütkepohl (2017, pp. 25–26), a VAR(p) can be given a VMA(∞) representation

$$Y_t = \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-j} \quad (3)$$

where $\Phi_0 = I_k$ and $\Phi_i = \sum_{j=1}^i \Phi_{i-j} A_j$. In other words, observations Y_t can be presented as a linear combination of past errors.

The error terms of a reduced form model can be viewed as comprising from a large number of economically meaningful shocks and are thus (potentially) contemporaneously correlated. Thus, making any economically meaningful inference requires structuralisation of the reduced form model into a SVAR model in order to extract the shocks of interest. This can be done by setting $u_t = Bv_t$ turning model (2) into

$$Y_t = \mathbf{v} + \sum_{i=1}^p A_i Y_{t-i} + Bv_t, \quad (4)$$

where $\Sigma_v = I_k$ and B is the matrix relating the contemporaneous effects of the orthogonal shocks v_t . In order to infer the structural model, one has to *identify* the orthogonal shocks of interest from the noisy error term. By substituting the structuralisation of the error term into the VMA(∞) model (3) we get

$$Y_t = \mu + \sum_{j=0}^{\infty} \Theta_j v_{t-j}, \quad (5)$$

where $\Theta_j = \Phi_j B$. The subject of identification will be examined more closely in Section 3.2.

2.3 Impulse response functions

As seen, VMA representations model the VAR process as a linear combination of its past shocks. As apparent from (5), Θ_j maps the impact of shocks v_{t-j} to Y_t . Following Kilian and Lütkepohl (2017, pp. 108–111), the effect of a given shock at t to variable i at horizon

j is acquired from the corresponding coefficient $\theta_{j,ik}$, or

$$\frac{\partial y_{i,t+j}}{\partial v_{kt}} = \theta_{j,ik}.$$

Here impact matrices Θ_j are acquired recursively

$$\Theta_0 = \Phi_0 B_0 = I_K B_0 = B_0$$

$$\Theta_1 = \Phi_1 B_0$$

$$\Theta_2 = \Phi_2 B_0$$

etc.

In these lines it is customary to plot the impulse responses of the structural shocks from these estimates. Naturally, these can also be cumulated.

2.4 The confidence interval problem

It is well-known that constructing reliable confidence intervals for applied VAR models is challenging. As is demonstrated by Benkwitz, Neumann, and Lütkepohl (2000, pp. 76–8), the limiting behaviour for the asymptotic distribution of the response estimate differs whether the response is zero or non-zero. Many of the methods based on asymptotic theory also require distributional assumptions, which often become questionable with finite samples (Kilian 1998, pp. 2–7).

Due to these problems, confidence intervals are often constructed by bootstrapping rather than asymptotic theory. The idea is to approximate the distributions of interest based on their sample analogues and then use these to simulate confidence intervals. Thus they tend to be more general and accurate in small samples than their more theory based counterparts. For many types of model specification, formal confidence interval methods also just don't exist as of yet (Kilian and Lütkepohl 2017, p. 340). It should be noted however, that bootstrap intervals come with their own problems. The performance of different methods may differ greatly depending on the model and sample sizes (Kilian and Lütkepohl 2017, pp. 368–369). The presence of heteroskedasticity of some sort does also play an important role, and as outlined by Lütkepohl and Netšunajev (2017a, p. 7), generally little is known about the actual properties of bootstrapping methods in these instances. Thus, it is often better to view bootstrapped or theory based intervals as indicators for sampling uncertainty rather than classic confidence intervals.

2.5 Fundamentalness

As we saw from model (4), economically meaningful shocks are acquired by assuming the VAR residuals to be a linear combination of them. More formally, let Z_t be a stochastic process with the VMA representation

$$Z_t = \sum_{i=0}^{\infty} \Theta_i v_{t-i}$$

where Z_t is $N \times 1$ vector of the observable variables and v_t is $q \times 1$ vector of structural shocks driving the process. As the estimate for this VMA process is acquired by estimating a VAR, the estimates for these structural shocks are *by construction* acquired as a part of the VAR residuals. Thus, it is easy to see that if $N < q$, acquiring the actual structural shocks is not possible. In this situation of course, the estimates for a single shock would, in general, be tainted by several shocks. Formally put; whenever $v_t \in \overline{\text{span}}\{Z_{t-i}, i \leq 0\}$, v_t is *fundamental* with respect to Z_t (Alessi, Barigozzi, and Capasso 2008, p. 7). Thus one can see fundamentalness as a close concept to invertibility, as nonfundamentalness would require inverting the VMA in the future to acquire the structural shocks, which VAR estimation does not allow for. Here it is easy to see that the central source of nonfundamentalness is due to agents possessing a larger information set than the econometrician (Alessi, Barigozzi, and Capasso 2008, p. 5). One should note that the concept of fundamentalness and invertibility are not identical, as the later rules out roots with modulus equal to 1 (Forni, Gambetti, and Sala 2019, pp. 226–227). In a more general sense, nonfundamentalness is a case of the VMA polynomial determinant having at least one root inside the unit circle (Alessi, Barigozzi, and Capasso 2008, p. 7). The cause for this strands from the underlying economic process including a moving average component. Thus, identifying structural shocks from a VAR would require either *a priori* knowing the real economic model or ruling out these kinds of nonfundamental representations (Alessi, Barigozzi, and Capasso 2008, p. 12).

At this point it should be pointed out that although SVARs by construction identify as many shocks as there are variables, all of these do not have to be attached with a particular economic meaning. The source for these could be from measurement errors or otherwise macroeconomically uninteresting shocks and/or their combination. Thus, when identifying a model, fully labelled identification should not always be seen as necessary. As is well put by Forni, Giannone, et al. (2009, p. 1327), "...the question *why assume fundamentalness*, which is legitimately asked when $n = q$, is replaced by *why should we care about nonfundamentalness* when $n > q$ ".

2.6 What to expect? Implications from the IS-MP framework

As we have seen, empirical structural impulse responses are dependent on the model specification of the reduced form, ie. estimates of A_i and Φ_j , and the identification of the structural form, ie. the estimate of B . As both of these hinge on some sort of assumptions of the underlying structure modelled by the VAR, a wide range of different impulse response shapes is possible depending on these assumptions. Thus having an idea of what kind of impulse responses can be generated by reasonable theoretical assumptions is useful when assessing the plausibility of their empirical counterparts. To this end we should start by constructing some theoretical frameworks.

Let's start with a straight forward IS-MP model modified from Romer (2019, pp. 265–267). As there, let's simplify the analysis by normalizing constants and potential output to zero, so the output gap equals the realized output. Thus we have

$$\pi_t = \pi_{t-1} + \lambda y_t \quad (6)$$

$$r_t = by_t + v_t^M \quad (7)$$

$$y_t = \mathbb{E}_t(y_{t+1}) - \frac{1}{\theta} r_t. \quad (8)$$

Here equation (6) is the so-called accelerationist Phillips curve with inflation π_t and the log of output y_t . Equation (7) is the monetary policy curve with the real interest rate r_t and an orthogonal monetary shock v_t^M which for now we are going to assume as mean zero and serially uncorrelated. Finally we have equation (8) as the forward looking IS curve. To solve inflation as a function of the monetary shock let's substitute (7) into (8) to get

$$\begin{aligned} y_t &= \phi \mathbb{E}_t(y_{t+1}) - \frac{\phi}{\theta} v_t^M \\ &= \phi(\phi \mathbb{E}_t(y_{t+2}) - \frac{\phi}{\theta} v_t^M) \\ &\vdots \\ &= \underbrace{\lim_{j \rightarrow \infty} \phi^j \mathbb{E}_t(y_{t+j})}_{\rightarrow 0} - \frac{\phi}{\theta} v_t^M \end{aligned}$$

where $\phi = \theta/(\theta + b)$. The iteration uses the fact that v_t^M is serially uncorrelated and has thus always an expected value of zero and $\phi < 0$ unless "the central bank followed the perverse policy of cutting real interest rate in response to increases in output" (Romer

(2019, p. 267)). Substituting this into (6) gives inflation as

$$\pi_t = \pi_{t-1} - \frac{\lambda\phi}{\theta} v_t^M = \pi_{t-1} - \frac{\lambda}{\theta+b} v_t^M \quad (9)$$

which is a simple random walk. As such it is easy to iterate forward

$$\begin{aligned} \pi_{t+1} &= \pi_{t-1} - \frac{\lambda}{\theta+b} v_t^M - \frac{\lambda}{\theta+b} v_{t+1}^M \\ \pi_{t+2} &= \pi_{t-1} - \frac{\lambda}{\theta+b} v_t^M - \frac{\lambda}{\theta+b} v_{t+1}^M - \frac{\lambda}{\theta+b} v_{t+2}^M \\ &\vdots \end{aligned}$$

from which we can easily take the derivative with respect to v_t^M to get

$$\frac{\partial \pi_{t+j}}{\partial v_t^M} = -\frac{\lambda}{\theta+b}, \quad j \geq 0.$$

In other words, as is well-known for random walks, shocks are infinitely persistent, as here a unit monetary shock at t lowers the rate of inflation by $\lambda/(\theta+b)$ from impact to the end of the world.

One could also entertain the idea of serially correlated shocks. This could be motivated by empirical VAR models sometimes exhibiting some remains of serial correlation in the residuals. In the same manner as above, it is straightforward to show that with $v_t^M = \rho_M v_{t-1}^M + e_t^M$, where $e^M \sim (iid)$,

$$\frac{\partial \pi_{t+j}}{\partial v_t^M} = -\frac{\lambda}{\theta+b-\rho_M}, \quad j \geq 0.$$

As we see, the effect of this shouldn't be dramatic, as the autocorrelation coefficient only serves to strengthen the response. The exception to this would be if $\theta+b \leq \rho_M$, as this would drive the response to minus infinity after it would jump to positive. Although an intriguing possibility, we will ignore this scenario.

Intuitively it should be quite obvious that the infinitely persistent shocks implied by the accelerationist Phillips curve is quite unreasonable when regarding an inflation targeting central bank, as this type of a random walk for inflation implicitly assumes the central bank not having credibility regarding its inflation targeting.

If one is to take the the first term of the Phillips curve as inherited from a past as is the case with the accelerationist version, a more reasonable assumption could be to assume past inflation rates to effect the current one with diminishing lags. Most straight forward

way to formulate this would be to replace the lagged inflation term in the accelerationist Phillips curve by a stationary VAR(∞) process, that is (9) becomes

$$\pi_t = \sum_{j=1}^{\infty} \rho^j \pi_{t-j} - \frac{\lambda}{\theta + b} v_t^M. \quad (10)$$

with $0 < \rho < 1$. Thus we get

$$\frac{\partial \pi_{t+j}}{\partial v_t^M} = -\frac{\rho^j \lambda}{\theta + b}, \quad j \geq 0 \quad (11)$$

which converges to zero as $j \rightarrow \infty$.

With the simple IS-MP we are already able to produce quite reasonable impulse responses with the effect being negative, lagged and with realistic applications considered converging to zero. Of course, with the framework considered here the peak of the effect is at impact. As will be touched upon in Section 3.1.5, more sophisticated DSGE models tend to produce impulse responses peaking with a lag. As these kinds of models do not tend to have explicit solutions, the shapes and even the sign of the responses depend on the model specifications and parameter calibration. Thus, these kinds of application will be left untouched for now.

3 Previous literature

3.1 Model specification

In this thesis model specification will be used to describe the informational set contained in the model. Other aspects of model specification, like asymptotic properties, though important, will not be covered. Proper specification is important since proper inference requires the model to capture relevant systematic dynamics of the variables. For example, since inflation and economic activity (for example GDP) are strongly linked to each other and reacting to inflation is one of the key mandates of modern central banks, excluding inflation would lead to the dynamics between it and the monetary policy instrument being captured as dynamics between GDP and the instrument by the model due to GDP and inflation being linked, and thus to faulty inference. This could be contrasted with omitted-variable bias in microeconometrics.

3.1.1 Indicators for future inflation

The original diagnosis for the price puzzle was conducted by Sims (1992, pp. 988–989) who postulated that since VAR models look into the past, a proper specification of the model should include variables useful for forecasting inflation, as central banks look also to the future. Hence, the model identifies an endogenous response to inflation expectations as an exogenous shock. In other words, in the context of the price puzzle, it is not prices rising in response to a contractionary monetary policy shock, but rather the monetary authority contracting because of expected inflation in the near future.

Hanson (2004, p. 1393) divides the possible indicator variables into two broad categories. First is the set of variables with a pass through effect to consumer goods. Certain variables like producer price indices can be viewed to have forecasting power for consumer good prices, since producers tend to eventually pass rising costs in production into the final goods. Second is the set of variables reacting to same signals as consumer goods, just faster. An example for this could be exchange rates, as in line with Dornbusch (1976) they react immediately by overshooting to inflationary pressures due to consumer prices reacting sluggishly. These two categories are not mutually exclusive however.

Sims (1992, p. 981) includes commodity prices and exchange rates in his models estimated for the US, the UK, Germany, France and Japan. The results are mixed however, as the

effect on the impulse responses varies a lot with the UK having a large mitigation of the price puzzle while Japan, France and Germany having barely. The price puzzle is not completely eliminated in any of the models.

The author also makes it clear that the models are not very satisfactory, as the same commodity price index is used in all of the models and hence, six different estimates for the modeling of the same variable are acquired. This problem can be solved if one takes commodity prices as exogenous. This of course is not feasible with large economies like the US with significant impact on global commodity prices but might be realistic enough in certain modelling instances. This approach was taken by Peersman and Smets (2001, p. 37) who conducted their VAR on the eurozone countries with synthetically aggregated data. Following the notation specified in Section 2.2, their model is as follows:

$$Y_t = v + \sum_{i=1}^p A_i Y_{t-i} + \sum_{i=0}^k D_i Z_{t-i} + u_t,$$

where $Y_t = (y_t \ \pi_t \ r_t \ s_t)'$ is the standard vector of endogenous variables, in this case the real GDP, the rate of inflation, the short term interest rate and the real exchange rate. $Z_t = (pcom_t \ y_t^{US} \ r_t^{US})'$ is the vector of exogenous variables, in this case a global commodity price index, the real GDP of the US and the short term interest rate in the US. Note, that there of course is no explicit reason for the lag lengths to be the same, ie. $p \neq k$ is a viable option. The rationalisation behind Z_t is to control for inflation expectations and the global business cycle. Taking Z_t as given means naturally that we are assuming no feedback from variables in Y_t to Z_t removing the need to model their dynamics. One could however ponder whether the euro area is too large in the global scale to assume this.

Peersman and Smets (2001, p. 39) get impulse responses which can't really be argued to exhibit the classic hump-like pattern so often seen with the price puzzle. Still, they do raise some questions. Looking at the estimated impulse responses, one has to ponder on why does it take a full year for inflation to react in any meaningful way? This is usually explained away by ushering the magic word "frictions", but this explanation should be examined more critically. If we take price frictions in the classic sense of Calvo (1983), surely inflation should start to adjust immediately or quite soon at some rate, albeit gradually. This is how we see pretty much every other variable in almost all cases behaving and it makes intuitive sense. For the exhibited impulse responses of inflation to make sense from the point of view of frictions, we would have to assume that during the first year after the shock, there are no significant supply or demand effects that would lead to re-optimization of any prices, but after this agents suddenly start to act. Though not impossible, this seems like a very strong assumption and thus Occam's razor would caution

us to make too strong conclusions.

The findings of Hanson (2004, pp. 1393–1407) shed some much needed light on the inclusion of variables forecasting future inflation. The author tests the forecasting power of a wide range of often used control variables, including different commodity prices and exchange rates. The author finds that the consistency of the forecasting power over different horizons is poor for all of the tested variables as none of them is able to produce superior forecasting power over others for all of the tested horizons. Moreover, none of the tested variables is able to mitigate the positive point estimate under 8 months or so. The relationship between mitigation of the price puzzle and forecasting power is also found to be small at best, implying that the forecastability of inflation is not the problem in the first place. One could even argue that these findings dispute the whole puzzle. This view will be considered further in Section 3.1.5.

Another interesting interpretation could also be given however: As the general price level aggregates all of the different price effects in the economy, it could be argued that different indicators have forecasting power over some of the underlying price movements but not the aggregate level itself. For example, a rise in the price of oil and the price of conductors might have unrelated pass through effects on prices. Thus including one does not account for the effects of the other though it is reasonable to assume the monetary authority to be aware of both. More formally, the monetary author acts upon general, but unobservable inflation expectations for which all of the observed variables are noisy indicators. This line of thought will be expanded upon in Section 3.1.4.

3.1.2 The output gap, Phillips curve and inflation

As is evident from the different forms of the Phillips curve that litter the literature (for example see the accelerationist, Lucas and new Keynesian Phillips curves from Romer 2019, p. 338 or the Phillips-type and NAIRU-type curves from Higo, Nakada, et al. 1999, pp. 131–133), mainstream economic theory doesn't really see a relationship between inflation and output, but rather inflation and the output gap. As outlined well by Fisher, Mahadeva, and Whitley (1997, pp. 68–69), "The output gap is generally used to measure the extent to which the economy is operating at an unsustainable level of resource utilisation...". As such one could see it reflecting the extent to which the economy deviates from equilibrium. In this view, equilibrium prices or inflation are constant, and movements in these are caused by the disequilibrium resulting from rigidities in the economy.

Due to this evident discrepancy between the traditional monetary VARs and theory, Gior-

dani (2001, p. 1) proposes the inclusion of the output gap or potential output to be paramount for the proper specification of monetary VARs. As outlined by the author, solving the apparent misspecification of the VAR with commodity prices is problematic also due to the fact that it makes it more difficult to interpret the structural shocks. As VARs need as many shocks as there are variables, a CPI and a commodity price index need their separate, orthogonal shocks. Again, these tend to be absent from theoretical models (Giordani 2001, p. 3). With these lines of thought as well as the conclusions made at the end of Section 3.1.1 one could easily argue that the inclusion of a commodity price seems to be a patch work solution at best.

The author constructs a simple backwards looking model which implies a Taylor rule both in the discretionary and commitment solutions. As such, the DGP is driven by the potential output, realized output, inflation and the interest rate. However, it turns out that giving this model a VAR representation and excluding the potential output implies a positive variance for the monetary policy shock even if the true variance is set to zero. In other words, a model with no output gap will overestimate the variance of the monetary shock (Giordani 2001, pp. 6–10). As is then shown by the author, the impulse responses implied by the theoretical model with a zero variance monetary shock (ie. the shock doesn't exist) when the output gap is excluded seem quite similar to the ones we see in the empirical literature: a decline in output, and an initial increase followed by a tapering of in the rate of inflation and the interest rate. Simulating the system with a stochastic, non-zero variance monetary shock reinforces these conclusions. These do not just exhibit the classic prize puzzle in the form of an initial rise in the rate of inflation (which does not appear in the correctly defined DGP), but also other impulse responses are misestimated: the response of output to a aggregate demand shock gets overestimated while the response of the interest rate get underestimated. In short, excluding the output gap misspecifies the whole system and subsequently the impulse responses. These observations seem to hold quite well also for the empirical counterparts (Giordani 2001, pp. 23–25).

3.1.3 Different monetary policy regimes

For obvious reasons relating to degrees of freedom in estimation, researchers prefer to have as much observations as possible. Within time series econometrics this naturally means longer and/or more frequent time series. Longer time series might bring their own curses however, as monetary authors might switch instruments over time. As pointed out by Hanson (2004, p. 1407), between 1979 and 1982 the FED explicitly targeted non-borrowed reserves, not the federal funds rate. Thus, identifying monetary policy through the federal

funds rate would be inappropriate through this period and lead to faulty estimates.

Monetary authors might also change their preferences and goals. As outlined by Taylor (1999, pp. 336–339), the federal funds rate had at the time of him writing more or less responded to the rates implied by the Taylor’s rule since the late 80s as well as being on point during Volcker’s disinflation years in 1979 and 1980. Before the disinflation years the rate was consistently too low and between 1982-1984 too high. This indicates shifts in FEDs goals. As outlined, in the 60s inflation was tolerated in the view of the inflation/unemployment-trade off and in 1982-1984 the FED tried to establish its credibility and keep inflation expectations low as it was essentially ”in a transition between policy rules” (Taylor 1999, p. 339). In other words, the FEDs response function has changed multiple times since the 60s. Thus estimating the whole period as a whole with one model would naturally lead to biased results.

This line of thought could be taken further by considering rational expectations. Let’s say that for the first α -part of the estimation period the monetary authority conducted it’s policy by an old rule such that $S_t^{old} = f^{old}(\Omega_t) + v_t$ and for the remainder $(1 - \alpha)$ -part by a new rule such that $S_t^{new} = f^{new}(\Omega_t) + v_t$. Without rational expectations, this would simply mean that estimation over the whole period would lead the monetary reaction function estimate to be a linear combination

$$\hat{S}_t = \alpha f^{old}(\Omega_t) + (1 - \alpha) f^{new}(\Omega_t) + e_t.$$

However, as outlined by Kydland and Prescott (1977, p. 480), with dynamic rational expectations the reaction function of economic agents is in itself a function of the policy function. In the case of a monetary VAR, this would mean that a change in monetary policy regime alters the expectations in the economy thus changing all of the other reaction functions modelled by the system as well. In other words, the whole model and its impulse responses gets misestimated, not just the monetary instrument. This also seems to be the case when looking at the impulse responses from Hanson (2004, p. 1408), as the responses from 1959 to 1979 seem to be more volatile than the ones from 1982 to 1998, while the responses from the whole period seem to be something in between.

As shown by Hanson (2004, p. 1408) the price puzzle is mostly mitigated from their models by simply restricting the data from 1982 onward. This idea is also taken up by Borys, Horváth, and Franta (2009, p. 424), who restrict their observations to relatively low 101 in order to acquire estimates only on the post-1997 period of inflation targeting by the Czech National Bank (a fixed exchange rate regime was in place prior to this). As a result, they do not observe a price puzzle. This is also reinforced by the meta-analysis of Rusnák, Havranek, and Horváth (2013, p. 42) who observe that compared to the whole pool of

studies of which 15% exhibited a significant price puzzle and 50% in some form, studies estimating a single monetary regime only faced this 8% and 38% of the time.

3.1.4 Factor augmentation

As outlined by Bernanke, Boivin, and Elias (2005, pp. 388–389), traditional VAR models exhibit at least three problems when applied to monetary policy research. First, due to each variable raising the amount of estimated coefficients by the number of lags used (plus the constant), having a large number of variables in the system is not generally feasible with the finite set of observations econometricians are limited by. Thus models with more than six to eight variables are rarely seen, as having more starts to pose some serious problems with degrees of freedom. This poses an obvious challenge from the point of view of proper model specification, as monetary authorities observe and weigh in hundreds of variables when conducting monetary policy. Thus, even with carefully chosen proxies, it is very likely that some relevant information used by the monetary authority and economic agents gets left out from a traditional VAR model.

Second, a lot of concepts and variables in economic theory have no precise counterparts in the observable data. For example, "economic activity" may not be a one-to-one measurement with variables such as GDP or industrial production. As these observable variables may not represent the actual variables of interest precisely, using them creates model misspecification and thus faulty estimates for the impulse responses. In addition, as mentioned in Section 2.1, these observable variables are subject to measurement errors which in time of policy conducting haven't been revised yet. Thus, it could be argued that these variables are actually (to a point) contemporaneously unobservable.

These two aspects alone paint a grim picture of traditional monetary VARs. As the observed variables are rather just noisy proxies for the actual driving forces of the DGP, including only a handful of them would lead to information insufficiency, while including enough to ensure a proper information set would lead to serious cross-correlation and overfitting issues.

Third, often the econometrician is not interested in the effects of monetary policy on just a few variables, but rather a wide range of them, which a traditional small dimension model doesn't allow for. Furthermore, one could argue that as contemplated by Sims (1992, p. 997), estimating several models to acquire estimates for a large amount of variables would lead to several different model equations for certain variables.

The first two aspects become quite clear when reading the accounts of monetary authors meetings. For example, in the monetary policy meeting of the Governing Council of the European Central Bank in October 2023, a large number of variables were discussed. To name a few; different interest rates and spreads, indicators for inflation and output expectations, disaggregated price indices, energy price volatility, employment, credit supply and demand changes, changes in the outlook for the US and Chinese economy, service sector activity, productivity trends, real estate transactions, the effect of monetary policy transmission etc. were discussed. It is also quite evident that monetary policy decisions are not made by explicitly applying some observed variables into a rule in the moment, but rather by extrapolating a general view on the outlook of the real economy and inflation expectations from a large amount of indicators (ECB 2023). This of course is to be expected as especially the output gap estimates are not in general available at their final form at the moment of policy conduct.

As a solution to these problems, Bernanke, Boivin, and Elias (2005) propose the use of FAVAR models. In this setup, the model of directly observable variables is augmented with unobserved factors, which are estimated from a large set of data. The general formulation of this looks like following:

$$\begin{pmatrix} F_t \\ Y_t \end{pmatrix} = v + \sum_{i=1}^p A_i \begin{pmatrix} F_{t-i} \\ Y_{t-i} \end{pmatrix} + u_t.$$

Here Y_t is the vector of observed variables and F_t is the vector of unobserved factors. As mentioned before, the amount of observable variables which the monetary authority reacts to are actually quite small in number. For example, Bernanke, Boivin, and Elias (2005, p. 397) argue that even inflation and output could be seen as unobservable, since these are (at least contemporaneously) subject to measurement errors and do not necessarily fully align with their theoretical counterparts. For example, measuring inflation is always subject to subjective choices not just between different price indices, but also the composition of these indices making it hard to justify one measurement to be the true one. Thus they conclude, that the most realistic form of Y_t might actually be a univariate vector consisting of only the policy instrument ie. $Y_t = r_t$. In this case, all the other available variables are noisy macroeconomic indicators from which common factors are extrapolated. Again, this interpretation is reinforced by reading of ECB (2023) for example. Of course, one could argue that the composition of Y_t depends on the monetary authority in question. For example, as the ECB considers "...that the Harmonised Index of Consumer Prices (HICP) remains the appropriate price measure for assessing the achievement of the price stability objective..." (ECB 2021), including it in Y_t for the euro area could be justified.

A factor model can be defined as

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + \eta_t, \quad (12)$$

where X_t is a large $N \times 1$ vector of variables, which are driven by m common factors in vector $F_t = (f_1, \dots, f_m)$ and the observable variables Y_t . This can also be extended to a DFM by including the lags of the factors in F_t (Bernanke, Boivin, and Eliasch 2005, p. 393). As outlined by Kilian and Lütkepohl (2017, p. 539) this (static) model can be estimated by principal components by maximizing the variance of X_t explained by m common factors, or in reverse minimizing the sum of squared errors ie.

$$\min_{\Lambda^{f^*}, f_1, \dots, f_T} \frac{1}{T} \sum_{t=1}^T (X_t - \Lambda^{f^*} F_t)' (X_t - \Lambda^{f^*} F_t). \quad (13)$$

Note, that following Bernanke, Boivin, and Eliasch (2005, pp. 398–399) we aren't taking to account the effect of Y_t to X_t at this point yet. Solution to (13) can be acquired by finding the m largest eigenvalues $\lambda_1 > \dots > \lambda_m$ and their corresponding eigenvectors $\lambda_1 \dots \lambda_m$ from $S_x = T^{-1} \sum_{t=1}^T X_t X_t'$. From this we set the estimates for factor loadings as $\hat{\Lambda}^{f^*} = (\lambda_1 \dots \lambda_m)$ and thus $\hat{F}_t^* = \hat{\Lambda}^{f^*} X_t$. Now the final estimate cleaned of Y_t can be acquired for example by regressing the estimates of the common factors with unobserved "cleaned factors" and Y_t , though the exact procedure would have to be chosen in accordance with identifying assumptions. For example Bernanke, Boivin, and Eliasch (2005, pp. 404–405) suggest extracting the common factors from the regression $\hat{F}_t = \hat{F}_t^* + \Lambda^y Y_t + e_t$ using the so-called slow-moving variables, as these aren't in their case assumed to be affected by $Y_t = r_t$ contemporaneously excluding problems of multicollinearity. As the series of common factors \hat{F}_t have been extracted, they can be used in the VAR model as usual.

The asymptotic properties of the estimates for the common factor are discussed in Stock and Watson (2002, p. 1167) who show that under mild assumptions \hat{F}_t is consistent as $N, T \rightarrow \infty$. However, as discussed by Boivin and Ng (2006, p. 171), as data sets X_t are always put together by the researcher using subjective judgment and due to many of the series used in practical applications being highly correlated, the usual asymptotic assumptions, even mild ones, become quite strong. They show by Monte Carlo simulation and study replication that limiting N into the range of 40-50 gives often better estimates than ones extracted from three-digit amount of series. This is attributed to the fact that in general asymptotic properties have most of their effects kicking in when sample size is at minimum around 30-40, but at the same time the disturbing effect of strongly cross-correlated idiosyncratic errors and dominating factors is minimized by having less series.

The results of Boivin and Ng (2006) can be used to the advantage of the econometrician.

Bernanke, Boivin, and Elias (2005, p. 403) extract common factors out of the whole data set X_t . An alternative to this would be to divide X_t into subsamples and extract factors out of these and use the first factor (the one accounting for largest amount of variation in X_t) of each subsample. Subsampling could be done by dividing X_t into categories like "economic activity" and "prices" etc. The advantage of this procedure is that it allows for the use of very large sets of data without the complications discussed by Boivin and Ng (2006). Although used in the literature before (see Fernald, Spiegel, and Swanson (2014) and Holguín and Uribe (2020) for example), to my knowledge this kind of subsampling hasn't been motivated by the asymptotic results discussed by Boivin and Ng (2006) but rather just to give the factors an economic interpretation.

Another clear advantage of a FAVAR framework can be seen, when circling back to the problem of nonfundamentalness discussed in Section 2.5. As is pointed out in Alessi, Barigozzi, and Capasso (2008, pp. 24–27) and Forni, Giannone, et al. (2009, p. 1329), nonfundamentalness is not an issue for factor models, as having a large cross-section of more variables than structural shocks and heterogeneous impulse responses in a factor model leads to fundamentalness.

Compared to their traditional benchmark models without factor augmentation, Bernanke, Boivin, and Elias (2005, p. 406) are able to considerably limit the price puzzle effect to less than a year. The initial positive effect does not disappear completely however. Other authors, such as Laine (2020, p. 2913) and Holguín and Uribe (2020, p. 2459) estimate very different shapes, the first being the intuitive convex shape and the second concave with the greatest effect on impact.

3.1.5 The cost channel explanation

What if the prize puzzle isn't a puzzle at all? Although often neglected from theoretical models, monetary policy has also "supply-side" effects, creating the so-called *cost channel*. As interest rates rise, some production costs rise as well which could lead to initial rise in prices (Rabanal 2007, pp. 907–908). The standard way is to take working capital into consideration by assuming a delay between sales and their payments, so that for example wages have to be paid before sales revenue is acquired. Thus firms would have to loan money to pay wages, leading to interest rates having also an effect labour demand and aggregate supply (Barth III and Ramey 2001, pp. 208–209).

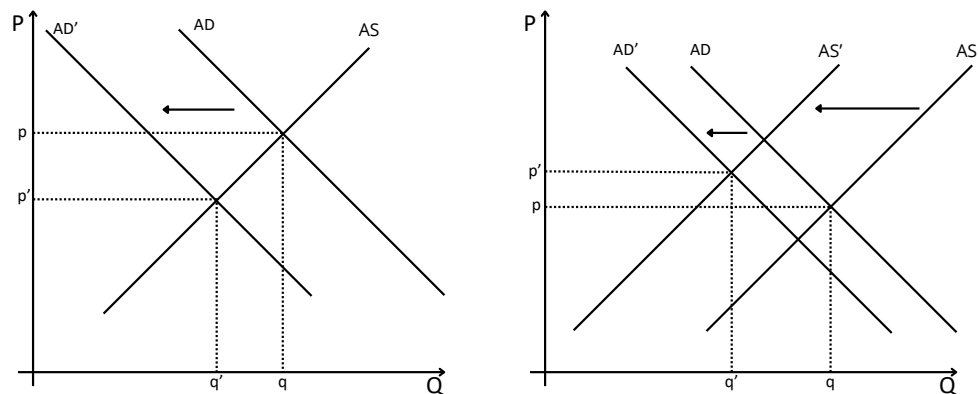


Figure 1: The cost channel in action

The empirical implications of the cost channel are discussed by Barth III and Ramey (2001), who study the cross-sectional dynamics of different industries. Their theoretical baseline is simple and sheds some light on the mechanism of the cost channel. As seen from the left side panel in Figure 1, if a monetary contraction has only traditional demand-side effects, the AD curve shifts to the left leading to a reduction in prices and output. However, if the AS curve is a decreasing function of interest rates it too shifts to the left as seen in the left panel of Figure 1. Thus the effect on prices becomes unambiguous (Barth III and Ramey 2001, p. 200). This standard way of modelling working capital also implies that demand for labour and thus the wage is a decreasing function of the interest rate. Thus, as discussed by Barth III and Ramey (2001, pp. 213–214), if an industry is primarily affected by the demand channel, prices should fall relative to wages as in the right panel of Figure 1. If the industry is primarily affected by the supply channel however, wages should fall relative to prices (or in other words prices rise relative to wages). If both channels are in effect equally, there should not be a change in the relation of prices and wages.

Out of the 21 industries studied, Barth III and Ramey (2001, pp. 216–219) find that 10 exhibit a rise in prices relative to wages, indicating a dominance of the supply channel. It is noteworthy however, that the magnitude of these relative price increases doesn't seem to be very large. Interestingly however, when subsampling the data reveals that the dominance of the supply channel is much more substantial in the pre-Volcker years, while it is very minimal during the post-disinflation years. One could even postulate one reason for the prevalence of the price puzzle in the pre-Volcker years compared to the post-

disinflation years to be this change in the transmission of monetary policy through the cost channel. Barth III and Ramey (2001, pp. 225–226) discuss numerous structural changes in the 70s and 80s from private-sector financial innovations to the switch from fixed to floating exchange rates, which could have contributed to the dampening of the cost channel. It should also be noted that these findings do not necessarily give a clear answer to how much one could expect the aggregate price level to rise due to the cost channel, but rather indicate that such a mechanism exists.

A DSGE approach to the cost channel with working capital is considered by several authors. CEE and Henzel et al. (2009) use a minimum distance estimation approach, where the model parameters are calibrated by minimizing the distance between empirical and theoretical impulse responses. Christiano, Eichenbaum, and Evans (2005, pp. 21–22) find that with US data from 1965 to 1995 this procedure is able to generate an initial rise in the price level following a positive monetary shock. Rather than sticky prices, sticky wages is found to be the key driver of this dynamic, as wages have an effect on marginal costs through the cost channel (Christiano, Eichenbaum, and Evans 2005, p. 30).

Using data from the euro area from 1997 to 2002, Henzel et al. (2009, p. 279) come to the same conclusion as CEE. Their model is able to generate a positive response in inflation following a negative monetary shock by allowing for low price stickiness and high wage stickiness. Their model also includes the same Calvo-type stickiness for loan rates set by the banking sector. However, this dampens the positive effect as only very low to non-existent stickiness is found to be able to contribute a positive initial effect, while the estimate for the Calvo-parameter is found to be 0,41 with a standard error of 0,03. The authors find that despite initial estimates for price and wage stickiness parameters (0,56 and 0,61) do not allow for a positive initial effect on inflation, re-calibrating them in line with previous literature (0,35 and 0,7) does. These values are not found to be rejected by the data. It should be noted however, that even with these restrictions, the positive effect on inflation is small and lasts only for two quarters (Henzel et al. 2009, pp. 280–281).

Rabanal (2007) follows the modelling assumptions of CEE as well as using US data from 1959 to 2004. Calibrating the initial parameters in line with Christiano, Eichenbaum, and Evans (2005, p. 17), including full indexation of wages and prices, all intermediate firms being affected by the cost channel and setting the elasticity of capital utilization with respect to the rental rate of capital high, Rabanal (2007, pp. 917–919) is able to produce the initial rise in inflation after a monetary contraction. From this baseline model, allowing for price flexibility increases this phenomenon, while allowing for wage flexibility or eliminating the variability of capital utilization leads to inflation falling from impact as one would expect in the traditional view. Thus the author concludes, that the mere ex-

istence of the cost channel, even when all firms are subject to it, is not enough for the positive response of inflation, but "... having a large elasticity of the nominal interest rate on the real marginal costs of production, high real wage stickiness and variable capital utilization rates (which implies low volatility in the rental rate of capital) are needed" (Rabanal 2007, p. 919). Instead of a minimum distance estimation, the author then uses a likelihood-based Bayesian approach to re-calibrate the parameters. These contradict the conditions set by the initial model for a positive inflation response, as price stickiness is found to be relatively high while wage stickiness relatively low. In addition, practically all of the probability mass of the posterior distribution for the estimate of firms subject to the cost channel lie under 0,5, contradicting Christiano, Eichenbaum, and Evans (2005, p. 10) who assume this to be 1. Thus the author concludes that "... the posterior probability of observing an increase of inflation after a monetary policy tightening is zero" (Rabanal 2007, p. 908).

As the conclusions of the papers discussed above are conflicting, it is worthwhile to interpret their findings a bit closer. As mentioned in Rabanal (2007, pp. 908–909), since CEE calibrate their parameters using minimum distance estimation, their model will naturally reflect the properties of the VAR used, which displays a small price puzzle. Thus, a misspecification of the VAR would lead to misspecified parameters in the DSGE model too. In the context of this theses it seems quite possible that this is the case, as their VAR spans over several distinct monetary regimes and does only include a handful of US domestic real variables (Christiano, Eichenbaum, and Evans 2005, p. 4). It should also be noted that the initial calibration of Henzel et al. (2009, pp. 275–277) does not lead to a price puzzle even though the underlying VAR does display one. It is only after re-calibration that a small, six month increase in inflation is produced without being rejected by the data. Even then, one should note that the impulse responses from the original calibration are confined within the 68% bootstrap intervals of the empirical impulse responses. Henzel et al. (2009, p. 269) do mention that unlike the US, the euro area financial system is bank based rather than market based which could explain the differing results of them and Rabanal (2007, pp. 908–909).

Based on the above discussed evidence it can be concluded that witnessing an increase in inflation following a negative monetary shock seems improbable. The evidence seems to indicate that in certain situations, like older data or a more bank-based financial system, an increase is possible, but even then it can be expected to be small in magnitude and brief, no more than a couple of quarters.

3.2 Identification

Identification refers to the procedure by which structure of the contemporaneous dynamics (which isn't observable from the raw data) of the estimated reduced form model is extrapolated in order to acquire the orthogonal shocks to the system. This can be contrasted with microeconomic methods by which one can orthogonalise residuals with the explanatory variables allowing for inference on the model.

The fundamental problem is that in general the covariance matrix of the reduced form errors cannot be assumed as an identity matrix (ie. the errors are usually contemporaneously correlated meaning that one shock of interest might have a contemporaneous effect on several variables). Thus acquiring orthogonal (ie. $\Sigma_v = I_k$) shocks v_t from model (4) leads to the following (this is analogous to Kilian and Lütkepohl 2017, p. 494) situation (for the sake of illustration let $k = 2$):

$$\Sigma_u = B_0 \Sigma_v B_0' \Rightarrow \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \quad (14)$$

From this we get the set of equations:

$$\begin{aligned} \sigma_1^2 &= b_{11}^2 + b_{12}^2 \\ \sigma_{12} &= b_{11}b_{21} + b_{12}b_{22} \\ \sigma_2^2 &= b_{21}^2 + b_{22}^2 \end{aligned}$$

As we can see, with covariances given by the data, there are three equations with four free parameters while a unique solution to a set of equations requires as many equations as free parameters.

The most commonly used identification method is to impose zero restrictions directly to the impact matrix B . As is well known, full identification requires $k(k-1)/2$ restrictions (see Kilian and Lütkepohl 2017, p. 215 for example). In practice this means assuming that certain shocks do not have a contemporaneous effect on certain variable. This is usually done by a recursive ordering using the Cholesky decomposition on the covariance matrix of the reduced form errors Σ_u . This is done by defining the lower-triangular matrix P such that $PP' = \Sigma_u$. As $\Sigma_u = B_0 B_0'$, this gives *one* possible identified structure for the model with $P = B$ (Kilian and Lütkepohl 2017, p. 216).

Identification can be achieved also by restricting the long run effects of shocks. That is, instead of imposing a recursive structure on the impact matrix B , rather a recursive

structure on the sum of Θ_i :s is imposed. More formally, zero restrictions are imposed on matrix

$$\Theta(1) = \sum_{i=0}^{\infty} \Theta_i.$$

For a classic example, Blanchard and Quah (1989) assume the bivariate system of US unemployment and the difference in log real GDP to be stationary and identify the shocks as aggregate supply and demand shocks. For identification, it is assumed that an aggregate demand shock will not have a long run effect on the *level* of real GDP. That is, the cumulative impulse responses on the difference of GDP are zero at limit.

The obvious problem with these kind of identification strategies is of course the plausibility of the made restrictions. Even as it could be argued that it is enough for the "true" value of the restricted parameter to be close enough to zero in order to get satisfying estimates for the identified shocks, these restriction have to be made often quite *ad hoc* purely to get full identification. This is evident from the growing literature of statistical identification which has made these traditional identification procedures testable with the plausibility of them being rejected repeatedly. For examples, see Normandin and Phaneuf (2004, p. 1231), Lütkepohl and Netšunajev (2017b, p. 52) and Lütkepohl and Netšunajev (2017a, p. 15). This of course doesn't automatically disqualify inference made upon these identification restrictions, but as seen in Section 2.3, impulse responses are derived from matrix B and thus misidentifying it naturally leads to faulty estimates for the impulse responses.

3.2.1 The long lags of monetary policy

Since a series of articles by Milton Friedman, the idea of monetary policy effecting with long lags has been well established. That is, monetary policy doesn't actually have a direct effect on prices, but works indirectly through output. Estrella (2015, p. 1883) takes this idea and formulates the following simple macroeconomic model:

$$\pi_t = a_1 \pi_{t-1} + a_2 y_{t-1} + \varepsilon_t \quad (15)$$

$$y_t = b_1 y_{t-1} + b_2 (r_{t-1} - \pi_t) + \mu_t \quad (16)$$

$$r_t = c_1 r_{t-1} + c_2 y_t + c_3 \pi_t + v_t \quad (17)$$

where (17) is the monetary policy rule. The short-term rate set by the monetary authority affects output directly only with a lag, which then in turn affects inflation with a lag. Thus a monetary policy shock at t effects inflation only at $t + 2$ and beyond. From these insights the author suggests restricting the monetary policy instrument coefficient for inflation to

zero both contemporaneously and on the first lag. In their three variable model with $Y_t = (\pi_t \ y_t \ r_t)$ this means restricting $a_{1,13} = 0$ on top of the recursive identification.

This procedure raises some questions however. First, it could be argued that if the economy follows the dynamics described by (15)-(17), all of the monetary policy instrument coefficients on inflation should be restricted to zero, as the effect of monetary policy on inflation is assumed to work only indirectly through output. Second, even when assuming the described dynamics of the economy, the applicability of this insight in a VAR context is very much affected by the data frequency used. Restricting the first lag to zero with monthly data is very different from doing the same with yearly frequency. The author uses quarterly data, but no discussion on the effects of data frequency or why quarterly data specifically fits the proposed period structure is given. Thus, the choice of data frequency can be deemed purely *ad hoc*. In the end, the question of how many lags on what kind of a frequency should be restricted is a purely empirical one. Third, one could argue that given sufficient model specification, these restrictions shouldn't be necessary as the model should estimate them as zeros anyway. The need for restricting lags thus implies information insufficiency to begin with.

Restricting the first lag seems to dampen the price puzzle effect leading the author to conclude that "the puzzle disappears". However, as with many other proposed solutions, the shape of the estimated curve still exhibits a small positive initial effect. The difference to the benchmark model seems to be the magnitude of the effect and the bootstrapped significance level of it. Apparently, the initial positive effect of the unrestricted model was deemed significant whereas the restricted model was not. It is noteworthy however, that the bootstrapped intervals used were 0,68 ones and thus quite small by normal standards of 0,9 or 0,95.

3.2.2 Long run restrictions in a VECM

Another pure identification solution is proposed by Krusec (2010, p. 147). They propose a model with $Y_t = (y_t \ \pi_t \ r_t)$ which gets identified with two long run restrictions and one short run restriction (this equals to $k(k-1)/2 = 3$ as $k = 3$). However, instead of a normal VAR model, one cointegration relationship is assumed (and supported with the Jansen trace test) and thus a VECM is estimated instead. With a few simplifications (removing constants and deterministic trends) for illustration proposes, this would look like following

(Lütkepohl 2005, p. 247):

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \sum_{i=1}^p \Gamma_i \Delta Y_{t-1} + u_t. \quad (18)$$

Intuitively this means that "...integrated variables share a common stochastic trend such that a linear combination of these variables is stationary..." Kilian and Lütkepohl (2017, pp. 75–76). For example the price of a particular commodity in different markets could deviate from each other according to the dynamics of the individual markets while the general trend of these prices is common, ie. there exists an equilibrium relation (Lütkepohl 2005, p. 246).

The author restricts the long run effect of monetary policy shocks to zero for all of the variables, which due to one cointegration relationship accounts for two independent restriction. Thus an additional contemporaneous restriction of inflation shocks (it is unclear what the author means by this but usually this is modelled as a cost-push shock) to output is made.

Although the author concludes the results to show a solution to the price puzzle, it is worth noting that for the post-1981 era the estimate shows again the classic initial hump shape in the response to inflation. Also unlike in the pre-1978 era, a considerable proportion of the bootstrapped confidence interval mass is on the positive side. According to these intervals, the results cannot be considered significant *at any point*. As the author seems to use the bootstrapped intervals for inference, a more suitable conclusion based on the post-1981 era would be "monetary policy doesn't seem to have any significant effect on inflation", which of course one could argue to be a price puzzle in itself. In the pre-1978 era the negative effect is more or less significant, though the bootstrap intervals are very much all over the place leaving the magnitude of the impact effect quite open.

3.3 Statistical identification

Statistical identification refers to a broad set of identification strategies employing statistical properties of the data in order to identify orthogonalized shocks. As pointed out by Rigobon (2003, p. 777), sometimes the traditional identification strategies just cannot be justified. The clear advantage of statistically identified shocks tend to be that unlike traditional identification strategies which require assumptions which have to be validated by external empirical research or that are outright *ad hoc*, assumptions about the statistical properties needed for proper identification are usually quite straightforward to validate on

the data itself and can thus be seen as milder ones. The clear downside is that since a broad set of identifications for the shocks is possible, statistically identified shock don't have a natural economic interpretation but might rather be combinations of economically interesting shocks. However, this leaves room for over-identification of the model, which can be used to validate more traditional identification assumptions. The field of statistical identification is relatively new and still evolving, and as such use of it in the empirical research has been quite sparse to my knowledge.

Statistical identification methods shouldn't be necessarily seen as better (or worse) than traditional ones. Rather they should be seen as useful additional tools in the toolbox of an econometrician. As their validity can often be directly tested, they can be useful to relax traditional identification assumptions and thus achieve proper identification. As outlined by Lütkepohl and Netšunajev (2017a, p. 17), understanding the particular features of the model and data helps making informed decisions on the type of identification to be used.

3.3.1 Structural breaks in variance regimes

A simple solution to the problem of more free parameters than equations was proposed by Rigobon (2003), who argued that a shift in the variance of the structural shocks v_t could be used to solve this. Lets use example (14) to illustrate this. However, now we have two volatility states such that

$$\Sigma_v = \begin{cases} I_k, & \text{when } t = 1 \dots T_B - 1 \\ \Lambda, & \text{when } t = T_B \dots T \end{cases}$$

where $\Lambda = \text{diag}(\lambda_1 \dots \lambda_k)$ and T_B is the moment of the break in variance. Now

$$\Sigma_u^1 = B_0 I_2 B_0' \Rightarrow \begin{pmatrix} \sigma_{1,1}^2 & \sigma_{1,12} \\ \sigma_{1,12} & \sigma_{1,2}^2 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \quad (19)$$

and

$$\Sigma_u^2 = B_0 \Lambda B_0' \Rightarrow \begin{pmatrix} \sigma_{2,1}^2 & \sigma_{2,12} \\ \sigma_{2,12} & \sigma_{2,2}^2 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \quad (20)$$

giving us the following set of equations

$$\begin{aligned}\sigma_{1,1}^2 &= b_{11}^2 + b_{12}^2 \\ \sigma_{1,12} &= b_{11}b_{21} + b_{12}b_{22} \\ \sigma_{1,1}^2 &= b_{21}^2 + b_{22}^2 \\ \sigma_{2,1}^2 &= \lambda_1 b_{11}^2 + \lambda_2 b_{12}^2 \\ \sigma_{2,12} &= \lambda_1 b_{11}b_{21} + \lambda_2 b_{12}b_{22} \\ \sigma_{2,1}^2 &= \lambda_1 b_{21}^2 + \lambda_2 b_{22}^2.\end{aligned}$$

As we can see, the number of free parameters has increased by two but the number of equations by three, giving us six free parameters in six equation, thus enabling a unique solution for the parameters. This procedure thus requires two assumptions. First, volatility parameters $\lambda_1 \dots \lambda_k$ have to be unique, as $\lambda_i = \lambda_j$ would imply less free parameters than needed for identification. Second, there has to be at least two different volatility states to begin with, the break point between which has to be specified. However, as shown by Rigobon (2003, p. 783), given that the volatility break exists, misspecification of it will still lead to unique estimates for parameters as linear combinations of the true parameters. Thus even a misspecified model is identified and the estimates are consistent.

In practise, identification boils down to whether parameters $\lambda_1 \dots \lambda_k$ can be interpreted to be unique. Lanne and Lütkepohl (2008, pp. 1137–1139) do this by conducting (quasi-)LR tests on the stylized log likelihood functions of Σ_u . However, although this tests the validity of the assumed heteroskedasticity of the covariance matrix of the structural shocks (as B being constant through the whole model implies heteroskedasticity of u_t resulting from heteroskedasticity in v_t), it is not obvious that this confirms the uniqueness of individual volatility parameters λ_i . Lütkepohl, Meitz, et al. (2021, pp. 5–8) develop formal Wald-tests for uniqueness of $\lambda_1 \dots \lambda_k$ for models with two volatility states, assuming elliptically symmetric distributions for the reduced form errors. However, through simulations they find the power of the test to be questionable with large models on small samples (their small samples where $T = 100$ with larger being $T = 250$ and $T = 500$) (Lütkepohl, Meitz, et al. 2021, pp. 9–15).

3.3.2 SVAR-GARCH

As is seen in many economic and financial applications, time series are often well approximated by GARCH processes. Thus assuming the shocks in the VAR system to have conditional variances might be a useful tool. As it turns out, if the shocks are generated by

a GARCH process, they become identifiable (Kilian and Lütkepohl 2017, pp. 517–518).

Let's follow the model setup of Kilian and Lütkepohl (2017, pp. 518–519) and assume the shocks to be generated by individual and orthogonal GARCH(1,1) processes. Higher-order processes could be considered as well, but in practice this is rarely done (Kilian and Lütkepohl 2017, p. 518). Now we get

$$\mathbb{E}(u_t u_t' | \mathcal{F}_{t-1}) = B \Sigma_{v_t | t-1} B'$$

as the conditional covariance matrix of reduced form errors u_t . Here \mathcal{F}_{t-1} is the information set up to $t - 1$ and the covariance matrix of structural shocks is $\Sigma_{v_t | t-1} = \text{diag}(\sigma_{1,t|t-1}^2, \dots, \sigma_{k,t|t-1}^2)$. The individual structural shock conditional variances are given by a GARCH(1,1) process of the form

$$\sigma_{k,t|t-1}^2 = (1 - \gamma_k - g_k) + \gamma_k v_{k,t-1}^2 + g_k \sigma_{k,t-1|t-2}^2, \quad k = 1, \dots, K, \quad (21)$$

where $\gamma_k, g_k \geq 0$. Note, that the unconditional covariance matrix of the structural shocks is $\mathbb{E}(v_t v_t') = I_k$ as usually. Now full identification is achieved if at least $k - 1$ of the processes in (21) are non-trivial, that is $\gamma_k \neq 0$. Tests for full identification are considered by Lütkepohl and Milunovich (2016, pp. 246–253), who based on Monte Carlo evidence conclude them to have relatively low power in small samples.

Identifying the model through SVAR-GARCH has a couple of advantages. First, it is easy to rationalize the conditional covariance structure of the structural shocks. Second, as outlined by Lütkepohl and Netšunajev (2017a, p. 17), SVAR-GARCH models provide quite a lot of flexibility. This doesn't come without a price however, as estimation tends to be difficult for larger models as well as constructing reliable bootstrap intervals.

3.3.3 Smooth transition in variance regimes

Rigobon (2003) assumes an exogenous and immediate change in the volatility state of v_t . These can be a bit of strong and *ad hoc* assumptions however. Thus Lütkepohl and Netšunajev (2017b, pp. 45–46) propose the idea of two volatility states of v_t with a smooth, endogenous transition between them according to a transition function with exogenously determined transition variable.

Let's take the two covariance functions (19) and (20) from Section 3.3.1. However, instead of the structural break marking the change from one to the another, let the covariance

matrix Σ_{u_t} be a time variant combination of Σ_u^1 and Σ_u^2 . In other words,

$$\Sigma_{u_t} = (1 - G(s_t))\Sigma_u^1 + G(s_t)\Sigma_u^2 \quad (22)$$

where $G(\cdot)$ is the transition function and s_t the transition variable. Lütkepohl and Netšunajev (2017b, p. 45) use a transition function of the form

$$G(\gamma, c, s_t) = (1 + \exp[-\exp(\gamma)(s_t - c)])^{-1}. \quad (23)$$

Due to its logistic form, $0 < G(\gamma, c, s_t) < 1$. Here γ is the slope parameter, which can be interpreted to determine the speed of the transition between states Σ_u^1 and Σ_u^2 . c is the location parameter.

Note that for $\gamma \rightarrow \infty$ and $s_t < c$, $\Sigma_{u_t} \rightarrow \Sigma_u^1$ while $s_t > c$, $\Sigma_{u_t} \rightarrow \Sigma_u^2$. Interestingly, from this angle one could see the formulation of Lütkepohl and Netšunajev (2017b) as a generalization of Rigobon (2003), as setting $s_t = t$ and γ as "large" would keep the volatility state approximately at Σ_u^1 until $t > c$ at which the volatility state "jumps" to approximately Σ_u^2 , just as in the formulation of Rigobon (2003).

The smooth transition framework is beneficial as it provides a happy middle ground between the structural break models and GARCH models. Whereas the former is computationally efficient but quite inflexible and the latter is computationally inefficient but flexible, smooth transition models retain the computational efficiency while allowing for (at least often) enough flexibility. The transition between volatility states can also be given economic justifications, as the transition is captured by the transition variable. For example, Lütkepohl and Netšunajev (2017b, p. 48) use lagged values of inflation. This could be motivated by higher inflation causing more price dispersion under Calvo (1983) style price rigidities which in turn causes more disturbances in resource allocation, and thus could potentially also cause more volatile shocks.

3.3.4 Non Gaussian maximum likelihood

A lot of the theory and practices concerning VAR modelling is based on assumptions of Gaussian shocks, be it explicit or implicit. As we have seen earlier, these models do not allow for a straightforward identification of shocks as the joint distribution of the reduced form errors is determined only by their covariances (Lanne, Meitz, and Saikkonen 2017, p. 288). However, as shown by Lanne, Meitz, and Saikkonen (2017) the Gaussian case is the exception, as full identification is always achieved when the shocks are non-Gaussian

and mutually independent. The technical details are beyond the scope of this thesis, but the intuition can be clearly shown by the next simple example which closely follows Jarociński (forthcoming).

Let us assume that the price and quantity of a particular commodity follows a VAR process

$$\begin{pmatrix} \Delta q_t \\ \Delta p_t \end{pmatrix} = \begin{pmatrix} 0,6 & 0,5 \\ -0,4 & 0,7 \end{pmatrix} \begin{pmatrix} v_t^S \\ v_t^D \end{pmatrix}.$$

In other words, the changes in the price and traded quantities of the said commodity depend on the contemporaneous supply and demand shocks. Thus it is easy to extrapolate the supply curve to have a slope of $0,5/0,7 \approx 0,7$ and the demand curve a slope of $0,6/-0,4 = -1,5$. Next, let us simulate two examples of this model. In the first example, seen in the left panel of Figure 2, both shocks are Gaussian with mean zero and unity standard deviation. This Gaussian nature leads the observations to form a circle like group with no visible structure.

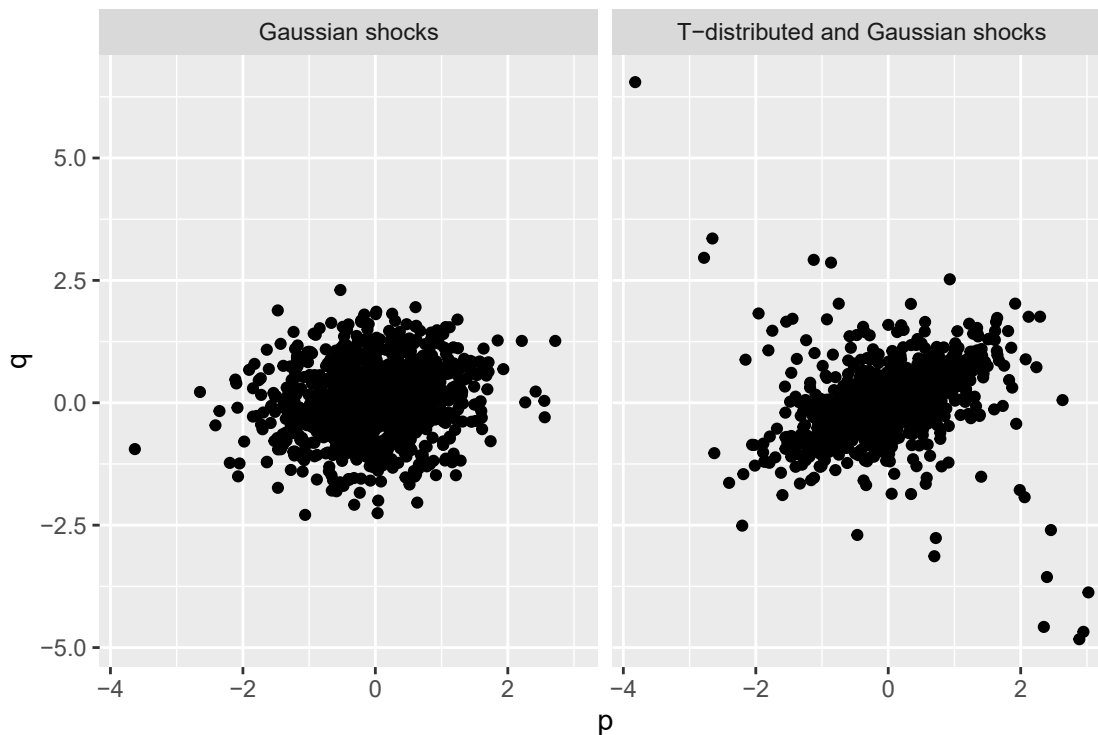


Figure 2: Simulated example of Gaussian and non-Gaussian shocks, $n=1000$.

To illustrate the fact that identification is achieved even if one of the shocks is Gaussian, now let the demand shock be as in the first example, but let the supply shocks be drawn from a Student-t distribution with the shape parameter set to 2 (for ease of illustration these shocks have been normalized to have unity standard deviation). Note, that the (causal) structure of the model stays the same, only the nature of the unforecastable

ie. from modelling point of view *random* disturbances changes. As we see from the right panel of Figure 2, the observations are spread around the supply and demand curves. This is due to the fact that the fat tailed nature of the t-distribution causes the shocks to cluster more heavily around their mean while large outliers become more prevalent than in the Gaussian case. To loosely paraphrase Jarociński (forthcoming), even an observer lacking any statistical training would have no problem identifying the structural model.

In layman terms, shocks with fat tailed distributions can be described as being almost always small but sometimes very large, while Gaussian shocks could be described as mostly small or medium and very large as often as one can spot a unicorn in the woods. Thus it is easy to see the attractiveness of the NGML identification procedure, as a lot of macroeconomically interesting shocks are clearly better characterized by the first description.

3.4 Conclusions from previous literature

As discussed in Section 3.1.1 and 3.1.2, the standard variable selection used in most monetary VARs leaves room for improvement, as most observed variables aren't necessarily the driving forces of the DGPs modelled. As central banks consider a wide range of variables, the usual models do not allow for sufficient information sets. As was discussed in Section 3.1.1, the usual procedure of adding commodity prices and exchange rates does not seem to be a reliable way to control for these shortcomings. Thus, it seems that proper model specification for monetary policy analysis requires at least some level of factor augmentation as well as considering a measure for the output gap rather than realized GDP as standard theory implies. An initial increase in inflation due to the presence of a cost channel might be possible, but cannot be expected to be large in magnitude or to last more than some months.

As discussed in Section 3.2, the standard identification procedures tend to fall short as they seem to be rejected by the data more often than not. Thus, some level of caution should be used when interpreting results acquired using these. As seen in Section 3.2, modifying these traditional identification schemes does not seem to give any additional and clear answers. Thus it seems that identification has to be thought as a case-by-case issue with great attention to the specifics of the application at hand taken.

4 Empirical evaluation

In this section I will be evaluating the performance of different model setups. For the sake of comparison, all of the models will be estimated on euro area data from January 2000 to December 2019 and will include a constant. To illustrate identification issues, models will be identified both by traditional means and a statistical procedure with a few exceptions where singularity problems were encountered. In order to make the interpretation and comparison of the impulse responses easier, a unit effect normalization in line of Stock and Watson (2016, pp. 451–453) will be utilized in the plots for impulse responses such that, a monetary shock increases the ECB interest rate by 25 basis points. Unless otherwise stated, the variables used in all of the models will be the same as in the benchmark model introduced in the next section. To illustrate estimation uncertainty, for the Cholesky identified impulse responses 0,9-percent bootstrap intervals will be drawn from 1000 simulations with the moving block bootstrap, as it is found to be quite robust to conditional heteroskedasticity, the existence of which can be seen quite reasonable (Brüggemann, Jentsch, and Trenkler 2016). For statistically identified impulse responses these will not be presented due to high computational needs.

For the sake of readability, only the impulse responses for HICP inflation will be presented in this section. The full set of impulse response estimates is available in the Appendix.

4.1 The Benchmark model

First, let's start by estimating a simple trivariate model with $Y_t = (prod_t, \pi_t, r_t)$, where the variables are the log of the ECB index for industrial production including construction (which will be used as a monthly proxy for GDP), the HICP inflation rate and the rate of the Main refinancing operations as the estimated monetary instrument. Selecting lag length for this model, Akaike, Hannan-Quinn and Bayes all suggest three. For now, a lag length of four is chosen, as it seems to better the residual diagnostics somewhat from three. Since the model is also estimated in levels, adding an extra lag might be useful in the case of integrated variables (Kilian and Lütkepohl 2017, p. 96). Although not particularly relevant for the discussion on the matter at hand, it provides a good starting point to assess the problems relating to monetary VARs.

Based on the LM , Q_1 and Q_2 tests formalized in Lütkepohl and Milunovich (2016, pp. 244–245), there seems to be evidence of at least $K - 1 = 2$ shocks with nontrivial GARCH structures implying full identification, especially since these tests tend to over-reject in practical

samples (Lütkepohl and Milunovich 2016, p. 257). This identification also heavily rejects the Cholesky restrictions. Thus identifying a viable monetary policy shock has to be made by interpreting a shock as a possible monetary shock based on its characteristics: if a shock has a major contemporaneous impact on the monetary instrument, it is a viable candidate for a monetary shock (this approach is taken for example in Lütkepohl and Netšunajev 2017b, pp. 54–55 and Lütkepohl and Netšunajev 2017a, pp. 15–16). Luckily, the model does identify such a shock. The same observations hold for the NGML identification, which identifies two non-Gaussian shocks.

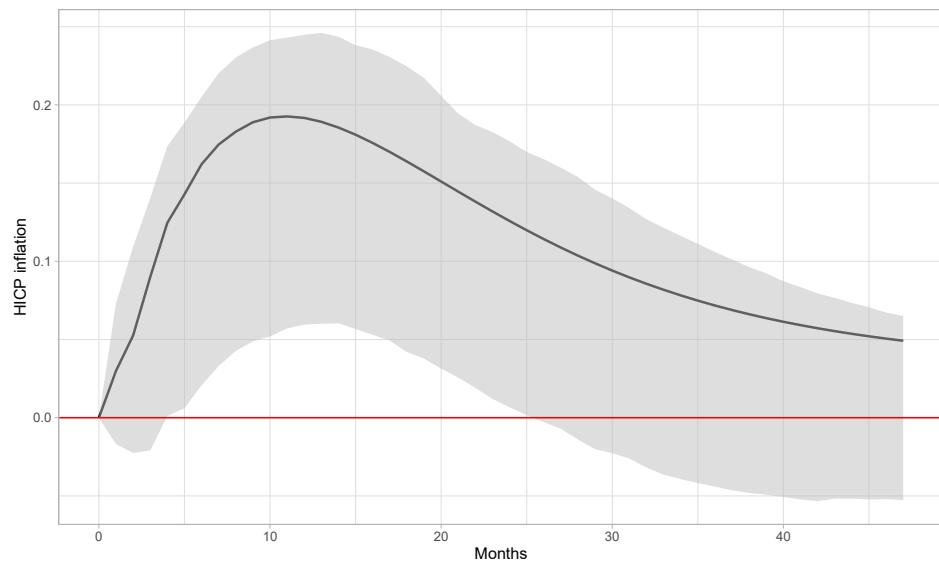


Figure 3: Benchmark impulse response estimates, Cholesky identified with 0.9-percent moving block bootstrap intervals (grey), 1000 simulations.

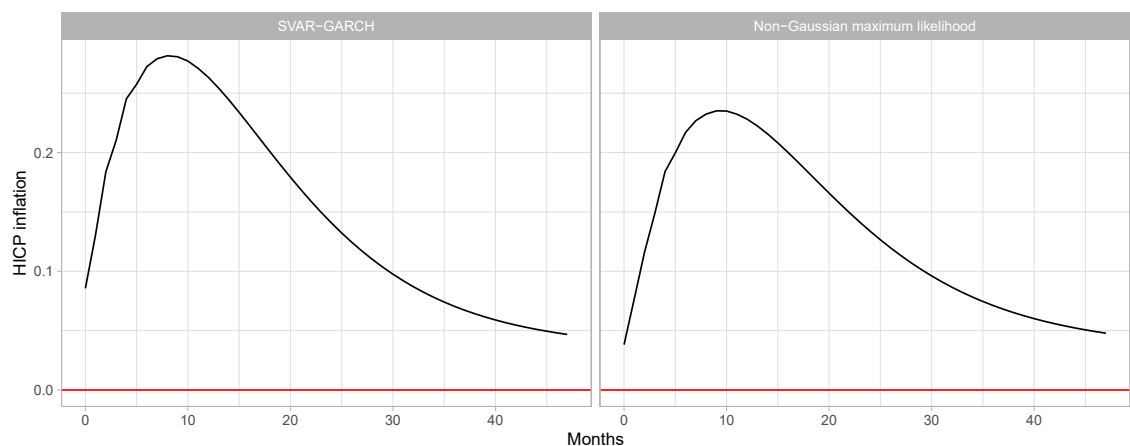


Figure 4: Benchmark impulse response estimates, statistical identification.

This benchmark model raises three interesting observations. First, as discussed in Section 3.2, the traditional Cholesky identification restrictions do not seem to hold against the data. Second, despite of this, the general shape of the impulse response does not seem

to be affected dramatically. The same observation is done by Kerssenfischer (2019, p. 22) who postulates that the identification scheme seems to play a relatively minor role compared to model specification. Third, and possibly countering the last point, although the general shape and later magnitude is the same with the Cholesky and statistically identified impulse responses, one should note the larger initial magnitude in the response. In a similar exercise, the same observation is made by Lanne, Lütkepohl, and Maciejowska (2010, p. 126), who hypothesise the (potentially) false Cholesky restrictions to hide omitted variable problems.

As a final curiosity one could mention that restricting the first interest rate lag in the inflation equation as discussed in Section 3.2.1 does reduce the magnitude of the response a little but not the general shape of it. Similarly, restricting all the interest rate lags reduces the general magnitude in about half, but again not the shape. These results are easily contributed to the fact that even in the unrestricted model the interest rate lag coefficients for inflation are small and statistically insignificant to begin with. It turns out, that in the later models, when additional variables are added (and model fits considerably bettered), these coefficients get such minuscule estimates, that it is hard to see any practical reason for manual restrictions. Thus, for our purposes this procedure will be deemed trivial.

4.2 A Mainstream model

To this day, adding control variables to account for inflation expectations remains the dominant way to address model specification. However, choosing a definite way to augment the benchmark model introduced in Section 4.1 turns out to be challenging. Based on information criteria, the single largest impact on model fit is achieved by introducing the real US to EUR exchange rate as an endogenous variable (Figure 5). After this, adding commodity prices (the log of World Bank Commodity Index) and the effective FFR as exogenous controls has only small to marginal benefits on the fit (Table 1). All models in this section were estimated with four endogenous lags and a constant (information criteria suggest either one or three), as this does a reasonable job at eliminating most of the autocorrelation from the residuals.

Table 1: Information criteria for different controlled model specifications.

	AIC	BIC
Benchmark	-1829,304	-1694,214
Exchange rates	-4976,781	-4741,240
Exchange rates, commodity prices, k=p	-5078,279	-4773,462
Exchange rates, commodity prices, FFR, k=p	-5113,056	-4738,962

As we see from Figure 5, the story for the exchange rate augmented model stays the same as for the benchmark model, as the general shapes and magnitudes stay the same. The GARCH identification is again successful, as the LM and Q_2 tests indicate significant evidence for at least three non-trivial GARCH shocks and a viable candidate for the monetary policy shock is identified. Unfortunately, ST and NGML do not indicate identification while testing Cholesky restrictions via GARCH runs into singularity problems. None of the model specifications seem to have an effect on the price puzzle.

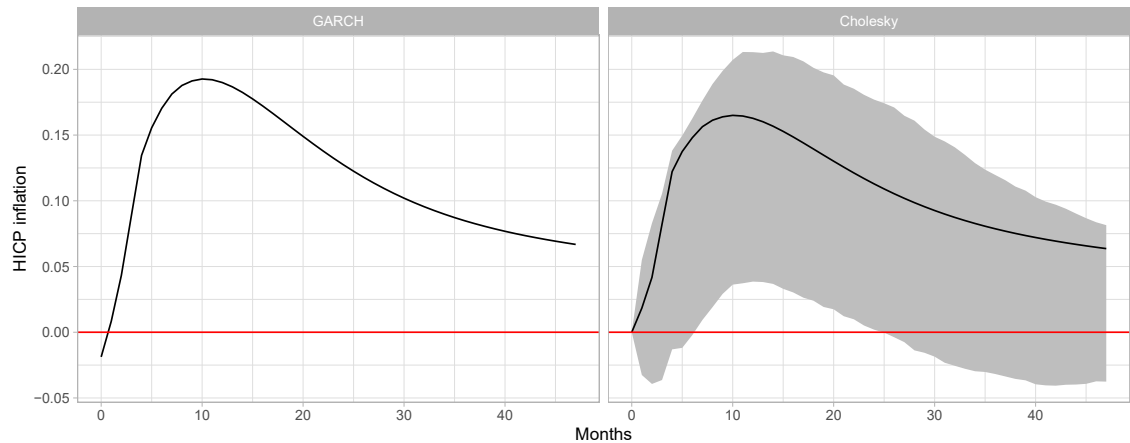


Figure 5: Exchange rates as inflation expectation proxy. Grey areas 0.9-percent moving block bootstrap intervals, 1000 simulations.

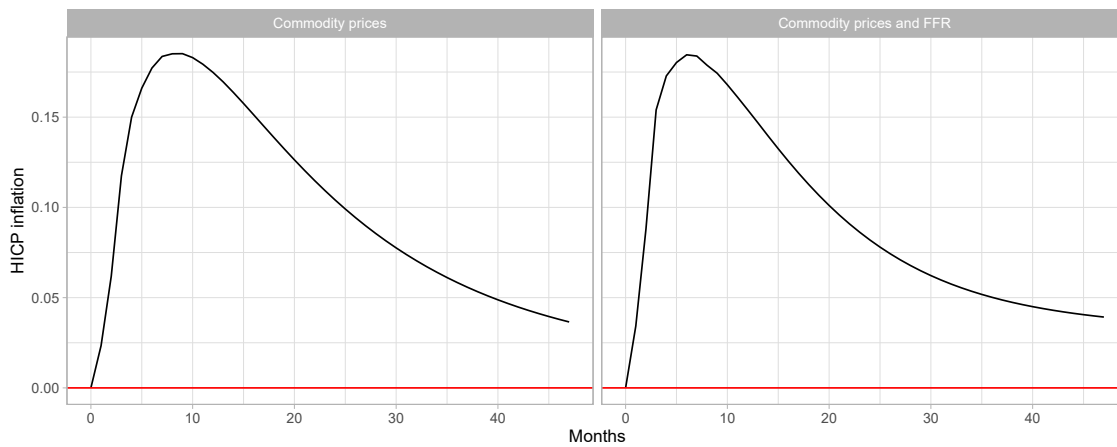


Figure 6: Commodity prices and effective Fed Funds rate as exogenous controls, Cholesky identified.

These results indicate that these usual controls are *ad hoc* at best and outright faulty at worst. It is hard to infer a clear reason why this, apparently time tested procedure, does not seem to have the expected effects. The most glaring difference to most models in the literature seems to be the lack of a trend component, though this will be touched on in Section 4.3.1. Of course, all of this should come as expected as discussed in Section 3.4.

4.3 Including the output gap

4.3.1 A note on linear trends

As seen from the exercises in previous sections, the results do not line with the ones from similar models in Peersman and Smets (2001, p. 39) for example. This is due to the fact that in the previous section we did not include deterministic trends in the model. Adding a linear trend gives us a new interpretation for industrial production as without one, it is estimated as an $I(1)$ variable. However, if we add a linear trend we are implicitly estimating the (potentially) $I(0)$ fluctuation around this trend and thus the trend could be seen as a linear approximation of potential output. In this view, adding a linear trend into the model is already a crude inclusion of the output gap.

To test this hypothesis, let us take the benchmark model of Section 4.1 and re-estimate it with linear trends. Not surprisingly, Cholesky identification is heavily rejected by the GARCH identification, for which full identification is strongly supported by Q_2 and LM tests. As we can see from Figure 7, the price puzzle is mitigated to the usual one year pos-

itive response so often seen in the literature. It also seems, that the Cholesky identification might again hide some omitted variable problems, as the GARCH identified response is circa double in magnitude during the first 12 months. Again as before and seen in the middle panel of Figure 7, adding endogenous exchange rates as well as exogenous commodity prices and Fed funds rate does not seem to have interesting effects on the results.

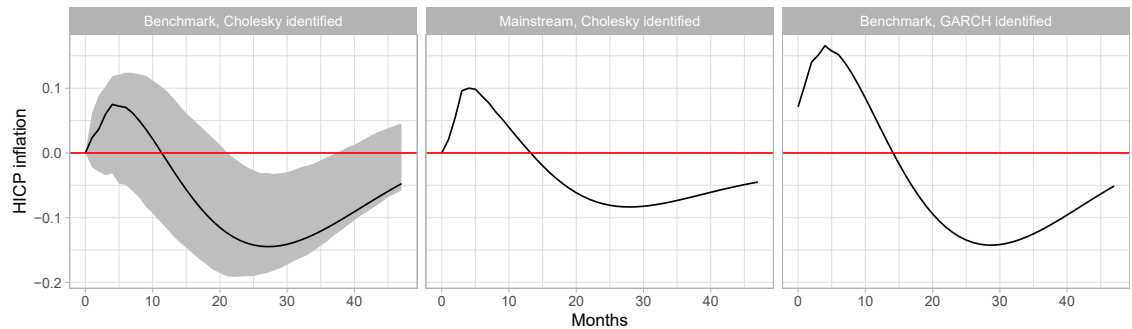


Figure 7: Benchmark and mainstream models estimated with linear trends. Moving block 90-percent bootstrap intervals in grey, 1000 simulations.

As an interesting note, repeating this exercise with quarterly data and replacing industrial production with real GDP yields different results. As seen from Figure 8, there is no price puzzle to be seen.

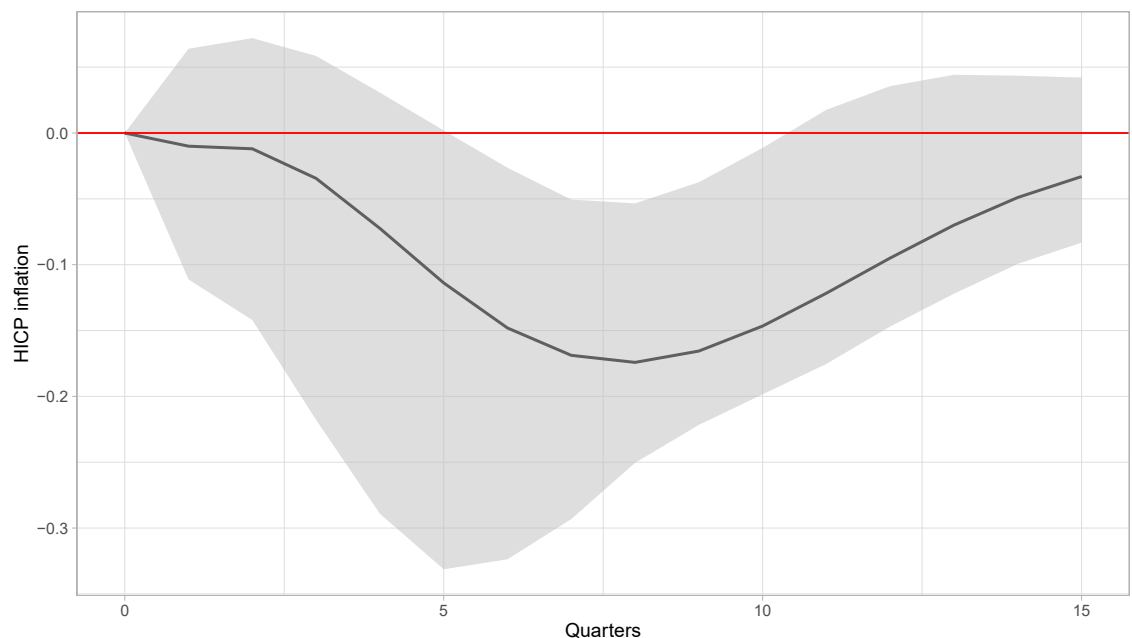


Figure 8: Benchmark model with quarterly data and the real GDP, Cholesky identified. Moving block 90-percent bootstrap intervals in grey, 1000 simulations.

Why is this? The most straightforward answer seems to be the fact that industrial production fluctuates more than GDP. One could also hypothesise industrial production to be

more affected by the cost channel and other rigidities than the service sector included in GDP.

4.3.2 A more sophisticated approach

In order to use more formal measures for the output gap we need to turn into quarterly data, as GDP in general is only reported as such. This poses an obvious drawback as it cuts the number of observations practically to less than 100 thus limiting degrees of freedom. As a result, despite several information criteria suggesting larger lag lengths (AIC six, Hannan-Quinn four), the following models are estimated with one and two lags (Bayes criteria suggests one). For the output gap I am using the baseline estimate by Morley et al. (2023, p. 6), who use BN on a BVAR to estimate the gap. Thus we get $Y_t = (gap_t, \pi_t, r_t)'$. Both models are estimated using a trend and a constant.

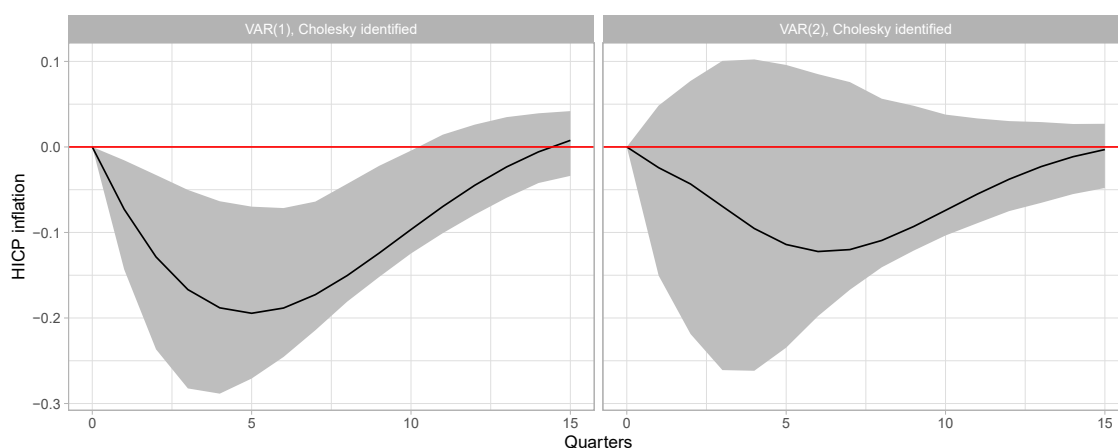


Figure 9: Benchmark and mainstream models estimated with linear trends, Cholesky identified. Moving block 90-percent bootstrap intervals in grey, 1000 simulations.

As we see from Figure 9, there is no trace of the price puzzle. It is not clear however, which estimate should be considered more reliable. Although the estimates somewhat diverge, their general shape is the same and their bootstrapped lower intervals follow very similar paths. The peak of the effect seems to also shift closer to impact compared to the linear trend model in the previous section. It should be noted that the residual diagnostics reveal major autocorrelation in the residuals and the Cholesky identification used here is strongly rejected by the GARCH identification. However, this identification is not able to identify a clear monetary shock. NGML and ST are also unsuccessful at reaching identification. Thus, Cholesky identification had to be used here.

4.4 The FAVAR approach

In what follows, common factors are estimated through principal components from a dataset of 67 time series of euro area macro indicators, price indices, commodity prices, the Eurostoxx index, Eurostoxx and oil futures and a few US macro indicators. As the usual diagnosis for the possible misspecification behind the price puzzle is the lack of information about future expectations, variables reacting to expectations, such as interest rates and stocks have been emphasized. The full list is provided in the Appendix. As can be seen from Table 2, the first principal component alone explains about 60% of the variance in the dataset, while the first five together explain about 91%. After this, individual principal components explain less than 5% of the variance in the data. Thus, in what follows, three FAVAR models will be considered: $Y_t = (\pi_t, r_t)$ augmented with one common factor, four common factors and a D-FAVAR with four common factors.

Table 2: Principal components estimated from a large data set of macroeconomic variables

	Proportion of variance	Cumulative
PC1	0,6008	0,6008
PC2	0,1849	0,7856
PC3	0,0683	0,8539
PC4	0,0545	0,9084
PC5	0,0246	0,9330

Before we continue, it is worth mentioning a few important details concerning the FAVAR approach. First, at this point any *a priori* Cholesky identification procedure should be seen as completely arbitrary. As the common factors are drawn from a large pool of data and are not attached to an obvious and concrete measurable variable, deciding which contemporaneous responses to restrict would be completely *ad hoc*. Of course, one could take the BBE route and divide the variables into "fast" and "slow" moving. With dozens or even hundreds of series it is easy to see that this procedure seems quite arbitrary too however. Of course, this is not to say these restrictions cannot be done, just that they have to be explicitly justified by over-identification for example.

Second, even if the practical pitfalls of constructing impulse responses to normal VAR models are disregarded, constructing any meaningful confidence intervals at the time of writing is - at least to my knowledge - not possible. As common factors are *estimated* from the data, they cannot be seen as given in the same sense as explicit variables in Y_t , but rather exhibiting estimation uncertainty in themselves. Thus there are three layers of uncertainty in the model: the one from common factor estimation, the one from VAR

estimation and the one from identification. Due to this, even if meaningful confidence intervals accounting for model specification and identification could be constructed, uncertainty from the common factor estimation would render these meaningless in the strict frequentist sense. In what follows, bootstrapping will be used for illustration purposes but won't be given major attention from the point of view of inference.

4.4.1 One common factor

As previously, information criteria suggest a lag length of one (Bayes) to three (Akaike). As previously, I am going with four lags, as this seems to more or less remove all relevant serial correlation and is in line with the previous models. As we see from Figure 10, there does not seem to be a price puzzle anymore.

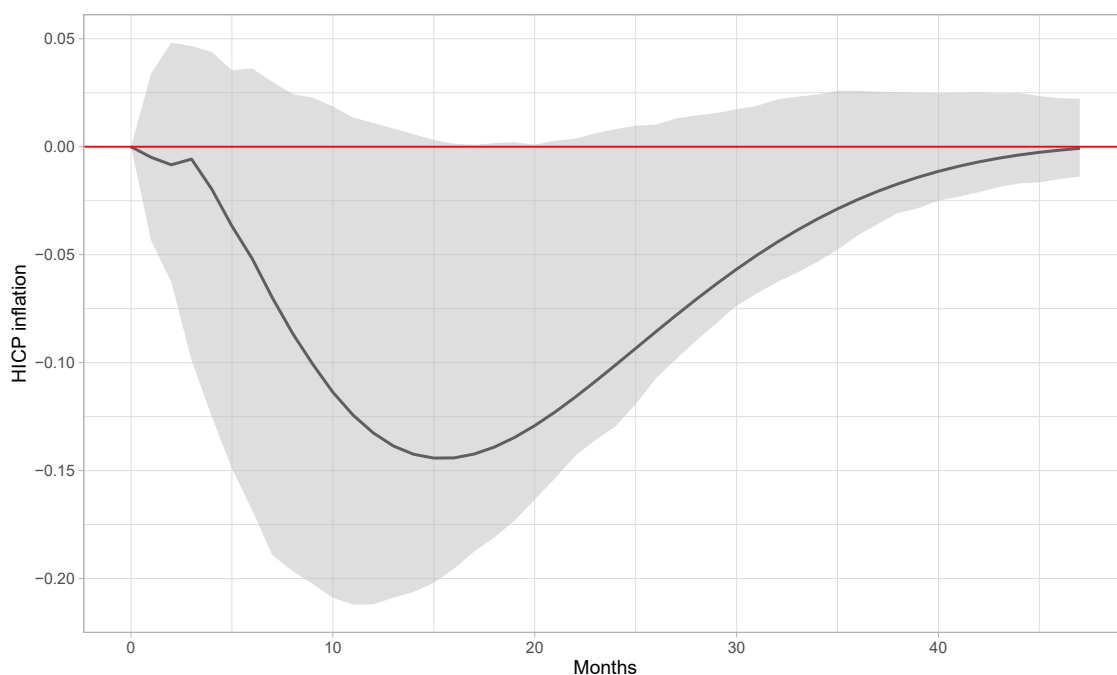


Figure 10: FAVAR with one common factor, Cholesky identified. Grey areas 0.9-percent moving block bootstrap intervals, 1000 simulations.

Turning again into statistical identification for answers, our earlier workhorse SVAR-GARCH does not help here, as there is no evidence for *any* GARCH shocks in the FAVAR. However, unlike earlier, NGML estimation identifies two shocks which seem to be non-Gaussian (Shapiro test p -values 0,16; 0,0065 and $1,03 \times 10^{-12}$) implying full identification. ST gives at least weak evidence for full identification, as the null hypothesis of $\lambda_1 = \lambda_2$ for the pairwise Wald tests is only rejected at the 0,1-confidence level with a p -value of 0,08. In addition, the estimates for these happen to be 1,72 and 0,87 making it

unclear whether they can be interpreted as differing from unity, which of course is needed for full identification. As we can see from Figure 11, the price puzzle has re-emerged indicating again that the Cholesky procedure might hide omitted variable biases.

Note, that whereas earlier the ST was identified by time t as the transition variable, here the lagged (ie. $t - 1$) values of the first common factor seem to produce the clearest evidence for identification. This highlights the usefulness of the ST procedure the FAVAR context, as a common factor includes much more information about the possible transitions in the variance structure than any single variable could.

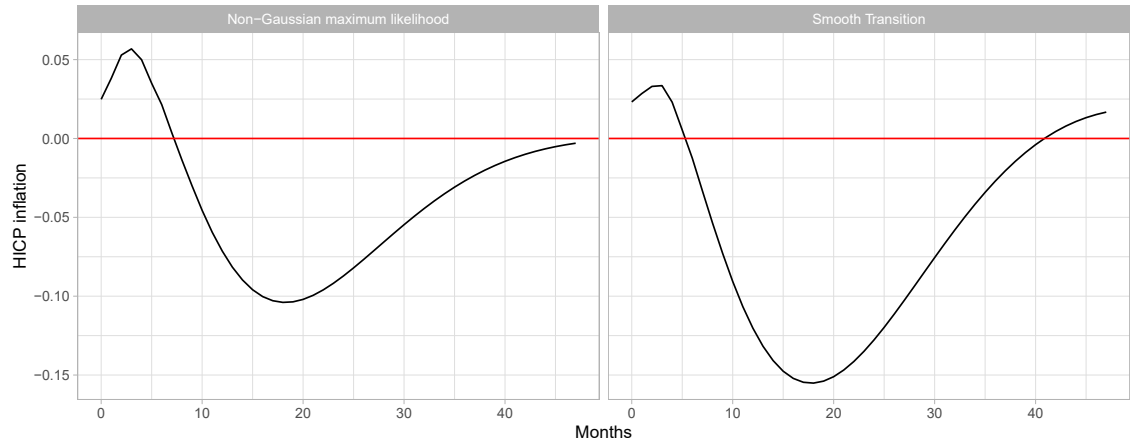


Figure 11: FAVAR with one common factor, unrestricted statistical identification.

Contemporaneous restrictions aren't completely rejected however, as restricting the contemporaneous effect of the candidate monetary shock to inflation isn't rejected. Statistical properties needed for identification stay practically the same. As seen from Figure 12 the price puzzle seems to be reduced.

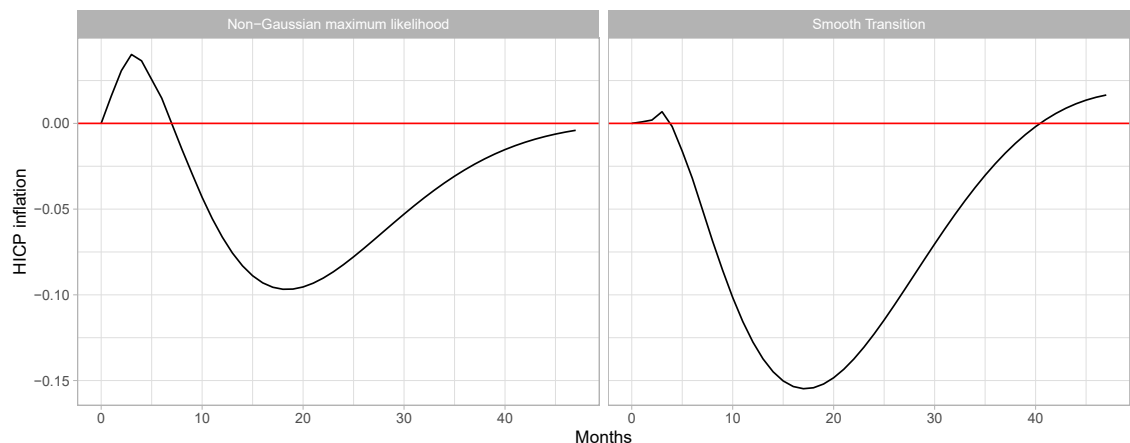


Figure 12: FAVAR with one common factor, restricted statistical identification.

4.4.2 Four common factors

Adding more common factors seems to have some conflicting results. In general, adding more common factors seems to worsen the model fits according to information criteria (as Akaike suggests 24 lags and Bayes one) while residual diagnostics get worse as well. However, at the same time the price puzzle seems to vanish practically completely. While this could be seen as pure chance, as we have seen throughout this thesis, the more information (in this case in form of common factors) is included, the more the price puzzle diminishes and the closer to impact the peak of the response seems to get. Thus, it seems likely that the model is getting better in general.

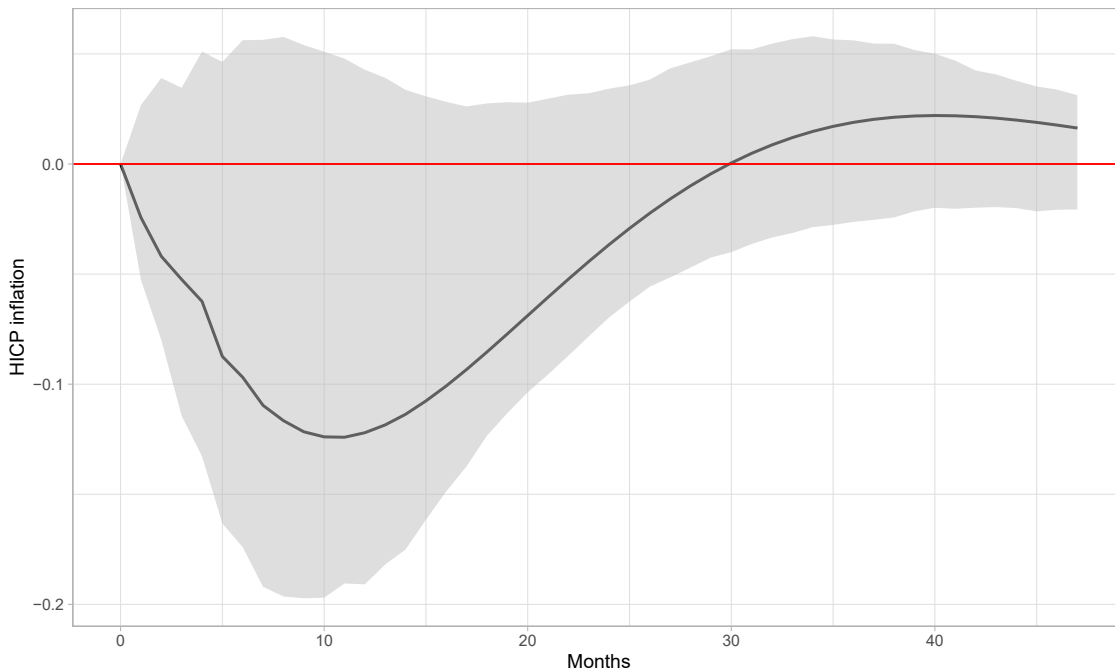


Figure 13: FAVAR with four common factors, Cholesky identified. Grey areas 0.9-percent moving block bootstrap intervals, 1000 simulations.

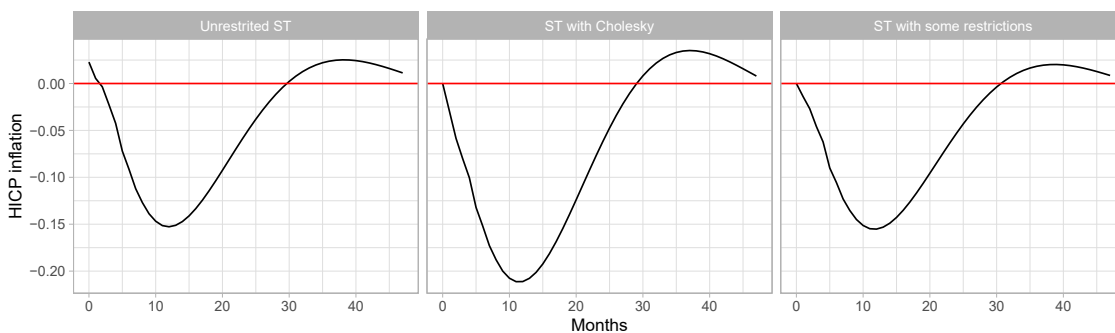


Figure 14: FAVAR with four common factors, statistical identification and additional restrictions.

Unlike with the previous FAVAR, NGML does not identify enough non-Gaussian shocks to imply full identification. For the unrestricted ST the case is the same, as out of the 15 pairwise Wald test, five produce p -values ranging from 0,15 to 0,43. What is especially interesting, is that the Cholesky identification does not get rejected (LR-test p -value 0,1047). A look at the estimated B -matrix and its standard deviations suggests that out of the restricted parameters only b_{34} and b_{16} seem to significantly differ from zero. Thus, unrestricting them means we need only two distinct, non-unity lambdas for full identification. Even a conservative interpretation of the pairwise Wald tests implies at least four distinct, non-unity lambdas which thus implies over-identification, which in turn is easily accepted by the data with a p -value of 0,8348 for the LR-test. This specification, as seen in the right panel of Figure 14, does not exhibit even the tiniest of price puzzles.

An interesting detail can also be observed by comparing the differences between impulse responses estimated the Cholesky and statistical identification or rather lack there of. Unlike earlier, there does not seem to be any major differences in shape nor peak. The magnitude might be still considered a bit smaller, but it is hard to infer whether this can be considered a result of omitted variables biases or pure chance.

As discussed earlier, FAVARs allow for impulse responses to be estimated for a wide range of variables. These impulse responses are easily acquired by a linear combination of the impulse responses of the common factors scaled by their individual loading for said variable. More formally, as a single variable x_{it} has the representation

$$x_{it} = \Lambda_i' F_t + e_t,$$

where Λ_i' is the i th row of the loading matrix Λ , the impulse response of x_i at horizon j to shock v_{kt} is

$$\frac{\partial x_{i,t+j}}{v_{kt}} = \Lambda_i' \frac{\partial F_{t+j}}{v_{kt}}$$

where F_t is the vector of common factors. It should be noted however, that reliable impulse responses would require that (almost) all of the important common factors effected by the shock of interest are included in the FAVAR, which is not as obvious as one could interpret at first glance: the number of common factors are in general chosen from the viewpoint of model performance as a whole, not the modelling performance of one noisy proxy. For now, I am going to continue with this assumption, though it could be regarded quite *ad hoc*.

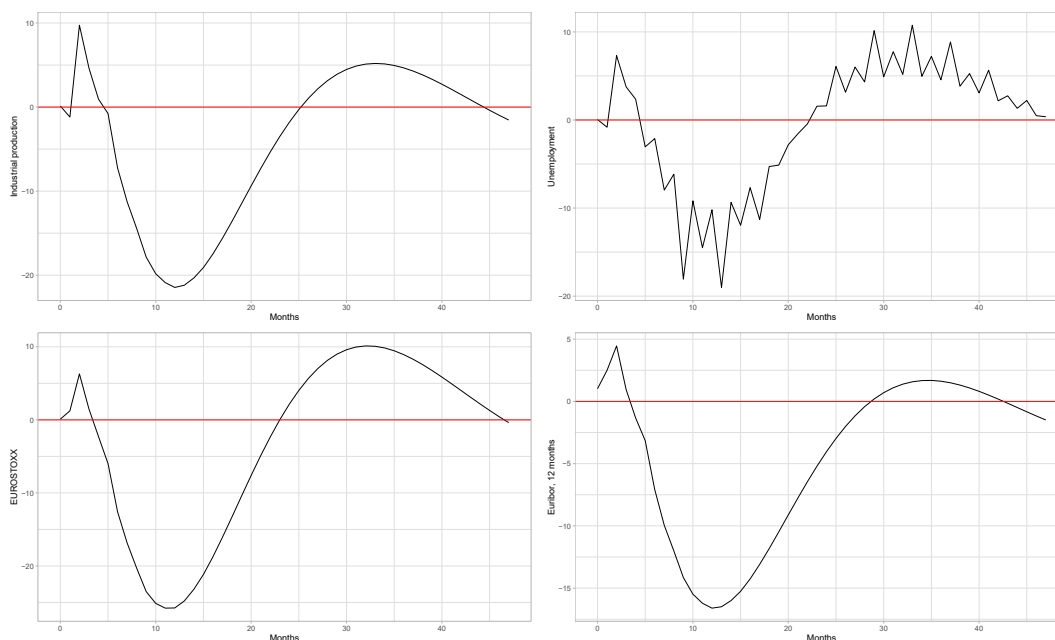


Figure 15: Impulse responses for selected variables in X_t estimated from FAVAR with four common factors.

As can be seen from Figure 15, all of the estimated impulse responses have the same general shapes. This can be attributed to the fact that all estimated common factors have very similar impulse response shapes for a monetary shock while all of the loading matrix coefficients are positive. Thus it is natural for linear combinations of these to exhibit the same kinds of shapes too. This is of course hard to comprehend for unemployment and the Euribor rate, which should by all logic go up with the monetary shock. Thus one has to conclude that there seems to be room for improvement for this model.

4.4.3 D-FAVAR expansion

Although the FAVAR with static factors in the last section seemed to perform fairly well in regards of identifying the monetary policy shock and its effects on inflation, it still seemed to leave some room for improvement. The obvious first step would be to use dynamic factors instead of static ones.

Using four common factors again, Akaike suggest estimation with six lags, Hannan-Quinn three and Bayes two. For now estimation is carried out with four lags, though the results seem to be very robust to differing lag lengths. Both static and dynamic factors are depicted in Figure 17 in normalized values.

Unlike earlier, ST procedures do not identify shocks easily interpreted as monetary shocks.

NGML however identifies four non-Gaussian shocks ie. only one less than needed for full identification. A look at the estimated B -matrix and its standard errors suggests that at least B_{16} , B_{25} , B_{46} and B_{56} can be restricted to zeros clearly over-identifying the model, which is accepted by the data with LR-test p -value of 0,2192; though the effects of different restrictions (or lack there of) on the identified monetary shock are minimal.

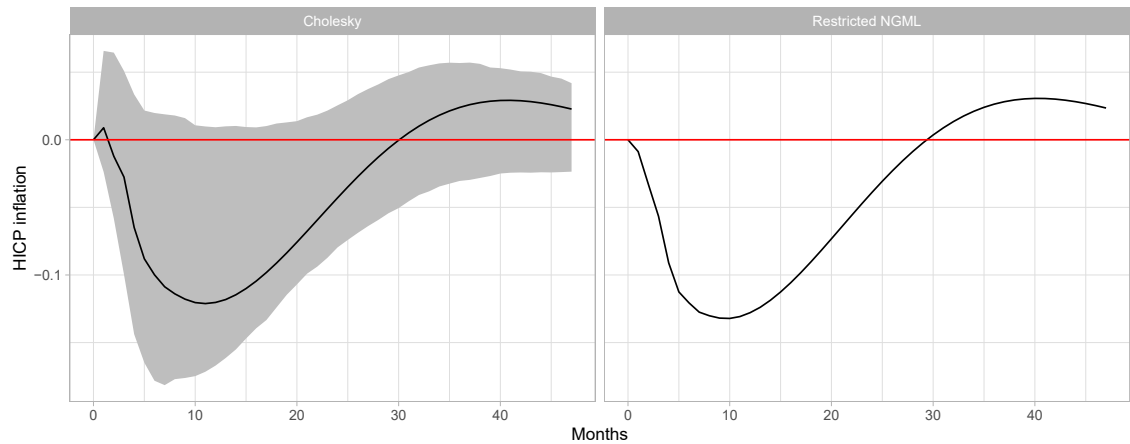


Figure 16: D-FAVAR with four common factors. Grey areas 90% moving block bootstrap intervals, 1000 simulations.

As we see from Figure 16, the estimated impulse responses align quite well with the ones estimated before. However, one interesting note can be seen in the right panel, where one can observe the NGML procedure to push the peak of the effect even closer, now peaking around 10 months after impact. This seems to continue the trend of less persistent monetary shock estimates causing shorter effects for the inflation response. This line of thought will be expanded upon in Section 4.6

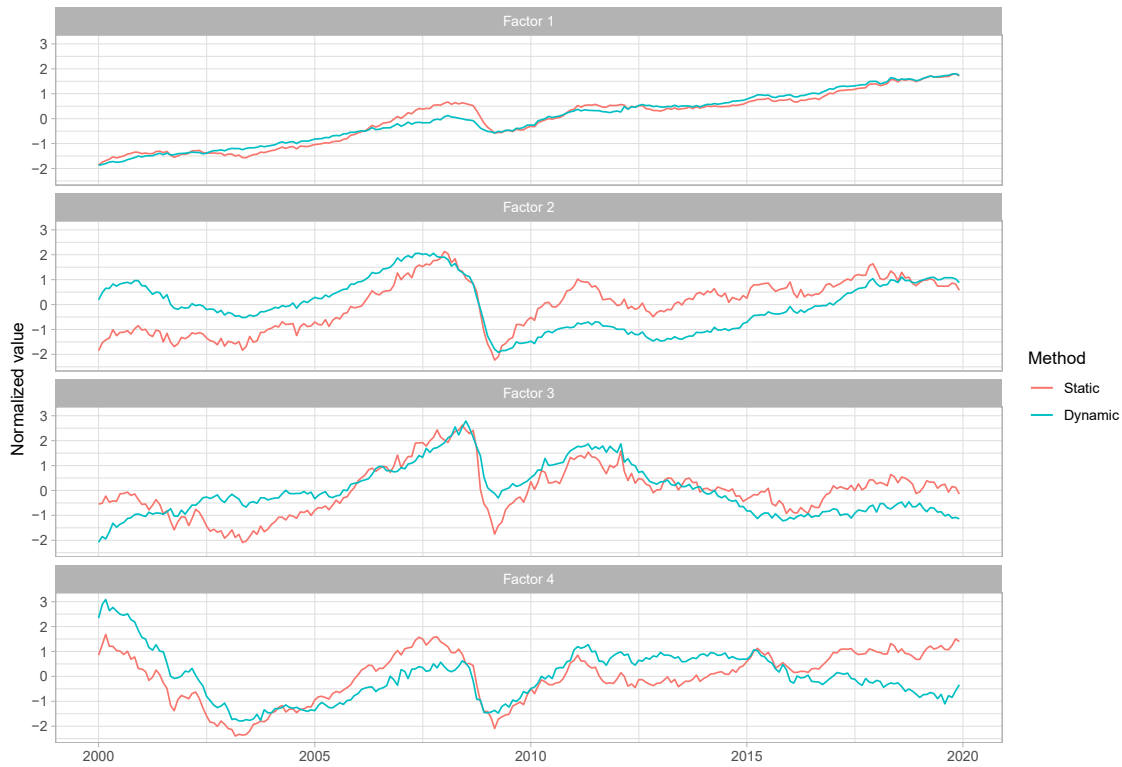


Figure 17: Comparison of static and dynamic estimates for four common factors, normalized to mean zero and unity standard deviation.

Returning to the question of the response of particular variables, as we see from Figure 18, the puzzling results from the static model seem to have improved a bit: the general response of unemployment seems to be positive while the positive values for the Euribor response have increased. Still, there seems to be room for significant improvement.

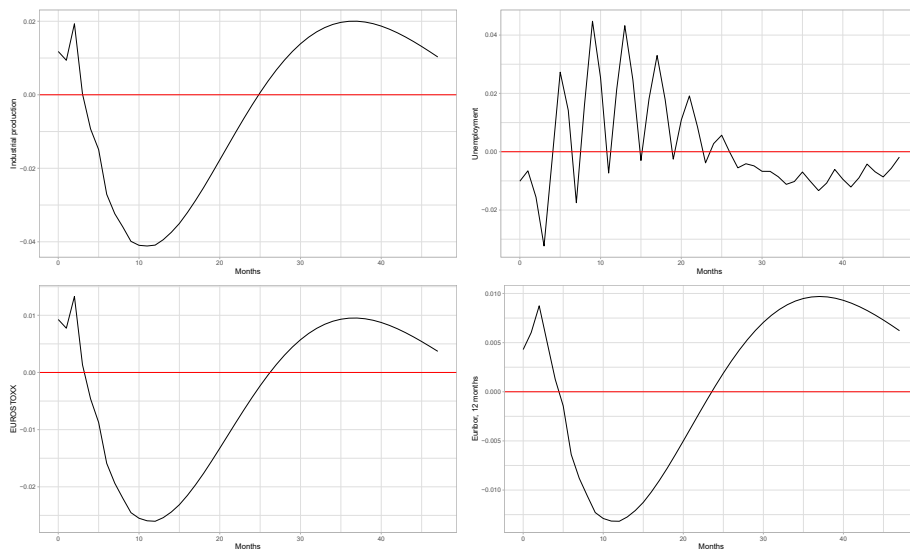


Figure 18: Impulse responses for selected variables in X_t estimated from D-FAVAR with four common factors.

4.4.4 Some notes on the estimation of common factors

In the case of both static and dynamic factors we saw two problematic features. First, the model diagnostics left room for improvement as in both cases the estimated FAVAR errors exhibited clear autocorrelation. This would indicate that systematic dynamics are left outside the model. Second, the estimated impulse responses of the variables in X_t seemed to leave some serious questions, indicating individual loading matrix weights to be somewhat off.

The most probable culprit for the puzzling impulse responses for variables in X_t is the fact that the common factors have not been cleaned of the variation caused by Y_t . That is, the estimation does not include the controlling of the second term in Equation (12). This cleaning procedure would be straight forward for a Cholesky identification as done in Bernanke, Boivin, and Elias (2005, pp. 404–405). However, this procedure requires the observed variables in X_t to be divided into "slow moving" and "fast moving", ie. ones reacting to a monetary shock contemporaneously and ones not reacting. This of course in the context of this thesis is not possible, as *a priori* assumptions on these have been avoided. Thus, other methods would be required, though coming up with such will be left outside the scope of this thesis.

The problems discussed above raise the obvious question of how problematic the lack of prior cleaning of the common factors is in practice. The results from both the static and dynamic models indicate this to be a non-issue for the estimation of the FAVAR itself, as they line up with both expectations and other models. This of course is expected, as one would not think of "cleaning" the output gap of the effects of other variables in the VAR: estimating these effects is the whole point. However, it should be quite easy to see that the lack of this controlling will have an obvious omitted variable bias on the estimates of the individual weights of the loading matrix Λ^f . In other words, the problem is not in the common factor estimates but rather the weights of individual variables.

4.5 Practical concerns regarding integrated variables

As discussed in Section 3.2.2, a VECM approach might be attractive. As VECM models explicitly count for long run relationships, they could give better insight into these than the iterated short run dynamics from a VAR model. For this reason, the VECM approach is quite often used in monetary analysis. However, before deciding to use this approach one should have clear reasons for this. Macroeconometric applications often exhibit very

complex DGP structures of unambiguous form making the parsimony principle especially important: adding more complexity into the model should always be strongly founded as applying modelling structures on data which does not actually exhibit these can lead to spurious inference at least in finite samples. Thus, there should be clear evidence for not just integrated variables (as most monetary VARs are estimated in levels to begin with) but co-integration relations. This raises the obvious question of which variables should be co-integrated? The answer will of course depend on the model specification, but especially with the kinds of models considered in this thesis, the question becomes a bit tricky.

The obvious "problem" variables from a co-integration point of view are the output gap and inflation, as it should be quite reasonable to assume both to be $I(0)$ (at least in the sense of mean convergence). By definition, the output gap should fluctuate *around* the output measure which should make it by construction $I(0)$. Inflation on the other hand should at least in the cases considered in this thesis be for the most part $I(0)$, as anything else would imply the ECB's inflation targeting to have no credibility. Thus assuming inflation to be integrated to begin with would implicitly assume the ECB to have no credibility or control of inflation (as the ECB explicitly targets inflation). Indeed, a sizable amount of empirical research indicates that at least in the developed countries the persistence of inflation has lowered significantly after the adoption of inflation targeting (see Bratsiotis, Madsen, and Martin 2015 for example). This naturally suggests that inflation should at least for the most part converge towards the rate anchored as inflation expectations.

In the context of this thesis the most potential model specification to yield any benefits from a co-integration analysis is probably the FAVAR approach, as it has more variables and no obvious reasons to exclude the possibility of co-integration.

In Table 3 we can see the results from the Johansen-procedure for determining the co-integration rank when applied to the data set used in the D-FAVAR model of Section 4.4.3. Following the suggestion of Hjalmarsson and Österholm (2010, p. 60), a "combined" approach of both the maximum eigenvalue test and the trace test (as especially the trace test tends to over-reject the null hypothesis of r or less co-integration relations) is used. This would thus suggest using $r = 2$.

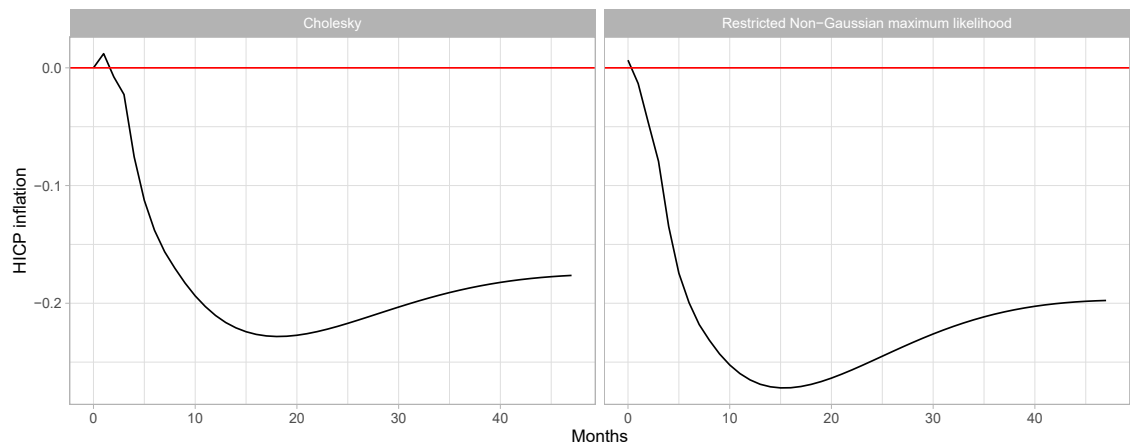


Figure 19: FAVEC with four common factors.

Table 3: Johansen procedure test statistics for the dataset used in the D-FAVAR model.

H_0	Maximum eigenvalue		Trace	
	Statistic	5% critical value	Statistic	5% critical value
$r \leq 3$	17,66	25,54	34,70	42,44
$r \leq 2$	29,30	31,46	64,00*	62,99
$r \leq 1$	41,63*	37,52	105,63*	87,31
$r \leq 0$	54,93*	43,97	160,56*	114,90

Test statistic values significant on the 5% level marked with *.

Identification-wise it has to be said that long run restrictions could not be run without singularity issues, so for now the two identification procedures used are the Cholesky and NGML procedures. The NGML identifies three non-Gaussian shocks and as such two restrictions on B are made, b_{16} and b_{56} (note that the candidate monetary shock is identified on the fourth row so these restrictions do not affect it). As seen from Figure 19, both procedures identify a persistent effect on inflation following a monetary shock, suggesting inflation to be $I(1)$. The estimated effect magnitude is also considerably larger than with the VAR estimates, which makes sense without a mean converging effect of mean stationarity.

The results from the SVEC analysis raise the obvious question of the degree of integration for inflation especially, as the estimated persistence obviously clashes with the zero converging estimates from VAR models. As is discussed in length by Hjalmarsson and Österholm (2010), the Johansen procedure is not without its fair share of problems and tends to over-reject especially in the presence of near-integrated variables, and as such "[t]he risk of concluding that completely unrelated series are co-integrated is therefore non-negligible" (Hjalmarsson and Österholm 2010, p. 51). It should also be noted, that the Augmented-Dickey-Fuller test rejects the null-hypothesis of a unit root for the infla-



Figure 20: Euro area HICP inflation between 2000 and 2020.

tion series against the alternative hypothesis of stationarity with a $p = 0,0299$. One could of course cite the critique of Cochrane (1991) against too far reaching conclusions about unit root tests in finite samples, which would underline even more the muddled waters one has to traverse through here. Thus it might be useful to take a step back and instead of relying too much on statistical procedures with questionable power to take a more intuitive look at the plotted series of inflation. It is easy to see from Figure 20 that even with the fluctuations following the financial crisis of 2008, until the introduction of ECBs unconventional policies in 2012 inflation seems to converge towards the 2% target. After this however, inflation starts to exhibit a more of a wandering dynamic. This could indicate that the rank of integration has changed.

Although the possibility of co-integration and especially inflation being $I(1)$ cannot be ignored, the evidence for such seem to be inconclusive. Thus, for now it seems more appropriate to follow the parsimony principle and not include more complexity into the model without substantial reasoning and evidence. Despite this and the discrepancies between the VAR and SVEC estimates regarding magnitude sizes and long term effects, it is interesting to note that the peak of the response seems to align with earlier models. This seems to reinforce the earlier conclusions of monetary transmission not being as lagged as usually credited.

4.6 Some notes on the identified monetary shocks

An interesting observation one can make from the many different models and the monetary shocks identified from them is that the persistence of the identified monetary shock seems

to vary quite a lot. As we can see in Figure 21, the identified monetary shocks from the traditional low-dimensional models vary greatly in shape and persistence.

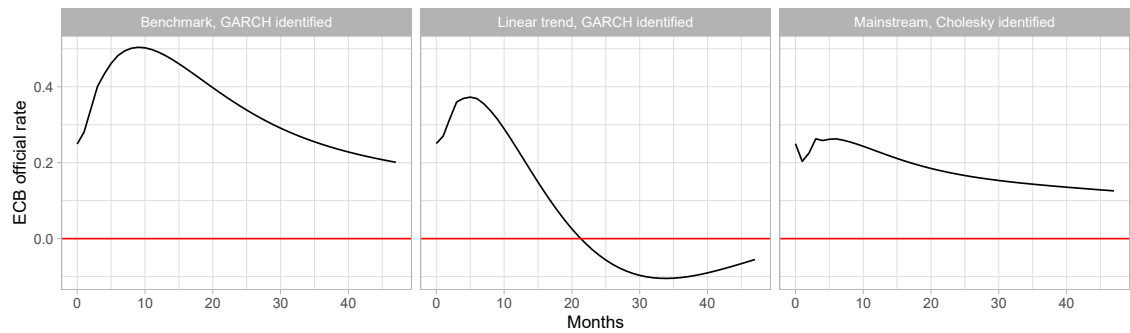


Figure 21: Some monetary shocks identified from traditional low-dimensional models.

This raises two obvious questions. First, how realistic is the high persistence of these shocks and second, how realistic are the shapes of these shocks? As was discussed in Section 2.1, the nature and source of monetary shocks should be such that high persistence should be improbable (unless the monetary instrument follows a random walk making the persistence infinite). If one assumes the policy instrument to follow a coherent rule, say a linear function of a set of macroeconomic indicators like in Equation (1), one should ponder why a shock should be positively persistent. If there is a shock in the policy rate, the rate would be too high given the rule of the central bank. This should slow down economic activity and inflation more than intended, which in turn should imply that the central bank should, according to its rule face downward pressure for the rate. If anything, one should thus expect the impulse response to the rate to exhibit a relatively steep decline, possibly below zero. After all, if a "too high" policy rate does not slow down the economy more than intended and thus does not create downward pressure to the policy rate, then why should the central bank bother with a policy rate as its instrument in the first place? Subsequently, this is exactly the kind of behaviour seen in the ECB official rate response to monetary shocks identified from the FAVAR and output gap models, as seen in Figure 22.

One could also ponder the shape of the benchmark and linear trend models. As seen in the left and right panel of Figure 21, the estimated impulse responses do not peak at impact but rather after it and in the case of the benchmark model the peak is circa twice as large as the initial impact. Due to the reasons discussed earlier, this seems quite unreasonable.

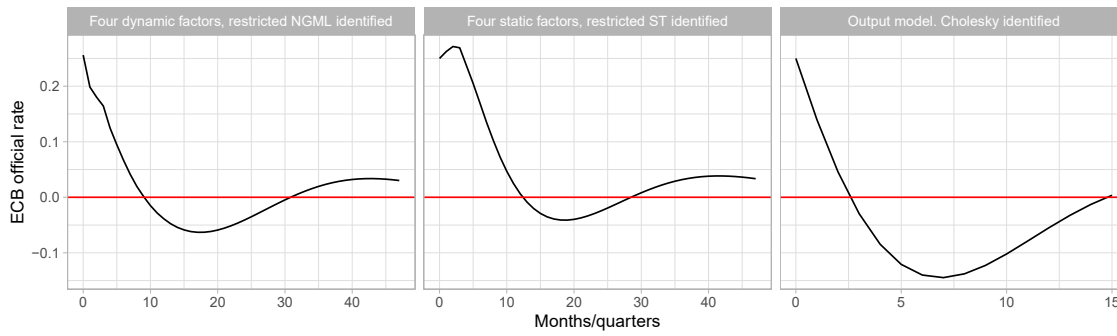


Figure 22: Some monetary shocks identified from better performing models.

As discussed in Section 2.5, identifying shocks from models with non-fundamental representations would require inverting the VMA in the future. As this is not possible in a VAR setup, identified shocks will be linear combinations of a larger set of shocks than the model can truly identify. Thus one could argue the reason for persistent monetary shocks to lie in non-fundamentalness: the shock labelled as a monetary shock is truly a combination of several shocks. Thus, assuming that monetary shocks exhibit the kinds of properties discussed earlier, one could use the shape of the identified monetary shock as an informal test for fundamentalness.

5 Concluding remarks

This thesis contributes to the existing econometric literature in several ways. First, I show evidence to support the conclusions from Kerssenfischer (2019, p. 24) who credits the price puzzle to be mostly a result of misspecification rather than identification. My results show that when a proper output gap measure is used in a low-dimensional VAR or a FAVAR is estimated, the resulting dynamic effects are not puzzling, and these results are qualitatively fairly robust across identification procedures. It could be added though, that specification seems to be a mandatory requisite for truly identifying and isolating monetary shocks and thus the question of identifying shocks can not be isolated from model specification.

Second, although the effect magnitudes of this are somewhat unclear, my results seem to support the hypothesis put out by Lanne, Lütkepohl, and Maciejowska (2010, p. 126). That is, unjustifiable Cholesky identification might hide omitted variable biases by reducing estimated effect sizes. This finding is still quite anecdotal however, and more formal research is warranted. In addition, supporting a growing literature on the matter, almost all of the models presented in this thesis reject the Cholesky identification procedure. In light of these two findings, it seems reasonable to conclude that for macroeconometric applications in general, Cholesky identification cannot be regarded as *a priori* justified. Statistical identification procedures might be used to offset this, but the specifics are left as a case-by-case matter.

Third, as FAVAR and output gap approaches converge in their findings regarding inflation response to a monetary shock, the empirical findings of this thesis seem quite clear: there is no price puzzle in the euro area, and the effect of a 25 basis point monetary policy shock is a reduction of 10 to 15 basis points in the rate of HICP inflation peaking around 12 months. Thus it seems likely that the transmission channels of monetary policy might not be as "long and variable" as often attributed. This finding might underline the importance of proper specification of variables and the underlying monetary regimes.

Fourth, as more complete information sets are introduced either in the form of factor augmentation or the output gap, the peak of the effect seems to get closer to impact. This seems to support the recent trends mentioned by Powell (2022): the lags of monetary policy aren't necessarily as long as previously thought. In light of the findings of this thesis, in the euro area the peak of the effect seems to happen around 12 months, though some evidence of even faster effect peaks can be found. This is substantial, as especially the older literature seemed to indicate that this is the moment when the effect only starts to

turn negative. The estimates from the VAR models also imply the effects of a monetary shock on inflation as a whole are contained within three years. Thus these results support the conclusion that ECBs policy rates are an effective instrument for inflation targeting.

Fifth, the many exercises conducted in Section 4 shed some light on the practical concerns for applied work. As seen in Sections 4.3 and 4.4, there doesn't seem to be a one-size-fits-all solution to model specification. In light of these findings it seems reasonable to recommend a FAVAR approach when working with monthly data frequencies, as suitable measures for the output gap are usually not available. Using FAVARs is also recommended when the effects of monetary policy on other variables than just output and inflation is of interest. When working with quarterly data, a low-dimensional approach might be more straight forward to execute when a suitable measure for the output gap is available. However, quarterly data might impose challenges when shorter sample sizes are of interest. As the results seen in this thesis seem to diverge from a lot of the literature especially with regard to the effect lags, it also seems that careful consideration regarding the span of the data and the underlying monetary regimes is of utmost importance: despite the seductive allure of the degrees of freedom gained, more data might lead to spurious inference.

As discussed in Section 4.5, the question of integration regarding inflation is one of great importance, as it has major implications on the long term effects of monetary policy and the credibility of the central banks inflation targeting. Thus the possibility of a structural break in the rank of integration with the introduction of ECBs unconventional monetary policies might provide fruitful soil for future research.

References

- Alessi, Lucia, Matteo Barigozzi, and Marco Capasso (2008). “A review of nonfundamentality and identification in structural VAR models”. In.
- Angrist, Joshua D and Victor Lavy (1999). “Using Maimonides’ rule to estimate the effect of class size on scholastic achievement”. In: *The Quarterly journal of economics* 114.2, pp. 533–575.
- Barth III, Marvin J and Valerie A Ramey (2001). “The cost channel of monetary transmission”. In: *NBER macroeconomics annual* 16, pp. 199–240.
- Benkwitz, Alexander, Michael H Neumann, and Helmut Lütkepohl (2000). “Problems related to confidence intervals for impulse responses of autoregressive processes”. In: *Econometric Reviews* 19.1, pp. 69–103.
- Bernanke, Ben S, Jean Boivin, and Piotr Elias (2005). “Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach”. In: *The Quarterly journal of economics* 120.1, pp. 387–422.
- Blanchard, Olivier Jean and Danny Quah (1989). “The Dynamic Effects of Aggregate Demand and Supply Disturbances”. In: *American Economic Review* 79.4, pp. 655–673.
- BoF (2023). *Suomen Pankin sijaisjärjestelyt Olli Rehnin presidenttiehdokkuuden johdosta*. url: <https://www.suomenpankki.fi/fi/media-ja-julkaisut/tiedotteet/2023/suomen-pankin-sijaisjarjestelyt-olli-rehnin-presidenttiehdokkuuden-johdosta/> (visited on 01/09/2024).
- Boivin, Jean and Serena Ng (2006). “Are more data always better for factor analysis?” In: *Journal of Econometrics* 132.1, pp. 169–194.
- Borys, Magdalena Morgese, Roman Horváth, and Michal Franta (2009). “The effects of monetary policy in the Czech Republic: an empirical study”. In: *Empirica* 36, pp. 419–443.
- Bratsiotis, George J, Jakob Madsen, and Christopher Martin (2015). “Inflation targeting and inflation persistence”. In: *Economic and Political Studies* 3.1, pp. 3–17.
- Brüggemann, Ralf, Carsten Jentsch, and Carsten Trenkler (2016). “Inference in VARs with conditional heteroskedasticity of unknown form”. In: *Journal of econometrics* 191.1, pp. 69–85.
- Calvo, Guillermo A (1983). “Staggered prices in a utility-maximizing framework”. In: *Journal of monetary Economics* 12.3, pp. 383–398.
- Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans (1999). “Monetary policy shocks: What have we learned and to what end?” In: *Handbook of macroeconomics* 1, pp. 65–148.

- Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans (2005). “Nominal rigidities and the dynamic effects of a shock to monetary policy”. In: *Journal of political Economy* 113.1, pp. 1–45.
- Cochrane, John H (1991). “A critique of the application of unit root tests”. In: *Journal of Economic Dynamics and Control* 15.2, pp. 275–284.
- (2023). *Interest rates and inflation part 2: Losing faith in VARs*. url: <https://johnhcochrane.blogspot.com/2023/08/interest-rates-and-inflation-part-2.html> (visited on 11/04/2023).
- Dornbusch, Rudiger (1976). “Expectations and exchange rate dynamics”. In: *Journal of political Economy* 84.6, pp. 1161–1176.
- ECB (2021). *The ECB’s monetary policy strategy statement*. url: https://www.ecb.europa.eu/home/search/review/html/ecb.strategyreview_monpol_strategy_statement.en.html (visited on 11/04/2023).
- (2023). *Account of the monetary policy meeting of the Governing Council of the European Central Bank held in Frankfurt am Main on Wednesday and Thursday, 13-14 September 2023*. url: <https://www.ecb.europa.eu/press/accounts/2023/html/ecb.mg231012~2f3d803d32.en.html> (visited on 11/04/2023).
- Estrella, Arturo (2015). “The price puzzle and var identification”. In: *Macroeconomic Dynamics* 19.8, pp. 1880–1887.
- Fernald, John G, Mark M Spiegel, and Eric T Swanson (2014). “Monetary policy effectiveness in China: Evidence from a FAVAR model”. In: *Journal of International Money and Finance* 49, pp. 83–103.
- Fisher, Paul G, Lavan Mahadeva, and John D Whitley (1997). “The output gap and inflation—Experience at the Bank of England”. In: *BIS Conference Papers*. Vol. 4, pp. 68–90.
- Forni, Mario, Luca Gambetti, and Luca Sala (2019). “Structural VARs and noninvertible macroeconomic models”. In: *Journal of Applied Econometrics* 34.2, pp. 221–246.
- Forni, Mario, Domenico Giannone, et al. (2009). “Opening the black box: Structural factor models with large cross sections”. In: *Econometric Theory* 25.5, pp. 1319–1347.
- Giordani, Paolo (2001). *An alternative explanation of the prize puzzle*. Tech. rep. SSE/EFI Working Paper Series in Economics and Finance.
- Hanson, Michael S (2004). “The “price puzzle” reconsidered”. In: *Journal of Monetary Economics* 51.7, pp. 1385–1413.
- Henzel, Steffen et al. (2009). “The price puzzle revisited: Can the cost channel explain a rise in inflation after a monetary policy shock?” In: *Journal of Macroeconomics* 31.2, pp. 268–289.
- Higo, Masahiro, Sachiko-Kuroda Nakada, et al. (1999). “What determines the relation between the output gap and inflation? An international comparison of inflation ex-

- pectations and staggered wage adjustment”. In: *Monetary and Economic Studies* 17.3, pp. 129–155.
- Hjalmarsson, Erik and Pär Österholm (2010). “Testing for cointegration using the Johansen methodology when variables are near-integrated: size distortions and partial remedies”. In: *Empirical Economics* 39, pp. 51–76.
- Holguín, Juan S and Jorge M Uribe (2020). “The credit supply channel of monetary policy: evidence from a FAVAR model with sign restrictions”. In: *Empirical Economics* 59.5, pp. 2443–2472.
- Jarociński, Marek (forthcoming). “Estimating the Fed’s unconventional policy shocks”. In: *Journal of Monetary Economics*.
- Kerssenfischer, Mark (2019). “The puzzling effects of monetary policy in VARs: Invalid identification or missing information?” In: *Journal of Applied Econometrics* 34.1, pp. 18–25.
- Kilian, Lutz (1998). “Confidence intervals for impulse responses under departures from normality”. In: *Econometric Reviews* 17.1, pp. 1–29.
- Kilian, Lutz and Helmut Lütkepohl (2017). *Structural vector autoregressive analysis*. Cambridge University Press.
- Krusec, Dejan (2010). “The “price puzzle” in the monetary transmission VARs with long-run restrictions”. In: *Economics Letters* 106.3, pp. 147–150.
- Kydland, Finn E and Edward C Prescott (1977). “Rules rather than discretion: The inconsistency of optimal plans”. In: *Journal of political economy* 85.3, pp. 473–491.
- Laine, Olli-Matti Juhani (2020). “The effect of the ECB’s conventional monetary policy on the real economy: FAVAR-approach”. In: *Empirical Economics* 59, pp. 2899–2924.
- Lanne, Markku and Helmut Lütkepohl (2008). “Identifying monetary policy shocks via changes in volatility”. In: *Journal of Money, Credit and Banking* 40.6, pp. 1131–1149.
- Lanne, Markku, Helmut Lütkepohl, and Katarzyna Maciejowska (2010). “Structural vector autoregressions with Markov switching”. In: *Journal of Economic Dynamics and Control* 34.2, pp. 121–131.
- Lanne, Markku, Mika Meitz, and Pentti Saikkonen (2017). “Identification and estimation of non-Gaussian structural vector autoregressions”. In: *Journal of Econometrics* 196.2, pp. 288–304.
- Lütkepohl, Helmut (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Lütkepohl, Helmut, Mika Meitz, et al. (2021). “Testing identification via heteroskedasticity in structural vector autoregressive models”. In: *The Econometrics Journal* 24.1, pp. 1–22.
- Lütkepohl, Helmut and George Milunovich (2016). “Testing for identification in SVAR-GARCH models”. In: *Journal of Economic Dynamics and Control* 73, pp. 241–258.

- Lütkepohl, Helmut and Aleksei Netšunajev (2017a). “Structural vector autoregressions with heteroskedasticity: A review of different volatility models”. In: *Econometrics and statistics* 1, pp. 2–18.
- (2017b). “Structural vector autoregressions with smooth transition in variances”. In: *Journal of Economic Dynamics and Control* 84, pp. 43–57.
- Morley, James et al. (2023). “Estimating the euro area output gap using multivariate information and addressing the COVID-19 pandemic”. In: *European Economic Review* 153, p. 104385.
- Normandin, Michel and Louis Phaneuf (2004). “Monetary policy shocks:: Testing identification conditions under time-varying conditional volatility”. In: *Journal of Monetary Economics* 51.6, pp. 1217–1243.
- Peersman, Gert and Frank Smets (2001). “The monetary transmission mechanism in the euro area: more evidence from var analysis (mtn conference paper)”. In: *Available at SSRN 356269*.
- Powell, Jerome (2022). *Transcript of Chair Powell’s Press Conference November 2, 2022*.
- Rabanal, Pau (2007). “Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model”. In: *Journal of Economic Dynamics and control* 31.3, pp. 906–937.
- Rigobon, Roberto (2003). “Identification through heteroskedasticity”. In: *Review of Economics and Statistics* 85.4, pp. 777–792.
- Romer, David (2019). *Advanced Macroeconomics*. McGraw-Hill Education.
- Rusnák, Marek, Tomas Havranek, and Roman Horváth (2013). “How to solve the price puzzle? A meta-analysis”. In: *Journal of Money, Credit and Banking* 45.1, pp. 37–70.
- Sims, Christopher A (1980). *Comparison of interwar and postwar business cycles: Monetarism reconsidered*.
- (1992). “Interpreting the macroeconomic time series facts: The effects of monetary policy”. In: *European economic review* 36.5, pp. 975–1000.
- Stock, James H and Mark W Watson (2002). “Forecasting using principal components from a large number of predictors”. In: *Journal of the American statistical association* 97.460, pp. 1167–1179.
- (2016). “Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics”. In: *Handbook of macroeconomics*. Vol. 2. Elsevier, pp. 415–525.
- Taylor, John B (1999). “A historical analysis of monetary policy rules”. In: *Monetary policy rules*. University of Chicago Press, pp. 319–348.

A Appendix: Estimated impulse responses

In what follows are the impulse responses estimates from the key models in this thesis. Note that these impulse shocks are the raw ones straight out of estimation and thus have not been standardized to 25 basis points. As such their magnitude and sign might vary from the ones presented earlier in this thesis. The results are still the same, as linear transformations do not change the properties of impulse responses from a linear model such as these VARs. Grey areas are the 90-percent moving block bootstrap intervals from 1000 simulations. Due to high computational needs bootstrapping has been conducted only to Cholesky identified models.

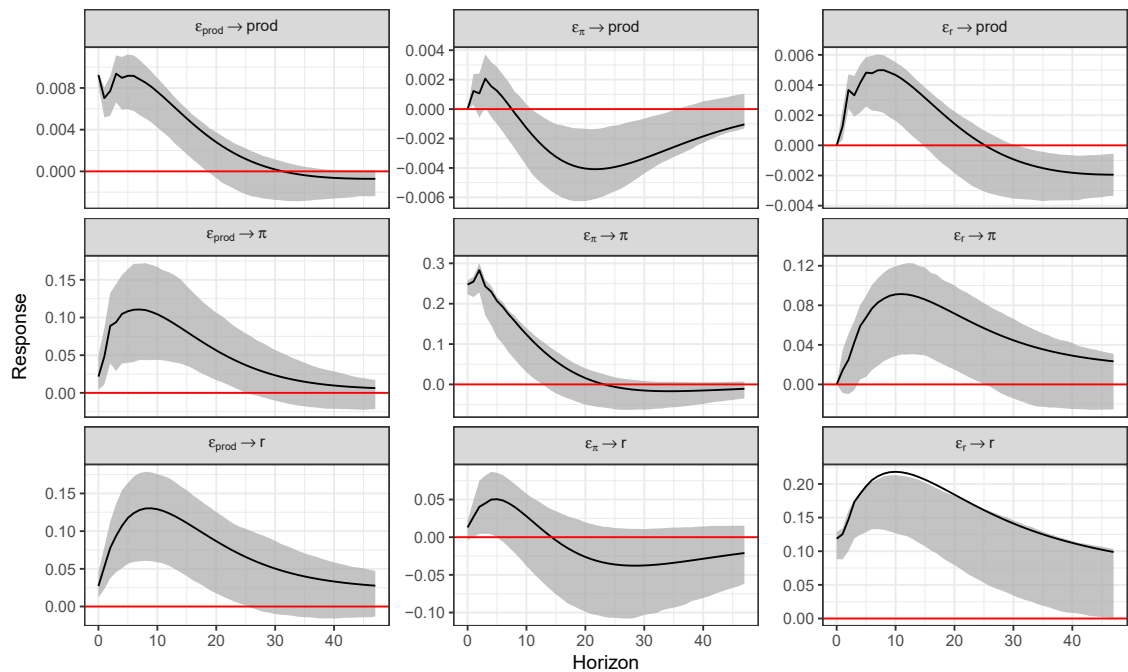


Figure 23: Benchmark model, Cholesky identified

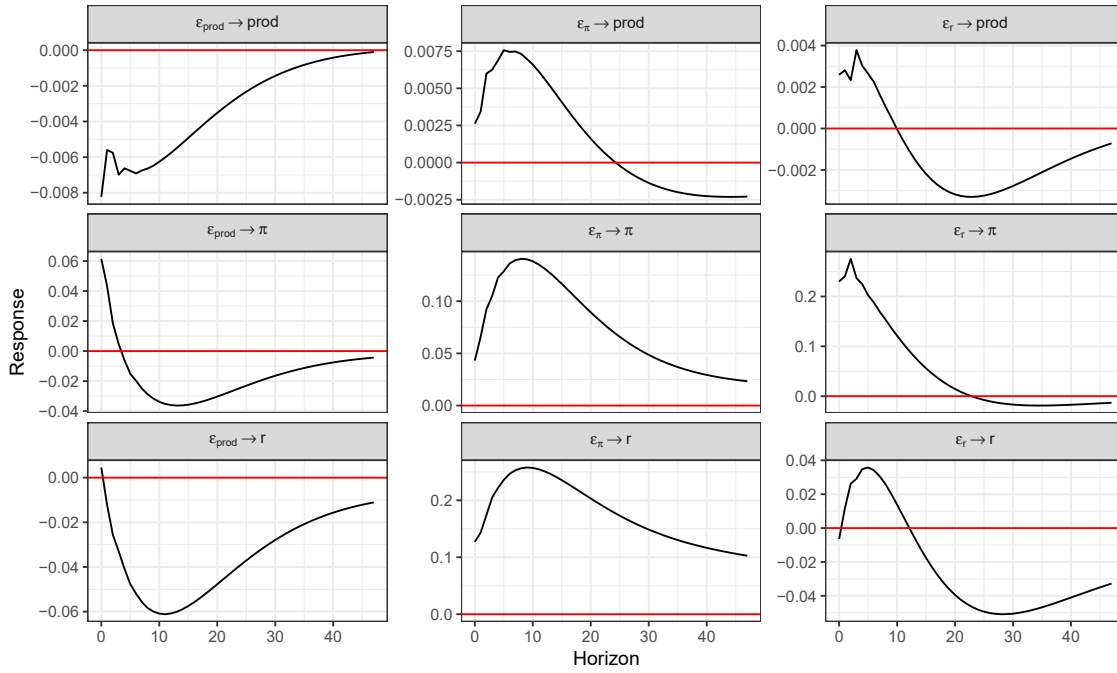


Figure 24: Benchmark model, GARCH identified

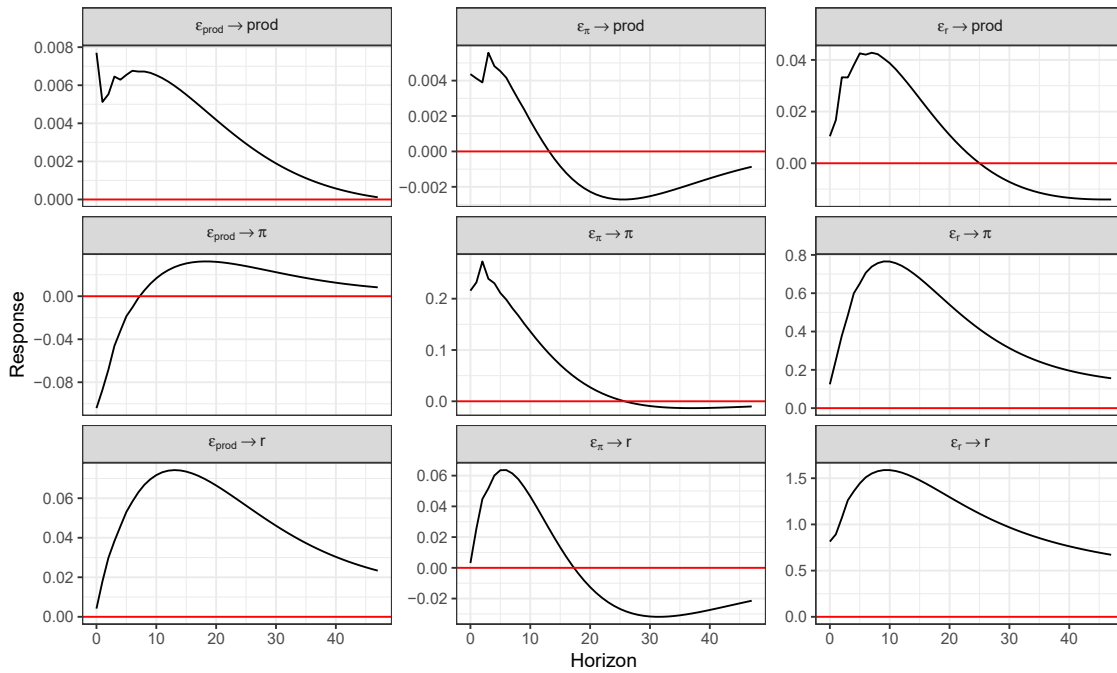


Figure 25: Benchmark model, NGML identified

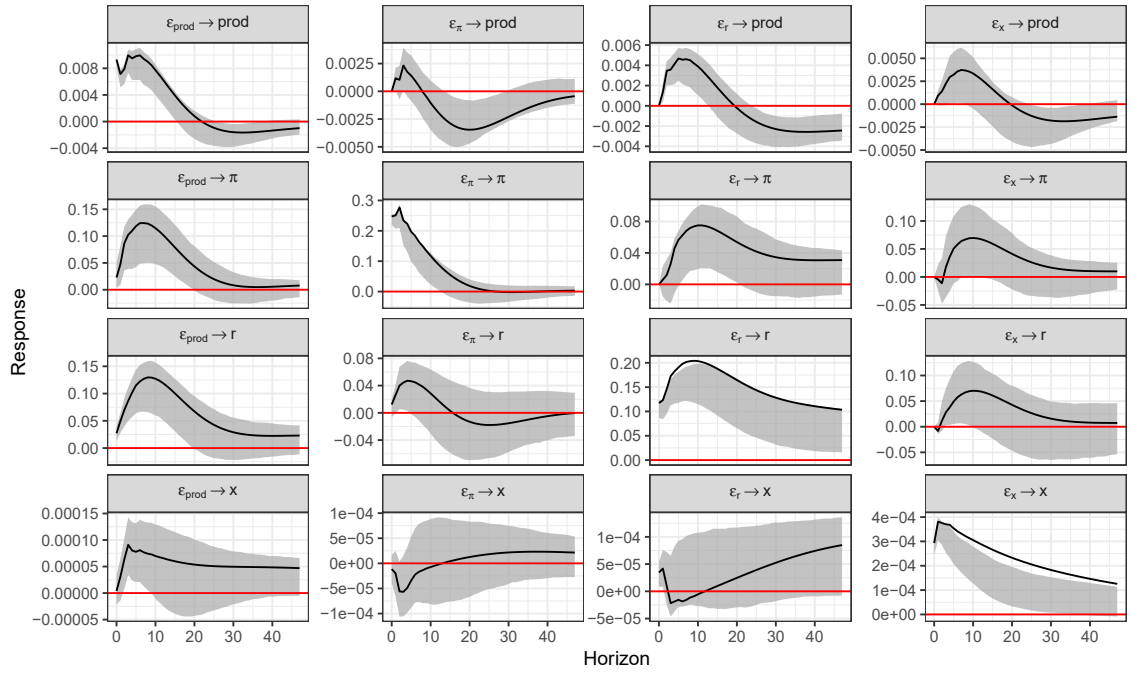


Figure 26: Exchange rate model, Cholesky identified

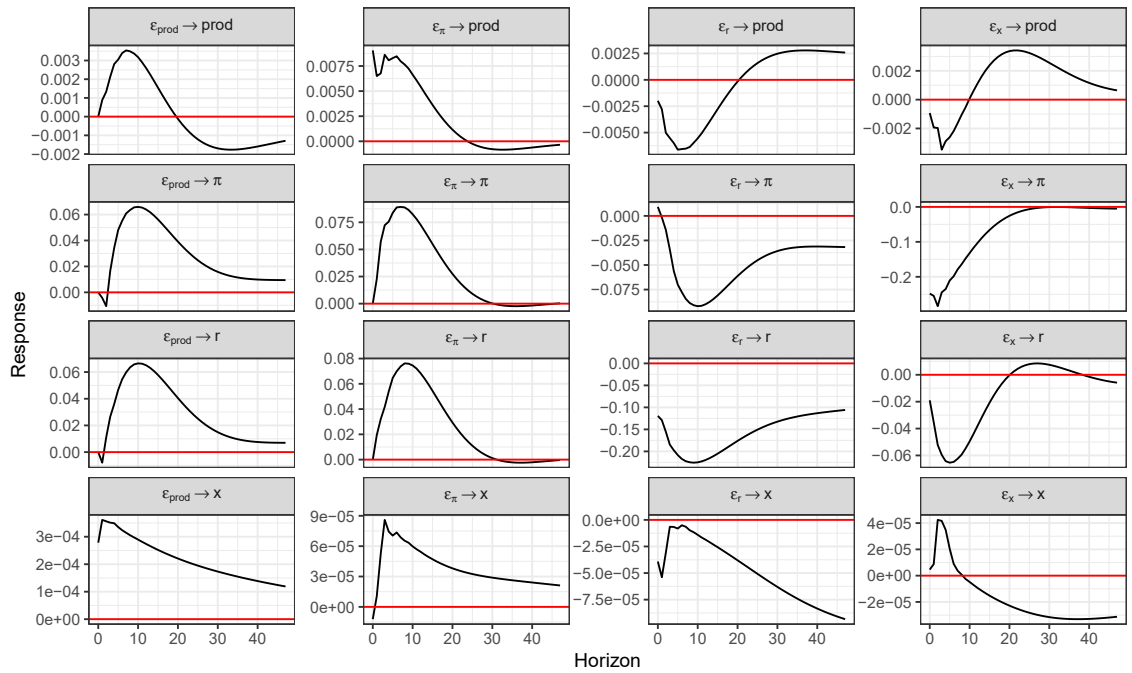


Figure 27: Exchange rate model, GARCH identified

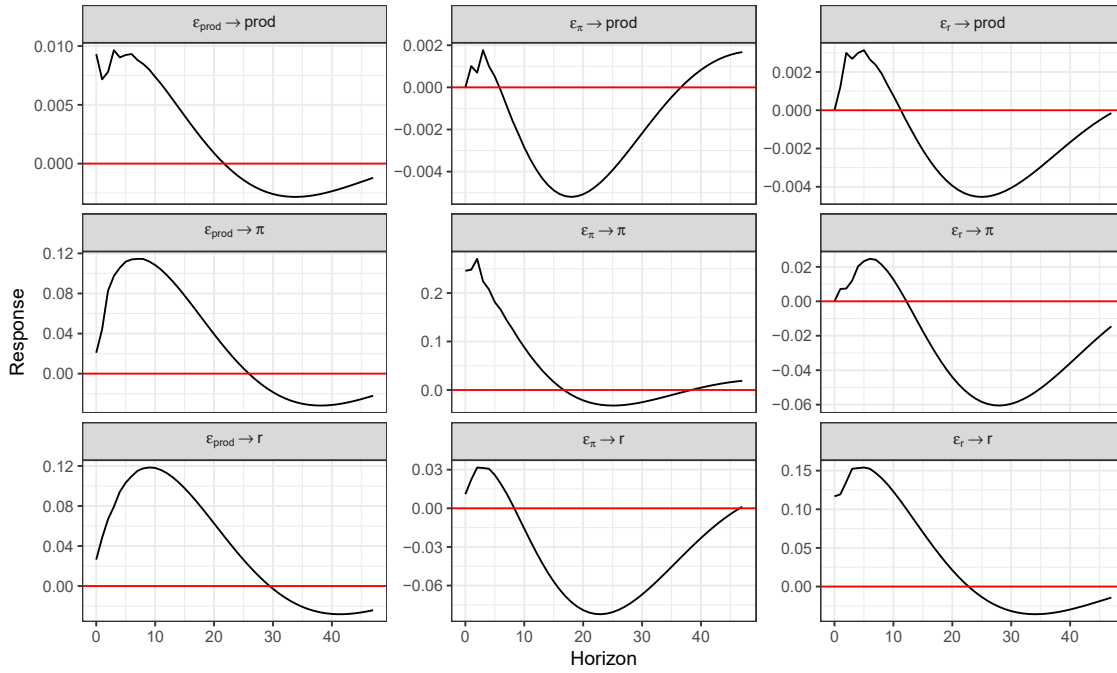


Figure 28: Model with linear estimate for the output gap, Cholesky identified

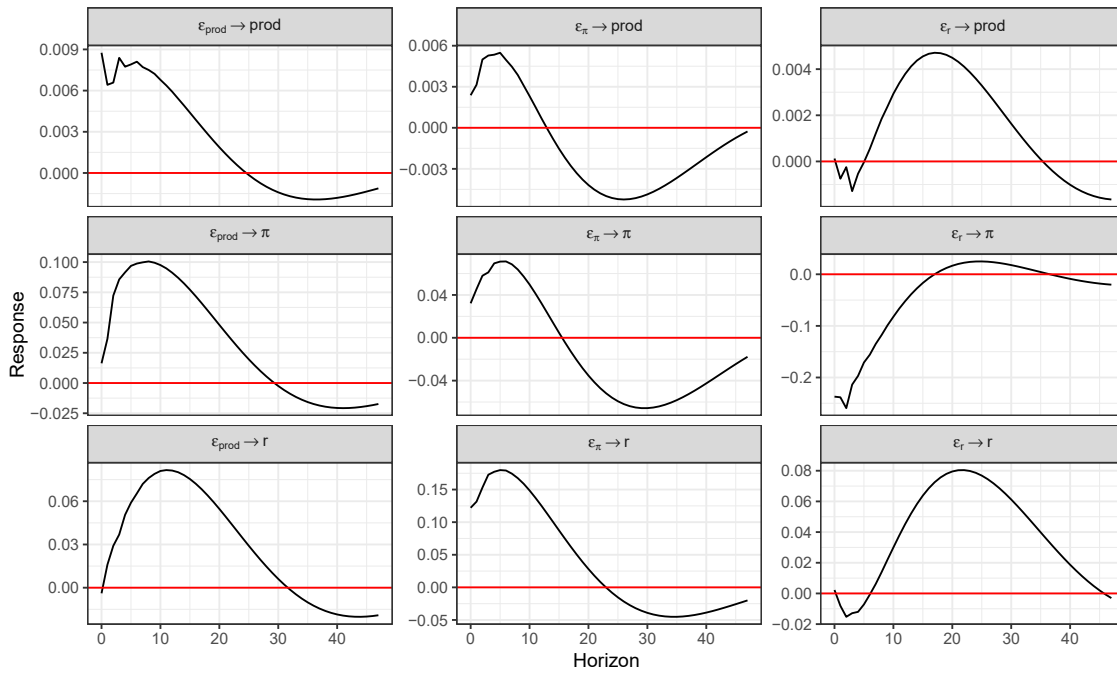


Figure 29: Model with linear estimate for the output gap, GARCH identified

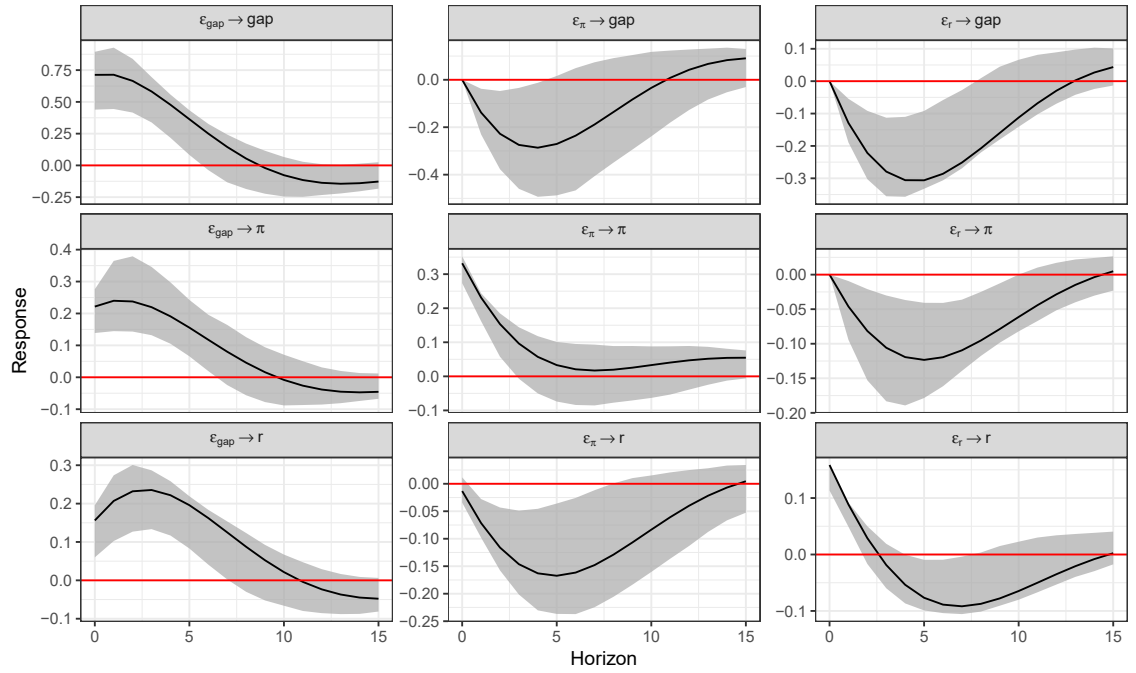


Figure 30: VAR(1) with BN-BVAR estimate for the output gap, Cholesky identified

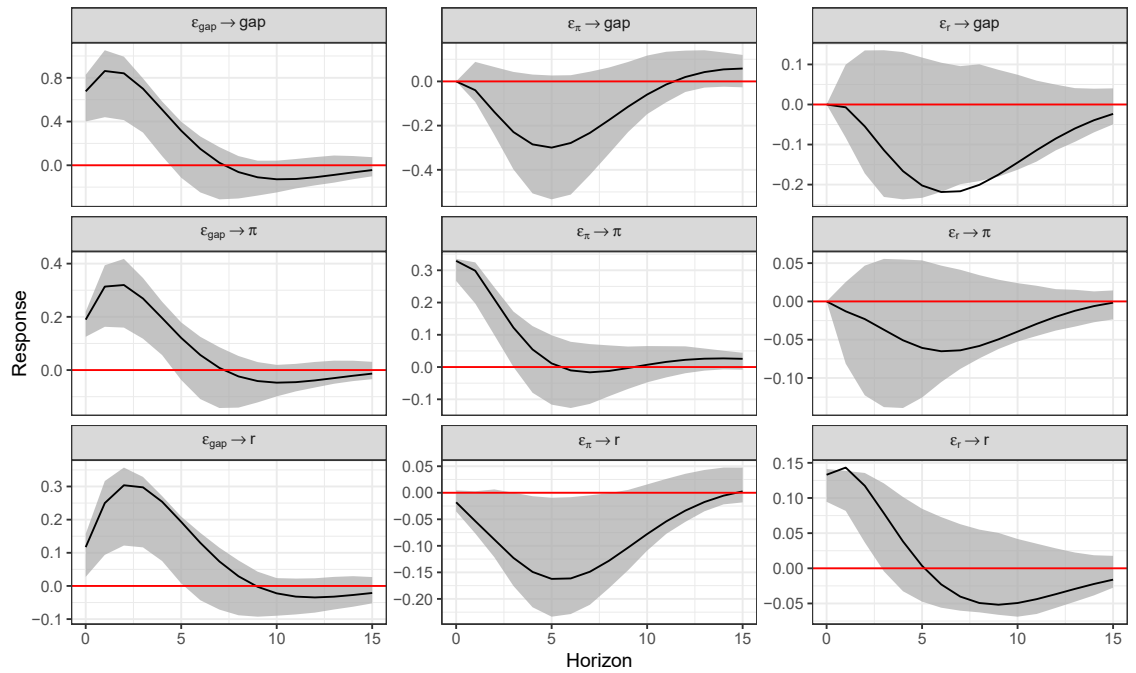


Figure 31: VAR(2) with BN-BVAR estimate for the output gap, Cholesky identified

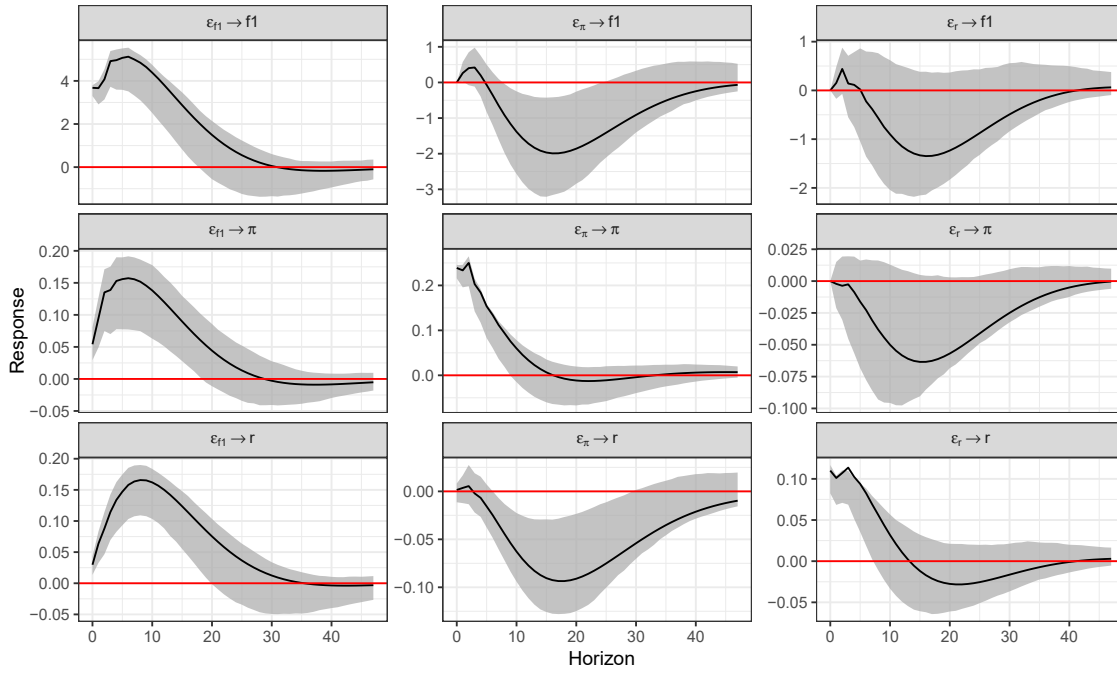


Figure 32: One common factor FAVAR, Cholesky identified

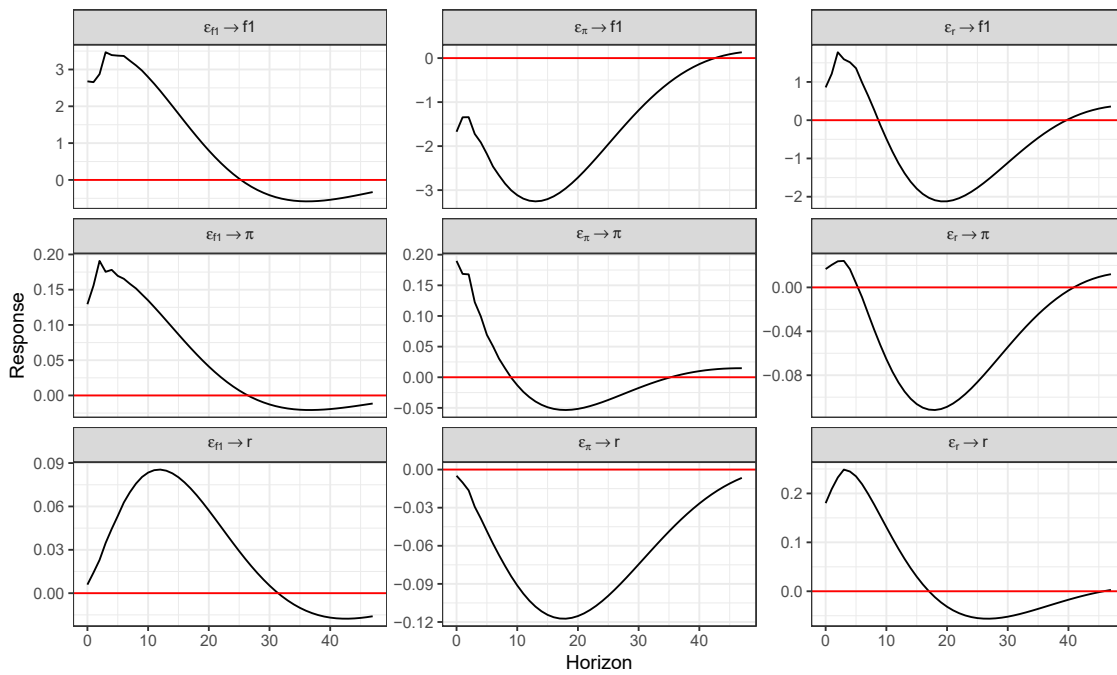


Figure 33: One common factor FAVAR, ST identified

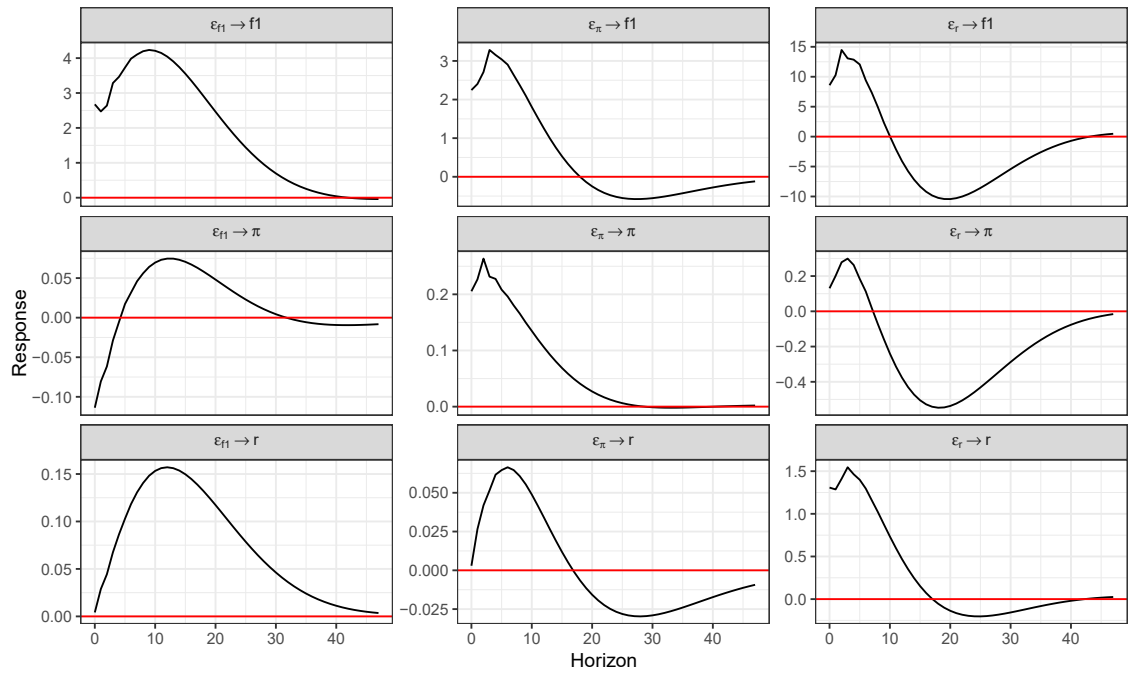


Figure 34: Four common factor FAVAR, NGML identified

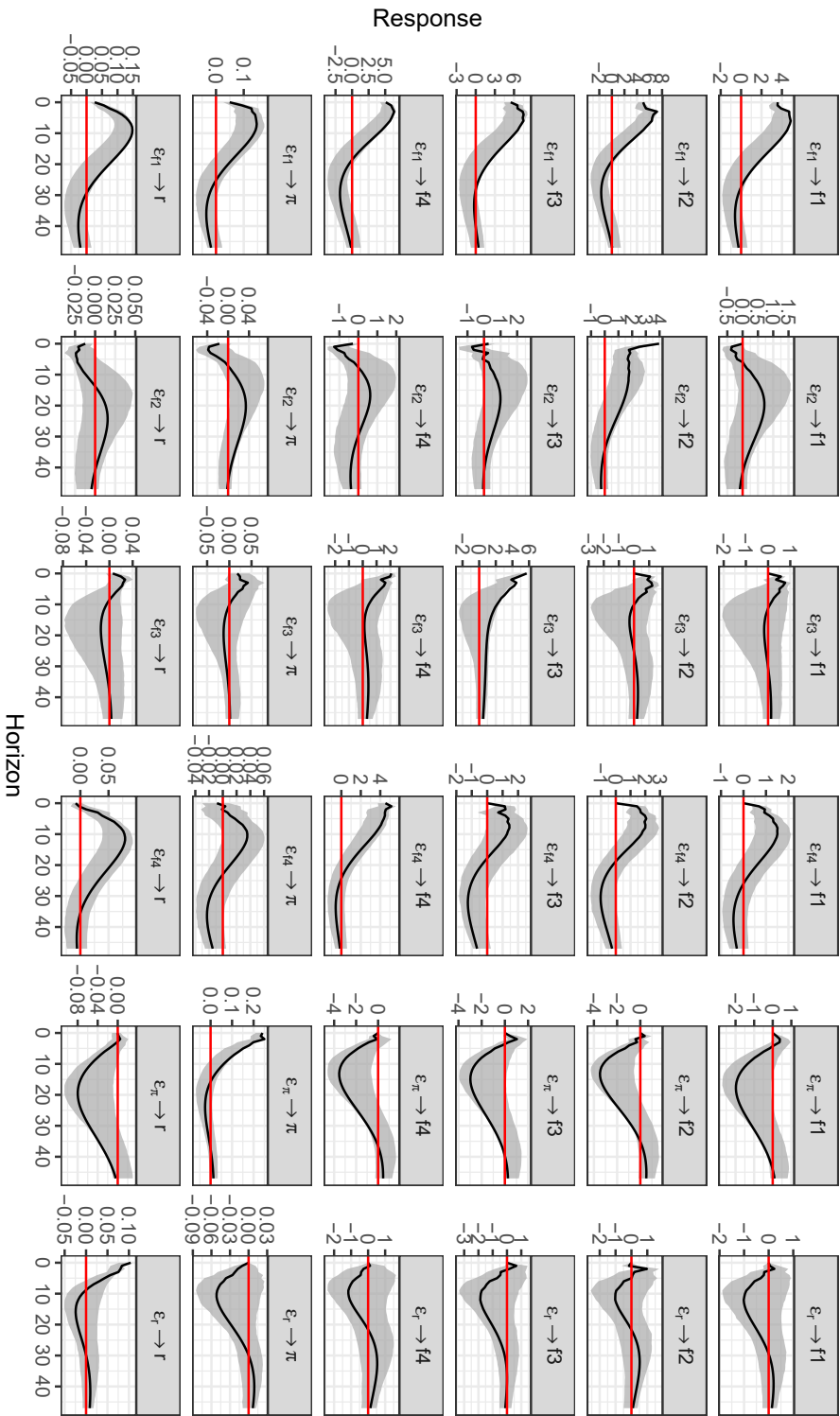


Figure 35: Four common factor FAVAR, Cholesky identified

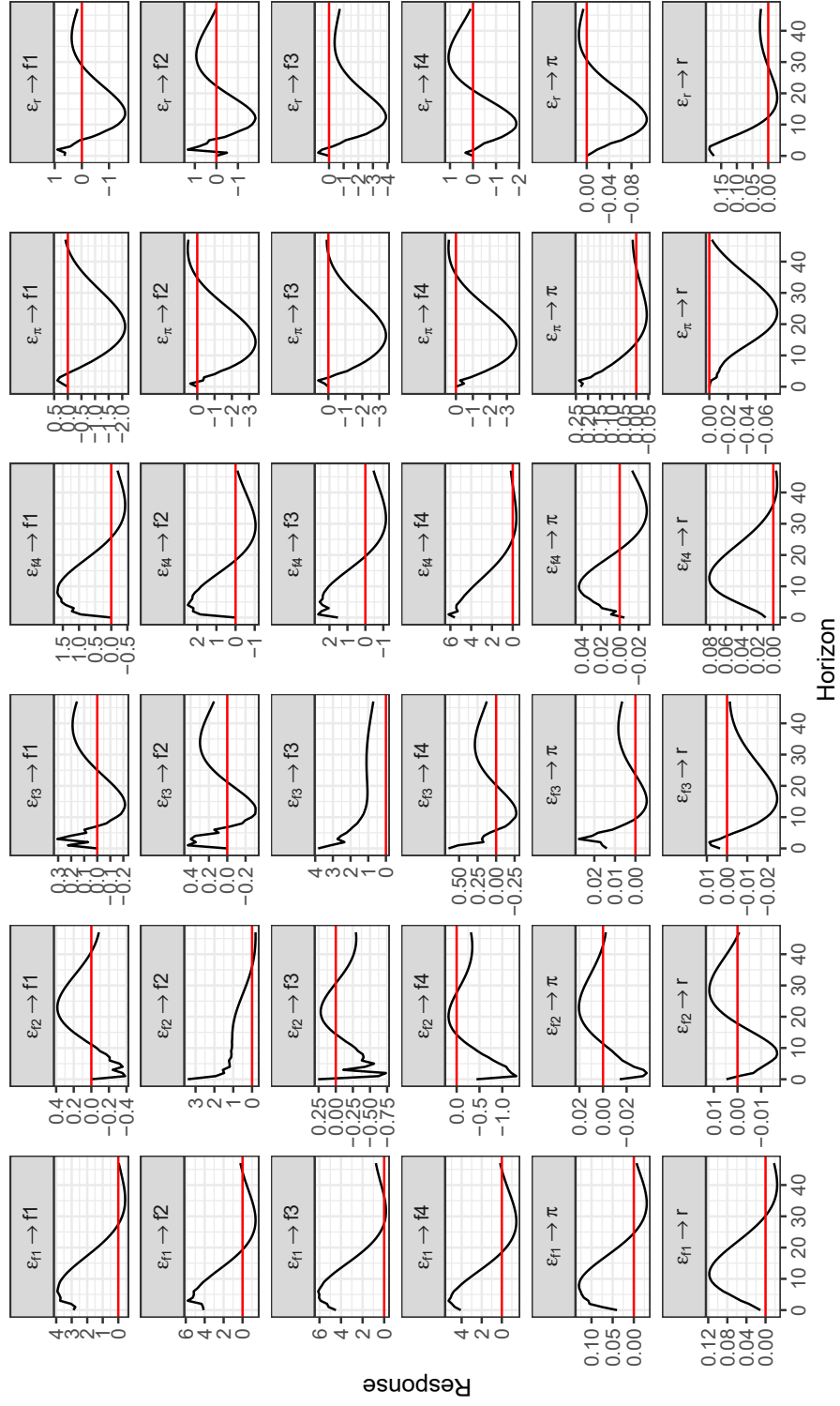


Figure 36: Four common factor FAVAR, partially restricted ST identified

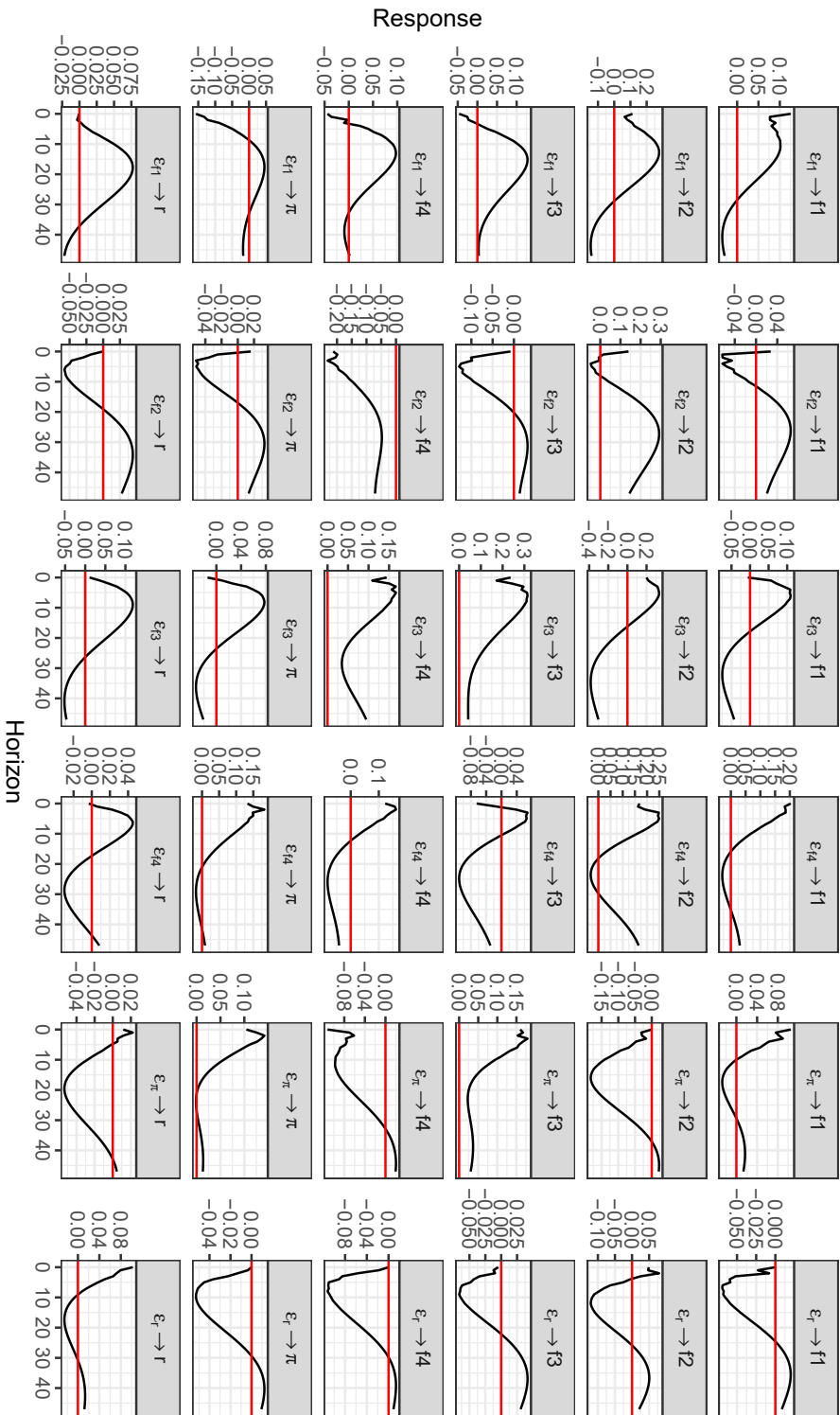


Figure 37: Four dynamic common factor FAVAR, partially restricted NGML identified

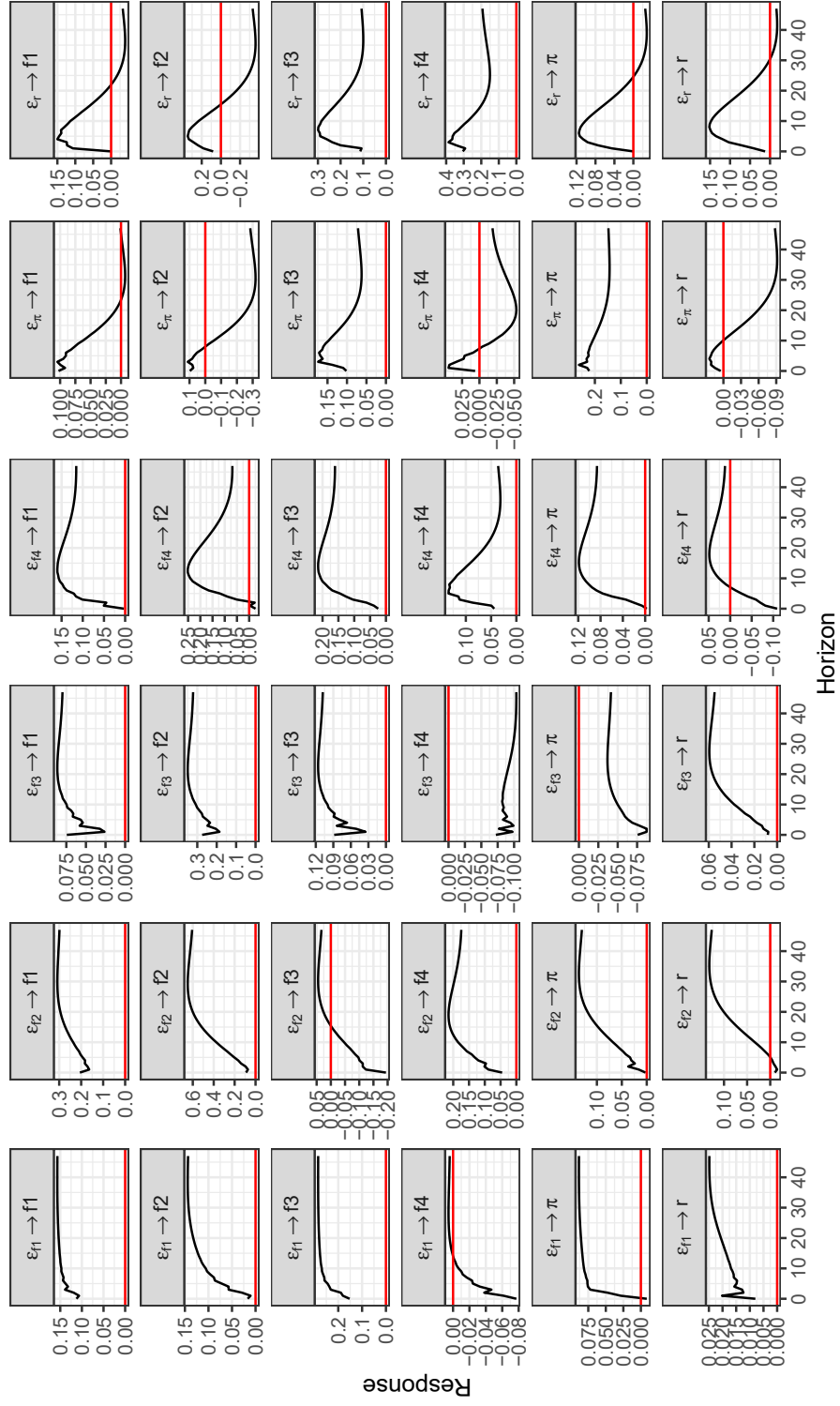


Figure 38: Four dynamic common factor FAVEC, partially restricted NGML identified