Aggregation in Large Dynamic Panels*

M. Hashem Pesaran†
Cambridge University and University of Southern California

Alexander Chudik‡
Federal Reserve Bank of Dallas and CIMF

March 1, 2012

Abstract

This paper investigates the problem of aggregation in the case of large linear dynamic panels, where each micro unit is potentially related to all other micro units, and where micro innovations are allowed to be cross sectionally dependent. Following Pesaran (2003), an optimal aggregate function is derived and used (i) to establish conditions under which Granger's (1980) conjecture regarding the long memory properties of aggregate variables from 'a very large scale dynamic, econometric model' holds, and (ii) to show which distributional features of micro parameters can be identified from the aggregate model. The paper also derives impulse response functions for the aggregate variables, distinguishing between the effects of composite macro and aggregated idiosyncratic shocks. Some of the findings of the paper are illustrated by Monte Carlo experiments. The paper also contains an empirical application to consumer price inflation in Germany, France and Italy, and re-examines the extent to which 'observed' inflation persistence at the aggregate level is due to aggregation and/or common unobserved factors. Our findings suggest that dynamic heterogeneity as well as persistent common factors are needed for explaining the observed persistence of the aggregate inflation.

Keywords: Aggregation, Large Dynamic Panels, Long Memory, Weak and Strong Cross Section Dependence, VAR Models, Impulse Responses, Factor Models, Inflation Persistence.
JEL Classification: C43, E31

---

*A preliminary version of this paper was presented as the 4^{th} Annual Granger Lecture delivered at the University of Nottingham, May 25, 2010, and at the Conference on High-Dimensional Econometric Modelling, Cass Business School, December 3-4, 2010. We thank participants at these venues for providing us with useful comments. This version has also benefited greatly from constructive comments and suggestions by the Editor (Robert Taylor) and two anonymous referees. We are also grateful to Benoit Mojon for providing us with the price data set used in Altissimo et al. (2009).

†Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. E-mail: mhp1@econ.cam.ac.uk; http://www.econ.cam.ac.uk/faculty/pesaran/

‡Federal Reserve Bank of Dallas, Research Department, 2200 N. Pearl Street, Dallas, TX 75201. E-mail: alexander.chudik@dal.frb.org. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

Nearly every study in economics must implicitly or explicitly aggregate: over time, individuals (consumers, firms, or agents), products, or space, and usually over most of these dimensions. It is therefore important that the consequences of aggregation for the analysis of economic problems of interest are adequately understood. It is widely acknowledged that aggregation can be problematic, but its implications for empirical research are often ignored either by adopting the concept of a 'representative agent', or by arguing that 'aggregation errors' are of second order importance. However, there are empirical studies where aggregation errors are shown to be quite important, including the contributions by Hsiao et al. (2005), Altissimo et al. (2009), and Imbs et al. (2005).[1]

There are several different aspects to the aggregation problem. One important issue is the conditions under which micro parameters or some of their distributional features can be identified and estimated from aggregate relations. Theil (1954) was the first to consider this problem in the context of static micro relations. Robinson (1978) considers the problem of estimating moments of the distribution of AR(1) micro coefficients, but excludes the possibility of a long memory when deriving the asymptotic distribution of his proposed estimator. Pesaran (2003) discusses estimating the average long-run micro effects and mean lags of the autoregressive distributed lag (ARDL) micro models from aggregate data.

A second closely related problem is derivation of an optimal aggregate function which could be used to compare persistence of shocks when aggregate and disaggregated models are considered. Theil (1954), Lewbel (1994), and Pesaran (2003) consider the problem of deriving an optimal aggregate function. The problem of aggregation of a finite number of independent autoregressive moving average (ARMA) processes is considered, for example, by Granger and Morris (1976), Rose (1977), and Lütkepohl (1984). The problem of aggregating a large number of independent time series processes was first addressed by Robinson (1978) and Granger (1980). Granger showed that aggregate variables can have fundamentally different time series properties as compared to those of the underlying micro units. Focusing on autoregressive models (AR) of order 1, he showed that aggregation can generate long memory even if the micro units follow stochastic processes with

[1] In addition to the empirical studies, Geweke (1985) develops a theoretical example, where he argues that ignoring the sensitivity of the aggregates to policy changes seems no more compelling than the Lucas critique of ignoring the dependence of expectations on the policy regime.

exponentially decaying autocovariances.

The aggregation problem has also been studied from the perspective of forecasting: is it better to forecast using aggregate or disaggregate data, if the primary objective is to forecast the aggregates? Pesaran, Pierse, and Kumar (1989) and Pesaran, Pierse, and Lee (1994), building on Grunfeld and Griliches (1960), develop selection criteria for a choice between aggregate and disaggregate specifications. Giacomini and Granger (2004) discuss forecasting of aggregates in the context of space-time autoregressive models. Cross-sectional aggregation of vector ARMA processes and a comprehensive bibliography is provided in Lütkepohl (1987).

A third issue of importance concerns the role of common factors and cross-sectional dependence in aggregation, which was first highlighted by Granger (1987), and further developed and discussed in Forni and Lippi (1997) and Zaffaroni (2004). Granger showed that the strength and pattern of cross-sectional dependence plays a central role in aggregation. Using a simple factor model, he argued that the factor dominates the aggregate relationship; and consequently, variables that may have very good explanatory power at the micro level might be unimportant at the macro level, and *vice versa*. Implications of Granger's finding that common factors dominate aggregate relationships have been explored in various papers in the literature.[2]

In this paper we investigate the problem of aggregation in the context of large linear dynamic panels, or high-dimensional VARs, where each micro unit is potentially related to all other micro units, and where micro innovations are allowed to be cross-sectionally dependent. In this way the earlier literature on aggregation of independent dynamic regressions is extended to aggregation of dynamic models with interactions and cross-sectional dependence. In particular, we allow for different degrees of interconnections across the individual units, relax the assumption that micro coefficients are independently distributed, and allow for a general pattern of cross-sectional dependence of micro innovations, which can be either strong or weak.[3] Using this generalized framework we re-visit two of the issues in the aggregation literature mentioned above. First, following Pesaran

[2] Granger also contributed to the discussion of temporal aggregation, aggregation of non-linear models, and small scale aggregation of space-time processes. See Granger (1993), Granger and Siklos (1995), Granger and Lee (1999) and Giacomini and Granger (2004). Other contributions to the theory of aggregation include the contributions of Kelejian (1980), Stoker (1984), Stoker (1986), and Garderen et al. (2000), on aggregation of static non-linear micro models, Pesaran and Smith (1995), Phillips and Moon (1999), and Trapani and Urga (2010) on the effects of aggregation on cointegration. Granger (1990) and Stoker (1993) provide early surveys.

[3] Concepts of strong and weak cross-sectional dependence are discussed and analysed in Chudik, Pesaran, and Tosetti (2011) and Bailey, Kapetanios, and Pesaran (2012).

(2003), we derive an optimal aggregate function and use it to establish links between parameters of the aggregate function (macro parameters) and the distributional moments of the underlying micro parameters. We examine the conditions under which the distributional features of micro parameters can be identified from aggregate relations. We also use the optimal aggregate function to establish the conditions under which Granger's (1980) conjecture about the long memory properties of aggregate variables from 'a very large scale dynamic, econometric model' is valid.[4]

Understanding the persistence of aggregate variables is the second main objective of this paper, where we compare impulse response functions of the aggregate variables derived using the optimal aggregate function with the impulse responses obtained using the disaggregated model. The combined aggregate shock in our set-up is defined as the sum of the composite macro and aggregated idiosyncratic shocks. The issue of persistence and the relative importance of the two components of the combined aggregate shock for the aggregate variable is also investigated by Monte Carlo experiments. The paper concludes with an empirical application to consumer price inflation in Germany, France and Italy, and re-examines the extent to which inflation persistence at the aggregate level is due to aggregation and/or common unobserved factors. We find that dynamic heterogeneity alone cannot explain the persistence of aggregate inflation, rather it is the combination of factor persistence and cross-sectional heterogeneity that seems to be responsible for the observed persistence of the aggregate inflation.

The remainder of the paper is organized as follows. We begin with the derivation of the optimal aggregate function in Section 2 for a factor augmented VAR model in N cross-sectional units. The optimal aggregate function is used in Section 3 to examine the relationship between micro and macro parameters. The impulse responses of the effects of aggregated idiosyncratic and composite macro shocks on the aggregate variable are derived and contrasted in Section 4. Monte Carlo experiments are presented in Section 5, and Section 6 reports on the empirical application. Section 7 concludes the paper. Some of the mathematical proofs are provided in an appendix.

A brief word on notations: $\|\mathbf{A}\|_1 \equiv \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$, and $\|\mathbf{A}\|_\infty \equiv \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$ denote the maximum absolute column and row sum norms of $\mathbf{A} \in \mathbb{M}^{n\times n}$, respectively, where $\mathbb{M}^{n\times n}$ is the space

[4] Granger uses different arguments to support his conjecture as compared to the formal analysis undertaken in this paper.

of real-valued $n \times n$ matrices. $\|\mathbf{A}\| = \sqrt{\varrho\left(\mathbf{A}'\mathbf{A}\right)}$ is the spectral norm of $\mathbf{A}$,[5] $\varrho\left(\mathbf{A}\right) \equiv \max_{1 \le i \le n}\left\{\left|\lambda_i\left(\mathbf{A}\right)\right|\right\}$ is the spectral radius of $\mathbf{A}$, and $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq ... \geq |\lambda_n(\mathbf{A})|$ are the eigenvalues of $\mathbf{A}$. All vectors are column vectors.

2 Aggregation of Factor Augmented VAR Models

Consider the following high-dimensional factor augmented VAR model in N cross-sectional units

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \mathbf{B}\mathbf{x}_t + \mathbf{\Gamma}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \text{ for } t = 1, 2, ..., T, \tag{1}$$

where $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})'$ is $N \times 1$ vector of cross-section specific regressors, $\mathbf{f}_t$ is $m \times 1$ vector of common factors, $\mathbf{\Phi}$ and $\mathbf{B}$ are $N \times N$ matrices of randomly distributed coefficients, and $\mathbf{\Gamma}$ is an $N \times m$ matrix of randomly distributed factor loadings with elements γ_{ij}, for $i = 1, 2, ..., N$, and $j = 1, 2, ..., m$. We denote the elements of $\mathbf{\Phi}$ by ϕ_{ij}, for $i, j = 1, 2, ..., N$, and assume that $\mathbf{B}$ is a diagonal matrix with elements β_i, also collected in the $N \times 1$ vector $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_N)'$.[6] The objective is to derive an optimal aggregate function for $\bar{y}_{wt} = \mathbf{w}'\mathbf{y}_t$ in terms of its lagged values, and current and lagged values of $\bar{x}_{wt} = \mathbf{w}'\mathbf{x}_t$ and $\mathbf{f}_t$, where $\mathbf{w} = (w_1, w_2, ..., w_N)'$ is a set of predetermined aggregation weights such that $\Sigma_{i=1}^N w_i = 1$. Throughout, it is assumed that $\mathbf{w}$ is known and the weights are granular, in the sense that

$$\frac{|w_i|}{\|\mathbf{w}\|} = O\left(N^{-1/2}\right), \text{ for any } i, \text{ and } \|\mathbf{w}\| = O\left(N^{-1/2}\right). \tag{2}$$

Denote the aggregate information set by $\Omega_t = (\bar{y}_{w,t-1},\ \bar{y}_{w,t-2}, ...; \bar{x}_{wt}, \bar{x}_{w,t-1}, ...; \mathbf{f}_t, \mathbf{f}_{t-1}, ...)$. When $\mathbf{f}_t$ is not observed the current and lagged values of $\mathbf{f}_t$ in Ω_t must be replaced by their fitted or forecast values obtained from an auxiliary model for $\mathbf{f}_t$, and possibly other variables, not included in (1). Consider the augmented information set, $\Upsilon_t = (\mathbf{y}_{t-M}; \mathbf{w}; \mathbf{x}_t, \mathbf{x}_{t-1}, ...; \mathbf{f}_t, \mathbf{f}_{t-1}, ...; \bar{y}_{w,t-1}, \bar{y}_{w,t-2}, ...)$, that includes the weights, $\mathbf{w}$, and the disaggregate observations on the regressors, x_{it}. Note that Ω_t is contained in Υ_t.

Now introduce the following assumptions on the eigenvalues of $\mathbf{\Phi}$ and the idiosyncratic errors,

[5] Note that if $\mathbf{x}$ is a vector, then $\|\mathbf{x}\| = \sqrt{\varrho\left(\mathbf{x}'\mathbf{x}\right)} = \sqrt{\mathbf{x}'\mathbf{x}}$ corresponds to the Euclidean length of vector $\mathbf{x}$.

[6] This specification can be readily generalized to allow for more than one cross-section specific regressor, by replacing $\mathbf{B}\mathbf{x}_t$ with $\mathbf{B}_1\mathbf{x}_{1t} + \mathbf{B}_2\mathbf{x}_{2t} + ... + \mathbf{B}_k\mathbf{x}_{kt}$.

$\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$.

ASSUMPTION 1 *The coefficient matrix,* $\mathbf{\Phi}$, *of the* VAR *model in (1) has distinct eigenvalues* $\lambda_i(\mathbf{\Phi})$, *for* $i = 1, 2, ..., N$, *and satisfy the following cross-sectionally invariant conditional moments*

$$\left.\begin{array}{l} E\left(\lambda_i^s(\mathbf{\Phi})\left|\Upsilon_t, \mathbf{P}, \boldsymbol{\varepsilon}_{t-s}\right.\right) = a_s, \\ E\left(\lambda_i^s(\mathbf{\Phi})\left|\Upsilon_t, \mathbf{P}, \boldsymbol{\beta}\right.\right) = b_s(\boldsymbol{\beta}), \\ E\left(\lambda_i^s(\mathbf{\Phi})\left|\Upsilon_t, \mathbf{P}, \mathbf{\Gamma}\right.\right) = c_s(\mathbf{\Gamma}), \end{array}\right\} \tag{3}$$

for all $s = 1, 2, ...,$ *and* $i = 1, 2, ..., N$, *where* $\Upsilon_t = (\mathbf{y}_{t-M}; \mathbf{w}; \mathbf{x}_t, \mathbf{x}_{t-1}, ...; \mathbf{f}_t, \mathbf{f}_{t-1}, ...; \bar{y}_{w,t-1}, \bar{y}_{w,t-2}, ...)$, *and* $\mathbf{P}$ *is* $N \times N$ *matrix containing the eigenvectors of* $\mathbf{\Phi}$ *as column vectors.*

ASSUMPTION 2 *The idiosyncratic shocks,* $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, *in (1) are serially uncorrelated with zero means and finite variances.*

Remark 1 *Assumption 1 is analytically convenient and can be viewed as a natural generalization of the simple AR(1) specifications considered by Robinson (1978), Granger (1980) and others. Using the spectral decomposition of* $\mathbf{\Phi} = \mathbf{P\Lambda P}^{-1}$, *where* $\mathbf{\Lambda} = diag\left[\lambda_1(\mathbf{\Phi}), \lambda_2(\mathbf{\Phi}), ..., \lambda_N(\mathbf{\Phi})\right]$ *is a diagonal matrix with eigenvalues of* $\mathbf{\Phi}$ *on its diagonal, the factor augmented VAR model can be written as*

$$y_{it}^* = \lambda_i(\mathbf{\Phi})\, y_{i,t-1}^* + z_{it}^*, \;\; i = 1, 2, ..., N, \;\; \text{and } t = 1, 2, ..., T; \tag{4}$$

where y_{it}^* *is the* i^{th} *element of* $\mathbf{y}_t^* = \mathbf{P}^{-1}\mathbf{y}_t$, *and* z_{it}^* *is the* i^{th} *element of* $\mathbf{z}_t^* = \mathbf{P}^{-1}(\mathbf{Bx}_t + \mathbf{\Gamma f}_t + \boldsymbol{\varepsilon}_t)$. *Consider now the conditions under which an optimal aggregate function exists for* $\bar{y}_{wt}^* = \mathbf{w}'\mathbf{y}_t^* = \mathbf{w}'\mathbf{P}^{-1}\mathbf{y}_t$. *We know from the existing literature that such an aggregate function exists if* $E\left(\lambda_i^s(\mathbf{\Phi})\left|z_{it}^*\right.\right) = a_s^*$, *for all* i. *Seen from this perspective, our assumption that conditional on* $\mathbf{P}$ *the eigenvalues have moments that do not depend on* i *seems sensible, and is likely to be essential for the validity of Granger's conjecture.*

Remark 2 *It is also worth noting that Assumption 1 does allow for possible dependence of* $\lambda_i(\mathbf{\Phi})$ *on the coefficients* β_i *and* γ_{ij}.

As shown in Pesaran (2003), the optimal aggregate function (in the mean squared error sense) is given by

$$\bar{y}_{wt} = E\left(\mathbf{w}'\mathbf{y}_t \,|\Omega_t\right) + v_{wt}, \tag{5}$$

where by construction $E\left(v_{wt}\,|\Omega_t\right) = 0$, and v_{wt}, $t = 1, 2, ...$ are serially uncorrelated, although they could be conditionally heteroskedastic.[7] Solving (1) recursively forward from the initial state, $\mathbf{y}_{-M}$, we have

$$\mathbf{y}_t = \mathbf{\Phi}^{t+M}\mathbf{y}_{-M} + \sum_{s=0}^{t+M-1} \mathbf{\Phi}^s \left(\mathbf{B}\mathbf{x}_{t-s} + \mathbf{\Gamma}\mathbf{f}_{t-s} + \boldsymbol{\varepsilon}_{t-s}\right). \tag{6}$$

Hence, using the spectral decomposition of $\mathbf{\Phi} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$, we obtain

$$\bar{y}_{wt} = \mathbf{w}'\mathbf{P}\mathbf{\Lambda}^{t+M}\mathbf{P}^{-1}\mathbf{y}_{-M} + \sum_{s=0}^{t+M-1} \mathbf{w}'\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1} \left(\mathbf{B}\mathbf{x}_{t-s} + \mathbf{\Gamma}\mathbf{f}_{t-s} + \boldsymbol{\varepsilon}_{t-s}\right). \tag{7}$$

Let $\mathcal{F}_{PB} = (\mathbf{P}, \mathbf{B})$ and $\mathcal{F}_{P\Gamma} = (\mathbf{P}, \mathbf{\Gamma})$. By the chain rule of expectations we obtain

$$E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{B}\,|\mathbf{P},\Upsilon_t\right) = E\left[E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{B}\,|\mathcal{F}_{PB},\Upsilon_t\right)|\mathbf{P},\Upsilon_t\right] = E\left[\mathbf{P}E\left(\mathbf{\Lambda}^s\,|\mathcal{F}_{PB},\Upsilon_t\right)\mathbf{P}^{-1}\mathbf{B}\,|\mathbf{P},\Upsilon_t\right].$$

Similarly,

$$E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma}\,|\mathbf{P},\Upsilon_t\right) = E\left[E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma}\,|\mathcal{F}_{P\Gamma},\Upsilon_t\right)|\mathbf{P},\Upsilon_t\right] = E\left[\mathbf{P}E\left(\mathbf{\Lambda}^s\,|\mathcal{F}_{P\Gamma},\Upsilon_t\right)\mathbf{P}^{-1}\mathbf{\Gamma}\,|\mathbf{P},\Upsilon_t\right].$$

But under (3) we have, $E\left(\mathbf{\Lambda}^s\,|\mathcal{F}_{PB},\Upsilon_t\right) = b_s(\boldsymbol{\beta})\mathbf{I}_N$, and $E\left(\mathbf{\Lambda}^s\,|\mathcal{F}_{P\Gamma},\Upsilon_t\right) = c_s(\mathbf{\Gamma})\mathbf{I}_N$. Hence

$$\begin{aligned} E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{B}\,|\mathbf{P},\Upsilon_t\right) &= E\left[\mathbf{P}b_s(\boldsymbol{\beta})\mathbf{P}^{-1}\mathbf{B}\,|\mathbf{P},\Upsilon_t\right] \\ &= E\left[b_s(\boldsymbol{\beta})\mathbf{B}\,|\mathbf{P},\Upsilon_t\right]. \end{aligned}$$

Similarly,

$$\begin{aligned} E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma}\,|\mathbf{P},\Upsilon_t\right) &= E\left[E\left(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma}\,|\mathcal{F}_{P\Gamma},\Upsilon_t\right)|\mathbf{P},\Upsilon_t\right] = E\left[\mathbf{P}c_s(\mathbf{\Gamma})\mathbf{P}^{-1}\mathbf{\Gamma}\,|\mathbf{P},\Upsilon_t\right] \\ &= E\left[c_s(\mathbf{\Gamma})\mathbf{\Gamma}\,|\mathbf{P},\Upsilon_t\right]. \end{aligned}$$

[7] Recall that under Assumption 2, we have $E(\mathbf{w}'\boldsymbol{\varepsilon}_t\,|\mathbf{\Phi},\mathbf{B},\mathbf{\Gamma},\Upsilon_t) = 0$, and hence $E(\mathbf{w}'\boldsymbol{\varepsilon}_t\,|\mathbf{P},\Upsilon_t) = 0$.

Finally,

$$
\begin{aligned}
E\left(\mathbf{P}\boldsymbol{\Lambda}^{s}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_{t-s}\left|\mathbf{P},\Upsilon_{t}\right.\right) &= E\left[E\left(\mathbf{P}\boldsymbol{\Lambda}^{s}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_{t-s}\left|\mathbf{P},\boldsymbol{\varepsilon}_{t-s},\Upsilon_{t}\right.\right)\left|\mathbf{P},\Upsilon_{t}\right.\right] = E\left[\mathbf{P}a_{s}\mathbf{I}_{N}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_{t-s}\left|\mathbf{P},\Upsilon_{t}\right.\right] \\
&= a_{s}E\left(\boldsymbol{\varepsilon}_{t-s}\left|\mathbf{P},\Upsilon_{t}\right.\right).
\end{aligned}
$$

Taking expectations of both sides of (7) conditional on $(\mathbf{P},\Upsilon_t)$, we now have

$$
\begin{aligned}
E\left(\bar{y}_{wt}\left|\mathbf{P},\Upsilon_{t}\right.\right) &= \mathbf{w}'E\left(\mathbf{P}\boldsymbol{\Lambda}^{t+M}\mathbf{P}^{-1}\left|\mathbf{P},\Upsilon_{t}\right.\right)\mathbf{y}_{-M} + \sum_{s=0}^{t+M-1}\mathbf{w}'E\left(\mathbf{P}\boldsymbol{\Lambda}^{s}\mathbf{P}^{-1}\mathbf{B}\left|\mathbf{P},\Upsilon_{t}\right.\right)\mathbf{x}_{t-s} + \\
&\quad \sum_{s=0}^{t+M-1}\mathbf{w}'E\left(\mathbf{P}\boldsymbol{\Lambda}^{s}\mathbf{P}^{-1}\boldsymbol{\Gamma}\left|\mathbf{P},\Upsilon_{t}\right.\right)\mathbf{f}_{t-s} + \\
&\quad \sum_{s=1}^{t+M-1}\mathbf{w}'E\left(\mathbf{P}\boldsymbol{\Lambda}^{s}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_{t-s}\left|\mathbf{P},\Upsilon_{t}\right.\right).
\end{aligned}
$$

Using the results derived above we obtain

$$
\begin{aligned}
E\left(\bar{y}_{wt}\left|\mathbf{P},\Upsilon_{t}\right.\right) &= \left(\mathbf{w}'\mathbf{y}_{-M}\right)a_{t+M} + \sum_{s=0}^{t+M-1}\mathbf{w}'E\left[b_{s}(\boldsymbol{\beta})\mathbf{B}\left|\mathbf{P},\Upsilon_{t}\right.\right]\mathbf{x}_{t-s} + \\
&\quad \sum_{s=0}^{t+M-1}\mathbf{w}'E\left[c_{s}(\boldsymbol{\Gamma})\boldsymbol{\Gamma}\left|\mathbf{P},\Upsilon_{t}\right.\right]\mathbf{f}_{t-s} + \sum_{s=1}^{t+M-1}a_{s}E\left(\mathbf{w}'\boldsymbol{\varepsilon}_{t-s}\left|\mathbf{P},\Upsilon_{t}\right.\right),
\end{aligned}
$$

and finally taking expectations conditional on the available aggregate information set (and noting that $\Omega_t \subset (\mathbf{P},\Upsilon_t)$)

$$
\begin{aligned}
E\left(\bar{y}_{wt}\left|\Omega_{t}\right.\right) &= \left(\mathbf{w}'\mathbf{y}_{-M}\right)E\left(a_{t+M}\left|\Omega_{t}\right.\right) + \sum_{s=0}^{t+M-1}\mathbf{w}'E\left[b_{s}(\boldsymbol{\beta})\mathbf{B}\mathbf{x}_{t-s}\left|\Omega_{t}\right.\right] \\
&\quad + \sum_{s=0}^{t+M-1}\mathbf{w}'E\left[c_{s}(\boldsymbol{\Gamma})\boldsymbol{\Gamma}\left|\Omega_{t}\right.\right]\mathbf{f}_{t-s} + \sum_{s=1}^{t+M-1}a_{s}E\left(\bar{\varepsilon}_{w,t-s}\left|\Omega_{t}\right.\right).
\end{aligned}
\tag{8}
$$

where $\bar{\varepsilon}_{wt} = \mathbf{w}'\boldsymbol{\varepsilon}_t$.

2.1 Aggregation of stationary micro relations with random coefficients

The optimal aggregate function derived in (8) is quite general and holds for any N, and does not require the underlying micro processes to be stationary. But its use in empirical applications is limited as it depends on unobserved initial state, $\mathbf{w}'\mathbf{y}_{-M}$, and the micro variables, $\mathbf{x}_t$. To derive empirically manageable aggregate functions in what follows we assume that the underlying processes are stationary and the micro parameters, β_i and γ_{ij}, are random draws from a common distribution. More specifically, we make the following assumptions:

ASSUMPTION 3 *The micro coefficients, β_i and γ_{ij}, are random draws from common distributions with finite moments such that*

$$E\left[b_s(\boldsymbol{\beta})\mathbf{B}\left|\Omega_t\right.\right] = b_s\mathbf{I}_N, \tag{9}$$

$$E\left[c_s(\boldsymbol{\Gamma})\boldsymbol{\Gamma}\left|\Omega_t\right.\right] = \boldsymbol{\tau}_N\mathbf{c}_s', \tag{10}$$

where $b_s(\boldsymbol{\beta})$ and $c_s(\boldsymbol{\Gamma})$ are defined in Assumption 1, $b_s = E\left[b_s(\boldsymbol{\beta})\beta_i\right]$, $\mathbf{c}_s = E\left[c_s(\boldsymbol{\Gamma})\boldsymbol{\gamma}_i\right]$, and $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones.

ASSUMPTION 4 *The eigenvalues of $\boldsymbol{\Phi}$, $\lambda_i(\boldsymbol{\Phi})$, are draws from a common distribution with support over the range $(-1, 1)$.*

Under Assumption 3, (8) simplifies to

$$\begin{aligned} E\left(\bar{y}_{wt}\left|\Omega_t\right.\right) &= \left(\mathbf{w}'\mathbf{y}_{-M}\right)E\left(a_{t+M}\left|\Omega_t\right.\right) + \sum_{s=0}^{t+M-1} b_s\bar{x}_{w,t-s} \\ &\quad + \sum_{s=0}^{t+M-1}\mathbf{c}_s'\mathbf{f}_{t-s} + \sum_{s=1}^{t+M-1} a_s E\left(\bar{\varepsilon}_{w,t-s}\left|\Omega_t\right.\right), \end{aligned}$$

where $\bar{x}_{wt} = \mathbf{w}'\mathbf{x}_t$, and $E\left(\bar{y}_{wt}\left|\Omega_t\right.\right)$ no longer depends on the individual specific regressors. Under the additional Assumption 4, and for M sufficiently large the initial states are also eliminated and we have

$$E\left(\bar{y}_{wt}\left|\Omega_t\right.\right) = \sum_{s=0}^{\infty} b_s\bar{x}_{w,t-s} + \sum_{s=0}^{\infty}\mathbf{c}_s'\mathbf{f}_{t-s} + \sum_{s=1}^{\infty} a_s\eta_{t-s}.$$

where $\eta_{t-s} = E\left(\bar{\varepsilon}_{w,t-s}\left|\Omega_t\right.\right)$. Note that $\sum_{s=1}^{\infty} a_s\eta_{t-s} = E\left[\sum_{s=1}^{\infty} a_s\bar{\varepsilon}_{w,t-s}\left|\Omega_t\right.\right]$. Using this result in (5) we obtain the optimal aggregate function

$$\bar{y}_{wt} = \sum_{s=0}^{\infty} b_s\bar{x}_{w,t-s} + \sum_{s=0}^{\infty}\mathbf{c}_s'\mathbf{f}_{t-s} + \sum_{s=1}^{\infty} a_s\eta_{t-s} + v_{wt}, \tag{11}$$

which holds for any finite N.

The dynamic properties of $\bar{y}_{wt}$ and its persistence to shocks depend on the decay rates of the distributed lag coefficients, $\{a_s\}, \{b_s\}$ and $\{\mathbf{c}_s\}$. If $|\lambda_i(\mathbf{\Phi})| < 1 - \epsilon$, for some strictly positive constant $\epsilon > 0$, then the distributed lagged coefficients, $\{a_s\}, \{b_s\}$ and $\{\mathbf{c}_s\}$ decay exponentially fast and the aggregate function will not exhibit long memory features. However, in the case where $\lambda_i(\mathbf{\Phi})'s$ are draws from distributions with supports covering -1 and/or 1, the rate of decay of the distributed lagged coefficients will be slower than exponential, typically the decay rate is given by $1/(1+s)$, and the resultant aggregate function will be subject to long memory effects. This result confirms Granger's conjecture in the case of large dimensional VAR models, and establishes sufficient conditions for its validity. Just to summarize, the conditions are as set out in Assumptions 1, 2, 3, and 4.

It is also worth noting that in general $\bar{y}_{wt}$ has an infinite order distributed lag representation even if the underlying micro relations have finite lag orders. This is an important consideration in empirical macro economic analysis where the macro variables under consideration are often constructed as aggregates of observations on a large number of micro units.

2.2 Limiting behavior of the optimal aggregate function

The aggregate function in (11) continues to hold even if $N \to \infty$, so long as the degree of cross-sectional dependence in the idiosyncratic errors, ε_{it}, $i = 1, 2, ..., N$, is sufficiently weak; otherwise there is no guarantee for the aggregation error, $\sum_{s=1}^{\infty} a_s \eta_{t-s} + v_{wt}$, to vanish as $N \to \infty$. To this end we introduce the following assumption that governs the degree of error cross-sectional dependence.

ASSUMPTION 5 *The idiosyncratic errors,* $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$ *in (1) are cross-sectionally weakly dependent in the sense that*

$$\left\|\mathbf{\Sigma}_{\varepsilon}\right\|_1 = \left\|\mathbf{\Sigma}_{\varepsilon}\right\|_{\infty} = O\left(N^{\alpha_{\varepsilon}}\right),$$

where $\mathbf{\Sigma}_{\varepsilon} = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$*, for some constant* $0 \leq \alpha_{\varepsilon} < 1$.

Remark 3 *Condition* $0 \leq \alpha_{\varepsilon} < 1$ *in Assumption 5 is sufficient and necessary for weak cross-sectional dependence of micro innovations. See Chudik, Pesaran, and Tosetti (2011). Following*

Bailey, Kapetanios, and Pesaran (2012), we shall refer to the constant α_ε as the exponent of cross-sectional dependence of the idiosyncratic shocks.

Since under Assumption 2 the errors, $\boldsymbol{\varepsilon}_t$ for all t are serially uncorrelated, we have

$$Var\left(\sum_{s=1}^{\infty} a_s \bar{\varepsilon}_{w,t-s}\right) = \sum_{s=1}^{\infty} a_s^2 Var\left(\bar{\varepsilon}_{w,t-s}\right) \leq \left(\sum_{s=1}^{\infty} a_s^2\right) \sup_t \left[Var(\bar{\varepsilon}_{wt})\right].$$

Furthermore

$$Var(\bar{\varepsilon}_{wt}) = \mathbf{w}'\boldsymbol{\Sigma}_\varepsilon \mathbf{w} \leq \|\mathbf{w}\|^2 \varrho\left(\boldsymbol{\Sigma}_\varepsilon\right),$$

and by Assumption 5, and the granularity conditions (2), we have[8]

$$\sup_t \left[Var(\bar{\varepsilon}_{wt})\right] = O\left(N^{\alpha_\varepsilon - 1}\right),$$

and $\sum_{s=1}^{\infty} a_s \bar{\varepsilon}_{w,t-s} \overset{q.m}{\rightarrow} 0$, so long as $\sum_{s=1}^{\infty} a_s^2 < K$, for some positive constant K.[9] Recall that under Assumption 5, $\alpha_\varepsilon < 1$, and $\sup_t \left[Var(\bar{\varepsilon}_{wt})\right] \rightarrow 0$, as $N \rightarrow \infty$. Also since $\sum_{s=1}^{\infty} a_s \eta_{t-s} = E\left(\sum_{s=1}^{\infty} a_s \bar{\varepsilon}_{w,t-s} \left|\Omega_t\right.\right)$, it follows that

$$\sum_{s=1}^{\infty} a_s \eta_{t-s} \overset{q.m}{\rightarrow} 0, \tag{12}$$

and hence for each t we have

$$\bar{y}_{wt} - \sum_{s=0}^{\infty} b_s \bar{x}_{w,t-s} - \sum_{s=0}^{\infty} \mathbf{c}_s' \mathbf{f}_{t-s} - v_{wt} \overset{q.m}{\rightarrow} 0, \text{ as } N \rightarrow \infty.$$

The limiting behavior of v_{wt}, as $N \rightarrow \infty$, depends on the nature of the processes generating x_{it}, $\mathbf{f}_t$, and ε_{it}, as well as the degree of cross-sectional dependence that arise from the non-zero off-diagonal elements of $\boldsymbol{\Phi}$. Sufficient conditions for $v_{wt} \overset{q.m}{\rightarrow} 0$ are not presented here due to space constraints, but can be found in Pesaran and Chudik (2011, Proposition 1). The key conditions for $v_{wt} \overset{q.m}{\rightarrow} 0$ are weak error cross-sectional dependence and sufficiently bounded dynamic interactions across the units. These conditions are satisfied, for example, if $\|\boldsymbol{\Sigma}_\varepsilon\| = \|E\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right)\| < K$, and $\sum_{s=1}^{\infty} E\left\|\boldsymbol{\Phi}^s\right\| \leq \sum_{s=1}^{\infty} E\left\|\boldsymbol{\Phi}\right\|^s < K$, for some finite positive constant, K. If on the other hand

[8] Note that $\varrho\left(\boldsymbol{\Sigma}_\varepsilon\right) \leq \left\|\boldsymbol{\Sigma}_\varepsilon\right\|_1 = O\left(N^{\alpha_\varepsilon}\right)$.

[9] A sufficient condition for $\sum_{s=1}^{\infty} a_s^2$ to be bounded is $|\lambda_i| < 1 - \epsilon$, where ϵ is a small strictly positive number.

$\sum_{s=1}^{\infty} E \left\| \mathbf{\Phi} \right\|^s$ is not bounded as $N \to \infty$, or $\boldsymbol{\varepsilon}_t$ is strongly cross-sectionally dependent, then the aggregation error v_{wt} does not necessarily converge to zero and could be sizeable.

3 Relationship between Micro and Macro Parameters

In this section we discuss the problem of identification of micro parameters, or some of their distributional features, from the aggregate function given by (11). Although it is not possible to recover all of the parameters of micro relations, there are a number of notable exceptions. An important example is the average long-run impact defined by,

$$\bar{\theta} = \frac{1}{N} \sum_{i=1}^{N} \theta_i = \frac{1}{N} \boldsymbol{\tau}_N' \boldsymbol{\theta} = \frac{1}{N} \boldsymbol{\tau}_N' \left(\mathbf{I}_N - \mathbf{\Phi} \right)^{-1} \boldsymbol{\beta}, \tag{13}$$

where $\boldsymbol{\theta} = \left(\mathbf{I}_N - \mathbf{\Phi} \right)^{-1} \boldsymbol{\beta} = \left(\boldsymbol{\beta} + \mathbf{\Phi}\boldsymbol{\beta} + \mathbf{\Phi}^2\boldsymbol{\beta} + ... \right)$ is the $N \times 1$ vector of individual long-run coefficients, and as before $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones. Suppose that Assumptions 3 and 4 are satisfied and denote the common mean of β_i by β. Using (9), we have $E\left(\mathbf{\Phi}^s \boldsymbol{\beta} \right) = E\left\{ E\left[b_s(\boldsymbol{\beta}) \mathbf{B} \left| \Omega_t \right. \right] \right\} = b_s \mathbf{I}_N$ for $s = 0, 1, ...$ Hence, the elements of $\boldsymbol{\theta}$ have a common mean, $E\left(\theta_i \right) = \theta = \sum_{\ell=0}^{\infty} b_s$, which does not depend on the elements of $\mathbf{P}$. If, in addition, the sequence of random variables θ_i is ergodic in mean, then for sufficiently large N, $\bar{\theta}$ is well approximated by its mean, $\sum_{\ell=0}^{\infty} b_s$, and the cross-sectional mean of the micro long-run effects can be estimated by the long-run coefficient of the associated optimal aggregate model. This result holds even if β_i and $\lambda_i\left(\mathbf{\Phi} \right)$ are not independently distributed, and irrespective of whether micro shocks contain a common factor.

Whether $\bar{\theta} \stackrel{p}{\to} \theta$ deserves a comment. A sufficient condition for $\bar{\theta}$ to converge to its mean (in probability) is given by

$$\left\| Var\left(\boldsymbol{\theta} \right) \right\| = O\left(N^{1-\epsilon} \right), \text{ for some } \epsilon > 0, \tag{14}$$

in which case $\left\| Var\left(\bar{\theta} \right) \right\| \leq N^{-1} \left\| Var\left(\boldsymbol{\theta} \right) \right\| = O\left(N^{-\epsilon} \right) \to 0$ as $N \to \infty$ and $\bar{\theta} \stackrel{q.m.}{\to} \theta$. Condition (14) need not always hold. This condition can be violated if there is a high degree of dependence of micro coefficients β_i across i, or if there is a dominant unit in the underlying model in which case the column norm of $\mathbf{\Phi}$ becomes unbounded in N.

The mean of β_i is straightforward to identify from the aggregate relation since $E\left(\beta_i \right) = b_0$. But further restrictions are needed for identification of $E[\lambda_i(\mathbf{\Phi})]$ from the aggregate model. Similarly to Pesaran (2003) and Lewbel (1994), independence of β_i and $\lambda_i(\mathbf{\Phi})$ would be sufficient for the identification of the moments of $\lambda_i(\mathbf{\Phi})$. Under the assumption that β_i and $\lambda_i(\mathbf{\Phi})$ are independently distributed, all moments of $\lambda_i(\mathbf{\Phi})$ can be identified by

$$E[\lambda_i^s(\mathbf{\Phi})] = \frac{b_s}{b_0}. \tag{15}$$

Another possibility is to adopt a parametric specification for the distribution of the micro coefficients and then identify the unknown parameters of the cross-sectional distribution of micro coefficients from the aggregate specification. For example, suppose β_i is independently distributed of $\lambda_i(\mathbf{\Phi})$, and $\lambda_i(\mathbf{\Phi})$ has a beta distribution over $(0,1)$,

$$f(\lambda) = \frac{\lambda^{p-1}\left(1-\lambda^{q-1}\right)}{B(p,q)},\ p>0,\ q>0,\ 0<\lambda<1.$$

Then as discussed in Robinson (1978) and Pesaran (2003), we have

$$p = \frac{b_1(b_1-b_2)}{b_2b_0-b_1^2},\ q = \frac{(b_0-b_1)(b_1-b_2)}{b_2b_0-b_1^2},$$

and $\theta = b_0(p+q-1)/(q-1)$. Another example is uniform distribution for $\lambda_i(\mathbf{\Phi})$ on interval $[\lambda_{\min}, \lambda_{\max}]$, $\lambda_{\min} > -1$, $\lambda_{\max} < 1$. Equation (15) can be solved to obtain (see Robinson, 1978),

$$\lambda_{\min} = \frac{b_1 - \sqrt{3\left(b_0b_2-b_1^2\right)}}{b_0},\ \text{and}\ \lambda_{\max} = \frac{b_1 + \sqrt{3\left(b_0b_2-b_1^2\right)}}{b_0}.$$

4 Impulse Responses of Macro and Aggregated Idiosyncratic Shocks

For the analysis of impulse responses we assume that the common factors in (1) follow the VAR(1) model

$$\mathbf{f}_t = \mathbf{\Psi}\mathbf{f}_{t-1} + \mathbf{v}_t, \tag{16}$$

where $\mathbf{\Psi}$ is an $m \times m$ matrix of coefficients, and $\mathbf{v}_t = (v_{1t}, v_{2t}, ..., v_{mt})'$ is the $m \times 1$ vector of macro shocks. To simplify the analysis we also set $\boldsymbol{\beta} = \mathbf{0}$, and write the micro relations as

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{y}_{t-1} + \mathbf{u}_t, \ \ \mathbf{u}_t = \mathbf{\Gamma}\mathbf{f}_t + \boldsymbol{\varepsilon}_t. \tag{17}$$

Including the exogenous variables, $\mathbf{x}_t$, in the model is relatively straightforward and does not affect the impulse responses of the shocks to macro factors, $\mathbf{v}_t$, or the idiosyncratic errors. The lag orders of the VAR models in (16) and (17) are set to unity only for expositional convenience.

We make the following additional assumption.

ASSUMPTION 6 *The $m \times 1$ macro shocks, $\mathbf{v}_t$, are distributed independently of $\boldsymbol{\varepsilon}_{t'}$, for all t and t'. They are also serially uncorrelated, with zero means, and a diagonal variance matrix, $\mathbf{\Sigma}_v = Diag(\sigma^2_{v_1}, \sigma^2_{v_2}, ..., \sigma^2_{v_m})$, where $0 < \sigma^2_{v_j} < \infty$, for all j.*

We are interested in effects of two types of shocks on the aggregate variable $\bar{y}_{wt} = \mathbf{w}'\mathbf{y}_t$, namely the composite macro shock, defined by $\bar{v}_{\bar{\gamma}t} = \mathbf{w}'\mathbf{\Gamma}\mathbf{v}_t = \bar{\boldsymbol{\gamma}}_w'\mathbf{v}_t$, and the aggregated idiosyncratic shock defined by $\bar{\varepsilon}_{wt} = \mathbf{w}'\boldsymbol{\varepsilon}_t$. We shall also consider the combined aggregate shock defined by

$$\bar{\xi}_{wt} = \mathbf{w}'\mathbf{\Gamma}\mathbf{v}_t + \mathbf{w}'\boldsymbol{\varepsilon}_t = \bar{\boldsymbol{\gamma}}_w'\mathbf{v}_t + \bar{\varepsilon}_{wt} = \bar{v}_{\bar{\gamma}t} + \bar{\varepsilon}_{wt},$$

and investigate the time profiles of the effects of these shocks on $\bar{y}_{w,t+s}$, for $s = 0, 1, ...$ The combined aggregate shock, $\bar{\xi}_{wt}$, can be identified from the aggregate equation in $\bar{y}_{wt}$, so long as an $AR(\infty)$ approximation for $\bar{y}_{wt}$ exists. Since by assumption $\boldsymbol{\varepsilon}_t$ and $\mathbf{v}_t$ are distributed independently then

$$Var\left(\bar{\xi}_{wt}\right) = \bar{\boldsymbol{\gamma}}_w'\mathbf{\Sigma}_v\bar{\boldsymbol{\gamma}}_w + \mathbf{w}'\mathbf{\Sigma}_\varepsilon\mathbf{w} = \sigma^2_{\bar{v}} + \sigma^2_{\bar{\varepsilon}} = \sigma^2_{\bar{\xi}},$$

where $\sigma^2_{\bar{v}} = \bar{\boldsymbol{\gamma}}_w'\mathbf{\Sigma}_v\bar{\boldsymbol{\gamma}}_w$ is the variance of the composite macro shock, and $\sigma^2_{\bar{\varepsilon}} = \mathbf{w}'\mathbf{\Sigma}_\varepsilon\mathbf{w}$ is the variance of the aggregated idiosyncratic shock. Note that when $\mathbf{f}_t$ is unobserved, the separate effects of composite macro shock, $\bar{v}_{\bar{\gamma}t}$, and aggregated idiosyncratic shock, $\bar{\varepsilon}_{wt}$, can only be identified under the disaggregated model, (17). Only the effects of $\bar{\xi}_{wt}$ on $\bar{y}_{w,t+h}$ can be identified if the aggregate specification is used.

Using the disaggregate model we obtain the following generalized impulse response functions

(GIRFs)

$$g_{\bar{\varepsilon}}\left(s\right)=E\left(\bar{y}_{w,t+s}\left|\bar{\varepsilon}_{wt}=\sigma_{\bar{\varepsilon}},\mathcal{I}_{t-1}\right.\right)-E\left(\bar{y}_{w,t+s}\left|\mathcal{I}_{t-1}\right.\right)=\frac{\mathbf{w}'\mathbf{\Phi}^{s}\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}}, \tag{18}$$

$$g_{v_j}\left(s\right)=E\left(\bar{y}_{w,t+s}\left|v_{jt}=\sigma_{v_j},\mathcal{I}_{t-1}\right.\right)-E\left(\bar{y}_{w,t+s}\left|\mathcal{I}_{t-1}\right.\right)=\frac{\mathbf{w}'\mathbf{C}_{s}\mathbf{\Sigma}_{v}\mathbf{e}_{j,v}}{\sqrt{\mathbf{e}_{j,v}'\mathbf{\Sigma}_{v}\mathbf{e}_{j,v}}}, \tag{19}$$

for $j=1,2,...,m$, where $\mathcal{I}_t$ is an information set consisting of all current and past available information at time t,

$$\mathbf{C}_s=\sum_{j=0}^{s}\mathbf{\Phi}^{s-j}\mathbf{\Gamma}\mathbf{\Psi}^{j}, \tag{20}$$

and $\mathbf{e}_{j,v}$ is an $m\times 1$ selection vector that selects the j-th element of $\mathbf{v}_t$. Hence

$$g_{\bar{v}}\left(s\right)=E\left(\bar{y}_{w,t+s}\left|\bar{v}_{\bar{\gamma}t}=\sigma_{\bar{v}},\mathcal{I}_{t-1}\right.\right)-E\left(\bar{y}_{w,t+s}\left|\mathcal{I}_{t-1}\right.\right)=\frac{\mathbf{w}'\mathbf{C}_{s}\mathbf{\Sigma}_{v}\bar{\boldsymbol{\gamma}}_{w}}{\sqrt{\bar{\boldsymbol{\gamma}}_{w}'\mathbf{\Sigma}_{v}\bar{\boldsymbol{\gamma}}_{w}}}. \tag{21}$$

Finally,

$$\begin{aligned} g_{\bar{\xi}}\left(s\right) &= E\left(\bar{y}_{w,t+s}\left|\bar{\xi}_{wt}=\sigma_{\bar{\xi}},\mathcal{I}_{t-1}\right.\right)-E\left(\bar{y}_{w,t+s}\left|\mathcal{I}_{t-1}\right.\right) \\ &= \frac{\mathbf{w}'\mathbf{C}_{s}\mathbf{\Sigma}_{v}\bar{\boldsymbol{\gamma}}_{w}+\mathbf{w}'\mathbf{\Phi}^{s}\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}{\sqrt{\bar{\boldsymbol{\gamma}}_{w}'\mathbf{\Sigma}_{v}\bar{\boldsymbol{\gamma}}_{w}+\mathbf{w}'\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}}. \end{aligned} \tag{22}$$

Note that $\mathbf{C}_0=\mathbf{\Gamma}$, and we have $g_{\bar{\xi}}\left(0\right)=\sqrt{\bar{\boldsymbol{\gamma}}_{w}'\mathbf{\Sigma}_{v}\bar{\boldsymbol{\gamma}}_{w}+\mathbf{w}'\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}=\sigma_{\bar{\xi}}$, as to be expected.

When N is finite, both, the combined aggregated idiosyncratic shock ($\bar{\varepsilon}_{wt}$) and the composite macro shock ($\bar{v}_{\bar{\gamma}t}$) are important; and the impulse response of the combined aggregate shock on the aggregate variable, given by (22), is a linear combination of $g_{\bar{\varepsilon}}\left(s\right)$ and $g_{\bar{v}}\left(s\right)$, namely

$$g_{\bar{\xi}}\left(s\right)=\omega_{\bar{v}}g_{\bar{v}}\left(s\right)+\omega_{\bar{\varepsilon}}g_{\bar{\varepsilon}}\left(s\right), \tag{23}$$

where $\omega_{\bar{\varepsilon}}=\sigma_{\bar{\varepsilon}}/\sigma_{\bar{\xi}}$, $\omega_{\bar{v}}=\sigma_{\bar{v}}/\sigma_{\bar{\xi}}$, and $\omega_{\bar{\varepsilon}}^2+\omega_{\bar{v}}^2=1$.

When $N\to\infty$, it is not necessarily true that both shocks are important, and $\lim_{N\to\infty}\sigma_{\bar{v}}^2/\sigma_{\bar{\xi}}^2$, if it exists, could be any value on the unit interval, including one or zero. We investigate the case when $N\to\infty$ below. First, we consider the impulse responses of the aggregated idiosyncratic shock on the aggregate variable in the next proposition.

Proposition 1 *Suppose that* $\|\mathbf{\Sigma}_{\varepsilon}\|_1 = O\left(N^{\alpha_{\varepsilon}}\right)$, *for some constant* $0 \leq \alpha_{\varepsilon} < 1$, $E\|\mathbf{\Phi}\|$ *is bounded in* N, *where* $\|\mathbf{\Phi}\| = \varrho\left(\mathbf{\Phi}\mathbf{\Phi}'\right)$, *and the aggregation weights satisfy* $\|\mathbf{w}\| = O\left(N^{-1/2}\right)$. *Then, for any given* $s = 0, 1, 2, ...,$ *we have*

$$E\left|g_{\bar{\varepsilon}}(s)\right| = O\left(N^{(\alpha_{\varepsilon}-1)/2}\right). \tag{24}$$

For a proof see the Appendix.

The aggregated idiosyncratic shock and its corresponding impulse response function vanishes as $N \to \infty$ at the rate which depends on the degree of cross-sectional dependence of idiosyncratic shocks. This rate could be very slow; and if the condition $\|\mathbf{w}\| = O\left(N^{-1/2}\right)$ is not satisfied, then the rate of convergence would depend also on the degree of granularity of the weights, w_i. The composite macro shock and its corresponding impulse-response function, on the other hand, does not necessarily vanish as $N \to \infty$, depending on the factor loadings. For the ease of exposition, we focus on the following model for factor loadings:

$$\begin{aligned} \boldsymbol{\gamma}_i &= \varkappa_i, \text{ for } i = 1, 2, ..., \left[N^{\alpha_\gamma}\right], \\ \boldsymbol{\gamma}_i &= \mathbf{0}, \text{ for } i = \left[N^{\alpha_\gamma}\right] + 1, ..., N, \end{aligned}$$

where $\varkappa_i \sim IID\left(\boldsymbol{\mu}_{\varkappa}, \mathbf{\Sigma}_{\varkappa}\right)$, $\left[N^{\alpha_\gamma}\right]$ denotes the integer part of N^{α_γ}, constant α_γ is the exponent of cross-sectional dependence of y_{it} due to factors, see Bailey, Kapetanios, and Pesaran (2012), and $0 < \alpha_\gamma \leq 1$. Note that the aggregated factor loadings satisfy $\text{plim}_{N\to\infty} N^{1-\alpha_\gamma}\bar{\boldsymbol{\gamma}}_w = \boldsymbol{\mu}_{\varkappa}$, and the variance of the composite macro shock, $\sigma^2_{\bar{v}} = \bar{\boldsymbol{\gamma}}'_w \mathbf{\Sigma}_v \bar{\boldsymbol{\gamma}}_w$, satisfies

$$\operatorname*{plim}_{N\to\infty} N^{2(1-\alpha_\gamma)}\sigma^2_{\bar{v}} = \boldsymbol{\mu}'_{\varkappa}\mathbf{\Sigma}_v\boldsymbol{\mu}_{\varkappa}. \tag{25}$$

The variance of the aggregated idiosyncratic shock, on the other hand, is bounded by

$$\sigma^2_{\bar{\varepsilon}} = \mathbf{w}'\mathbf{\Sigma}_{\varepsilon}\mathbf{w} \leq O\left(N^{\alpha_{\varepsilon}-1}\right). \tag{26}$$

It follows from (25)-(26) that only when $\alpha_\gamma > (\alpha_{\varepsilon} + 1)/2$ and $\boldsymbol{\mu}_{\varkappa} \neq \mathbf{0}$, the variance of the composite macro shock dominates, in which case $\text{plim}_{N\to\infty}\sigma^2_{\bar{v}}/\sigma^2_{\bar{\xi}} = 1$, and the combined aggregate shock, $\bar{\xi}_{wt} = \bar{v}_{\bar{\gamma}t} + \bar{\varepsilon}_{wt}$ converges in quadratic mean to the composite macro shock as $N \to \infty$. It is then possible to scale $g_{\bar{\xi}}(s)$ by $\sigma_{\bar{v}}^{-1}$, and for any given $s = 0, 1, 2, ...$, we can obtain

$$\underset{N\to\infty}{\text{plim}} \left[\sigma_{\bar{v}}^{-1} g_{\bar{\xi}}(s)\right] = \underset{N\to\infty}{\text{plim}} \left[\sigma_{\bar{v}}^{-1} g_{\bar{v}}(s)\right].$$

When $\alpha_\gamma \leq (\alpha_\varepsilon + 1)/2$ and/or $\boldsymbol{\mu}_\varkappa = \mathbf{0}$, the macro shocks do not necessarily dominate the aggregated idiosyncratic shock (as $N \to \infty$), and the latter shock can be as important as macro shocks, or even dominate the macro shocks as $N \to \infty$.

5 A Monte Carlo Investigation

We consider a first-order VAR model with a single unobserved factor to examine the response of $\bar{y}_t = N^{-1}\sum_{i=1}^N y_{it}$, to the combined aggregate shock, $\bar{\xi}_t = \bar{\gamma}v_t + \bar{\varepsilon}_t$, where $\bar{\gamma} = N^{-1}\sum_{i=1}^N \gamma_i$ and $\bar{\varepsilon}_t = N^{-1}\sum_{i=1}^N \varepsilon_{it}$. As before, we decompose the effects into the contribution due to a macro shock, v_t, and the aggregated idiosyncratic shock, $\bar{\varepsilon}_t$. Using (23), we have

$$g_{\bar{\xi}}^d(s) = m_v^d(s) + m_{\bar{\varepsilon}}^d(s), \tag{27}$$

where $m_v^d(s) = \omega_v g_v^d(s)$, and $m_{\bar{\varepsilon}}^d(s) = \omega_{\bar{\varepsilon}} g_{\bar{\varepsilon}}^d(s)$ are the respective contributions of the macro and aggregated idiosyncratic shocks, and the weights ω_v and $\omega_{\bar{\varepsilon}}$ are defined below (23).

Aggregation weights are set equal to N^{-1} in all simulations. The subscript d is introduced to highlight the fact that these impulse responses are based on the disaggregate model. We know from the theoretical results that in cases where the optimal aggregate function exists, the common factor is strong (i.e. $\alpha_\gamma = 1$), and the idiosyncratic shocks are weakly correlated (i.e. $\alpha_\varepsilon = 0$), then $g_{\bar{\xi}}^d(s)$ converges to $g_v^d(s)$ as $N \to \infty$, for all s. But it would be of interest to investigate the contributions of macro and aggregated idiosyncratic shocks to the aggregate impulse response functions, when N is finite, as well as when α_γ takes intermediate values between 0 and 1.

We also use the Monte Carlo experiments to investigate persistence properties of the aggregate variable. The degree and sources of persistence in macro variables, such as consumer price inflation, output and real exchange rates, have been of considerable interest in economics. We know from the theoretical results that there are two key components affecting the persistence of the aggregate variables: distribution of the eigenvalues of lagged micro coefficients matrix, $\mathbf{\Phi}$, which we refer to as dynamic heterogeneity, and the persistence of common factor itself, which we refer to as the factor persistence. Our aim is to investigate how these two sources of persistence combine and get amplified in the process of aggregation.

Finally, a related issue of practical significance is the effects of estimation uncertainty on the above comparisons. To this end, we estimate disaggregated models using observations on individual micro units, y_{it}, as well as an aggregate model that only make use of the aggregate observations, $\bar{y}_t$. We denote the estimated impulse responses of the combined aggregate shock on the aggregate variable by $\hat{g}^d_{\bar{\xi}}(s)$ when based on the disaggregate model, and by $\hat{g}^a_{\bar{\xi}}(s)$ when based on an aggregate autoregressive model fitted to $\bar{y}_t$. It is important to recall that the effects of macro and aggregated idiosyncratic shocks cannot be identified from the aggregate model.

The remainder of this section is organized as follows. The next subsection outlines the Monte Carlo design. Subsection 5.2 describes the estimation of $g^d_{\bar{\xi}}(s)$ using aggregate and disaggregate data, and the last subsection discusses the main findings.

5.1 Monte Carlo design

To allow for neighborhood effects as well as an unobserved common factor we used the following data generating process (DGP)

$$y_{it} = \lambda_i y_{i,t-1} + \gamma_i f_t + \varepsilon_{it}, \text{ for } i = 1, \tag{28}$$

and

$$y_{it} = d_i y_{i-1,t-1} + \lambda_i y_{i,t-1} + \gamma_i f_t + \varepsilon_{it}, \text{ for } i = 2, 3, ..., N, \tag{29}$$

where each unit, except the first, has one left neighbor ($y_{i-1,t-1}$). The micro model given by (28)-(29) can be written conveniently in vector notations as

$$\mathbf{y}_t = \mathbf{\Phi} \mathbf{y}_{t-1} + \boldsymbol{\gamma} f_t + \boldsymbol{\varepsilon}_t, \tag{30}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})'$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_N)'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, and

$$\mathbf{\Phi} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ d_2 & \lambda_2 & 0 & \cdots & 0 \\ 0 & d_3 & \lambda_3 & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & & d_N & \lambda_N \end{pmatrix}.$$

The autoregressive micro coefficients, λ_i, are generated as $\lambda_i \sim IIDU\,(0, \lambda_{\max})$, for $i = 1, 2, ..., N$, with $\lambda_{\max} = 0.9$ or 1. Recall $\bar{y}_t$ will exhibit long memory features when $\lambda_{max} = 1$, but not when $\lambda_{\max} = 0.9$. The neighborhood coefficients, d_i, are generated as $IIDU\,(0, 1 - \lambda_i)$, for $i = 2, 3, ..., N$, to ensure bounded variances as $N \to \infty$. Specifically, $\|\mathbf{\Phi}\|_\infty \leq \max_i \{|\lambda_i| + |d_i|\} < 1$, see Chudik and Pesaran (2011).

The idiosyncratic errors, $\boldsymbol{\varepsilon}_t$, are generated according to the following spatial autoregressive process,

$$\boldsymbol{\varepsilon}_t = \delta \mathbf{S} \boldsymbol{\varepsilon}_t + \boldsymbol{\varsigma}_t,\ 0 < \delta < 1,$$

where $\boldsymbol{\varsigma}_t = (\varsigma_{1t}, \varsigma_{2t}, ..., \varsigma_{Nt})'$, $\boldsymbol{\varsigma}_t \sim IIDN\left(\mathbf{0}, \sigma_\varsigma^2 \mathbf{I}_N\right)$, and the $N \times N$ dimensional spatial weights matrix $\mathbf{S}$ is given by

$$\mathbf{S} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

To ensure that the idiosyncratic errors are weakly correlated, the spatial autoregressive parameter, δ, must lie in the range $[0, 1)$. We set $\delta = 0.4$. The variance σ_ς^2 is set equal to $N/(\boldsymbol{\tau}_N' \mathbf{R}\mathbf{R}' \boldsymbol{\tau}_N)$, where $\boldsymbol{\tau}_N = (1, 1, ..., 1)'$ and $\mathbf{R} = (\mathbf{I}_N - \delta \mathbf{S})^{-1}$, so that $Var\,(\bar{\varepsilon}_t) = N^{-1}$.

The common factor, f_t, is generated as

$$f_t = \psi f_{t-1} + v_t,\ v_t \sim IIDN\left(0, 1-\psi^2\right),\ |\psi| < 1,$$

for $t = -49, -48, ..., 1, 2, ..., T$, with $f_{-50} = 0$. We consider three values for $\psi = 0, 0.5$ and 0.8. By construction, $Var\left(f_t\right) = 1$.

Finally, the factor loadings are generated as

$$\begin{aligned}\gamma_i &= \varkappa_i, \text{ for } i = 1, 2, ..., \left[N^{\alpha_\gamma}\right], \\ \gamma_i &= 0, \text{ for } i = \left[N^{\alpha_\gamma}\right] + 1, \left[N^{\alpha_\gamma}\right] + 2, ..., N,\end{aligned}$$

where $\left[N^{\alpha_\gamma}\right]$ denotes the integer part of N^{α_γ}, $0 < \alpha_\gamma \leq 1$ is the exponent of cross-sectional dependence of y_{it} due to the common factor, see Bailey, Kapetanios, and Pesaran (2012), and $\varkappa_i \sim IIDN\left(1, 0.5^2\right)$. The unobserved common factor therefore affects a fraction $\left[N^{\alpha_\gamma}\right]/N$ of the units, with this fraction tending to zero if $\alpha_\gamma < 1$. It is easily seen that $\bar{\gamma} = N^{-1}\sum_{i=1}^{N}\gamma_i = O\left(N^{\alpha_\gamma}\right)$. We consider four values for $\alpha_\gamma \in \{0.25, 0.5, 0.75, 1\}$, representing different degrees of cross-sectional dependence due to the common factor. Note that for $\alpha_\gamma = 1$, we have $p\lim_{N\to\infty}\bar{\gamma} = 1$, whereas $p\lim_{N\to\infty}\bar{\gamma} = 0$ for $\alpha_\gamma < 1$. Note also that $\lim_{N\to\infty} NVar\left(\bar{\gamma}f_t\right) = 1$ for $\alpha_\gamma = 0.5$, in which case we would expect the macro shock and the aggregated idiosyncratic shock to be of equal importance for $g_{\bar{\xi}}^{d}(s)$.

5.2 Estimation of $g_{\bar{\xi}}(s)$ using aggregate and disaggregate data

The estimate of $g_{\bar{\xi}}(s)$ based on the aggregate data, which we denote by $\hat{g}_{\bar{\xi}}^{a}(s)$, is straightforward to compute and can be based on the following autoregression, (intercepts are included in all regressions below but not shown)

$$\bar{y}_t = \sum_{\ell=1}^{p_a}\pi_\ell \bar{y}_{t-\ell} + \zeta_{at}.$$

To estimate $g_{\bar{\xi}}(s)$ using disaggregated data is much more complicated and requires estimates of the micro coefficients. In terms of the micro parameters, using (22), we have

$$\begin{aligned} g_{\bar{\xi}}^{d}(s) &= E\left(\bar{y}_{w,t+s}\left|\bar{\xi}_{wt}=\sigma_{\bar{\xi}},\mathcal{I}_{t-1}\right.\right)-E\left(\bar{y}_{w,t+s}\left|\mathcal{I}_{t-1}\right.\right) \\ &= \sum_{\ell=0}^{s}\mathbf{\Phi}^{\ell}\left[E\left(\mathbf{u}_{t+s-\ell}\left|\bar{\xi}_{wt}=\sigma_{\bar{\xi}},\mathcal{I}_{t-1}\right.\right)-E\left(\mathbf{u}_{t+s-\ell}\left|\mathcal{I}_{t-1}\right.\right)\right]. \end{aligned} \tag{31}$$

Following Chudik and Pesaran (2011), we first estimate the nonzero elements of $\mathbf{\Phi}$, namely λ_i and d_i, using the cross-sectional augmented least squares regressions,

$$y_{it}=\lambda_i y_{i,t-1}+d_i y_{i-1,t-1}+h_i\left(L,p_{hi}\right)\bar{y}_t+\zeta_{it},\text{ for } i=2,3,...,N, \tag{32}$$

where $h_i\left(L,p_i\right)=\sum_{\ell=0}^{p_{hi}}h_{i\ell}L^{\ell}$, and p_{hi} is the lag order. The equation for the first micro unit is the same except that it does not feature any neighborhood effects.[10] These estimates are denoted by $\hat{\lambda}_i$ and $\hat{d}_i$, and an estimate of u_{it} is computed as

$$\hat{u}_{it} = y_{it}-\hat{\lambda}_i y_{i,t-1},\text{ for } i=1,\text{ and} \tag{33}$$

$$\hat{u}_{it} = y_{it}-\hat{\lambda}_i y_{i,t-1}-\hat{d}_i y_{i-1,t-1},\text{ for } i=2,3,...,N. \tag{34}$$

To obtain an estimate of $\xi_{it}=\gamma_i v_t+\varepsilon_{it}$, we fit the following conditional models

$$\widehat{u}_{it}=r_i\widehat{\bar{u}}_t+\epsilon_{it},\text{ for } i=1,2,...,N, \tag{35}$$

where $\widehat{\bar{u}}_t=N^{-1}\sum_{i=1}^{N}\hat{u}_{it}$; and the following marginal model,

$$\widehat{\bar{u}}_t=\psi_{\bar{u}}\widehat{\bar{u}}_{t-1}+\vartheta_t. \tag{36}$$

An estimate of ξ_{it} is computed as $\hat{\xi}_{it}=\widehat{u}_{it}-\hat{r}_i\hat{\psi}_{\bar{u}}\widehat{\bar{u}}_{t-1}$, for $i=1,2,...,N$, where $\hat{r}_i$ and $\hat{\psi}_{\bar{u}}$ are the estimates of r_i and $\psi_{\bar{u}}$, respectively. When $\alpha_\gamma=1$, $\hat{\psi}_{\bar{u}}$ is a consistent estimator (as $N,T\overset{j}{\to}\infty$) of the autoregressive parameter ψ that characterizes the persistence of the factor, $\hat{r}_i$ is a consistent estimator of the scaled factor loading, $\gamma_i/\bar{\gamma}$, and the regression residuals from (36), denoted by $\hat{\vartheta}_t$,

[10] Chudik and Pesaran (2011) show that if $\left\|\mathbf{\Phi}\right\|_{\infty}<1$, these augmented least squares estimates of the micro lagged coefficients are consistent and asymptotically normal when $\alpha_\gamma=1$ (as $N,T\overset{j}{\to}\infty$), and also when there is no factor, i.e. $\gamma=\mathbf{0}$.

are consistent estimates of the macro shock, v_t. But, when $\boldsymbol{\gamma} = \mathbf{0}$, $\bar{u}_t = N^{-1}\sum_{i=1}^{N} u_{it}$ is serially uncorrelated and $\hat{\psi}_{\bar{u}} \xrightarrow{p} 0$ as $N, T \xrightarrow{j} \infty$.

To compute the remaining terms in (31), we note that for $s = \ell = 0$, $E\left(\mathbf{u}_t \left|\bar{\xi}_{wt} = \hat{\sigma}_{\bar{\xi}}, \mathcal{I}_{t-1}\right.\right) - E\left(\mathbf{u}_t \left|\mathcal{I}_{t-1}\right.\right) = E\left(\boldsymbol{\xi}_t \left|\bar{\xi}_{wt} = \hat{\sigma}_{\bar{\xi}}, \mathcal{I}_{t-1}\right.\right)$ can be consistently estimated by $\hat{\boldsymbol{\Sigma}}_{\xi}\mathbf{w}/\hat{\sigma}_{\bar{\xi}}$, where $\hat{\sigma}_{\bar{\xi}} = \left(\mathbf{w}'\hat{\boldsymbol{\Sigma}}_{\xi}\mathbf{w}\right)^{1/2}$, $\hat{\boldsymbol{\Sigma}}_{\xi} = T^{-1}\sum_{t=p_h+1}^{T}\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t'$, $\hat{\boldsymbol{\xi}}_t = \left(\hat{\xi}_{1t}, \hat{\xi}_{2t}, ..., \hat{\xi}_{Nt}\right)'$, and $p_h = \max_i p_{hi}$. Similarly, for $s - \ell > 0$, $E\left(\mathbf{u}_{t+s-\ell} \left|\bar{\xi}_{wt} = \hat{\sigma}_{\bar{\xi}}, \mathcal{I}_{t-1}\right.\right) - E\left(\mathbf{u}_{t+s-\ell} \left|\mathcal{I}_{t-1}\right.\right)$ can be consistently estimated by $\hat{\psi}_u^{s-\ell}\hat{\sigma}_{\vartheta}^2\hat{\mathbf{r}}/\left(\mathbf{w}'\hat{\boldsymbol{\Sigma}}_{\xi}\mathbf{w}\right)^{1/2}$, where $\hat{\mathbf{r}} = (\hat{r}_1, \hat{r}_2, ..., \hat{r}_N)'$, and $\hat{\sigma}_{\vartheta}^2 = T^{-1}\sum_{t=p_h+1}^{T}\hat{\vartheta}_t^2$. All lag orders are selected by AIC with the maximum lag order set to $[T^{1/2}]$.

5.3 Monte Carlo Results

Figure 1 plots the relative contributions of macro and aggregated idiosyncratic shocks to the GIRF of the aggregate variable for the sample of $N = 200$ micro units. See (27). There are four panels, corresponding to different choices of cross-sectional exponents, α_{γ}, with the plots on the left of each panel relating to $\lambda_{\max} = 0.9$ and the ones on the right to $\lambda_{\max} = 1$. As expected, when $\alpha_{\gamma} = 0.25$ the macro shock is not 'strong enough' and the aggregated idiosyncratic shock dominates. When $\alpha_{\gamma} = 0.5$ (Panel B), the macro shock is equally important as the aggregated idiosyncratic shock. As α_{γ} is increased to 0.75 (Panel C), the aggregated idiosyncratic shock starts to play only a minor role; and when $\alpha_{\gamma} = 1$ (Panel D), the macro shock completely dominates the aggregate relationship. Similar results are obtained for N as small as 25 (not reported). Whether the support of the distribution of the eigenvalues λ_i covers unity or not does not seem to make any difference to the relative importance of the macro shock. Table 1 reports the weights ω_v and $\omega_{\bar{\varepsilon}}$ for different values of N, and complements what can be seen from the plots in Figure 1. Note that these weights do not depend on the choice of $\lambda_{\max}$ and by constructions $\omega_v^2 + \omega_{\bar{\varepsilon}}^2 = 1$. We see in Table 1 that for $\alpha_{\gamma} = 1$, ω_v is very close to unity for all values of N considered, and $g_{\xi}^d(s)$ is mainly explained by the macro shock, regardless of the shape of the impulse response functions.

Next we examine how dynamic heterogeneity and factor persistence affect the persistence of the aggregate variable. Figure 2 plots the GIRF of the combined aggregate shock on the aggregate variable, $g_{\bar{\varepsilon}}^d(s)$, for $N = 200$ and different values of $\lambda_{\max}$ and ψ, that control the dynamic heterogeneity and the persistence of the factor, respectively. Similarly to Figure 1, the plot on the left relates to $\lambda_{\max} = 0.9$ and the one on the right to $\lambda_{\max} = 1$. It is interesting that $g^d_{\bar{\xi}}(s)$ looks very different when we allow for serial correlation in the common factor. Even for a moderate value of ψ, say 0.5, the factor contributes significantly to the overall persistence of the aggregate. In contrast, the effects of long memory on persistence (comparing the plots on the left and the right of the panels in Figure 2), are rather modest. Common factor persistence tends to become accentuated by the individual-specific dynamics.

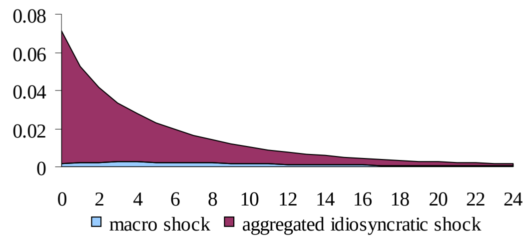
Finally, we consider the estimates of $g_{\bar{\xi}}(s)$ based on the disaggregate and the aggregate models, namely $\hat{g}^d_{\bar{\xi}}(s)$ and $\hat{g}^a_{\bar{\xi}}(s)$. Table 2 reports the root mean square error (RMSE×100) of these estimates averaged over horizons $s = 0$ to 12 and $s = 13$ to 24, for the parameter values $\alpha_\gamma = 0.5$, 1, and $\psi = 0.5$, using 2000 Monte Carlo replications.[11] The estimator based on the disaggregate model, $\hat{g}^d_{\bar{\xi}}(s)$, performs about $50 - 200\%$ better than its counterpart based on the aggregate model. The difference between the two estimators is slightly smaller when $\alpha_\gamma = 0.5$. As to be expected, an increase in the time dimension improves the precision of the estimates considerably. Also, $\hat{g}^d_{\bar{\xi}}(s)$ improves with an increase in N, whereas the RMSE of $\hat{g}^a_{\bar{\xi}}(s)$ is little affected by increasing N when $\alpha_\gamma = 1$, but improves with N when $\alpha_\gamma = 0.5$.

[11] The bias statistics are not reported due to space constraint.

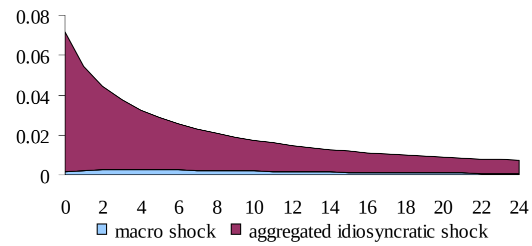
Figure 1: Contribution of the macro and aggregated idiosyncratic shocks to GIRF of one unit (1 s.e.) combined aggregate shock on the aggregate variable; $N = 200$.

Panel A. Experiments with $\alpha_\gamma = 0.25$.

$\lambda_{\max} = 0.9$

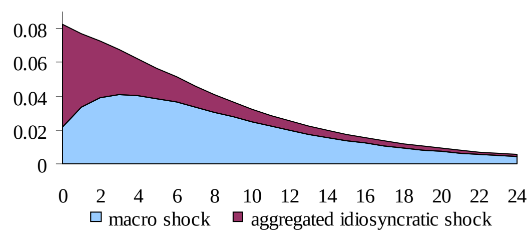


$\lambda_{\max} = 1$

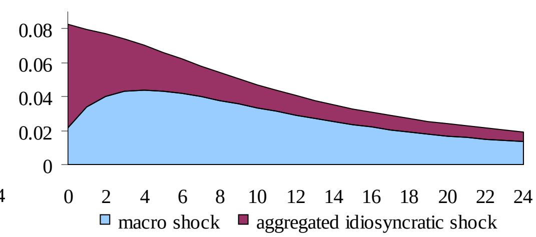


Panel B. Experiments with $\alpha_\gamma = 0.5$.

$\lambda_{\max} = 0.9$

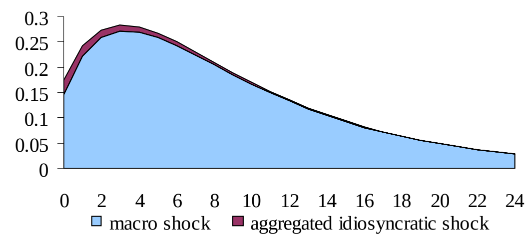


$\lambda_{\max} = 1$

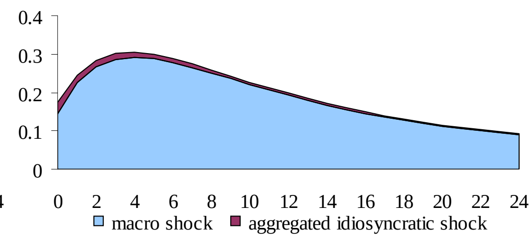


Panel C. Experiments with $\alpha_\gamma = 0.75$.

$\lambda_{\max} = 0.9$

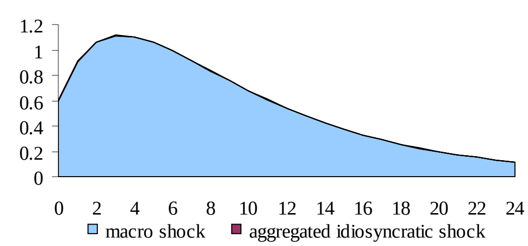


$\lambda_{\max} = 1$



Panel D. Experiments with $\alpha_\gamma = 1$.

$\lambda_{\max} = 0.9$



$\lambda_{\max} = 1$

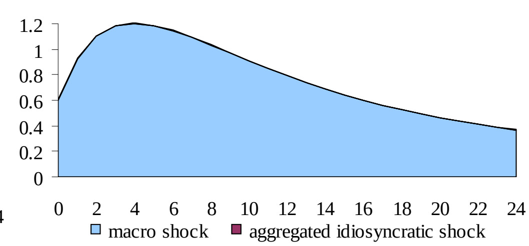
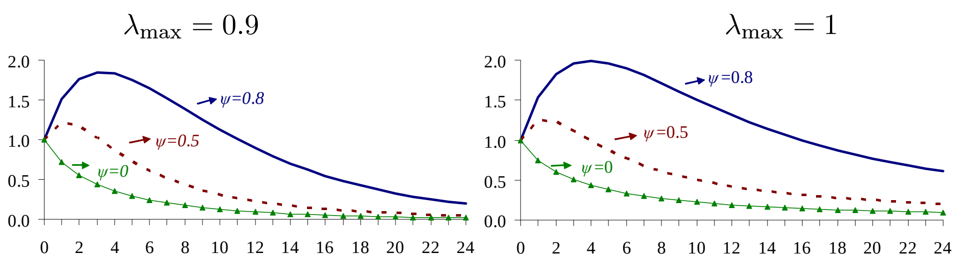


Table 1: Weights ω_v and $\omega_{\bar{\varepsilon}}$ in experiments with $\psi = 0.5$

	$\alpha_\gamma = 0.25$		$\alpha_\gamma = 0.5$		$\alpha_\gamma = 0.75$		$\alpha_\gamma = 1$	
N	ω_v	$\omega_{\bar{\varepsilon}}$	ω_v	$\omega_{\bar{\varepsilon}}$	ω_v	$\omega_{\bar{\varepsilon}}$	ω_v	$\omega_{\bar{\varepsilon}}$
25	0.33	0.93	0.63	0.76	0.88	0.47	0.97	0.23
50	0.24	0.96	0.63	0.76	0.90	0.42	0.99	0.16
100	0.25	0.96	0.64	0.76	0.93	0.35	0.99	0.12
200	0.18	0.98	0.64	0.76	0.95	0.30	1.00	0.08

Notes: Weights $\omega_v = \sigma_v / \sigma_{\bar{\xi}}$ and $\omega_{\bar{\varepsilon}} = \sigma_{\bar{\varepsilon}} / \sigma_{\bar{\xi}}$ do not depend on the parameter $\lambda_{\max}$.

Figure 2: GIRFs of one unit combined aggregate shock on the aggregate variable, $g_{\bar{\xi}}(s)$, for different persistence of common factor, $\psi = 0, 0.5$ and 0.8.



Notes: The vertical axis shows units of the shock. $N = 200$ and $\alpha_\gamma = 1$.

Table 2: RMSE ($\times 100$) of estimating GIRF of one unit (1 s.e.) combined aggregate shock on the aggregate variable, averaged over horizons $s = 0$ to 12 and $s = 13$ to 24.

	Estimates averaged over horizons from $s = 0$ to 12				Estimates averaged over horizons from $s = 13$ to 24			
N\T	**100**		**200**		**100**		**200**	
	$\hat{g}^a_{\bar{\xi}}$	$\hat{g}^d_{\bar{\xi}}$	$\hat{g}^a_{\bar{\xi}}$	$\hat{g}^d_{\bar{\xi}}$	$\hat{g}^a_{\bar{\xi}}$	$\hat{g}^d_{\bar{\xi}}$	$\hat{g}^a_{\bar{\xi}}$	$\hat{g}^d_{\bar{\xi}}$
	Experiments with $\alpha_\gamma = 1$							
	(a) $\lambda_{\max} = 0.9$							
50	20.18	12.81	13.50	8.70	10.39	4.38	8.22	3.20
100	20.00	12.41	13.49	8.32	10.76	3.89	8.39	2.76
200	20.45	12.39	13.61	8.30	10.27	3.61	8.17	2.62
	(b) $\lambda_{\max} = 1$							
50	24.13	15.23	15.95	10.41	21.15	12.55	16.34	8.66
100	23.92	14.76	16.44	9.96	20.36	11.37	16.96	7.34
200	24.34	14.65	15.99	9.70	20.75	10.58	16.36	6.56
	Experiments with $\alpha_\gamma = 0.5$							
	(c) $\lambda_{\max} = 0.9$							
50	3.24	2.21	2.31	1.57	1.87	0.96	1.48	0.72
100	2.24	1.50	1.62	1.06	1.24	0.59	1.02	0.45
200	1.55	0.99	1.11	0.72	0.88	0.36	0.69	0.28
	(d) $\lambda_{\max} = 1$							
50	3.66	2.86	2.84	1.99	3.38	2.86	2.64	2.04
100	2.71	1.96	1.96	1.30	2.54	1.77	1.90	1.25
200	1.78	1.27	1.36	0.88	1.56	1.09	1.29	0.78

Notes: Experiments with $\psi = 0.5$.

6 Inflation Persistence: Aggregation or Common Factor Persistence

Persistence of aggregate inflation and its sources have attracted a great deal of attention in the literature. Prices at the micro level are known to be relatively flexible, whereas at the aggregate level the overall rate of inflation seems to be quite persistent. In a recent paper, using individual category price series, Altissimo et al. (2009) conclude that "...the aggregation mechanism explains a significant amount of aggregate inflation persistence." (p.231). In this section, we investigate the robustness of this conclusion by estimating a factor augmented high dimensional VAR model in disaggregate inflation series, where the relative contributions of aggregation and common factor persistence can be evaluated. We also consider the way the two sources of persistence interact and get amplified in the process. We use the same data set as the one used by Altissimo et al. (2009), so that our respective conclusions can be compared more readily.[12] We find that persistence due to dynamic heterogeneity alone does not explain the persistence of the aggregate inflation, rather it is the combination of factor persistence and dynamic heterogeneity that is responsible for the high persistence of aggregate inflation as compared to the persistence of the underlying individual inflation series.

6.1 Data

The inflation series for the i-th price category is computed as $y_{it} = 400 \cdot [\ln(q_{it}) - \ln(q_{i,t-1})]$, where q_{it} is the seasonally adjusted consumer price index of unit i at time t.[13] Units are individual categories of the consumer price index (e.g. bread, wine, medical services,...) and the time dimension is quarterly covering the period 1985Q1 to 2004Q2, altogether 78 observations per price category. We have data on 85 categories in Germany, 145 in France and 168 in Italy. The aggregate inflation measure is computed as $\overline{y}_{wt} = \sum_{i=1}^{N} w_i y_{it}$, where N is the number of price categories and w_i is the weight of the i^{th} category in the consumer price index. The empirical analysis is conducted for each of the three countries separately. Country subscripts are, however, omitted to simplify the notations. No micro regressors are included in the analysis, and all measures of persistence

[12] We are grateful to Altissimo *et al.* for providing us with their data set.

[13] Descriptive statistics of the individual price categories are provided in Altissimo *et al.* (2009, Table 2).

reported below are therefore unconditional.

6.2 Micro model of consumer prices

Following Chudik and Pesaran (2011), we investigate the possibility that there are unobserved factors or neighborhood effects in the micro relations. Selecting neighboring units tends to be subjective. Here we categorize individual units into a small sets of products that are close substitutes and are generally close in terms of their characteristics. For example, spirits, wine and beer are assumed to be 'neighbors'. A complete list of 'neighbors' for Germany is provided in Pesaran and Chudik (2011). An alternative possibility would be to define neighbors in terms of their proximity as measured by flows of transactions between different commodity categories using input-output tables. But the misspecifications of neighboring units might not be that serious if the object of the exercise is to estimate the persistence of shocks on the aggregates. With this in mind we shall not pursue the input-output metric, although we acknowledge that it might be worth further investigation.

Let $\mathbb{C}_i$ be the index set defining the neighbors of unit i, and consider the following local averages

$$y_{it}^{\diamond} = \frac{1}{|\mathbb{C}_i|} \sum_{j \in \mathbb{C}_i} y_{jt} = \mathbf{s}_i' \mathbf{y}_t, \ i = 1, 2, ..., N, \tag{37}$$

where $|\mathbb{C}_i|$ is the number of neighbors of unit i, assumed to be small and fixed as $N \to \infty$, $\mathbf{s}_i$ is the corresponding $N \times 1$ sparse weights vector with $|\mathbb{C}_i|$ nonzero elements. $y_{it}^{\diamond}$ represents the local average of unit i. No unit is assumed to be dominant in the sense discussed by Chudik and Pesaran (2012).[14]

We follow Pesaran (2006) and its extension to dynamic panels in Chudik and Pesaran (2011), and model the effects of unobserved common factors by means of cross-sectional averages, at the national and sectoral levels. Accordingly, we use the economy wide average, $\bar{y}_t = N^{-1} \sum_{j=1}^{N} y_{jt}$, and the three sectoral averages

$$\bar{y}_{kt} = \frac{1}{|\mathbb{Q}_k|} \sum_{j \in \mathbb{Q}_k} y_{jt} = \mathbf{w}_k' \mathbf{y}_t, \ \text{for } k \in \{f, g, s\}, \tag{38}$$

[14] We have also estimated high dimensional VAR models of consumer price categories with the consumer energy category treated as a dominant unit, but found little empirical support for the dominance of consumer energy prices.

where $\mathbb{Q}_k$ for $k = \{f, g, s\}$ defines the set of units belonging food and beverages sector (f), goods sector (g), and services sector (s). $|\mathbb{Q}_k|$ is the number of units in sector k, and $\mathbf{w}_k$ is the corresponding vector of sectoral weights. This set up allows us to accommodate up to four common factors.

The following regressions are estimated by least squares for the price category i belonging to sector k, (intercepts are included but not shown)

$$y_{it} = \sum_{\ell=1}^{p_{i\phi}} \phi_{ii\ell} y_{i,t-\ell} + \sum_{\ell=1}^{p_{id}} d_{i\ell} y^{\diamond}_{i,t-\ell} + \sum_{\ell=0}^{p_{ih}} h_{i\ell} \bar{y}_{t-\ell} + \sum_{\ell=0}^{p_{ik}} h_{ki\ell} \bar{y}_{k,t-\ell} + \zeta_{it}, \text{ for } i \in \mathbb{Q}_k \text{ and } k \in \{f, g, s\}. \quad (39)$$

The same equations are also estimated for the energy price category, but without sectoral averages. Impulse response function of the combined aggregate shock on the aggregate variable in a disaggregate model is computed in the same way as in Section 5, with the exception that higher lag orders for the lagged micro coefficients are considered in (39) and we allow also for sectoral cross-sectional averages in addition to the country cross-sectional averages. The lag orders for the individual price equations are chosen by AIC with the maximum lag order set to 2 (to keep the number of unknown parameters to be estimated at a reasonable level). In line with the theoretical derivations, a higher maximum lag order is selected when estimating the aggregate inflation equations. See Footnote 15 below.

6.2.1 Estimation results

Table 3 summarizes the statistical significance of the various coefficients in the price equations, (39), for Germany, France and Italy. The parameters are grouped into own lagged effects ($\phi_{ii\ell}$), lagged neighborhood effects ($d_{i\ell}$), country effects ($h_{i\ell}$), and sectoral effects ($h_{ki\ell}$, for $k = f, g, s$). All four types of effects are statistically important, although own lagged effects, perhaps not surprisingly, are more important statistically as compared to the other effects. At the 5% significance level, own lagged effects are significant in 90 cases out of 112 in Germany, 111 cases out of 169 in France, and 158 out of 209 cases in Italy, representing 65%-80% share of all estimated own lagged effects. Local and cross-sectional averages are statistically significant in about 12-25% of cases, which is above the 5% nominal size of the tests. These results suggest that the micro relations that ignore common factors and the neighborhood effects are most likely misspecified. Idiosyncratic shocks are likely to dominate the micro relations, which could explain the lower rejection rate for the cross-sectional averages, compared to the own lagged coefficients. The fit is relatively high in most cases. The average $\overline{R}^2$ is 56% in Germany, 48% in France, and 51% in Italy (median values are 61%, 52%, and 54%, respectively).

Table 3: Summary statistics for individual price relations for Germany, France and Italy (equation (39))

	No. of estimated coef.	No. of significant coef. (*at the 5% nominal level*)	Share
	Results for Germany		
Own lagged effects	112	90	80.4%
Lagged neighborhood effects	66	16	24.2%
Sectoral effects	182	34	18.7%
Country effects	190	33	17.4%
	Results for France		
Own lagged effects	169	111	65.7%
Lagged neighborhood effects	166	23	13.9%
Sectoral effects	302	57	18.9%
Country effects	314	38	12.1%
	Results for Italy		
Own lagged effects	209	158	75.6%
Lagged neighborhood effects	173	38	22.0%
Sectoral effects	335	54	16.1%
Country effects	345	73	21.2%

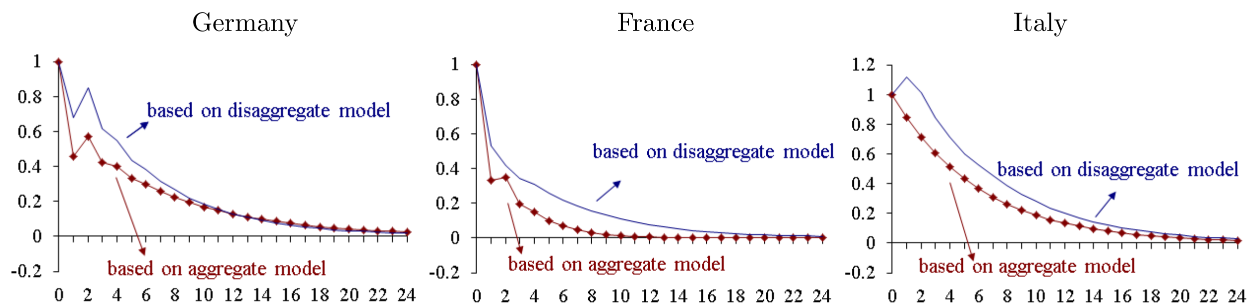
6.3 Sources of aggregate inflation persistence

For each of the three countries, we compute and report in Figure 3 the GIRF of a unit combined aggregate shock on the aggregate variable, using aggregate and disaggregate models, as explained in Section 4. We also provide 90% bootstrap confidence bounds together with the bootstrap means.[15] These impulse responses are quite persistent. The estimates based on the disaggregate model show a higher degree of persistence in the case of France and Italy.

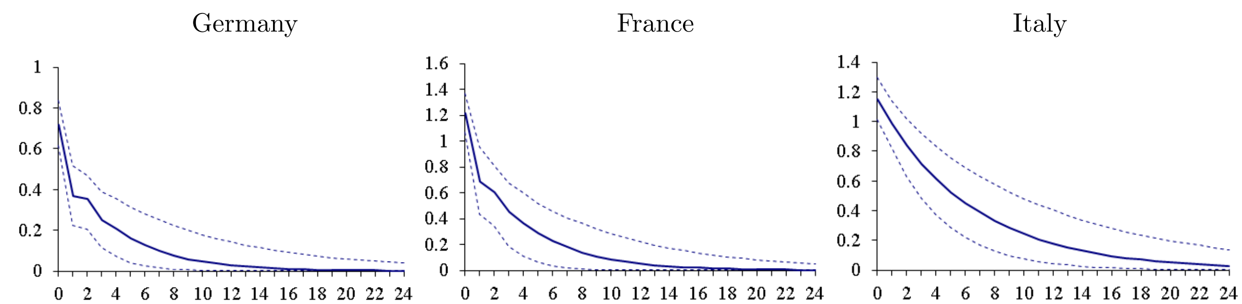
[15] The aggregate model is assumed to follow the AR(p) process estimated using $\bar{y}_t$. The lag order is chosen by AIC in the case of Italy and France with the maximum lag order set to $\left[T^{1/2}\right]$. In the case of Germany, both AIC and SBC chose $p = 3$, but the corresponding GIRFs were erratic and volatile. Therefore, we set the lag order to 2, to generate a less erratic GIRF for Germany.

Figure 3: GIRFs of one unit combined aggregate shock on the aggregate variable

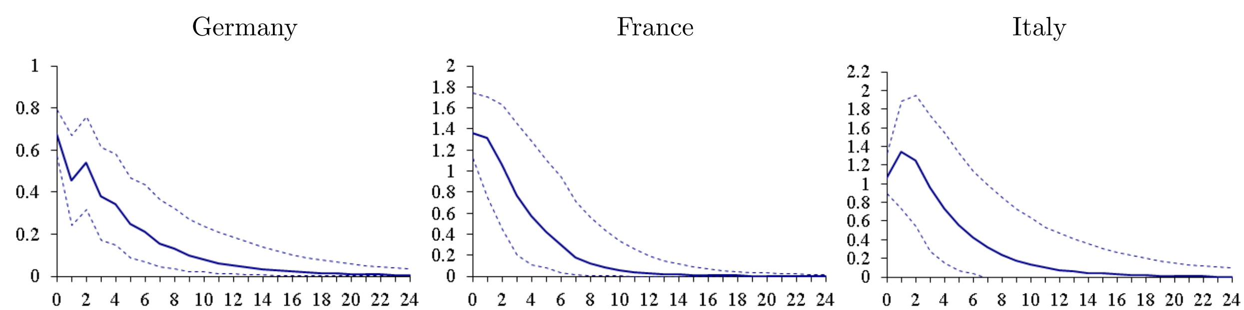
Panel A. Point estimates, y-axis shows units of the shock.



Panel B. Bootstrap means and 90% confidence bounds based on aggregate model; y-axis shows the estimated size of the shock.



Panel C. Bootstrap means and 90% confidence bounds based on disaggregate model; y-axis shows the estimated size of the shock.

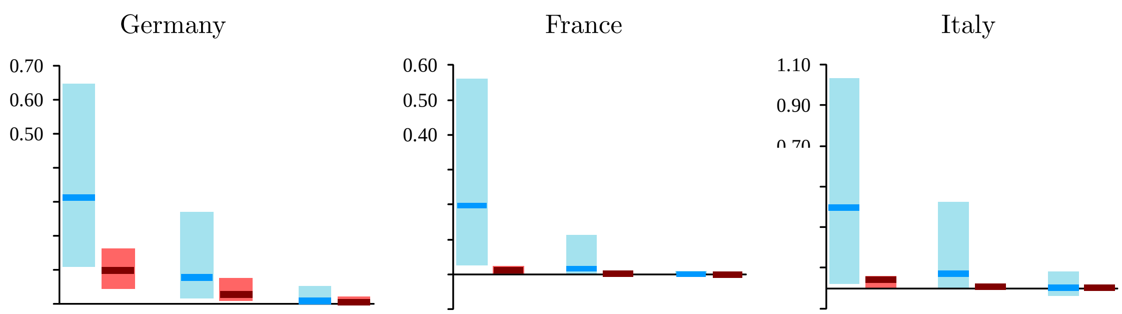


Using the estimates of micro lagged coefficients in (39), for $i = 1, 2, ..., N$, we compute eigenvalues of the companion matrix corresponding to the VAR polynomial matrix $\hat{\mathbf{\Phi}}(L)$,

$$\hat{\mathbf{\Phi}}(L) = \begin{pmatrix} \hat{\phi}_{11}(L) & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \hat{\phi}_{NN}(L) \end{pmatrix} + \begin{pmatrix} \hat{d}_1(L)\mathbf{s}_1' \\ \vdots \\ d_N(L)\mathbf{s}_N' \end{pmatrix},$$

where $\hat{\phi}_{ii}(L) = \sum_{\ell=1}^{p_{i\phi}} \phi_{ii\ell} L^{\ell-1}$, $\hat{d}_i(L) = \sum_{\ell=1}^{p_{id}} d_{i\ell} L^{\ell-1}$, and $\hat{\phi}_{ii\ell}$ and $\hat{d}_{i\ell}$ denote estimates of $\phi_{ii\ell}$ and $d_{i\ell}$, respectively. The modulus of the largest eigenvalue is 0.94 in Germany and Italy, and 0.89 in France, and do no cover unity. Hence, given the theory advanced in the paper, it is unlikely that the dynamic heterogeneity alone could generate the degree of persistence observed in Figure 3. This conclusion is further investigated in Figure 4, which compares the estimates of GIRFs for the combined aggregate shock on the aggregate variable with $\hat{a}_s = \mathbf{w}'\hat{\mathbf{G}}_s \boldsymbol{\tau}_N$ at horizons $s = 6, 12$ and 24, where the matrix $\hat{\mathbf{G}}_s$is defined by $\hat{\mathbf{\Phi}}^{-1}(L) = \hat{\mathbf{G}}(L) = \sum_{s=0}^{\infty} \hat{\mathbf{G}}_s L^{\ell}$. $\hat{a}_s$ shows the effects of dynamic heterogeneity on the persistence of the aggregate variable, whereas the GIRFs of the combined aggregate shock on the aggregate variable is determined by factor persistence as well as dynamic heterogeneity. $\hat{a}_s$ is found to die out much faster as compared to the effects of the combined aggregate shock in the case of all the three countries. Thus, dynamic heterogeneity alone does not seem sufficient for explaining the observed persistence of the aggregate inflation. In the case of France and Italy, $\hat{a}_s$ is close to zero for $s \geq 6$ months horizon.

Figure 4: GIRFs of one unit combined aggregate shocks on the aggregate variable (light/blue color) and estimates of a_s (dark/red color); bootstrap means and 90% confidence bounds, $s = 6, 12$ and 24.



Notes: The vertical axis shows units of the shock.

Altissimo et al. (2009) reach a similar conclusion in terms of the importance of common factor for the behavior of the aggregate inflation, albeit using a different set of techniques. They find one unobserved common factor and estimate the following model in order to study the implications of aggregation for the persistence of aggregate inflation, $y_{it} = \psi_i(L) v_t + \varphi_i(L) \varepsilon_{it}$, where $\psi_i(L)$ and $\varphi_i(L)$ are unit-specific polynomials, v_t is a serially *uncorrelated* unobserved common factor innovation orthogonal to ε_{it}, and ε_{it} is $IID\left(0, \sigma_i^2\right)$. Altissimo et al. (2009) find that the persistence of aggregate inflation originates from the unobserved common component, $\psi_i(L) v_t$, and that the persistence of the aggregate idiosyncratic component, $\sum_{i=1}^{N} w_i \varphi_i(L) \varepsilon_{it}$, is relatively small. The latter finding is in line with our results, which shows that $\hat{a}_s = \mathbf{w}' \hat{\mathbf{G}}_s \boldsymbol{\tau}_N$ seems to decline at a geometric rate. Their analysis focuses on the roots of $\psi_i(L)$, but does not study whether one could decompose $\psi_i(L)$ into the products $\gamma_i(L) \mu(L)$, in which case one could write $\psi_i(L) v_t = \gamma_i(L) \mu(L) v_t = \gamma_i(L) f_t$ where $f_t = \mu(L) v_t$ could be viewed as a serially correlated unobserved common factor. Thus, by assuming that the common factor is serially uncorrelated, they end up attributing the observed persistence of inflation to the aggregation process. Accordingly, they find that the empirical distribution of the maximal autoregressive roots (the modulus of the roots of $\psi_i(L)$) peaks at one, which leads them to argue that the aggregate inflation presents a long memory behavior and that the aggregation mechanism explains a significant part of aggregate inflation persistence.

Our exercise allows us to evaluate how the two sources of persistence - dynamic heterogeneity and the unobserved common factor persistence - combine and get amplified in the process. Figure 4 shows that the interaction of the persistence in common factors and dynamic heterogeneity of the underlying processes is likely the key to understanding the slow response of the aggregate inflation to macro shocks. As pointed out by Granger (1987), a relatively benign common factor at the micro level becomes pertinent by aggregation at the macro level, and therefore understanding where this common factor comes from and why it is (or is not) persistent would be important for a proper understanding of consumer price inflation behavior.

7 Conclusion

In this paper we extend the literature on aggregation of linear dynamic models in a number of directions. We derive conditions under which an optimal aggregate equation exists in the case of large dynamic panels with individual specific regressors and common factors. We also derive conditions under which aggregation errors are of second order importance in empirical analysis, and show how these conditions are related to the long memory property of aggregate time series models highlighted by Granger. We also consider the problem of identification of some of the distributional features of micro parameters from aggregate relations, and derive impulse response functions for the analysis of the effects of the composite macro and aggregated idiosyncratic shocks on the aggregate variable, allowing for weak cross-sectional dependence in the errors of the underlying dynamic panel data model. Some of the theoretical findings are illustrated by a series of Monte Carlo simulations. An empirical application investigating the sources of the persistence of aggregate inflation is also presented. It is shown that the observed persistence of aggregate inflation could be due to a combination of factor persistence and dynamic heterogeneity in the underlying micro model of inflation. It is hoped that the present paper initiates further research in the area of aggregation in economics. There are clearly important links between aggregation and pooling of information in dynamic heterogenous panels which are worthy of further investigations. The present paper should be seen as a small step in this direction.

A Mathematical Appendix

Proof of Proposition 1. Taking the absolute values of (18) and applying the matrix norm inequality yields

$$\left|g_{\bar{\varepsilon}}\left(s\right)\right| \leq \left\|\mathbf{w}\right\| \left\|\mathbf{\Phi}^{s}\right\| \left\|\frac{\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}}\right\|, \text{ for } s = 1, 2, ...,$$

and for every possible realization of the random elements in $\mathbf{\Phi}$. The matrix $\mathbf{\Sigma}_{\varepsilon} = Var\left(\boldsymbol{\varepsilon}_{t}\right)$ is symmetric and positive definite and therefore there exists a matrix $\mathbf{Z}_{\varepsilon}$ such that $\mathbf{\Sigma}_{\varepsilon} = \mathbf{Z}_{\varepsilon}\mathbf{Z}_{\varepsilon}'$. Therefore,[16]

$$\left\|\frac{\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}_{\varepsilon}\mathbf{w}}}\right\| = \left\|\mathbf{Z}_{\varepsilon}\frac{\mathbf{Z}_{\varepsilon}'\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{Z}_{\varepsilon}\mathbf{Z}_{\varepsilon}'\mathbf{w}}}\right\| \leq \left\|\mathbf{Z}_{\varepsilon}\right\| \frac{\left\|\mathbf{Z}_{\varepsilon}'\mathbf{w}\right\|}{\left\|\mathbf{Z}_{\varepsilon}'\mathbf{w}\right\|} \leq \left\|\mathbf{Z}_{\varepsilon}\right\|,$$

and hence

$$\left|g_{\bar{\varepsilon}}\left(s\right)\right| \leq \left\|\mathbf{w}\right\| \left\|\mathbf{\Phi}^{s}\right\| \left\|\mathbf{Z}_{\varepsilon}\right\|. \tag{A.1}$$

Taking expectations of the both sides of (A.1), and noting that $\left\|\mathbf{w}\right\|$ and $\left\|\mathbf{Z}_{\varepsilon}\right\|$ are non-stochastic we have

$$E\left|g_{\bar{\varepsilon}}\left(s\right)\right| \leq \left\|\mathbf{w}\right\| \left\|\mathbf{Z}_{\varepsilon}\right\| E\left\|\mathbf{\Phi}^{s}\right\|.$$

But $E\left\|\mathbf{\Phi}^{s}\right\| \leq \left[E\left\|\mathbf{\Phi}\right\|\right]^{s}$, and since by assumption $E\left\|\mathbf{\Phi}\right\| < K$, then $E\left\|\mathbf{\Phi}^{s}\right\|$ is also bounded in N and since by assumption $\left\|\mathbf{w}\right\| = O\left(N^{-1/2}\right)$, and $\left\|\mathbf{Z}_{\varepsilon}\right\| = \left\|\mathbf{\Sigma}_{\varepsilon}\right\|^{1/2} \leq \left(\left\|\mathbf{\Sigma}_{\varepsilon}\right\|_{1}\left\|\mathbf{\Sigma}_{\varepsilon}\right\|_{\infty}\right)^{1/4} = \left\|\mathbf{\Sigma}_{\varepsilon}\right\|_{1}^{1/2} = O\left(N^{\alpha_{\varepsilon}/2}\right)$, then it follows that

$$E\left|g_{\bar{\varepsilon}}\left(s\right)\right| = O\left(N^{\alpha_{\varepsilon}/2-1/2}\right),$$

as required. ■

[16] Note that $\sqrt{\mathbf{w}'\mathbf{Z}_{\varepsilon}\mathbf{Z}_{\varepsilon}'\mathbf{w}} = \left\|\mathbf{Z}_{\varepsilon}'\mathbf{w}\right\|$.

References

Altissimo, F., B. Mojon, and P. Zaffaroni (2009). Can aggregation explain the persistence of inflation? *Journal of Monetary Economics 56*, 231–241.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2012). Exponents of cross-sectional dependence: Estimation and inference. CESifo Working Paper No. 3722.

Chudik, A. and M. H. Pesaran (2011). Infinite dimensional VARs and factor models. *Journal of Econometrics 163*, 4–22.

Chudik, A. and M. H. Pesaran (2012). Econometric analysis of high dimensional VARs featuring a dominant unit. *Forthcomming in Econometric Reviews*.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal 14*, C45–C90.

Forni, M. and M. Lippi (1997). *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford University Press, Oxford.

Garderen, K. J., K. Lee, and M. H. Pesaran (2000). Cross-sectional aggregation of non-linear models. *Journal of Econometrics 95*, 285–331.

Geweke, J. (1985). Macroeconometric modeling and the theory of the representative agent. *The American Economic Review 75*, 206–210.

Giacomini, R. and C. W. J. Granger (2004). Aggregation of space-time processes. *Journal of Econometrics 118*, 7–26.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*, 227–238.

Granger, C. W. J. (1987). Implications of aggregation with common factors. *Econometric Theory 3*, 208–222.

Granger, C. W. J. (1990). Aggregation of time-series variables: A survey. In T. Barker and M. H. Pesaran (Eds.), *Disaggregation in Econometric Modelling*, Chapter 2, pp. 17–34. Routlege, London and New York.

Granger, C. W. J. (1993). Implications of seeing economic variables through an aggregation window. *Ricerche Economiche 47*, 269–279.

Granger, C. W. J. and T.-H. Lee (1999). The effect of aggregation on nonlinearity. *Econometric Reviews 18*(3), 259–269.

Granger, C. W. J. and M. J. Morris (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society A 139*, 246–257.

Granger, C. W. J. and P. L. Siklos (1995). Systematic sampling, temporal aggregation, seasonal adjustment, and cointegration: Theory and evidence. *Journal of Econometrics 66*, 357–369.

Grunfeld, Y. and Z. Griliches (1960). Is aggregation necessarily bad? *Review of Economics and Statistics 42*, 1–13.

Hsiao, C., Y. Shen, and H. Fujiki (2005). Aggregate vs disaggregate data analysis - A paradox in the estimation of a money demand function of Japan under the low interest rate policy. *Journal of Applied Econometrics 20*, 579–601.

Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2005). PPP strikes back: Aggregation and the real exchange rate. *Quarterly Journal of Economics 120*, 1–43.

Kelejian, H. H. (1980). Aggregation and disaggregation of non-linear equations. In J. Kmenta and J. B. Ramsay (Eds.), *Evaluation of econometric models*. Academic Press, New York.

Lewbel, A. (1994). Aggregation and simple dynamics. *American Economic Review 84*, 905–918.

Lütkepohl, H. (1984). Linear transformation of vector ARMA processes. *Journal of Econometrics 26*, 283–293.

Lütkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*. Springer, Berlin, Heidelberg.

Pesaran, M. H. (2003). Aggregation of linear dynamic models: An application to life-cycle consumption models under habit formation. *Economic Modelling 20*, 383–415.

Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica 74*, 967–1012.

Pesaran, M. H. and A. Chudik (2011). Aggregation in large dynamic panels. IZA Discussion Paper No. 5478.

Pesaran, M. H., R. Pierse, and K. Lee (1994). Choice between disaggregate and aggregate specifications estimated by IV method. *Journal of Business and Economic Statistics 12*, 111–121.

Pesaran, M. H., R. G. Pierse, and M. S. Kumar (1989). Econometric analysis of aggregation in the context of linear prediction models. *Econometrica 57*, 861–888.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*, 1057–1112.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics 5*, 163–168.

Rose, D. E. (1977). Forecasting aggregates of independent ARIMA processes. *Journal of Econometrics 5*, 323–345.

Stoker, T. (1984). Completeness, distribution restrictions, and the form of aggregate functions. *Econometrica 52*, 887–907.

Stoker, T. (1986). Simple tests of distributional effects on macroeconomic equations. *Journal of Political Economy 94*, 763–795.

Stoker, T. (1993). Empirical approaches to the problem of aggregation over individuals. *Journal of Economic Literature 31*, 1827–1874.

Theil, H. (1954). *Linear Aggregation of Economic Relations*. North-Holland, Amsterdam.

Trapani, L. and G. Urga (2010). Micro versus macro cointegration in heterogeneous panels. *Journal of Econometrics 155*, 1–18.

Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics 120*, 75–102.