Mind the gap! Stylized dynamic facts and structural models.

Fabio Canova *
Norwegian Business School, CAMP and CEPR
and
Filippo Ferroni
Federal Reserve Bank of Chicago

April 12, 2019

Abstract

We study what happens to identified shocks and to dynamic responses when the data generating process features q disturbances but only $q_1 < q$ variables are used in the empirical model. Identified shocks are mongrels: they are linear combinations of current and past values of all structural disturbances and do not necessarily combine disturbances of the same type. Sound restrictions may be insufficient to obtain structural dynamics. The theory used to interpret the data and the disturbances it features determine whether an empirical model is too small. An example shows the magnitude of the distortions and the steps needed to reduce them. We revisit Iacoviello [2005]'s evidence regarding the transmission of house price shocks.

Key words: Aggregation, state variables, dynamic responses, dynamic structural models.

JEL Classification: C31, E27, E32.

*We thank Stefano Neri, Jesper Linde', Adrian Pagan, Makram el-Shagi and the participants of the Bank of Finland Workshop on Empirical Macroeconomics, Norges Bank NBR spring institute, and of seminars at Henan University, the Riksbank for comments and suggestions. Canova acknowledges the financial support from the Spanish Ministerio de Economia y Competitividad through the grants ECO2015-68136-P and FEDER, UE. The views presented in this paper do not reflect those of the Federal Reserve Bank of Chicago, or the Federal Reserve system.

1 Introduction

It is common in macroeconomics to collect stylized facts about the dynamic transmission of certain structural shocks using (small scale) vector autoregressive (VAR) models and then build (larger scale) Dynamic Stochastic General Equilibrium (DSGE) models to explain the patterns found in the data (see e.g. Galí [1999]; Iacoviello [2005], Basu and Bundick [2017] among many others).

Several authors, including Ravenna [2007], Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson [2007], Giacomini [2013], have emphasized that the matching exercise is imperfect as the linear solution of a DSGE model has a vector autoregressive-moving average (VARMA) format. To reduce the mismatch, the VAR should feature a large number of lags; but even a generous lag length may be insufficient in relevant cases. When long lags can not be used due to short data, the *invertibility* problem is typically taken care by i) simulating data from the linear decision rules of the same length as the actual data, ii) running the same VAR on both actual and simulated data, and iii) comparing the dynamics of the endogenous variables in the two systems after shocks are conventionally identified (see Chari, Kehoe, and McGrattan [2005]).

This paper studies a different mismatch problem, largely disregarded in the literature, which could be more important than invertibility for deciding which theory is consistent with the data. We call it *aggregation*, for lack of better name. It is generated when the process generating the data features q shocks, but $\bar{q} < q$ variables are used in the empirical model. Aggregation distortions make shocks identified in the empirical system mongrels with little economic interpretation for two reasons.

First, the shocks recovered in a system with $\bar{q}$ variables do not necessarily combine only structural disturbances of the same type, making it difficult to relate, say, an identified technology shocks to TFP or other supply disturbances present in a structural model. Second, the shocks recovered in a system with $\bar{q}$ variables are, in general, linear combinations of *current and past* structural disturbances. Because of this time deformation, shocks identified in a small scale empirical system may display a stronger propagation mechanism.

The first problem (we name it cross sectional aggregation) emerges when the data generating process (DGP) is such that several structural disturbances contemporaneously affect the variables of the empirical model. The second problem (we name it time aggregation) instead occurs whenever the empirical model is specified without paying sufficient attention to the theory used to explain the data and is exacerbated when the small scale empirical model i) does not respect the relationship between the endogenous variables and the states or ii) alters the law of motion of the states. Cross sectional aggregation makes sound theoretical restrictions insufficient to obtain meaningful structural disturbances. Time aggregation deforms the information contained in the structural disturbances.

We derive these results formally in section 2 assuming that the DGP is a linear state space model. Even though throughout the paper we focus attention on general equilibrium models, aggregation problems have identical implications in partial equilibrium settings, since the linear solution of such models also has a state space representation. In section 3 we use a standard New Keynesian model to illustrate the issues at stake. We show how to match the theory to a small scale empirical model; the problems occurring when the empirical model is too small; and how to reduce the time distortions linking the theory and the empirical model more explicitly. The reader should take away three points from the exercises we present. First, if a SVAR is too small, identified shocks may not be interpretable. Second, when the DGP features more shocks than the empirical model, the theory should be reduced to the same observables used in the empirical model prior to the computation of dynamic responses. Third, the theory used to interpret the data and the disturbances it features must guide the choice of observables and the minimal dimension of the empirical model. The empirical model used to derive dynamic facts is not theory-free when $\bar{q} < q$. For example, a VAR used to identify monetary shocks may feature different variables if the theoretical counterpart has financial disturbances or not; and identifying permanent or transitory technology shocks may require empirical models of different dimensions and with different variables.

Section 4 provides suggestions to users who want to avoid falling into the aggregation trap. In section 5 we take a version of Iacoviello [2005]'s model with seven disturbances (the four originally used plus disturbances to the borrowing constraints and the wealth constraint of households) and use a four variable VAR to construct responses to house price shocks. We show that when additional disturbances are allowed, the match with the data is weaker than previously thought; that biases are severe, and that cross sectional aggregation is largely responsible for the distortions we observe. Section 6 extends the analysis to DGPs displaying higher order terms (such as those generated by higher order erturbed solutions of equilibrium models). We show that the results derived in section 2 hold, that aggregation distortions are likely to be more severe, and use Basu and Bundick [2017]'s uncertainty disturbance model to highlight them.

Our analysis abstracts from invertibility issues (recently studied in, e.g. Beaudry, Feve, Guay, and Portier [2016], Forni, Gambetti, and Sala [2016], Plagborg Moller [2017], Pagan and Robinson [2018], Chahrour and Jurado [2018]). Both aggregation and invertibility produce time deformation problems, making identified shocks filtered versions of the structural disturbances. However, invertibility does not create cross sectional aggregation. Thus, the interpretation problems we consider are distinct, and matter even when invertibility is not an issue. Also, while one may choose an empirical model with only q_1 variables because certain theoretical quantities are latent, the cross sectional aggregation problem we discuss is relevant even when all theoretical quantities are observables but short samples or identification convenience make applied researchers work with small scale empirical models.

The current literature is silent about aggregation issues. Apart from Canova and Hamidi Sahneh [2018], who analyze the effects of cross sectional aggregation on the properties of Granger causality tests, we are aware only of early work by Hansen and Sargent [1991], Marcet [1991], Lutkepohl [1984], Braun and Mittnik [1991] and Faust and Leeper [1988]. While former two examine the effects of time aggregating on the decision rules of a model, the latter two papers analyze the distortions due to cross sectional aggregation of structural shocks. However, they take the DGP to be a larger scale VAR rather than a structural model and thus have no insight about the role of theory in guiding the choice of the empirical model. Some of results we present have similar flavor to Wolf [2018]. However, they are produced by aggregation rather than insufficient identification restrictions.

2 A Few analytical results

We assume that the DGP is of the form:

$$x_t = A(\theta)x_{t-1} + B(\theta)e_t \quad (1)$$

$$y_t = C(\theta)x_{t-1} + D(\theta)e_t \quad (2)$$

where x_t is a $k \times 1$ vector of endogenous and exogenous states, y_t is a $m \times 1$ vector of endogenous controls, $e_t \sim (0, \Sigma(\theta))$ is a $q \times 1$ vector of disturbances, $\Sigma(\theta)$ a diagonal matrix and θ a vector of structural parameters; $A(\theta)$ is $k \times k$, $B(\theta)$ is $k \times q$, $C(\theta)$ is $m \times k$, $D(\theta)$ is $m \times q$. For convenience, we let the eigenvalues of $A(\theta)$ to be all less than one in absolute value. Thus, if there are disturbances with permanent effects, (1)-(2) represent a properly scaled version of the process generating the data. In our applications, (1)-(2) is the (log-)linear solution of the optimality conditions of a structural macroeconomic model.

In general, $m \geq q$ and some of the x_t's may be latent. For this reason, it is typical to assume that the variables entering the empirical model are $z_t = S[x_t, y_t]'$ where S is a selection matrix. In the literature, there are two different choices. For example Fernández-Villaverde et al. [2007] assume $S = [0, I]$ (which implies that m=q), while Ravenna [2007] and Pagan and Robinson [2018] assume that either S=I (so that $m + k = q$), or $S = [0, I]$. In general, S is chosen so that dim(z_t)=dim(e_t), i.e. the number of empirical variables and of structural disturbances match.

The reduced form (innovation representation) corresponding to (1)-(2) is

$$x_t = A(\theta)x_{t-1} + K_x(\theta)u_t \quad (3)$$
$$y_t = C(\theta)x_{t-1} + K_y(\theta)u_t \quad (4)$$

where $u_t = z_t - E_t[z_t|\Omega_{t-1}]$ is a $q \times 1$ vector of innovations, Ω_{t-1} includes (at least) lags of z_t, $K_x(\theta)$ and $K_y(\theta)$ are steady state Kalman gain matrices, and for those x_t and y_t belonging to z_t, $K_i(\theta)$ has a row with zero entries except in one position.

Given (3)-(4), identification of structural shocks and of structural responses requires the mapping from u_t into e_t. When $S = I$, this requires inverting $\begin{pmatrix} B(\theta) \\ D(\theta) \end{pmatrix} e_t = u_t$; when $S = [0, I]$, we need to invert $D(\theta)e_t = u_t$. In both cases, standard order and rank conditions apply.

In the identification exercises two assumptions are implicitly made. First, there is no misspecification in (1)-(2), at least, as far as sources of disturbances are concerned. If disturbances are left out, the identification exercises becomes problematic even when excluded disturbances are orthogonal to included ones, and the included disturbances account for a large portion of the variability of z_t. Second, when $z_t = y_t$, that is, when only the controls enter the empirical model, Ω_{t-1} must include long lags of z_t to take care of omitted states. When disturbances are left out from (1)-(2), having a rich Ω_{t-1} is generally insufficient to make the identification problem well behaved.

Our setup accounts for the possibility that the DGP has more disturbances than the variables entering the empirical system. This mimics, for example, the situation when a researcher runs a small scale empirical system (say, a two variable VAR) but the DGP features more than two disturbances. Typically, a researcher who wants to interpret the dynamics of the small scale empirical system employs a theoretical model that is less complex than the DGP and specifies only enough disturbances to match the number of empirical variables. This paper shows that the dynamics produced by such model are not relevant for the comparison and omitted disturbances play a crucial role.

For the rest of the paper, the empirical system uses $z_{it} \equiv S_i[x_t, y_t]'$ where S_i is a $q_i \times q$ selection matrix, and $dim(z_{it}) = q_i < dim(e_t) = q, \forall i$. We will consider three specific S_i matrix.

• Case 1: $S_1 = [I, S_{12}]$. This choice generates an observable system which retains the states but eliminates part of the controls. The DGP in terms of $z_{1t} = [x_t, y_{1t}]', y_{1t} \equiv S_{12}y_t$ is:

$$x_t = A(\theta)x_{t-1} + B(\theta)e_t \quad (5)$$
$$y_{1t} = C_1(\theta)x_{t-1} + D_1(\theta)e_t \quad (6)$$

or $z_{1t} = F_1(\theta)z_{1t-1} + G_1(\theta)e_t$, where $F_1(\theta) = \begin{pmatrix} A(\theta) & 0 \\ C_1(\theta) & 0 \end{pmatrix}$ and $G_1(\theta) = \begin{pmatrix} B(\theta) \\ D_1(\theta) \end{pmatrix}$.

• Case 2: $S_2 = [S_{21}, S_{22}]$. This choice generates an observable system which eliminates part of the states and part of the controls. Let $x_t = (x_{1t}, x_{2t}), y_t = (y_{1t}, y_{2t})$, where (x_{1t}, y_{1t}) are the variables excluded from the empirical system. The DGP in terms of $z_{2t} = [x_{2t}, y_{2t}], x_{2t} \equiv S_{21}x_t, y_{2t} \equiv S_{22}y_t$, is

$$x_{2t} = A_2(\theta)x_{2t-1} + B_2(\theta)e_t + w_{1t-1} \quad (7)$$
$$y_{2t} = C_2(\theta)x_{2t-1} + D_2(\theta)e_t + w_{2t-1} \quad (8)$$

where $w_{1t-1} = A_{21}(\theta)x_{1t-1}; w_{2t-1} = C_{21}(\theta)x_{1t-1}$ or $z_{2t} = F_2(\theta)z_{2t-1} + G_2(\theta)e_t + w_{2t-1}$, where $F_2(\theta) = \begin{pmatrix} A_2(\theta) & 0 \\ C_2(\theta) & 0 \end{pmatrix}$ and $G_2(\theta) = \begin{pmatrix} B_2(\theta) \\ D_2(\theta) \end{pmatrix}$. Alternatively, using (1) to separate observable and non-observable states, and integrating x_{1t} out, the DGP for z_{2t} is

$$x_{2t} = \tilde{A}_{21}(\theta)x_{2t-1} + \tilde{A}_{22}(\theta)x_{2t-2} + \tilde{B}_{20}(\theta)e_t + \tilde{B}_{21}(\theta)e_{t-1} \quad (9)$$
$$y_{2t} = \tilde{C}_{21}(\theta)x_{2t-1} + \tilde{C}_{22}(\theta)x_{2t-2} + \tilde{D}_{20}(\theta)e_t + \tilde{D}_{21}(\theta)e_{t-1} \quad (10)$$

(7)-(8) point out the misspecification present using a first order empirical model for z_{2t}. (9)-(10) shows that DGP for the observables is a VARMA(2,1).

• Case 3: $S_3 = [S_{31}, 0]$. This choice generates an empirical system which repackages the states and eliminates the controls. The DGP in terms of $z_{3t} = x_{3t} = S_{31}x_t$ is

$$x_{3t} = A_3(\theta)x_{3t-1} + B_3(\theta)e_t + w_{3t-1} \quad (11)$$

where w_{3t-1} is a function of the repackaged states. Analogously with case 2, one may write (11) as

$$z_{3t} = \bar{A}_{31}(\theta)z_{3t-1} + \bar{A}_{32}(\theta)z_{3t-2} + \bar{B}_{30}(\theta)e_t + \bar{B}_{31}e_{3t-1} \quad (12)$$

Intuitively, the processes for z_{it} displayed in cases 1-3 are obtained substituting optimality conditions into others, prior to the computation of the decision rules. Note that the matrices of these solutions generally differ from those obtained solving the original model and crossing out the rows corresponding to the variables absent from z_{it} because in our case not all the original states may be used in the computation of the decision rules. Section 3 provides examples of smaller scale empirical systems which produce (5)-(6), (9)-(10), and (12) for a specific DGP.

The innovation representation of (1)-(2) when z_{it} are observables is

$$x_t = A(\theta)x_{t-1} + \hat{K}_{ix}(\theta)u_{it} \quad (13)$$
$$y_t = C(\theta)x_{t-1} + \hat{K}_{iy}(\theta)u_{it} \quad (14)$$

where $u_{it} = z_{it} - E_t[z_{it}|\Omega_{it-1}]$ is a $q_i \times 1$ vector of innovations, $\hat{K}_{ix}(\theta)$, $\hat{K}_{iy}(\theta)$ are steady state Kalman gain matrices featuring some rows with zero entries except in one position.

2.1 The relationship between innovations and structural disturbances

The empirical system eliminates only theoretical controls We analyze the relationship between u_{1t} and e_t, when $E[z_{1t}|\Omega_{1t-1}] = \tilde{F}_1 z_{1t-1}$ and thus

$$u_{1t} = z_{1t} - \tilde{F}_1 z_{1t-1} \quad (15)$$

Proposition 1 *i) If* $\tilde{F}_1 = S_1F(\theta) \equiv F_1(\theta)$, $u_{1t} = \lambda_1(\theta)e_t$, *where* $\lambda_1(\theta)$ *is a* $q_1 \times q$ *matrix. Unless* $G_1(\theta)$ *has at most one non-zero element in each row* u_{1t}^k *will not load on* e_t^j *only, for some* k *and* j.
ii)If $\tilde{F}_1 \neq S_1F(\theta)$, $u_{1t} = \lambda_1(\theta, L)e_t$, *where* $\lambda_1(\theta, L)$ *is a* $q_1 \times q$ *matrix for every* L *and, in general, is an infinite dimensional function of* L.

The proof of the proposition is obtained matching (15) with (5)-(6). The first part considers the case $\tilde{F}_1 = S_1F(\theta)$. In this situation the innovations u_{1t} respect the timing protocol of the structural disturbances e_t, but cross sectionally aggregate them because $q_1 < q$. Thus, for example, if there are four structural disturbances in the DGP and only two elements in z_{1t}, we can at most identify two time t structural shocks from the empirical system. Because $G_1(\theta)$ is a rectangular matrix, u_{1t} compresses the information present in e_t and one may ask when u_{1t} carries enough information to recover some e_t. It turns out that u_{1t}^k aggregates certain types structural of disturbances only if $G_1(\theta)$ has a block structure. Furthermore, u_{1t}^k will carry information about one e_t^j if and only if $G_1(\theta)$ has at most one non-zero element in row j , i.e., if at most one structural disturbance enters the decision rule of each variable. Both restrictions are strong and even the block structure condition for $G_1(\theta)$ is unlikely to be satisfied in the majority of general equilibrium models nowadays considered. It requires that the theory features many "conveniently" placed delay restrictions.

When $\tilde{F}_1 \neq S_1F(\theta)$, time aggregation also occurs and u_{1t} becomes a one-sided infinite moving average of the structural disturbances, $u_{1t} = \lambda_1(\theta, L)e_t \equiv (\tilde{F}_1 - S_1F(\theta))(I - S_1F(\theta))^{-1}G_1(\theta)e_{t-1} + G_1(\theta)e_t$. In this situation, even if $G_1(\theta)$ has at most one non-zero element in row j, current information about u_{1t}^k may not be enough to obtain information about some of the current e_t^j. In general, the way u_{1t} compresses $\{e_{t-s}^j\}_{j=1}^q, s = 1, 2, \dots$ depends on the structure of the model as encoded in the $\lambda_1(\theta, L)$ polynomial.

Proposition 1 determines the properties of u_{1t}, given e_t. Thus, u_{1t} will be a mean zero process and its autocovariance function will be restricted by

$$E(u_{1t}u_{1t-s}') = E(\lambda_1(\theta, L)e_te_{t-s}'\lambda_1(\theta, L)'), \quad s \geq 0 \tag{16}$$

When e_t are iid, the variance of the u_{1t} differs from the variance of e_t and the magnitude of the amplification depends on the properties of $\lambda_1(\theta, L)$. Note that a e_t disturbance with a small variance or small initial loadings $\lambda_1(\theta, L = 0) \equiv \lambda_{10}(\theta)$ will be hard to identify from the u_{1t}. Similarly, the serial correlation properties of u_{1t} depend on the structure and magnitude of the $\lambda_1(\theta, L)$ polynomial and its dimension. However, even when $\lambda_1(\theta, L) \equiv \lambda_{10}(\theta) = G_1(\theta)$, cross sectional aggregation makes the autocovariance function of u_{1t} insufficient to recover the autocovariance of e_t. Moreover, invertibility of $\lambda_1(\theta, L)$ is insufficient to back out some e_t^j because $\lambda_1(\theta, L)$ is not a square matrix for each L.

The states in the empirical and the theoretical models differ We analyze the relationship between $u_{it}, i = 2, 3$ and e_t when $E[z_{it}|\Omega_{it-1}] = \tilde{F}_iz_{it-1}, i = 2, 3$ so that

$$u_{it} = z_{it} - \tilde{F}_iz_{it-1} \tag{17}$$

Proposition 2 *i)* $u_{it} = \lambda_i(\theta, L)e_t, i = 2, 3$, *where* λ_i *is* $q_i \times q$ *for each* L.
ii) $u_{it} = \psi_i(\theta, L)u_{1t}, i = 2, 3$.

To prove part i), we first match (15) and (7)-(8). Then $u_{2t} = (S_2F(\theta) - \tilde{F}_2)(I - S_2F(\theta)L)^{-1}(G_2(\theta)e_{t-1} + H_2(\theta)x_{1t-2}) + G_2(\theta)e_t + H_2(\theta)x_{1t-1}$. Because x_{1t} has a VARMA(2,1) format: $M(\theta, L)x_{1t} = N(\theta, L)e_t$, where $M(\theta, L)$ is invertible, we have $u_{2t} = \lambda_2(\theta, L)e_t$, where $\lambda_2(\theta, L) = G_2(\theta) + (S_2F(\theta) - \tilde{F}_2)(I - S_2F(\theta)L)^{-1}(G_2(\theta) + H_2(\theta)M(\theta, L)^{-1}N(\theta, L)L + H_2(\theta)M(\theta, L)^{-1}N(\theta, L)L^2$. Matching (17) with (11) one similarly obtains that $u_{3t} = \lambda_3(\theta, L)e_t$. Thus, an empirical system including only the states of the DGP will not solve time aggregation problems since their law of motion may be altered. Note that in this case $S_2F(\theta) = \tilde{F}_2$, or $S_{31}A(\theta) = \tilde{F}_3$ will be insufficient to avoid time aggregation and $u_{it}, i = 2, 3$ will always cross sectionally and time aggregate the structural disturbances.

In general $u_{it} \neq u_{1t}, i = 2, 3$ and the timing of information they contain differ even when $S_iF(\theta) = \tilde{F}_i(\theta), \forall i$. Letting $\lambda_1(\theta, L)^+$ be the generalized inverse of $\lambda_1(\theta, L)$, one can write

$$u_{it} = \lambda_i(\theta, L)\lambda_1(\theta, L)^+u_{1t} \equiv \psi_i(\theta, L)u_{1t} \tag{18}$$

By construction $\psi_{i0}(\theta) = I$. Thus, an impulse in u_{1t} and $u_{it}, i = 2, 3$ has identical effects on the variables present in both z_{1t} and z_{it} but will last longer when z_{it} are the observables - persistence will be altered. Clearly, the gaps between u_{it} and u_{1t} in terms of timing and cross sectional distortions depend on how different x_{it} and x_t, $\tilde{A}_2(\theta)(\bar{A}_3(\theta))$ and $A(\theta)$, $\tilde{C}_2(\theta)$ and $C(\theta)$ are.

(17) is misspecified when states are omitted or repackaged. What would happen if the innovations u_{it} are constructed using a larger information set, e.g.,

$$u_{it} = z_{it} - \tilde{F}_i(L)z_{it-1} \quad L = 1, 2, \dots \tag{19}$$

Because both z_{2t} and z_{3t} are VARMA processes, the standard non-invertibility and truncation issues discussed in the literature apply. Thus, in principle, $\tilde{F}_i(L)$ must be non-zero for $L \to \infty$ for time aggregation biases to disappear, which requires a alrge sample size. However, even when a very large sample is available, time aggregation may still be present due to non-invertibility of $N_i(\theta, L)$.

Proposition 1 is related to the aggregation results of Faust and Leeper [1988]. They show the conditions needed for VAR shocks to recover classes of structural disturbances (demand, supply, etc.). Because of their DGP is a VAR, they can not analyze the consequences of omitting states or altering their law of motion. Proposition 2 is, to the best of our knowledge, new. It has the same flavor as the result stated in Fernández-Villaverde et al. [2007]. The main difference is that here $u_{it}, i = 2, 3$ are reduced ranked moving averages of e_t and the reason for the time deformation is aggregation rather than non-invertibility.

The two propositions highlight that the variables entering in the empirical model determine the quality of the approximation of identified shocks to the structural disturbances. Eliminating theoretical controls creates innovations that cross sectionally aggregate the structural disturbances, but eliminating states or repackaging their law of motion may create both cross sectional and time aggregation distortions. Failure to include the theoretical states in an empirical model makes the innovations computed from a finite order empirical system correlated over time. However, having an empirical model with all the theoretical states may not be enough for proper inference because aggregation may alter their law of motion. In section 3 we discuss how the use of proxies may reduce time deformations when the empirical model omits or repackages some of the states.

2.2 Dynamic responses

Consider the computation of z_{it} responses to an impulse in the shocks. In the DGP they are:

$$\begin{aligned} z_{it} &= S_i \begin{pmatrix} B(\theta) \\ D(\theta) \end{pmatrix} e_t \\ z_{it+h} &= S_i \begin{pmatrix} A(\theta)^h B(\theta) \\ C(\theta)A(\theta)^{h-1}B(\theta) \end{pmatrix} e_t \quad i = 1,2,3; h = 1,2,\ldots \end{aligned} \tag{20}$$

In the empirical system with z_{1t} as observables, they are:

$$\begin{aligned} z_{1t} &= u_{1t} \\ z_{1t+h} &= \tilde{F}_1(\theta)^h u_{1t} \end{aligned} \tag{21}$$

The impact effect differs because $u_t = G_1(\theta)e_t$ and $G_1(\theta)$ is not a square matrix. Having the correct $B(\theta), D(\theta)$ matrices is insufficient to recover some e_t^j via $\Sigma_u = G_1(\theta)\Sigma(\theta)G_1(\theta)'$, unless $G_1(\theta)$ only has one non-zero element each row. Clearly, since $q_1 < q$, not all e_t disturbances can be recovered. However, if $\tilde{F}_1 = \begin{pmatrix} A(\theta) \\ S_{12}C(\theta) \end{pmatrix}$ responses at longer horizons to a properly identified shock in the empirical system are proportional to those of the DGP. Thus, at least qualitatively, (21) provides a good approximation to (20), if some structural disturbances could be recovered from u_{1t}.

The responses computed in systems with $z_{it}, i = 2,3$ as observables are instead:

$$\begin{aligned} z_{it} &= u_{it} \\ z_{it+h} &= \nu_{ij}u_{it} + \tilde{F}_i(\theta)^h u_{it} \end{aligned} \tag{22}$$

Here, both the instantaneous and the dynamic responses of z_{it} will be distorted; and their pattern may have nothing to do with those produced in the DGP. We summarize the discussion in a proposition.

Proposition 3 *i) Structural impulse responses constructed in a z_{1t} system could match those of the structural model if $\tilde{F}_1(\theta) = \begin{pmatrix} A(\theta) \\ S_{12}C(\theta) \end{pmatrix}$ and $G_1(\theta)$ has at most one non-zero element in each row. ii) Even if the conditions in i) hold, the dynamic responses obtained from properly identified shocks in a z_{it} system, $i = 2,3$, differ from those of the DGP.*

(21)-(22) provide an analytic approach to compute the biases in impulse responses due to aggregation. Braun and Mittnik [1991] derived an expression of these biases when the empirical model and the DGP are both VARs.

3 An example

To illustrate how aggregation may affect inference, we use a standard New Keynesian setup featuring five structural disturbances: a permanent a_t and a transitory ζ_t TFP shock, a preference χ_t shock, a cost push μ_t shock and a monetary policy ε_t shock (see Canova and Ferroni [2011] for details). The optimality conditions are (conditional expectations are omitted):

$$\chi_t = \chi_{t+1} - \frac{1}{1-h}\, g_{t+1} + \frac{h}{1-h}\, g_t + r_t - \pi_{t+1} \tag{23}$$

$$\pi_t = \pi_{t+1}\, \beta + k_p \left(\frac{h}{1-h}\, g_t + (1+\sigma_n)\; n_t \right) + k_p\; (\mu_t - \chi_t) \tag{24}$$

$$o_t = \zeta_t + (1-\alpha)\; n_t \tag{25}$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\, (\phi_y\, g_t + \phi_p\, \pi_t) + \varepsilon_t \tag{26}$$

$$g_t = a_t + o_t - o_{t-1} \tag{27}$$

$$c_t = o_t \tag{28}$$

(23) is the Euler equation, (24) is the Phillips curve, (25) is the production function, (26) is the Taylor rule, (27) is the definition of output growth, and (28) is the resource constraint. o_t is output and g_t its growth rate, n_t is hours worked, π_t is the inflation rate, r_t the nominal interest rate and c_t consumption. h is the coefficient of (external) consumption habit, β the discount factor, σ_n the inverse of the Frish elasticity of labor supply, κ_p the slope of the Phillips curve, α the labor share in production, ϕ_y, ϕ_π the coefficients of the Taylor rule. The disturbances evolve as:

$$\zeta_t = \rho_z\, \zeta_{t-1} + e_{zt} \tag{29}$$

$$a_t = \rho_a\, a_{t-1} + e_{at} \tag{30}$$

$$\chi_t = \rho_\chi\, \chi_{t-1} + e_{\chi t} \tag{31}$$

$$\mu_t = \rho_\mu\, \mu_{t-1} + e_{\mu t} \tag{32}$$

$$\epsilon_t = e_{mp_t} \tag{33}$$

where $0 < \rho_j < 1, j = z, a, \chi, \mu$. We solve the model using a first order perturbation setting $\alpha = 0.33; \beta = 0.99; \sigma_n = 1.5; h = 0.9; k_p = 0.05; \phi_y = 0.1; \phi_p = 1.5; \rho_r = 0.8; \rho_z = 0.5; \rho_a = 0.2; \rho_\chi = 0.5; \rho_\mu = 0.0$. The minimal state vector is $x_{t-1} = [o_{t-1}, r_{t-1}, \zeta_{t-1}, a_{t-1}, \mu_{t-1}, \chi_{t-1}]'$, and the control vector is $y_t = [g_t, c_t, o_t, \pi_t, n_t, r_t]'$. We obtain decision rules of the form (1)-(2) where $A(\theta)$ is 6×6, $B(\theta)$ is 6×5, $C(\theta)$ is 6×6 and $D(\theta)$ is 6×5.

To illustrate the effects of aggregation, the differences produced eliminating states and controls, and the importance of carefully selecting the variables entering the empirical model when q_1 becomes small, we consider three alternative systems. In the first $z_t = (o_t, \pi_t, n_t, r_t)$; it is obtained dropping (28) and using (27) in (23)-(26):

$$\chi_t = \chi_{t+1} - \frac{1}{1-h}\, (a_{t+1} + o_{t+1} - o_t) + \frac{h}{1-h}\, (a_t + o_t - o_{t-1}) + r_t - \pi_{t+1} \tag{34}$$

$$\pi_t = \pi_{t+1}\, \beta + k_p \left(\frac{h}{1-h}\, (a_t + o_t - o_{t-1}) + (1+\sigma_n)\; n_t \right) + k_p\; (\mu_t - \chi_t) \tag{35}$$

$$o_t = \zeta_t + (1-\alpha)\; n_t \tag{36}$$

$$r_t = \rho_r\, r_{t-1} + (1-\rho_r)\; (\phi_y\, (a_t + o_t - o_{t-1}) + \phi_p \pi_t\,) + \varepsilon_{mp_t} \tag{37}$$

Here the state vector is still $x_{t-1} = [o_{t-1}, r_{t-1}, \zeta_{t-1}, a_{t-1}, \mu_{t-1}, \chi_{t-1}]'$ and its law of motion is unaltered. Since we dropped controls, this system corresponds to case 1 of section 2. By proposition 1, there will be no time aggregation, but the innovations cross sectionally aggregate the structural disturbances.

The second system uses $z_t = (o_t, \pi_t, n_t)$. It is obtained using (37) into the other equations:

$$
\begin{aligned}
(1+\rho_r)\chi_t - \rho_r \chi_{t-1} &= \chi_{t+1} - \frac{1}{1-h}\,(a_{t+1} + o_{t+1} - o_t) + (\frac{h+\rho_r}{1-h} + (1-\rho_r)\phi_y)\,(a_t + o_t - o_{t-1}) \\
&- (\frac{h\rho_r}{1-h})\,(a_{t-1} + o_{t-1} - o_{t-2}) + (\rho_r + (1-\rho_r)\phi_p)\,\pi_t + e_{mp_t} - \pi_{t+1} \qquad (38) \\
\pi_t &= \pi_{t+1}\,\beta + k_p\left(\frac{h}{1-h}\,(a_t + o_t - o_{t-1}) + (1+\sigma_n)\,n_t\right) + k_p\,(\mu_t - \chi_t) \qquad (39) \\
o_t &= \zeta_t + (1-\alpha) n_t \qquad (40)
\end{aligned}
$$

Here an endogenous state is eliminated. As (38) indicates, leaving r_{t-1} out makes the Euler equation a second order difference equation. Thus, we loose one state, r_{t-1}, but acquire another one, o_{t-2}. Because both states and controls are eliminated, this system corresponds to case 2 of section 2. Proposition 2 then tells us that the innovations will cross-sectionally and time aggregate $e_{t-s}, s \geq 0$. By proposition 3, we expect identified responses to be more distorted than in the four variables system. Clearly, with three observables, at most three shocks are identifiable from the u_t's.

Intuitively, aggregation distortion occur for the following reasons. First, notice that (38) is a dynamic aggregate demand equation in output and inflation while (39)-(40) define a dynamic aggregate supply equation in the same variables and that both are instantaneously moved by TFP and preference disturbances. Thus, it will be impossible to separate these disturbances in such a system giving rise to cross-sectional aggregation problems. Second, (40) depend on a_{t-1}, ζ_{t-1}, and because of the presence of o_{t-2} in the equation, also on ζ_{t-2}. Hence, a moving average structure is created making the aggregate demand equation evolving more persistently in response to shocks than in the original model.

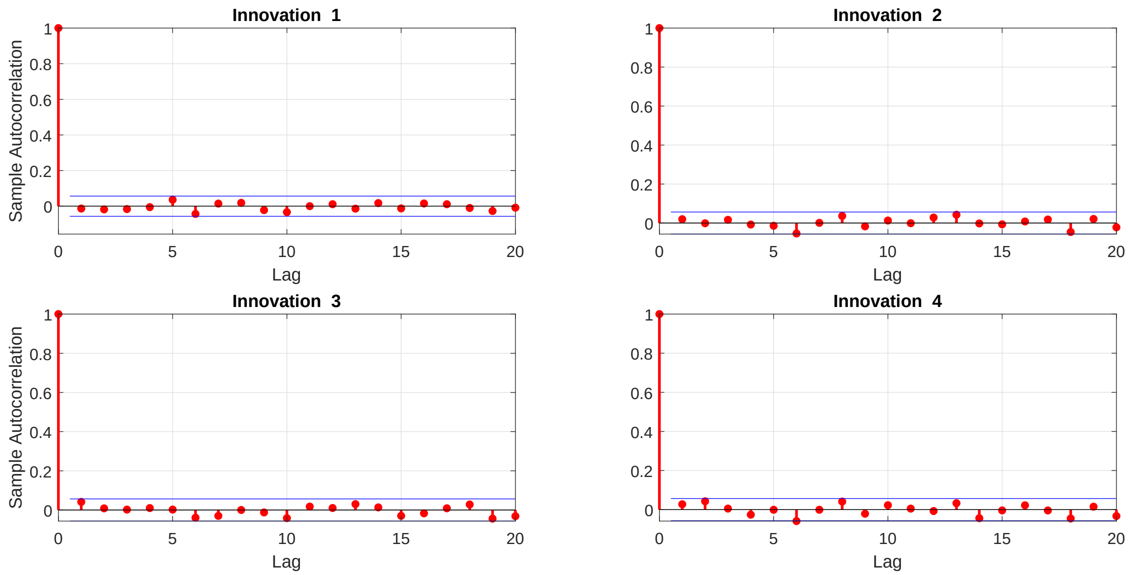
The third system uses $z_t = (\pi_t, n_t, r_t)$. It is obtained using (36) in the other equations:

$$
\begin{aligned}
\chi_t &= \chi_{t+1} - \frac{1}{1-h}\,(a_{t+1} + \zeta_{t+1} - \zeta_t + (1-\alpha)\,(n_{t+1} - n_t)) + \frac{h}{1-h}\,(a_t + \zeta_t - \zeta_{t-1} + (1-\alpha)\,(n_t - n_{t-1})) \\
&+ r_t - \pi_{t+1} \qquad (41) \\
\pi_t &= \pi_{t+1}\,\beta + k_p\left(\frac{h}{1-h}\,(a_t + \zeta_t - \zeta_{t-1} + (1-\alpha)\,(n_t - n_{t-1})) + (1+\sigma_n)\,n_t\right) + k_p\,(\mu_t - \chi_t) \qquad (42) \\
r_t &= \rho_r\, r_{t-1} + (1-\rho_r)\,(\phi_y\,(a_t + \zeta_t - \zeta_{t-1} + (1-\alpha)\,(n_t - n_{t-1})) + \phi_p\,\pi_t) + \varepsilon_{mp_t} \qquad (43)
\end{aligned}
$$

In this system a state variable, o_{t-1}, is lost. However, the optimality conditions remain a set of first order difference equations. The reason is that n_{t-1} becomes a state variable and, given the production function, it closely proxy for o_{t-1}. Because the states are repackaged and controls omitted, cross sectional and time aggregation will be present. However, because given ζ_{t-1}, n_{t-1} closely proxy for o_{t-1}, time deformations will be small. Thus, we expect the relationship between u_t and e_t and the impulse responses to be less distorted then in the (o_t, π_t, n_t) system.

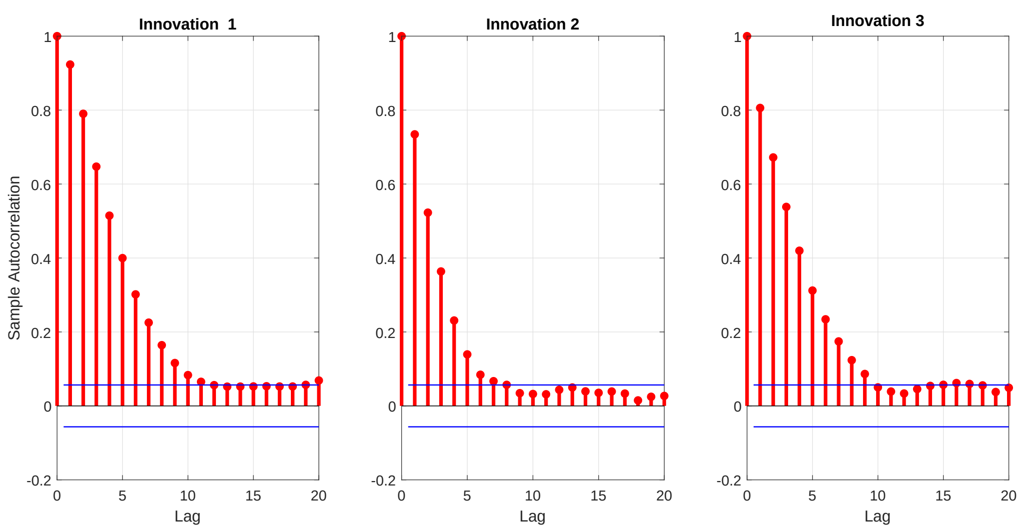
The properties of the reduced form innovations To confirm the intuition, we first analytically compute the autocorrelation function of the innovations in the three systems. With the value obtained, we report a 95% asymptotic tunnel for the hypothesis that the autocorrelation at each horizon is zero - which would hold if time aggregation is absent. The innovations of the (o_t, π_t, n_t, r_t) system are, as expected, white noise, see figure 1; those of the (o_t, π_t, n_t) system display serial correlation and numerous lags are significant, see figure 2. The innovations of the (π_t, n_t, r_t) system are instead serially uncorrelated, see figure 3.

Figure 1: Autocorrelation function, innovations in (o_t, π_t, n_t, r_t) system.



Note: Parallel lines describe 95 % asymptotic tunnel for the hypothesis of zero autocorrelations.

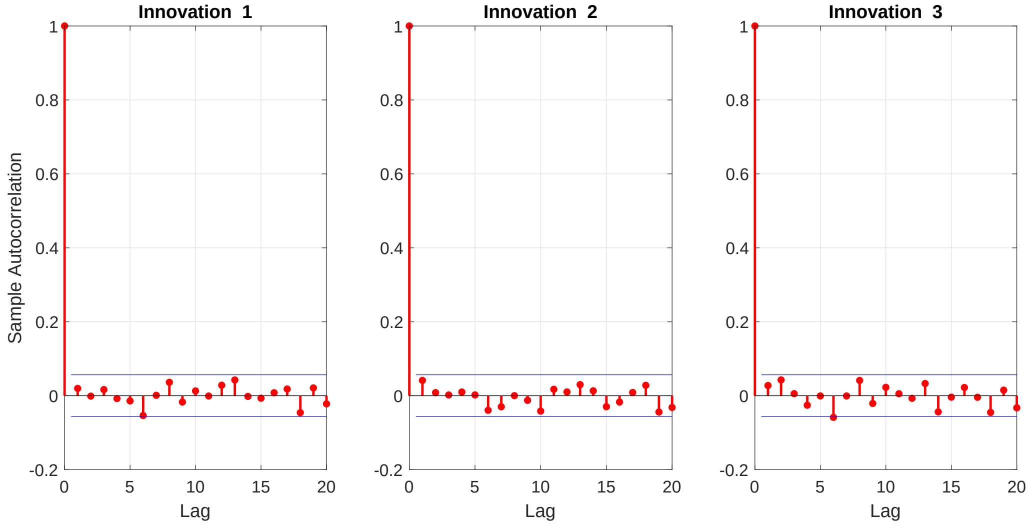
Figure 2: Autocorrelation function, innovations in (o_t, π_t, n_t) system.



Note: Parallel lines describe 95 % asymptotic tunnel for the hypothesis of zero autocorrelations.

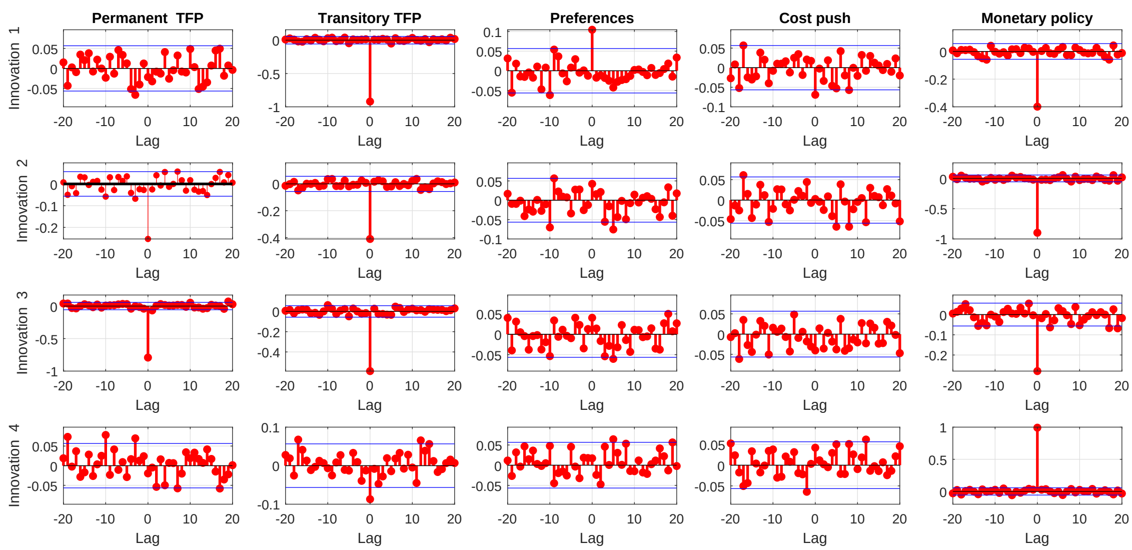
Next, we compute the cross correlation function between the innovations and the structural disturbances. We report the value obtained and an 95% asymptotic tunnel for the hypothesis that the cross correlation at each horizon is zero. If time aggregation is absent, only contemporaneous correlations should be significantly different from zero. In the four variable system, u_t and e_t are only contemporaneously linked, see figure 4. This is not the case for the innovations of the (o_t, π_t, n_t) system: u_t significantly correlates with several lags of e_t, see figure 5. The innovations of the (n_t, π_t, r_t) system instead show weak evidence of time aggregation, see figure 6.

Figure 3: Autocorrelation function, innovations in (π_t, n_t, r_t) system.



Note: Parallel lines delimit the 95 % asymptotic tunnel for the hypothesis of zero autocorrelations.

Figure 4: Cross correlation function, innovations in the (o_t, π_t, n_t, r_t) system and structural shocks.

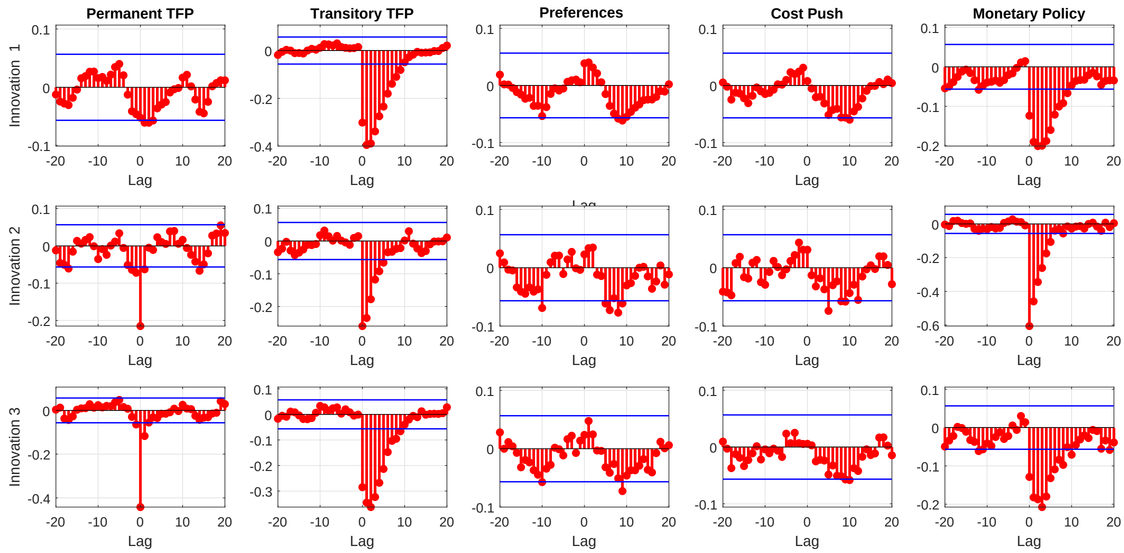


Note: Parallel line delimit the 95 % asymptotic tunnel for the hypothesis of zero cross correlations.

Shock identification and dynamic responses To examine the extent of cross sectional aggregation we present $\lambda(\theta, L)$ (for the (o_t, π_t, n_t, r_t) and (π_t, n_t, r_t) systems only $\lambda_0(\theta)$ is relevant), the Cholesky factor of the covariance matrix of innovations, and responses to identified shocks.

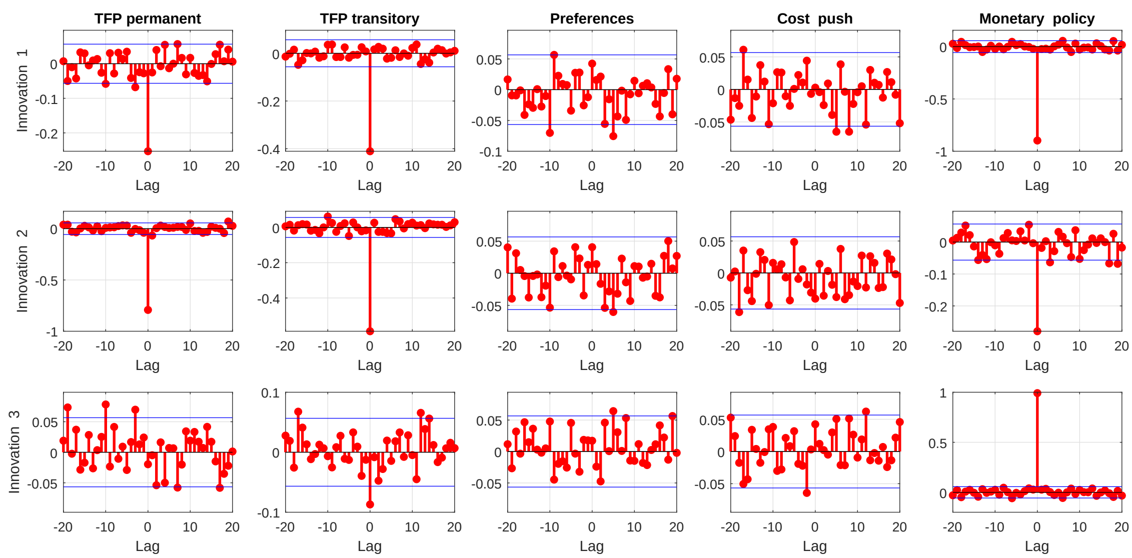
Cross sectional aggregation matters in all systems (see table 1). With four observables, transitory TFP and monetary policy disturbances receive the largest weights in the innovation and cost push disturbance the smallest. Thus, identification of cost push disturbances is difficult, even when the correct restrictions are used; their variability has to be of an order of magnitude larger for innovations to carry information about them. In addition, while the four variable system preserves the sign of the

Figure 5: Cross correlation function, innovations in the (o_t, π_t, n_t) system and structural shocks.



Note: Parallel lines delimit 95 % asymptotic tunnel for the hypothesis of zero cross correlations.

Figure 6: Cross correlation function, innovations in the (π_t, n_t, r_t) system and structural shocks.



Note: Parallel lines delimit 95 % asymptotic tunnel for the hypothesis of zero cross correlations.

contemporaneous responses to monetary policy disturbances (an increases interest rates and a fall in output, inflation, and hours), positive stationary TFP and negative preference disturbances will be confused, when sign restrictions are used for identification as they both produce an instantaneous fall in (o_t, π_t, n_t, r_t).

In the (o_t, π_t, n_t) system, structural disturbances enter the innovations for a number of time periods ($\lambda_0(\theta), \lambda_1(\theta)$ and $\lambda_2(\theta)$ are reported for illustration). Note also that sign restrictions can not now separate TFP, preference, and monetary policy disturbances. In fact, positive TFP, negative preference and contractionary policy disturbances all have negative effects on (o_t, π_t, n_t).

In the (π_t, n_t, r_t) system, the sign and the magnitude of the loadings of structural disturbances are the same as in the four variable system. As compared with the (o_t, π_t, n_t, r_t) system, we loose the possibility to distinguish stationary TFP, permanent TFP and preference shocks. However, there is no change in the ability to recover monetary policy disturbances.

Table 1: Entries of the $\lambda(L)$ matrix

	Structural shocks				
	a_t	ζ_t	χ_t	μ_t	ϵ_t
	Innovations in (y_t, π_t, n_t, r_t) system				
			$\lambda_0(\theta)$		
u_{1t}	0.018	-0.722	0.087	-0.005	-0.303
u_{2t}	-0.158	-0.306	0.042	0.042	-0.716
u_{3t}	-1.464	-1.078	0.131	-0.007	-0.452
u_{4t}	-0.047	-0.086	0.014	0.012	0.778
	Innovations in (π_t, n_t, r_t) system				
			$\lambda_0(\theta)$		
u_{1t}	-0.158	-0.306	0.042	0.042	-0.716
u_{2t}	-1.464	-1.078	0.131	-0.007	-0.452
u_{3t}	-0.047	-0.086	0.014	0.012	0.778
	Innovations in (y_t, π_t, n_t) system				
			$\lambda_0(\theta)$		
u_{1t}	-0.05	0.71	0.11	0.03	-0.29
u_{2t}	-0.19	-0.30	0.05	0.05	-0.70
u_{3t}	-1.57	-1.06	-0.17	0.05	-0.43
			$\lambda_1(\theta)$		
u_{1t}	-0.07	-0.92	0.12	0.04	-0.41
u_{2t}	-0.01	-0.28	0.03	0.01	-0.52
u_{3t}	-0.25	-1.37	0.18	0.06	-0.61
			$\lambda_2(\theta)$		
u_{1t}	-0.05	-0.90	0.11	0.04	-0.46
u_{2t}	-0.01	0.28	0.03	-0.01	-0.52
u_{3t}	-0.09	-1.35	0.16	-0.07	-0.69

Cholesky factors Table 2 displays the Cholesky factors of the covariance matrix of the innovations of original model (assuming disturbances have unit variance and with the rows and columns corresponding to the variables solved out eliminated) and of the three reduced systems. While the entries of $\lambda_0(\theta)$ are such that standard zero restrictions are unlikely to identify structural disturbances, applying the same recursive restrictions to the innovations of the original and of the reduced systems makes the comparison meaningful.

The Cholesky factor of the (o_t, π_t, n_t, r_t) system retains the signs of the Cholesky factor of the original model, but magnitudes are altered, sometimes substantially (see the (3,2) or (4,2) elements). A similar picture emerges in the (π, n_t, r_t) system: the signs don't change, but the magnitude are off (see the (3,1) element). Thus, the pattern of responses to orthogonal shocks in these two systems should mimic those of the original model but magnitude distortions could be important.

For the innovations of the (o_t, π_t, n_t) system the story is different: the signs are affected and magnitude differences are large. For example, while in the original system an orthogonal unitary shock to n_t implies a roughly similar instantaneous effect on o_t and π_t, the same shocks in the

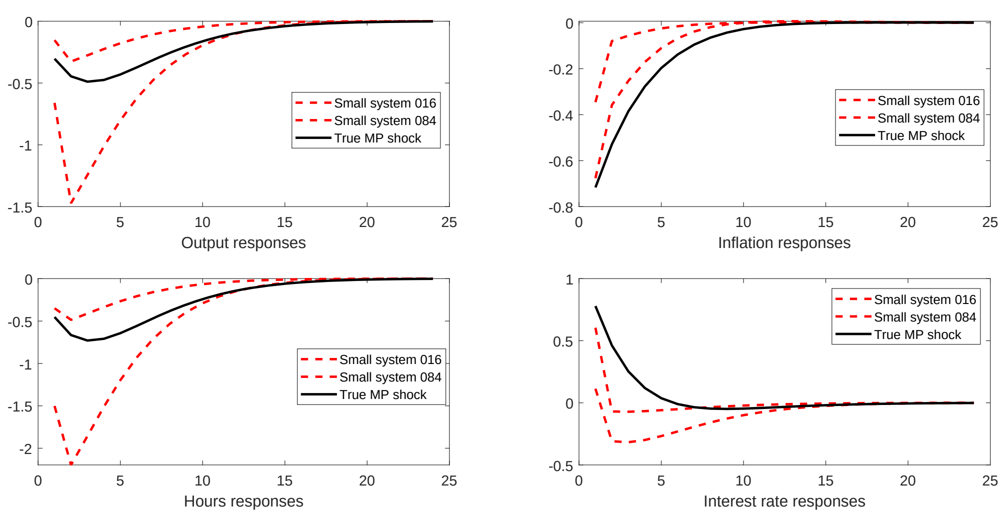
(o_t, π_t, n_t) system has a 15 times larger effect on o_t and a negative effect on π_t. As we have seen, these distortions will remain at longer horizons.

Table 2: Cholesky factors

Observables	Original system				Reduced system			
(o_t, π_t, n_t, r_t)	0.75				0.78			
	0.68	0.26			0.55	0.57		
	1.06	1.14	0.95		1.14	0.44	1.14	
	-0.42	-0.13	0.16	0.07	-0.22	-0.70	0.26	0.07
(π_t, n_t, r_t)	0.26				0.79			
	1.14	0.95			1.11	1.50		
	-0.13	0.16	0.07		-0.65	0.36	0.23	
(o_t, π_t, n_t)	0.75				9.55			
	0.68	0.26			5.16	1.50		
	1.06	1.14	0.95		15.36	-0.02	1.52	

Impulse responses We measure the dynamic distortions induced by aggregation when we identify disturbances via sign restrictions. Because propositions 3 tells us that magnitude of the distortions depends on the number and the type of variables present in the empirical model, we expect different systems to have different properties.

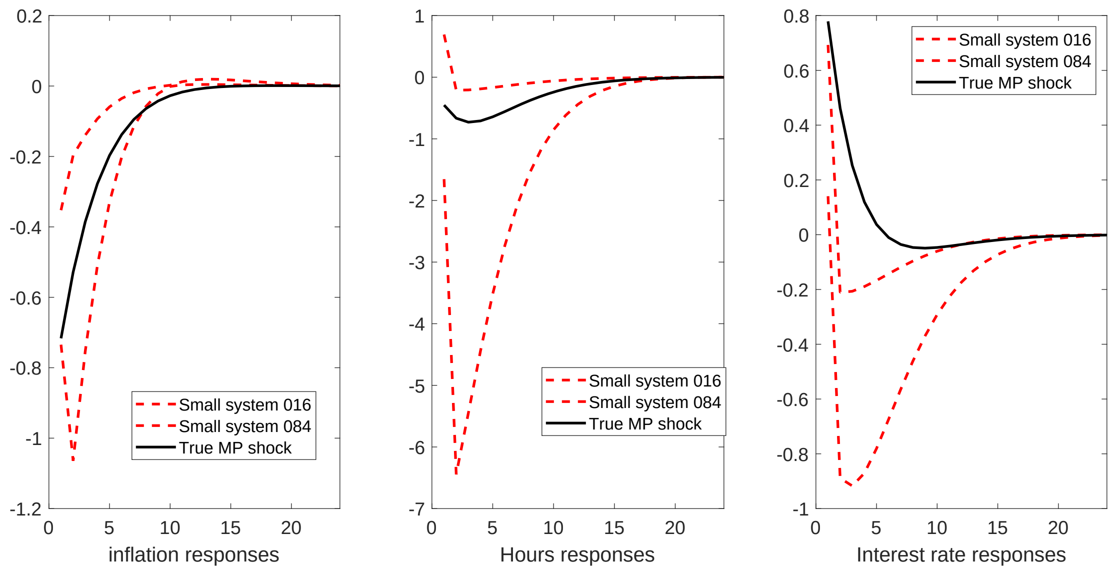
Figure 7: Responses to monetary policy shocks, (y_t, π_t, n_t, r_t) system



Note: The dashed regions report 68 % interval obtained accounting for rotation uncertainty. The solid line reports the responses in the DGP.

Figures 7 and 8 present the responses to a monetary policy shock in the (o_t, π_t, n_t, r_t) and the (π_t, n_t, r_t) systems when policy disturbances are identified assuming that an increase in r_t lead to a

Figure 8: Responses to identified monetary policy shocks, (π_t, n_t, r_t) system



Note: The dashed regions report 68 % interval obtained accounting for rotation uncertainty. The solid line reports the responses in the DGP.

contemporaneous fall in the other variables. Dotted lines represent 68% credible sets across rotations satisfying the restrictions. Superimposed as continuous lines are the responses of the original model. Clearly, even the (π_t, n_t, r_t) system encodes enough information to recover monetary policy disturbances. Thus, omitting consumption, output and its growth rate does not affect our ability to interpret the responses to identified monetary shocks, provided hours enter the empirical system.

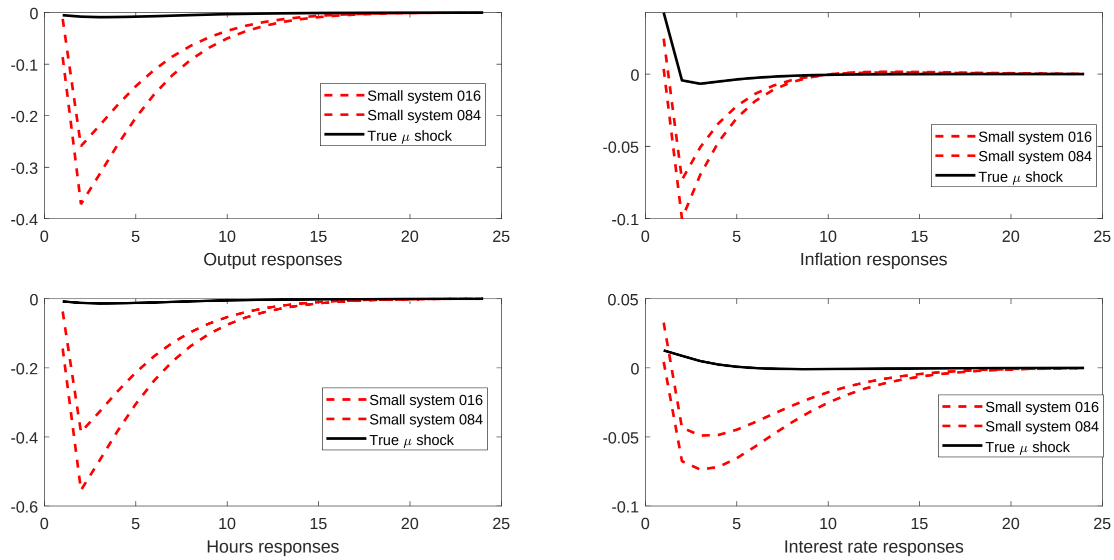
The conclusion is different if a researcher wants to measure the effects of cost push disturbances. As discussed, cost push disturbances are hard to obtain, even in the four variables, because their signal is weak. In agreement with Canova and Paustian [2011], figure 9 shows that the dynamics produced by identified post push shocks poorly approximate the dynamics induced by cost push disturbances in the original model, even when the correct sign restrictions are employed.

Recall that the entries of $\lambda_0(\theta)$ imply that positive stationary TFP and negative preference disturbances have the same sign implications on the four observables. Thus, imposing theoretically sound sign restrictions only identifies a linear combination of these two disturbances, a reminiscent of the masquerading effect discussed in Wolf [2018]. Figure 10 shows that the misspecification cross sectional aggregation produce in this case is large: rhe size of estimated impact responses is off by a large amount; and dynamic responses are more persistent in the smaller system.

An empirical model with only the theoretical states Omission of the theoretical states or failure to proxy for them generates time aggregation problems in small scale empirical systems. However, as discussed in case 3 of section 2, an empirical system with only the states (and none of the controls) will not necessarily produce interpretable identified shocks.

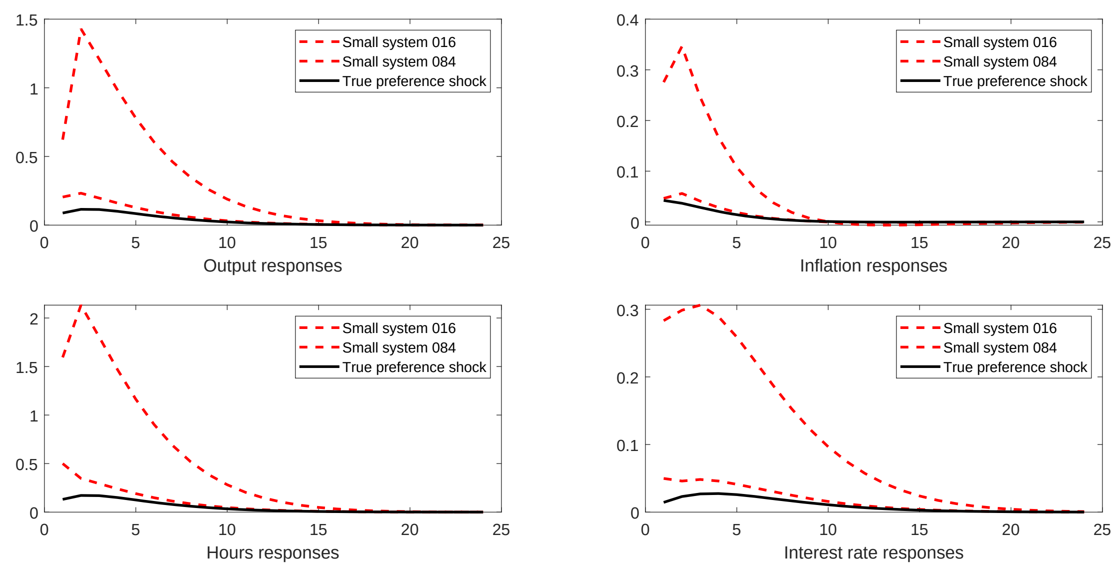
To show this, we take the (o_t, π_t, n_t, r_t) system and use the production function and the Phillips

Figure 9: Responses to identified cost push shocks, (o_t, π_t, n_t, r_t) system



Note: The dashed regions report 68 % interval obtained accounting for rotation uncertainty. The solid line reports the responses in the DGP.

Figure 10: Responses to identified preference shocks, (o_t, π_t, n_t, r_t) system



Note: The dashed regions report 68 % interval obtained accounting for rotation uncertainty. The solid line reports the responses in the DGP.

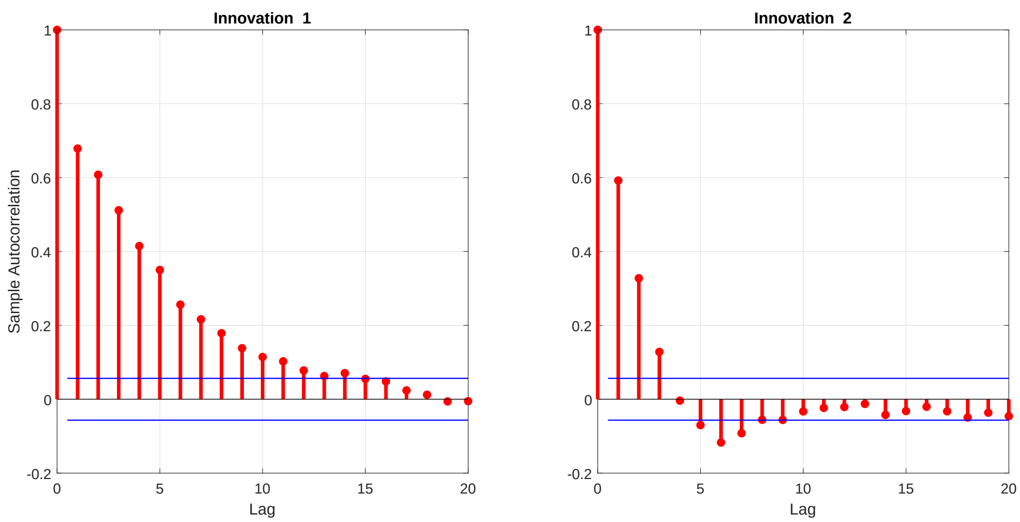
curve into the remaining two equations. The optimality conditions for $z_t = (o_t, r_t)$ are

$$\begin{aligned} \chi_t &= (1+\beta)\chi_{t+1} - \beta\chi_{t+2} - \frac{1}{1-h}(a_{t+1} + o_{t+1} - o_t) + \frac{\beta}{1-h}(a_{t+2} + o_{t+2} - o_{t+1}) \\ &+ (\frac{h}{1-h})(a_t + o_t - o_{t-1}) - (\frac{h\beta}{1-h})(a_{t+1} + o_{t+1} - o_t) + r_t - \beta r_{t+1} \\ &- k_p\left(\frac{h}{1-h}(a_{t+1} + o_{t+1} - o_t) + (1+\sigma_n)\frac{1}{1-\alpha}(o_{t+1} - \zeta_{t+1})\right) - k_p(\mu_{t+1} - \chi_{t+1}) \end{aligned} \tag{44}$$

$$
\begin{aligned}
r_t &= \beta r_{t+1} + \rho_r r_{t-1} - \beta \rho_r r_t + (1-\rho_r)\phi_y((a_t + o_t - o_{t-1}) - \beta(a_{t+1} + o_{t+1} - o_t)) \\
&+ (1-\rho_r)\phi_\pi \left(k_p \left(\frac{h}{1-h}(a_t + o_t - o_{t-1}) + (1+\sigma_n)\frac{1}{1-\alpha}(o_t - \zeta_t) \right) + k_p\,(\mu_t - \chi_t) \right) \\
&+ \epsilon_{mp_t} - \beta \epsilon_{mp_{t+1}} \qquad (45)
\end{aligned}
$$

Here o_{t-1}, r_{t-1} are still the endogenous states. However, inspection of (44)-(45) indicates that the optimization problem has changed and, for example, o_{t+2} and r_{t+1} now appear in the optimality conditions. Since the $(\bar{A}, \bar{B})$ matrices differs from the (A,B) matrices of the original system, this system will also feature time aggregation. Figures 11 and 12, which report the autocorrelation function of the innovations and their cross correlation with the five structural disturbances, indicate that u_t are serially correlated and load on e_{t-s} for $s \neq 0$. In particular, on a number of lags of the monetary policy disturbance.

Figure 11: Autocorrelation function, innovations in (o_t, r_t) system.

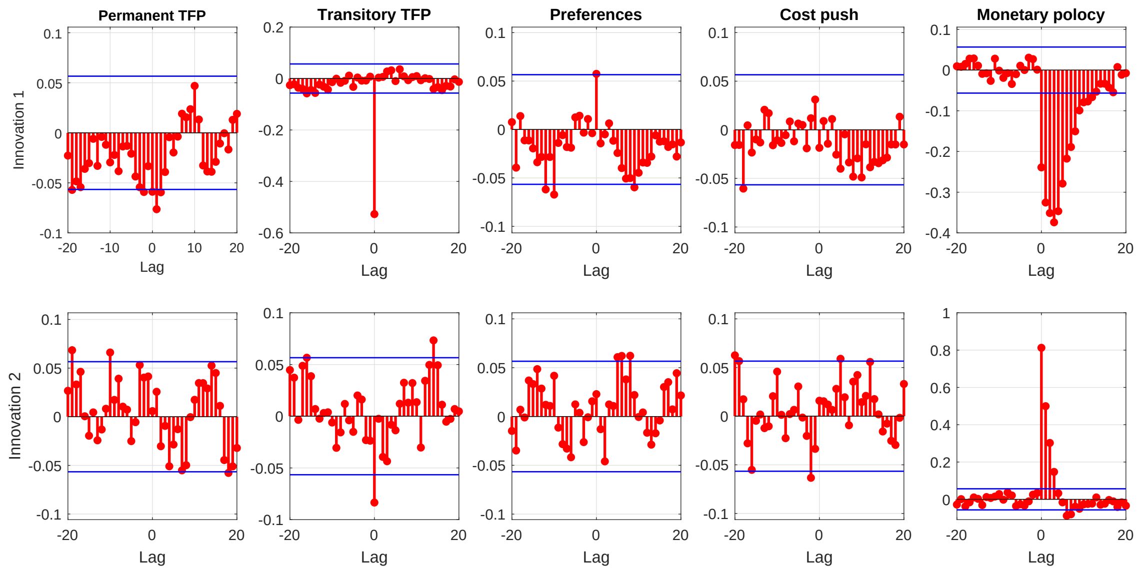


Note: Parallel lines describe the 95 % asymptotic tunnel for the hypothesis of zero autocorrelations.

Cross sectional aggregation is also important. With $z_t = (o_t, r_t)$, technology, monetary policy and cost push shocks are not separately identifiable with sign restrictions (they all have the same instantaneous effects on z_t). Figure 13 shows, monetary policy shocks identified with contemporaneous sign restrictions (r_t up and o_t down) is a a weighted average of the three underlying disturbances.

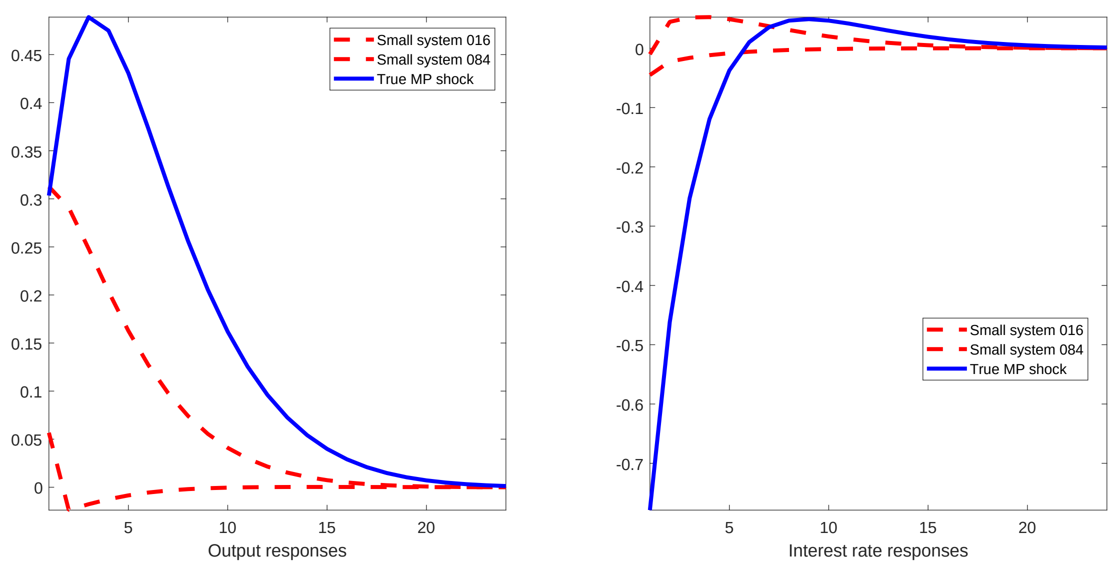
Permanent technology shocks and hours worked In the literature it has been common to use an empirical model with output growth (or labor productivity) and hours to identify permanent TFP shocks. The dynamics that are generated are then compared with the dynamics permanent TFP disturbances produce in standard RBC or new Keynesian models, see e.g. Galí [1999]. While the comparison could be meaningful when the DGP features only two disturbances (say, a permanent TFP and a demand shock), it may be inappropriate when the model used this section has generated the observed data. When only output growth and hours enter the empirical model, there will be both cross sectional and time aggregation problems since i) the five disturbances are compressed into two identified shocks; and (ii) the states of the original model are repackaged and their law of motion

Figure 12: Cross correlation function, innovations in (o_t, r_t) system and structural shocks.



Note: Parallel lines describe the 95 % asymptotic tunnel for the hypothesis of zero cross correlations.

Figure 13: Responses to identified monetary policy shocks, (o_t, r_t) system.



Note: The dashed regions report 68 % interval obtained accounting for rotation uncertainty. The solid line reports the responses in the DGP.

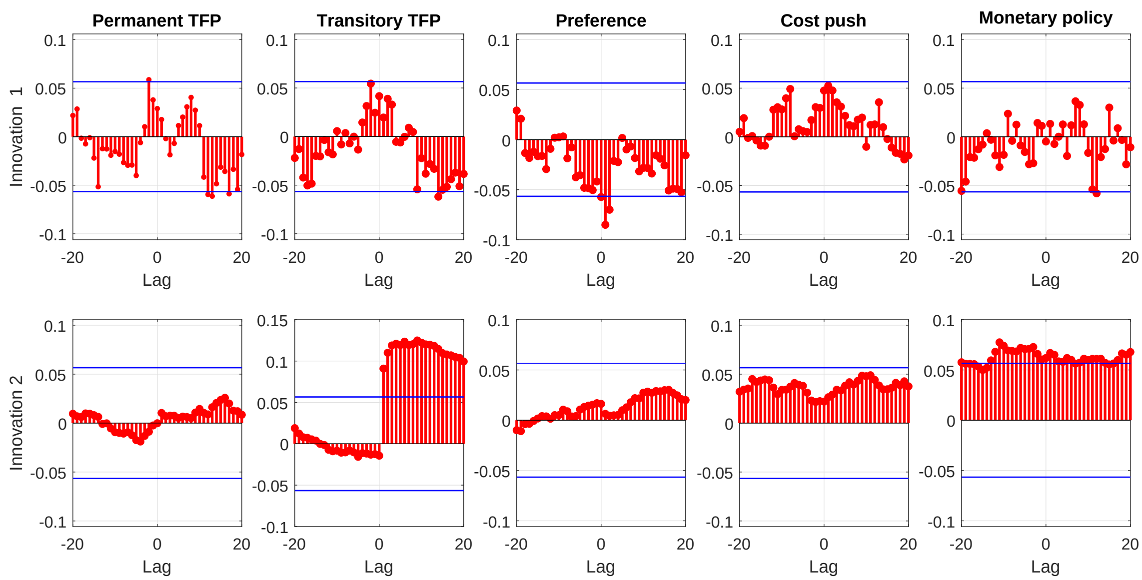
altered. To demonstrate these facts, we reduce the optimality conditions to contain $z_t = (g_t, n_t)$ [1]

$$\begin{aligned}(1+\rho_r)\chi_t &= \rho_r\chi_{t-1} + \chi_{t+1} + \frac{1}{1-h}g_{t+1} + (\frac{\rho_r + h}{1-h} + (1-\rho_r)\phi_y)g_t \\ &- \frac{h\rho_r}{1-h}g_{t-1} + \epsilon_{mp_t} + \kappa_p(\frac{h}{1-h}g_t + (1+\sigma_n)n_t) + \kappa_p(\mu_t - \chi_t) \qquad (46)\\ g_t &= a_t + \zeta_t + (1-\alpha)n_t - \zeta_{t-1} - (1-\alpha)n_{t-1} \qquad (47)\end{aligned}$$

[1]To obtain these equations one has to assume that $\beta^{-1} = (1-\rho_r)\phi_\pi\pi + \rho_r$. Such an assumption is retained also in the two shock system when comparisons are performed.

Note that lagged output growth and lagged hours are now endogenous states. To measure the extent of time aggregation we relate the innovations to the structural disturbances (see figure 14). Both innovations are moving averages of the five disturbances: lags of the stationary TFP and of the monetary policy disturbances enter the second innovation; lags and leads of the permanent TFP disturbance and lags of the preference disturbance load significantly on the first innovation.

Figure 14: Cross correlation function, innovations in (g_t, n_t) system and structural shocks.



Note: Parallel lines describe the 95 % asymptotic tunnel for the hypothesis of zero cross correlations.

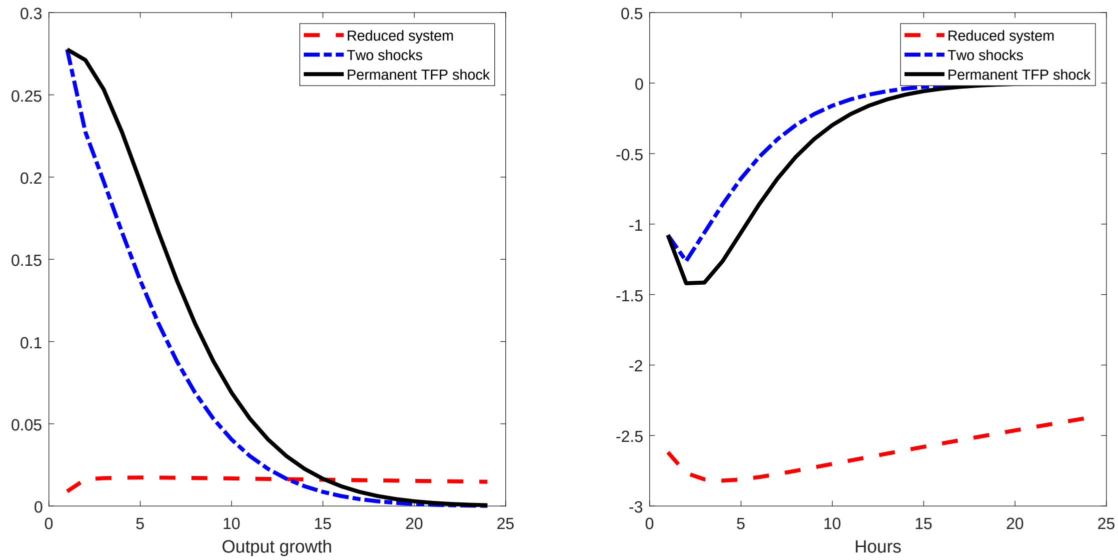
We identify a permanent TFP shock using standard long run restrictions and the data generated by (46)-(47) and compare the responses with those of the original model and those of the original model featuring with just two disturbances: a permanent TFP and monetary policy disturbance. Figure 15 shows that if the DGP has two disturbances, the responses obtained identifying a permanent shock in a VAR with $z_t = (g_t, n_t)$ capture well the features of the original responses. Instead, when the model of the section has generated the data, the magnitude and the persistence of the original responses is poorly captured.

In conclusion, the model we consider can not be reduced to a bivariate system with output growth and hours and meaningful innovations. Identified permanent technology shocks combine current and lagged values of permanent TFP as well as other stationary demand disturbances making the dynamics they induce are hard to interpret.

4 Summary and implications for practice

Small scale empirical models are easy to estimate and identify but problematic for interpretation and inference. If the DGP features more disturbances than the empirical model, dimensionality reductions may lead to aggregation biases. Cross sectional aggregation makes shock identification hard because the optimality conditions of the generating economy need not to be such that "classes" of disturbances will be properly compressed into the identified shocks. This is true even when the identification restrictions are sound and the empirical model is correctly specified.

Figure 15: Responses to identified permanent TFP shocks, (g_t, n_t) system.



Time aggregation dramatically complicates the identification process. When variable elimination leads to omission of states, alteration of their law of motion, or improper marginalization of the relationship between controls and states, a small scale empirical system becomes a very poor approximation to the DGP. Thus, the dynamics one obtains from identified shocks may have little to do with those generated by structural disturbances.

It is tempting to associate cross sectional aggregation with the elimination of theoretical controls and time aggregation with the elimination of theoretical states, but such an association is imperfect. As we have seen, time aggregation emerges also when the empirical system contains all the endogenous states and eliminating controls may induce both cross sectional and time aggregation biases, if the relationship between the remaining controls and the states is altered.

Aggregation problems have been generally ignored in the recent literature.Canova and Hamidi Sahneh [2018] showed that they may lead to spurious testing results when examining, e.g. fundamentalness issues, because they create time deformation in the innovations of the empirical system. Time aggregation problems have been discussed by Hansen and Sargent [1991], Marcet [1991], Fernández-Villaverde et al. [2007]. To the best of our knowledge, we are the first to show that the extent of time distortions depend on the dimensionality and the variables entering the empirical system and that a researcher can limit, to some extent, the magnitude of certain biases. Note that while aggregation and non-invertibility generate similar time deformations, a solution of the latter problem is insufficient to eliminate aggregation issues. To be clear, assuming away all the standard pile up, cancellation and identification problems, the estimation of a VARMA model can go a long way to reduce the gap between a DGP and the empirical model due to non-invertibility. However, estimating a small scale VARMA model will not solve aggregation problems if the DGP features a larger number of disturbances than the empirical model.

Our analysis has important implications for practice. If time aggregation problems are to be avoided, the empirical system needs to be sufficiently large. While is nowadays possible to estimate larger scale empirical models, even with relatively short datasets, their identification is still an issue.

Thus, small scale empirical models are likely to be preferred by macroeconomists for some time in the future. When the dimensionality of shock vector in the DGP and the empirical system may differ, two conditions need to be met for the matching exercise to be meaningful. First, because cross sectional aggregation makes identified shocks difficult to interpret, the size of an empirical system should be tailored to the disturbances of interest. In section 3, a monetary shock an be recovered from a trivariate system with (π_t, n_t, r_t) using meaningful restrictions but a cost push shock can not even in four variables system. Without guidance from theory, identified shocks may pick up the dynamics of structural disturbances which have distinct implications when a larger set of variables enter the empirical model. In section 3 we have provided a way to systematically explore dimensionality reductions: we started from a six variable structural model and analyzed whether interesting disturbances could be identified and the dynamics they produce well characterized when the empirical system includes only certain variables. We recommend applied researchers to do the same as routine practice, prior to the estimation of the empirical system. Second, by carefully choosing the variables entering the empirical system one can limit the magnitude of the aggregation distortions. Shrewd choices may dramatically change the quality of the inference. But for this to happen, empirical models can not be too small: a two variable system is likely to produce uninterpretable shocks and convoluted dynamics. In addition, one needs to be upfront about the structural model used to interpret the data. Canova and Paustian [2011] showed that shock identification is better anchored when business cycle measurement is tied up with robust identification restrictions. Our results indicate that the connection with theory is even more important if aggregation is present and an empirical system must be specified only after the structural model used to interpret the data has been selected. Stylized facts in a small scale empirical model are not theory-free. If two researchers use two theoretical models with the same (New Keynesian) features but with different number or type of disturbances to interpret the data, they ought to use different empirical models to identify disturbances and trace out their dynamics, even if they care about the same impulses.

Applied investigators who disregard aggregation issues should be aware that their analysis may be affected in a number of ways. On the one hand, the empirical impact responses may be off-mark and their sign may poorly characterize what happens in the DGP. On the other, identified and structural dynamics may have little to do with each other. Finally, viarance and historical decomposition exercises may be distorted. Researchers should also realize that an abundant number of lags may limit time aggregation, but it will do nothing to reduce cross sectional aggregation.

While it is common to sweep aggregation problems under the rug, assuming that the theory only features q_1 shocks, misspecification may be pervasive in the literature. For example, Central Banks use structural models with dozens of disturbances to interpret the data and academic researchers often twist standard models in estimation so that structural parameters become exogenous disturbances (e.g an elasticity of substitution becomes a markup disturbance) to improve the fit of their specifications. Furthermore, there is nothing that guides researchers in choosing both how large q_1 should be and which disturbances to include in the theory. The argument that it should include disturbances which are important for business cycle fluctuations is, unfortunately, a catch-22 proposition because their relevance depends on the choice and the number of disturbances included.

In general, the practice of comparing small scale SVAR and larger scale DSGE responses should be considerably refined. Showing that the qualitative pattern of responses to interesting impulses is similar is neither necessary nor sufficient for a structural model to be considered successful if aggregation is present. To make the gap smaller one should compare responses obtained from identified shocks in the small scale empirical system with the responses obtained in the theory, once it is reduced to the same variables as the empirical system. While the process may be analytically complicated, it is feasible even in large scale models and helps to quantify and interpret the distortions produced in specific empirical systems. Alternatively, one should compare theoretical and identified responses in the mimimum size empirical system, which preserves the dynamics and the state-control links of the theory. If a theory is satisfactory in both dimensions, evidence in its favor becomes stronger.

It has become popular recently to use IV approaches to identify certain shocks and local projection techniques to compute dynamic responses (see e.g. Rossi [2019] for a survey) Would such methods reduce the aggregation gap? Local projection techniques and IV estimation may help to limit both cross sectional and time aggregation problems. But for this to happen a number of strong conditions need to be met. Take for example case 2 of section 2, where the states are eliminated from the empirical model. In this case the DGP for the observables is a VARMA(2,1) which, in a companion from, can be written as $W_t = QW_{t-1} + Rv_t$ where $W_t = [y_t, y_{t-1}]'$ $v_t = [e_t, e_{t-1}]'$, $Q = \begin{pmatrix} F_{21} & F_{22} \\ I & 0 \end{pmatrix}$ and $R = \begin{pmatrix} G_{20} & G_{21} \\ 0 & 0 \end{pmatrix}$. Projecting $W_{t+h},\ h = 1, 2, \ldots$ on t-1 information:

$$W_{t+h} = Q^{h+1}W_{t-1} + Q^h Rv_{jt} + u_{t+h} \tag{48}$$

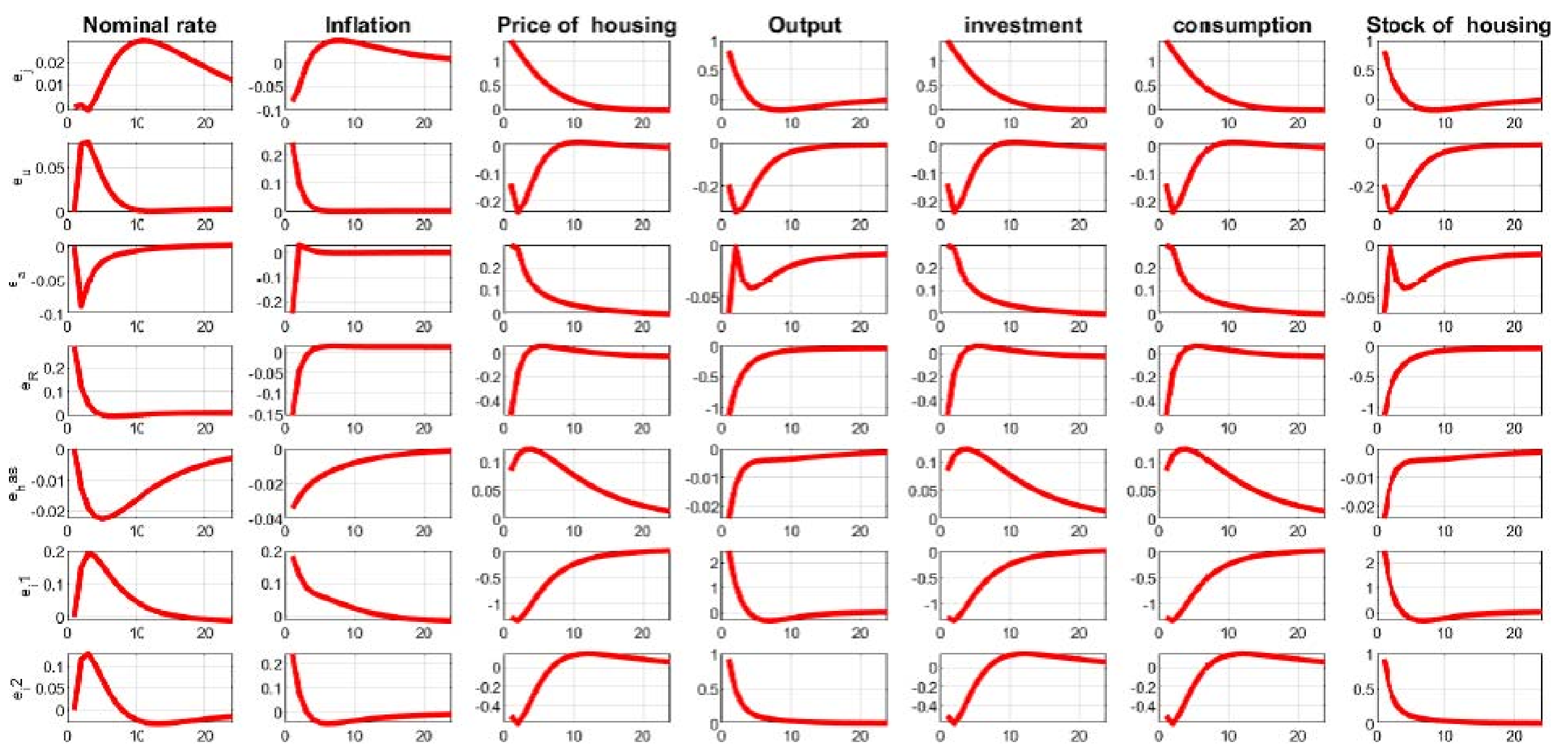
where v_{jt} is the disturbance of interest, $u_{t+t} = Q^h Rv_{-jt} + Q^{h-1}Rv_{t+1} + \ldots + Rv_{t+h}$, and v_{-jt} are all the disturbances at t except the $j-th$ one. Because local projections do not employ the residuals of a VAR in the exercise, they are less prone to cross sectional aggregation when $q_i < q$. However, for local projections to be successful in capturing $Q^h R$ the regressors of the projection equation should be W_{t-1} and v_{jt}. If e_{jt} is instead used in the projection equation, the right hand side variables will be correlated with the error term making OLS invalid. When v_{jt} is not observable, we need proxies that capture the effect of both e_{jt} and e_{jt-1}. Similarly, if an IV approach is used after normalization, the instruments have to be strictly exogenous and capture only the variations in W_{jt} which are due to v_{jt}. Predetermined instruments are insufficient in this case unless the conditioning set of the projection equation is extended to include an infinite number of lags of W_t. Thus, alternatives to SVARs could work in making the match with the theory tighter. However, to the best of our knowledge local projections and IV estimation have not yet come into the mainstream of stylized fact production. Furthermore, they have to be appropriately rigged to deliver results which are superior to those standard SVARs when $q_2 < q$.

5 The effect of house price disturbances

The dynamics of output and inflation following house price disturbances are of primary policy importance following the 2008 financial crisis. Starting with Iacoviello [2005] many authors have tried to understand whether the responses obtained in the data can be rationalized with a structural model featuring housing, leveraged agents, and standard macroeconomic frictions. Since house price disturbances are not necessarily the major source of fluctuations in macroeconomic variables, at least in normal times, the theoretical models one employs to interpret the data typically contain several other disturbances, see e.g. Rabanal [2018], Linde' [2018] for recent examples. However, apart from obvious core choices, it is not clear what one should include and, depending on the focus of the investigation, alternative disturbances may be considered. For example, for monetary policy the interaction between house price and other demand disturbances is important; for financial stability house price and leverage disturbances are at the center of attention.

Iacoviello [2005] sets the problem aside by selecting the minimum number of disturbances to map the empirical evidence into the structural model: he uses a four variable VAR model to construct stylized facts about the transmission of house price shocks and a model with preferences, monetary policy, technology and cost push disturbances to estimate the structural parameters and interpret the data. In this section, we work with Iacoviello [2005] model but, for illustration purposes, add disturbances to the borrowing constraints of entrepreneurs and impatient consumers and a wealth disturbance to the budget constraint of impatient consumers. These disturbances have been used in many exercises and by including them, we try to account for the fact that identified house price shocks may also be capturing the effect of disturbances affecting borrowers decisions, for example, because of taxation.

Figure 16: Impulse responses theory

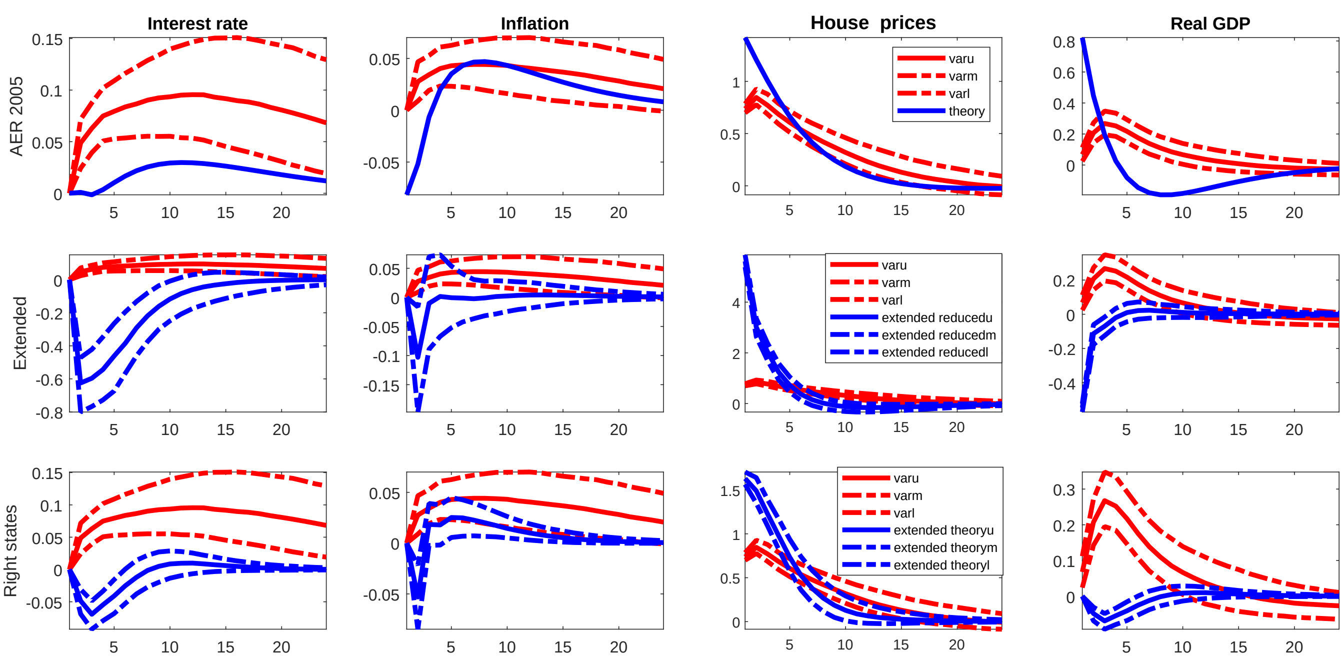


The optimality conditions and the law of motion of the shocks are in Appendix A. The model features 7 disturbances, 8 endogenous states (lagged house holdings of impatient consumers and of entrepreneurs, lagged bond holdings of patients and impatient consumers, lagged capital shock, lagged output, lagged nominal interest rate, and lagged inflation) and 15 endogenous controls. The theoretical responses of the nominal rate, inflation, house prices, output, consumption, investment, housing of constrained consumers to the disturbances are in figure 17. Positive preference disturbances increase all variables - the nominal rate and inflation only after a few quarters (see first row) and the qualitative dynamics produced by preferences disturbances (e_j) need not be confused with those produced by other disturbances once we restrict attention to the four endogenous variables used by Iacoviello (nominal rate, inflation, house price and output). Nevertheless, since the seven disturbances are compressed into four innovations, it is hard to predict a-priori what identified house price shocks may capture. In addition, since a number of states are excluded from the empirical model, time aggregation may matter.

We take data for real GDP, the nominal interest rate, inflation, and real house prices from the FRED data base for the period 1975:1-2018:3 and identify house price shocks using the same lag setting, the same data transformation, and the same identification scheme of Iacoviello [2005] [2].

The first row of figure 18, which plots the posterior 68% intervals to an identified house price shock in the data and the responses to preference disturbances in the theory, reproduces Iacoviello's conclusion: following a temporary house price increase, output, inflation and the nominal interest rate persistently rise, even though in the data, the maximum response of output is delayed by 4-5 quarters. Thus, the theory seems appropriate in qualitatively matching the data dynamics. Unfortunately, the first row of figure 18 is misleading: it displays theoretical dynamics when all states are used to calculate responses; and disregards that there are only four recoverable shocks in the VAR.

Figure 17: Models and data, q_t innovations



Note: The first row reports the response to preference shocks in Iacoviello (2005) model and the 68% highest posterior interval data; the second row median responses of the extended Iacoviello model with 7 shocks compressed to four observables and the 68% highest posterior interval in the data; the third row the median responses of the extended Iacoviello model, when data is generated with the right states and the 68% highest posterior interval in the data; the fourth row the theoretical responses to preference shocks in the extended Iacoviello model and the the 68% highest posterior interval produced when the data is generated with the right states.

To appreciate the effects of aggregation we solve equations out and reduce the first order conditions of the model to have same four endogenous variables used in the VAR as unknown. The second row of figure 16 still plots the posterior 68% interval responses to an identified house price shock in the data but now reports the posterior 68% interval responses to an identified house price shock using

[2] Iacoviello HP filters real GDP and house prices prior to their use in the VAR. While this choice has important implication for the timing of house price shocks and for the responses it generates, we decided to stick to this transformation since the purpose of the exercise is to show the effects of aggregation, rather than those of filtering.

data simulated from this reduced system [3]. The sign and the persistence are now altered: output and the nominal interest rate now respond negatively; and the response of inflation is insignificant after a few quarters.

Thus, if the extended model approximates well the dynamics of identified house price shocks, we should see output, inflation and the nominal rate responses in the data to be different. The responses generated by the theoretical decision rule are not relevant because the empirical system disregards states and has innovations cross sectionally combining structural disturbances. To restate the same concept differently aggregation matters: a four variable VAR is too small to be able to produce identified house price shocks that have the same interpretation as preference disturbances when the theory features six other disturbances.

What is it the cause of the drastic change in the dynamic responses? Is it cross sectional aggregation? Is it time series aggregation? Is it truncation lags? The third row of figure 18 evaluates the contribution of time aggregation to the changes. We use the decision rules of the extended model with seven disturbances, simulate data for the four relevant endogenous variables, and identify house price shocks as in the first two rows. Because the innovations contain information about all model states, only cross sectional aggregation is present.

The responses in rows 2 and 3 are qualitatively similar Thus, time series aggregation seems relative unimportant. Truncation problems are also minor: the lag length of the estimated VAR produced by the theory is irrelevant for the qualitative pattern we present. The differences between rows 1 and 2 of figure 18 are then due to cross sectional aggregation: the sign of output and interest rate responses changes because seven structural disturbances are compressed into four VAR innovations.

To understand what house price shocks capture, we compute the matrix of loadings of each innovation on the seven structural disturbances. If no contamination is present, we should expect a row of zeros for q_t, except in the position corresponding to the preference disturbance.

Table 3: Loading of innovations in (R, π, q, Y) on disturbances

	Disturbances						
	e_R	e_j	e_u	e_a	e_{has}	e_{i1}	e_{i2}
R_t	1.0	0	0	0	0	0	0
π_t	-0.53	-0.003	1.43	-0.11	-0.13	0.18	0.24
q_t	-1.83	0.05	-0.80	0.13	0.33	-1.27	-0.51
y_t	-3.92	0.03	-1.14	-0.02	-0.09	2.46	0.92

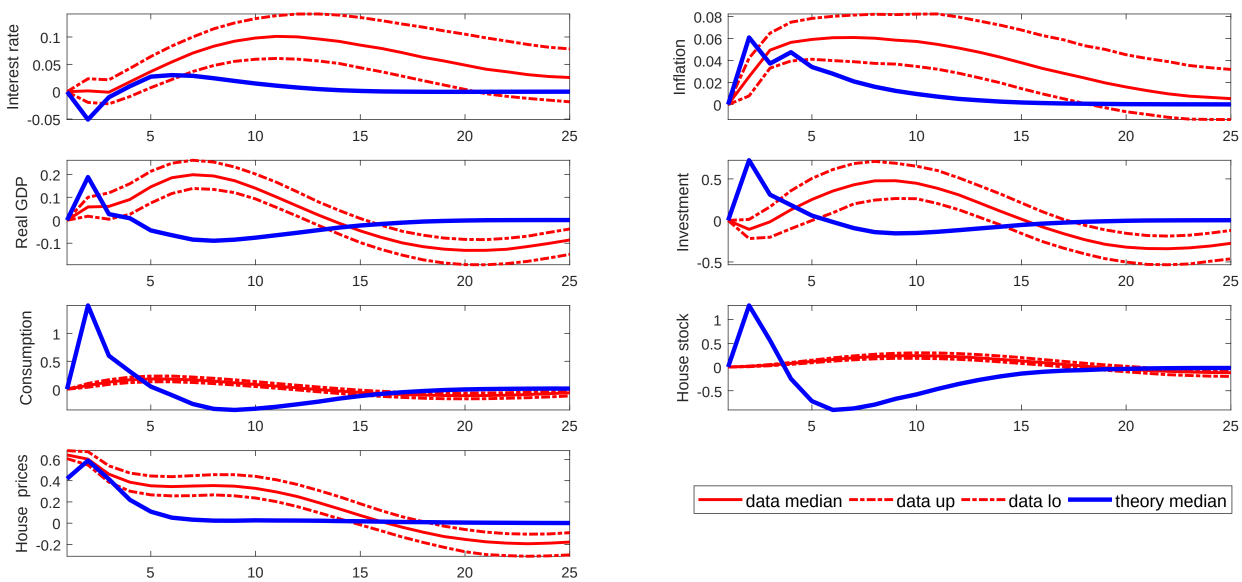
Interestingly, and confirming the results of section 3, the contamination present when characterizing monetary policy disturbances in a four variable system is small. Thus, comparing monetary policy disturbances in the theory and an identified monetary policy shock in the data is meaningful. On the other hand, house price innovations are strongly contaminated by cross sectional aggregation: house price innovations heavily load on monetary policy disturbances (-1.83) and on borrowing constraint disturbance of the impatient household (-1.27), while the loading on the preference shock

[3] The three new disturbances have persistence equal 0.75 and standard deviation 1.0, 1.0, 0.25, respectively. Since we normalize the impulse to unity, the magnitude of the standard deviations is irrelevant for the comparison.

is small (0.05). Hence, if the model correctly represents the DGP, identified house price shocks in the data are a mixture of monetary policy, borrowing constraints, and cost push disturbances while preference disturbances play a minor role.

The sixth row of figure 17 shows that positive borrowing constraint shocks imply a positive reaction of output and of the nominal rate. Thus, the negative output and interest rate responses we observe in rows 2 and 3 of figure 18 are produced by the large negative loading that borrowing constraint disturbances have on identified house price shocks. Note also that the preference disturbance e_j has small loadings in all innovations making it hard to recover them form any innovation.

Figure 18: Seven variable system, q_t innovations



Note: The red lines report the median responses and the 68% posterior interval in the data; the blue line the median response in the theory.

One may be interested in knowing what is the minimal dimension of the VAR needed to make sense of identified house price shocks, if the theory used in this section is a good approximation to the DGP. It turns out that cross sectional aggregation is important in all systems with less than seven observables. If one has to avoid cross sectional aggregation, a seven variable VAR is needed. The question then becomes how do responses to identified house price shocks in a seven variable VAR look like. Figure 19 shows the data responses when we add investment, consumption and stock of housing to the VAR together with the median responses produced by identified house price shocks in the theory. Two facts stand out. First, identified house price shocks in a seven variable empirical model produce dynamics that look similar to those of preference disturbances (compare with first row of figure 17). Thus, cross sectional aggregation is reduced. Second, the data responses are now different, and the match with the theory is still poor.

To summarize, if the theory features more than four structural disturbances, a VAR with four endogenous variables is too small to make the comparison between preference disturbances and identified house price shocks meaningful. When the theory is reduced to have the same number of innovations as the data, the match is less than ideal because identified house price shocks capture a number of theoretical disturbances. Thus, the responses they generate are different from those produced by preference disturbances. To make responses to identified theoretical house price shocks look like those of preference disturbances we need a VAR with at least seven variables. Still, even when a seven variable VAR is used, the match with the theory is poor. If we stick to a four variable system, we can compare the theory and the data only for those disturbances recoverable from such a system in theory. As we have seen monetary policy disturbances satisfy such a requirement; preference disturbances do not.

6 Extensions

The process assumed to generate the data in (1)-(2) is linear. This may be restrictive in certain situations. For example, when analyzing uncertainty disturbances (see e.g. Basu and Bundick [2017]), the model used for comparison is solved using higher order methods. Hence, a non-linear specification for the DGP is needed. This section analyzes how the conclusions of section 2 change when the process generating the data is non-linear.

As Andreasan, Fernandez Villaverde, and Rubio Ramirez [2018] have shown the pruned solution of a nonlinear state space model approximated with higher order perturbations has a linear representation of the form:

$$X_t = \mu_x(\theta) + \nu_1(\theta)X_{t-1} + \nu_2(\theta)E_t \tag{49}$$

$$Y_t = \mu_y(\theta) + \nu_3(\theta)X_t \tag{50}$$

where, for example in the case of a second order approximation, $X_t = ((x_t^f)', (x_t^s)', (x_t^f \otimes x_t^f)')'$, and x_t^f are the states of the first order system, x_t^s are the states of the second order system; $E_t = (e_t', (e_t \otimes e_t - vec(I_{n_e}))', (e_t \otimes x_{t-1}^f)'(x_{t-1}^f \otimes e_t)')'$ where e_t are the structural disturbances and I_{n_e} the identify matrix of dimension n_e; Y_t are the controls of the problem and the matrices $\mu_x(\theta), \mu_y(\theta), \nu_1(\theta), \nu_2(\theta), \nu_3(\theta)$ are given in the appendix A of Andreasan et al. [2018]. Comparing $(49) - (50)$ and $(1) - (2)$ one can immediately see that a higher order DGP features a larger number of states and of structural disturbances. Thus, if the empirical system is specified to be linear and features $\tilde{Z}_t = \tilde{S}[X_t, Y_t]$ as observables, where $\tilde{S} = [\tilde{S}_1, \tilde{S}_2]$, the conclusions derived in (1)-(3) still hold unchanged. However, the reduction to $\tilde{Z}_t$ observables is potentially more damaging because the dimension of E_t is likely to be much larger than the dimension of $\tilde{Z}_t$, making cross sectional aggregation more severe, and a larger number of states is eliminated (all those involving higher order and cross terms), making time aggregation more important.

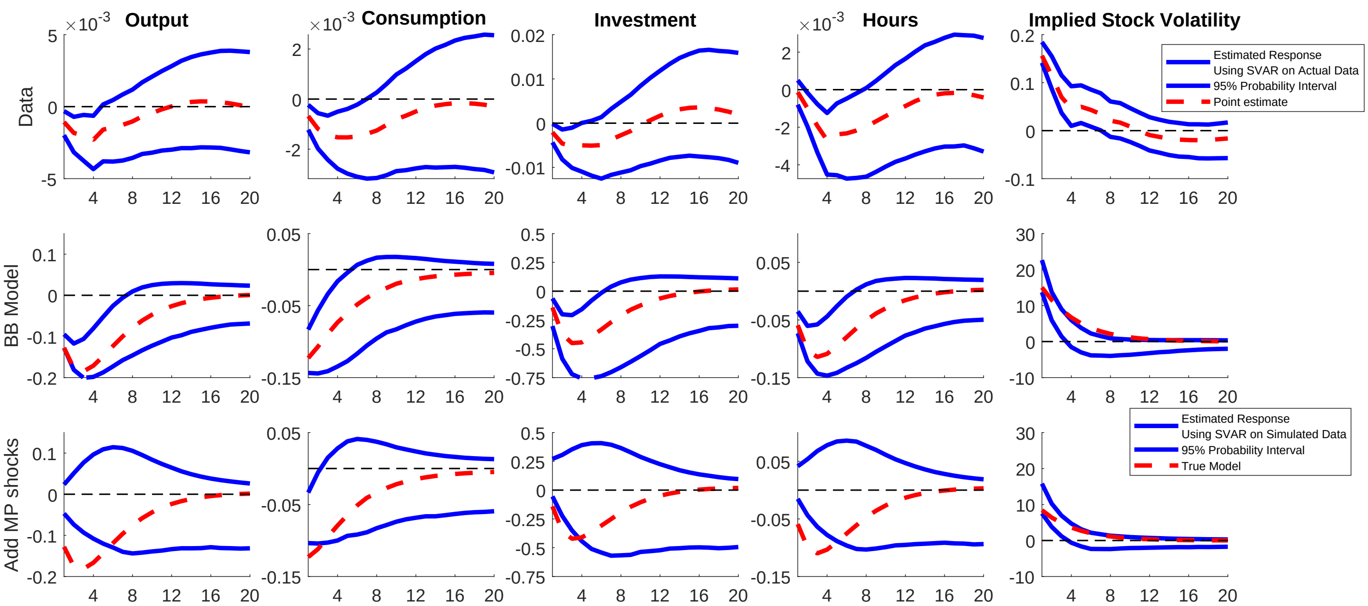
To highlight the effects of aggregation when the DGP features higher order terms, we take the model of Basu and Bundick [2017], which features a disturbance to the volatility of a preference, and two first moment disturbances: to the level of technology and to the level of preferences. The model is solved with a third order perturbation so that $E_t = [E_{1t}', E_{2t}']'$ where

$$E_{1t} = (e_t', (e_t \otimes e_t - vec(I_{n_e}))', (e_t \otimes x_{t-1}^f)'(x_{t-1}^f \otimes e_t)'(e_t \otimes x_{t-1}^s)')' \tag{51}$$

$$\begin{aligned} E_{2t} = {} & ((e_t \otimes x_{t-1}^f \otimes x_{t-1}^f)'(x_{t-1}^f \otimes x_{t-1}^f \otimes e_t)'(x_{t-1}^f \otimes e_t \otimes x_{t-1}^s)'(x_{t-1}^f \otimes e_t \otimes e_t)'(e_t \otimes x_{t-1}^f \otimes e_t)' \\ & (e_t \otimes e_t \otimes x_{t-1}^f)'((e_t \otimes e_t \otimes e_t) - E(e_t \otimes e_t \otimes e_t)))' \end{aligned} \tag{52}$$

Since e_t is a 3×1 vector, and x_t^f a 9×1 vector including lagged values of consumption, capital, hours, output, the nominal rate, of expected utility and of the three disturbances, X_t is a 432×1 vector and E_t is a 1112×1 vector. They use an eight variables VAR to trace out the effects of uncertainty shocks, which are identified with a Cholesky decomposition having the VXO index ordered first. The VAR variables include four of the endogenous states (output, consumption, hours and nominal rate), a proxy for the capital state (investment), two controls (inflation, and a volatility measure) and money supply variable, which is not present in the model.

Figure 19: Data and Models, VXO innovations



Note: The solid lines in the first row report 95% responses interval and the dashed line the point estimate using the actual data; the solid lines in the second and third row report the 95% response interval in the simulated data and the dashed line the conditional response in the theory.

The first row of figure 19 presents the point estimate and the 95% response intervals of output, consumption, investment, hours and VXO to an uncertainty shock in the data. The second row presents the responses to an uncertainty shock obtained from Basu and Bundick [2017]'s model: the dashed line reports theoretical responses, and the solid lines the estimated 95% response interval obtained running a VAR on simulated data, once an uncertainty shock is identified as in the first row. The match between the theory and the data appears to be good. Furthermore, there is no qualitative difference between the theoretical responses and those produced by an identified uncertainty shock in the simulated data.

Two features of the authors' specification are however puzzling. Despite the fact that the nominal interest rate is used in the VAR of the actual data, the model has little to say about interest rate dynamics because it posts a deterministic Taylor rule with no persistence (see equation (7), page 945). This goes against a large body of literature which has examined how the US policy rate has been determined over the last 20 years and, even away from the zero bound, implies that monetary policy has no role in affecting the economy. Second, while the dynamics induced by uncertainty shocks are clear, it is not obvious why changes in uncertainty of the economy are only demand driven. In principle, second moment shocks to technology could generate similar dynamics in real aggregate variables via a precautionary saving channel. Thus, the data potentially features more disturbances than those used in the model and that the restrictions used to identify uncertainty shocks may be insufficient to make the match between the theory and the data sound. Put it differently, aggregation may be important when analyzing the effect of identified uncertainty shocks. For illustration, we examine what happens when a monetary policy disturbance is included in the model. As row 3 of figure 19 shows the theoretical response and the estimated response intervals differ significantly. Moreover, VAR responses in the data and in the theory do not line up.

The reason for why rows 2 and 3 differ is that monetary policy disturbances get mixed up with uncertainty disturbances in the identification process since the both have the same qualitative effects on the variables of interest (an increase in the nominal rate makes all other variables fall). While the theoretical responses are constructed conditional on the monetary policy shocks being equal to zero, in the VAR with simulated data, the monetary policy disturbance takes both positive and negative values. Because disturbances are uncorrelated, output, consumption, investment and hours responses to uncertainty shocks could be both positive or negative depending on the relative importance of uncertainty and monetary policy disturbances for these variables and the sign of monetary policy disturbance at each t. The fact that responses are insignificant in the VAR indicates that Cholesky identified shocks pick up positive uncertainty disturbances and positive and negative monetary policy disturbances [4].

7 Conclusions

It is common in macroeconomics to collect stylized facts about the transmission of certain structural shocks using SVAR models and then build DSGE models to interpret the dynamics found in the data. However, DSGE models are typically of larger scale and may features more shocks than a SVAR. This paper argues that this dimensionality gap may create important inferential distortions.

When the structural model features q shocks, but only $q_1 < q$ variables are used in the empirical model, cross sectional and time aggregation biases make identified shocks and the dynamics they generate mongrels with little economic interpretation.

Cross sectional aggregation emerges when several structural disturbances contemporaneously affect the variables of the empirical model. Time aggregation occurs whenever the empirical model is specified without paying sufficient attention to the theory used to explain the data. Cross sectional aggregation makes sound theoretical restrictions insufficient to obtain meaningful disturbances. Time aggregation makes identified shocks distributed lags of the structural disturbances.

We use a standard New Keynesian model to show how to properly match the theory to a small scale empirical model, the problems that occur when the empirical model is too small, and how to reduce time distortions linking the theory and the empirical model more explicitly. We argue that the theory used to interpret the data and the disturbances of interest must guide both the choice of observables and the minimal dimension of the empirical model. Thus, the empirical model used to derive dynamic facts is not theory-free when $q_1 < q$.

We provide suggestions on how to avoid the aggregation trap when one insists in matching the dynamics produced by identified shocks in small scale empirical models and larger scale DSGE models. We revisit Iacoviello [2005]'s evidence about the transmission of house price shocks and show that the gap between the theory and the data may be larger than previously thought.

---

[4] It is worth emphasizing that a linear VAR is not the right empirical counterpart of (49)-(50). When the class of models suggested by Arouba, Boccola, and Schorfeide [2017] is used, some of the masquerading problems discussed here are eased.

Because there is no guidance to choose the number and the type of disturbances entering a theoretical model and because the interaction of different sets of disturbances may be crucial to understand the data and to formulate policy prescriptions, it should be clear that the problems we study in this paper are pervasive in applied macroeconomics. Furthermore, since small scale VAR models will remain the basic empirical tool to examine shock transmission for a while, researchers ought to be aware of the problems they face in practice and of ways to minimize the impact of aggregation on the results they present. Finally, we would like to reiterate that aggregation is distinct from invertibility, even though they both imply time deformation problems. The interpretation problems we emphasize are different and the distortions potentially more important.

References

Martin Andreasan, Jesus Fernandez Villaverde, and Juan Rubio Ramirez. The pruned state space system of a non-linear dsge model. *Review of Economic Studies*, 85:1–49, 2018.

Borogan Arouba, Luigi Boccola, and Frank Schorfeide. Assessing dsge nonlinearities. *Journal of Economic Dynamics and Control*, 83:34–54, 2017.

Susanto Basu and Brent Bundick. Uncertainty shocks in a model of effective demand. *Econometrica*, 85:937–958, 2017.

Paul Beaudry, Patrick Feve, Alain Guay, and Frank Portier. When is non-fundamentalness in var a real problem? Technical report, University of Toulouse, working paper 16-738, 2016.

Philip Braun and Stephan Mittnik. Misspecification in vector autoregressions and their effects on impulse responses and variance decompositions. *Journal of Econometrics*, 59:319–341, 1991.

Fabio Canova and Filippo Ferroni. Multiple filtering devices for the estimation of cyclical dsge models. *Quantitative Economics*, 2:73–98, 2011.

Fabio Canova and Mehdi Hamidi Sahneh. Are small scale var useful for business cycle analysis? revisiting non-fundamentalness. *Journal of the European Economic Association*, 16:1069–1093, 2018.

Fabio Canova and Matthias Paustian. Business cycle measurement with some theory. *Journal of Monetary Economics*, 58:345–361, 2011.

Ryan Chahrour and Kyle Jurado. Recoverability. Technical report, Boston College., 2018.

Varadarajan Chari, Patrick Kehoe, and Ellen McGrattan. A critique of structural vars using real business cycle theory. Technical report, Federal Reserve Bank of Minneapolis, 2005.

Jon Faust and Eric Leeper. When do long run restrictions give reliable results. *Journal of Business and Economic Statistics*, 15:345–353, 1988.

Jesus Fernández-Villaverde, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson. ABCs (and Ds) of Understanding VARs. *American Economic Review*, 97:1021–1026, 2007.

Mario Forni, Luca Gambetti, and Luca Sala. Reassessing structural vars: beyond abc (and d's). Technical report, Bocconi University, manuscript, 2016.

Jordi Galí. Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review*, 89:249–271, 1999.

Raffaella Giacomini. The relationship between dsge and var models. *in T.B. Fomby, L. Kilian, A. Murphy, (eds.), Advances in Econometrics, 32, Emerald Press*, pages 1–25, 2013.

Lars Peter Hansen and Thomas J Sargent. Two difficulties in interpreting vector autoregressions. *in L.P. Hansen and T.J. Sargent, (eds). Rational expectations econometrics, Westview Press Boulder, CO*, pages 77–120, 1991.

Matteo Iacoviello. House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review*, 95:739–764, 2005.

Jesper Linde'. Dsge models: still useful in policy analysis? *Oxford Review of Economic Policy*, 34: 269–286, 2018.

Helmuth Lutkepohl. Linear transformation of vector arma processes. *Journal of Econometrics*, 26: 283–293, 1984.

Albert Marcet. Time aggregation of economic time series. *in L.P. Hansen and T. J. Sargent, (eds). Rational expectations econometrics, Westview Press Boulder, CO*, pages 237–281, 1991.

Adrian Pagan and Tim Robinson. Implications of partial information for econometric modelling of macroeconomic systems. Technical report, University of Sydney., 2018.

Mikkel Plagborg Moller. Bayesian inference of structural impulse responses. Technical report, Princeton University, 2017.

Pau Rabanal. An estimated dsge model to analysize housing market policies in hong kong sar. Technical report, IMF working paper, 18-90, 2018.

Federico Ravenna. Vector autoregressions and reduced form representations of dsge models. *Journal of Monetary Economics*, 54:2048–2064, 2007.

Barbara Rossi. Identifying and estimating the effects of unconventional monetary policy in the data: How to do it and what have we learned? Technical report, Universitat Pompeu Fabra., 2019.

Christian Wolf. Svar (mis-)identification and the real effects of monetary policy. Technical report, Princeton University, 2018.

Appendix

The (linearized) equation of Iacoviello [2005]'s model are

$$rr_t = r_t - pi_{t+1} \tag{53}$$

$$\begin{aligned} y_t &= c_y c_t + (1 - c_y - cii_y - i_y)ci_t + cii_y cii_t + i_y i_t && (54)\\ ci_t &= ci_{t+1} - rr_t && (55)\\ i_t - k_{t-1} &= \gamma(i_{t+1} - k_t) + \frac{(1-\gamma(1-\delta)}{\psi}(y_{t+1} - x_{t+1} - k_t) + (\frac{1}{\psi}(c_t - c_{t+1}) && (56) \end{aligned}$$

$$\begin{aligned} q_t &= \gamma_E q_{t+1} + (1-\gamma_E)(y_{t+1} - x_{t+1} - h_t) - m\beta\ rr_t - i_{1,t} - (1-m\beta)(c_{t+1} - c_t)\\ &- \phi_E(h_t - h_{t-1} - \gamma(h_{t+1} - h_t)) && (57)\\ q_t &= \gamma_H q_{t+1} + (1-\gamma_H)(j_t - hii_t) - mii\beta\ rr - i_{2,t} + (1 - mii\beta)(cii_t - \omega cii_{t+1})\\ &- \phi_H(hii_t - hii_{t-1} - \beta_{ii}(hii_{t+1} - hii_t) && (58)\\ q_t &= \beta q_{t+1} + (1-\beta)j_t + \iota h_t + \iota_{ii} hii_t + ci_t - betaci_{t+1} + \frac{phi_H}{hi}(h(h_t - h_{t-1})\\ &+ hii(hii_t - hii_{t-1}) - \beta h(h_{t+1} - h_t) - \beta hii(hii_{t+1} - hii_t)) && (59) \end{aligned}$$

$$\begin{aligned} b_t &= q_{t+1} + h_t - rr_t + i_{1,t} && (60)\\ bii_t &= q_{t+1} + hii_t - rr_t + i_{2,t} && (61) \end{aligned}$$

$$\begin{aligned} y_t &= \frac{\eta}{\eta - (1-\nu-\mu)}(a_t + \nu h_{t-1} + \mu k_{t-1}) - \frac{1-\nu-\mu}{\eta-(1-\nu-\mu)}(x_t + \alpha ci_t + (1-\alpha)cii_t) && (62)\\ \pi_t &= \beta\pi_{t+1} - \kappa x_t + u_t && (63) \end{aligned}$$

$$\begin{aligned} k_t &= \delta i_t + (1-\delta)k_{t-1} && (64)\\ b_y b_t &= c_y c_t + qh_y(h_t - h_{t-1}) + i_y i_t + \frac{b_y}{\beta}(r_{t-1} + b_{t-1} - \pi_t) - (1 - si - sii)(y_t - x_t) && (65)\\ bii_y bii_t &= cii_y cii_t + qhii_y(hii_t - hii_{t-1}) + \frac{bii_y}{\beta}(bii_{t-1} + r_{t-1} - \pi_t) - sii(y_t - x_t) + w_t && (66) \end{aligned}$$

$$\begin{aligned} r_t &= (1-\rho_R)(1+\rho_\pi)\pi_{t-1} + \rho_y(1-\rho_R)y_{t-1} + \rho_R r_{t-1} + e_R && (67)\\ j_t &= \rho_j j_{t-1} + e_j && (68)\\ u_t &= \rho_u u_{t-1} + e_u && (69)\\ a_t &= \rho_a a_{t-1} + e_a && (70)\\ i_{1,t} &= \rho_1 i_{1,t-1} + e_b c1 && (71)\\ i_{2,t} &= \rho_2 i_{2,t-1} + e_b c2 && (72)\\ w_t &= \rho_w w_{t-1} + e_h as && (73)\\ tc_t &= c_y c_t + (1 - c_y - cii_y - i_y)ci_t + cii_y cii_t && (74)\\ th_t &= h_t + hii_t && (75) \end{aligned}$$