

Aggregation in Large Dynamic Panels

M. Hashem Pesaran
Alexander Chudik

CESIFO WORKING PAPER NO. 3346
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
FEBRUARY 2011

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

Aggregation in Large Dynamic Panels

Abstract

This paper considers 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 the limiting behavior of the aggregation error is investigated as N (the number of cross section units) increases. Certain distributional features of micro parameters are also identified from the aggregate function. The paper then establishes Granger's (1980) conjecture regarding the long memory properties of aggregate variables from a very large scale dynamic, econometric model, and considers the time profiles of the effects of macro and micro shocks on the aggregate and disaggregate variables. Some of these findings are illustrated in Monte Carlo experiments, where we also study the estimation of the aggregate effects of micro and macro shocks. The paper concludes with 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.

JEL-Code: C430, E310.

Keywords: aggregation, large dynamic panels, long memory, weak and strong cross section dependence, VAR models, impulse responses, factor models, inflation persistence.

M. Hashem Pesaran
Faculty of Economics
Cambridge University
UK – Cambridge, CB3 9DD
mhp1@econ.cam.ac.uk

Alexander Chudik
European Central Bank
Kaiserstrasse 29
Germany – 60311 Frankfurt am Main
alexander.chudik@ecb.europa.eu

January 2011

A preliminary version of this paper was presented as the 4th 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. We are also grateful to Benoit Mojon for providing us with the price data set used in Altissimo et al. (2009). The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank.

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 it is often ignored either by resorting to 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. For example, Hsiao et al. (2005) using Japanese aggregate and disaggregate money demand data show that they can obtain stable money demand equations only if they work with disaggregate data, and trace the fundamentally different conclusions reached using aggregate versus the disaggregate data to the prevalence of parameter heterogeneity. Altissimo et al. (2009) find that aggregation can explain a significant part of observed persistence in the consumer price inflation. Other examples include Imbs et al. (2005) who show that the estimated persistence of real exchange rates falls dramatically when the aggregation is taken into account, contributing to a long standing debate on the empirical validity of the Purchasing Power Parity hypothesis. Similar to the Lucas critique, Geweke (1985) argues that ignoring the sensitivity of the aggregates to policy changes seems no more compelling than ignoring the dependence of expectations on the policy regime.

There are several different, but related aspects of the aggregation problem that have been studied in the literature. Granger (1990) and Stoker (1993) provide early surveys. Theil (1954), Lewbel (1994), and Pesaran (2003) consider the problem of deriving an optimal aggregate function. In addition, Pesaran (2003) discusses estimating the average long-run micro effects and mean lags of the autoregressive distributed lag (ARDL) micro models from aggregate data. 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 exponentially decaying autocovariances. The identification and estimation of micro parameters or some of their distributional features from aggregate relations is another key issue of concern in the aggregation literature. 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. The role of common factors and cross-section dependence in aggregation was first highlighted by Granger (1987), and further developed and discussed in Forni and Lippi (1997) and Zaffaroni (2004). Another important issue of concern is the possible effects of aggregation on cointegration. The problem was initially considered in Pesaran and Smith (1995) and Phillips and Moon (1999). Trapani and Urga (2010) consider a more general setting and provide necessary and sufficient conditions for the aggregate cointegration to hold when the underlying micro units cointegrate. Lastly, 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 section aggregation of vector ARMA processes and a comprehensive bibliography is provided in Lütkepohl (1987). Our literature review is by no means comprehensive and it highlights only the main aspects of the aggregation problem.¹

In this paper we consider 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. In this way the earlier literature on aggregation of independent dynamic regressions is extended to aggregation of dynamic models with interactions and cross section dependence. In particular, we allow for various interconnections across the individual units, relax the assumption that micro coefficients are independently distributed, and allow for a general pattern of cross section dependence of micro innovations, which can be either strong or weak. Following Pesaran (2003), an optimal aggregate model is derived, and the limiting behavior

¹There are also a number of papers on the aggregation of nonlinear models: Kelejian (1980), Stoker (1984), Stoker (1986), and Garderen et al. (2000), all focussing on the aggregation of static non-linear micro models.

of aggregation error is investigated as N (the number of cross section units) increases. Certain distributional features of micro parameters are also identified from the aggregate relation. The paper then establishes Granger’s (1980) conjecture about the long memory properties of aggregate variables from ‘a very large scale dynamic, econometric model’, and considers the time profiles of the effects of macro and micro shocks on the aggregate and disaggregate variables. The extent to which aggregation can generate excessive persistence is investigated by Monte Carlo experiments in the context of large dynamic panel data models with and without unobserved common factors. 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 section heterogeneity that seems to be responsible for the observed persistence of the aggregate inflation.

The remainder of the paper is organized as follows. To place the contribution of this paper in the context of the literature, in Section 2 we begin with an overview of Granger’s main results on aggregation and persistence. Section 3 derives the optimal aggregate model for a factor augmented VAR in N cross section units and discusses the main implications of theoretical results. Relationship between micro and macro parameters are discussed in Section 4. The impulse response effects of micro and macro shocks on disaggregate and aggregate variables are derived and contrasted in Section 5. Monte Carlo experiments are presented in Section 6, and Section 7 reports on the empirical application. Section 8 concludes the paper. Some of the mathematical proofs are provided in an Appendix.

A brief word on notations: $\|\mathbf{A}\|_1 \equiv \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$ denotes the column sum matrix norm of $\mathbf{A} \in \mathbb{M}^{n \times n}$, where $\mathbb{M}^{n \times n}$ is the space of real-valued $n \times n$ matrices. $\|\mathbf{A}\|_\infty \equiv \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$ is the row matrix norm of \mathbf{A} . $\|\mathbf{A}\| = \sqrt{\varrho(\mathbf{A}'\mathbf{A})}$ is the spectral norm of \mathbf{A} ,² $\varrho(\mathbf{A}) \equiv \max_{1 \leq i \leq n} \{|\lambda_i(\mathbf{A})|\}$ is the spectral radius of \mathbf{A} , and $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq \dots \geq |\lambda_n(\mathbf{A})|$ are the eigenvalues of \mathbf{A} . All vectors are column vectors.

²Note that if \mathbf{x} is a vector, then $\|\mathbf{x}\| = \sqrt{\varrho(\mathbf{x}'\mathbf{x})} = \sqrt{\mathbf{x}'\mathbf{x}}$ corresponds to the Euclidean length of vector \mathbf{x} .

2 Granger's Contributions to Aggregation and Persistence

Clive Granger has contributed to and shaped many aspects of the literature on aggregation. His first paper on aggregation, Granger (1980), showed that aggregation over a large number of stationary AR(1) processes can generate long memory in the aggregated series. This finding together with an earlier work by Mandelbrot and van Ness (1968) on fractional Brownian motions started an extensive literature on fractionally integrated and cointegrated processes. The second seminal contribution of Granger to aggregation literature is on the role of cross section dependence in aggregation. In Granger (1987), he focused on the implications of cross sectional aggregation with common factors, and showed that a high degree of statistical fit at the micro level could be compatible with almost no fit at the macro level, and *vice versa*. Granger's finding that common factors dominate aggregate relationships has been explored in various papers in the literature. Some of these issues were further discussed in Granger (1990). Granger also contributed to the discussion of aggregation and cointegration, 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). Given the focus of our paper on aggregation and persistence in what follows we only consider Granger's result on memory properties of the cross sectionally aggregated dynamic processes, and the role of cross section dependence in a large scale cross section aggregation.

2.1 Aggregation of independent AR(1) models

Consider the following AR(1) disaggregate relations,

$$y_{it} = \lambda_i y_{i,t-1} + u_{it},$$

for $i = 1, 2, \dots, N$, and $t = \dots - 1, 0, 1, 2, \dots$, where $|\lambda_i| < 1$. Suppose these relations are independent, and in addition λ_i and $Var(u_{it}) = \sigma_i^2$ are independently and identically distributed (iid) random draws with the distribution function $F(\lambda)$ for λ on the range $[0, 1)$. Granger's objective was the memory properties of the aggregate variable $S_{t,N}(y) = \sum_{i=1}^N y_{it}$. The same set-up was considered also in an earlier work by Robinson (1978) with a different focus on the estimation of moments of

$F(\lambda)$.³ To study the persistence properties of the aggregates, Granger considered the spectrum of $\bar{y}_{Nt} = N^{-1} \sum_{i=1}^N y_{it}$,

$$\bar{f}_N(\omega) = N^{-1} \sum_{i=1}^N f_i(\omega) \approx \frac{1}{2\pi} E[Var(u_{it})] \int \frac{1}{|1 - \lambda e^{-i\omega}|^2} dF(\lambda).$$

Then assuming that λ is type II Beta distributed with parameters $p > 0$ and $q > 0$, he showed that for sufficiently large N , the s^{th} order autocovariance of $S_{t,N}(y) = N\bar{y}_{Nt}$, is $O(s^{1-q})$, and therefore the aggregate variable behaves as a fractionally integrated process of order $1 - q/2$. In fact the long memory property holds more generally, so long as the support of the distribution of λ covers 1.

2.2 Role of cross section dependence in aggregation

The second important area of Granger's contribution to the aggregation literature is on the role of cross section dependence in aggregation of a large population of micro units. Granger (1987) considered a simple factor model to illustrate the main issues,

$$y_{it} = x_{it} + \gamma_i f_t,$$

where x_{it} is unit-specific explanatory variable, f_t is common factor with loading γ_i , and as before y_{it} is observation for the unit i at time t . Suppose x_{it} and f_t have zero means, bounded variances, and x_{it} is independently distributed of f_t and of x_{jt} for all $j \neq i$. Consider the variance of the aggregate variable $S_{t,N}(y)$,

$$Var[S_{t,N}(y)] = \sum_{i=1}^N Var(x_{it}) + N^2 \bar{\gamma}_N^2 Var(f_t),$$

where $\bar{\gamma}_N = N^{-1} \sum_{i=1}^N \gamma_i$. The first summand is at most of order N , denoted as $O(N)$, and, provided that $\lim_{N \rightarrow \infty} \bar{\gamma}_N \neq 0$, the second summand is of order N^2 . The second term will therefore generally dominate the aggregate relationship. Granger demonstrated striking implications of this finding in terms of the fit of the aggregate (macro) relationship, where common factor prevails when

³Robinson (1978) identified moments of $F(\lambda)$ in terms of autocovariances $\gamma_\ell = E(y_{it}y_{it+\ell})$, established sufficient and necessary conditions for $y_t = E(y_{it})$ to have continuous spectral density, and considered the problem of estimation of the moments of $F(\lambda)$ using disaggregate data, where he excluded some cases with long-memory of the aggregate variable. In particular, he required $E(y_{it}^2)$ to exist for consistency, and $E(y_{it}^4)$ to exist for asymptotic normality of his proposed estimator.

N is sufficiently large, and disaggregate (micro) relationships, where micro regressor could play a leading role. If the common factor was unobserved, then the aggregate relation would have zero fit (for N large) whereas the fit of disaggregate relations could be quite high, being driven by the micro regressor, x_{it} . On the other hand, if f_t was observed and x_{it} was unobserved then the macro relation would have a perfect fit (for N large), whereas the micro relation may have very poor fit due to the missing micro regressor, x_{it} . Hence variables that may have very good explanatory power at the micro level might be unimportant at the macro level, and *vice versa*. Granger showed that the strength and pattern of cross section dependence thus plays a central role in aggregation and the components with weaker cross section dependence typically do not matter for the behavior of aggregate variables.

3 Aggregation of Factor Augmented VAR Models

Granger (1980) also discussed extensions of the basic AR(1) set-up outlined in the previous section, including what he called as ‘a very large scale dynamic, econometric model’ given by $\mathbf{y}_t = \Phi \mathbf{y}_{t-1} + \mathbf{u}_t$, where $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$, and $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$. Granger (1980), p. 237, conjectured that the distribution of eigenvalues of Φ could be pertinent to the long memory properties of the aggregate series. We study aggregation of high-dimensional VARs and establish conditions under which Granger’s conjecture turns out to be correct using the following augmented VAR model in N cross section units

$$\mathbf{y}_t = \Phi \mathbf{y}_{t-1} + \mathbf{B} \mathbf{x}_t + \mathbf{\Gamma} \mathbf{f}_t + \mathbf{u}_t, \quad (1)$$

where $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$, \mathbf{f}_t is $m \times 1$ vector of common factors, Φ and \mathbf{B} are $N \times N$ matrices of coefficients, and $\mathbf{\Gamma}$ is an $N \times m$ vector of factor loadings with elements denoted by γ_{ij} , for $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, m$. We denote the elements of Φ by ϕ_{ij} , for $i, j = 1, 2, \dots, N$, and assume that \mathbf{B} is a diagonal matrix with elements given by the elements of $\beta = (\beta_1, \beta_2, \dots, \beta_N)'$. This specification can be readily generalized to allow for more cross section specific regressors.

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} and \mathbf{f}_t , where $\mathbf{w} = (w_1, w_2, \dots, w_N)'$ is a set of pre-determined aggregation weights such that $\sum_{i=1}^N w_i = 1$. Throughout it is assumed that \mathbf{w} is known

and the weights are granular, in the sense that

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

Denote the aggregate information set by $\Omega_t = (\bar{y}_{w,t-1}, \bar{y}_{w,t-2}, \dots, \bar{x}_{wt}, \bar{x}_{w,t-1}, \dots, \mathbf{f}_t, \mathbf{f}_{t-1}, \dots)$. When \mathbf{f}_t is not observed the values of \mathbf{f}_t in Ω_t must be replaced by their fitted or forecast values obtained from an auxiliary model in \mathbf{f}_t and possibly other variables, not included in (1).

Introduce the extended information set $\Upsilon_t = (\mathbf{y}_{t-M}; \mathbf{x}_t, \mathbf{x}_{t-1}, \dots; \mathbf{f}_t, \mathbf{f}_{t-1}, \dots; \bar{y}_{w,t-1}, \bar{y}_{w,t-2}, \dots)$, which contains Ω_t . The following assumptions on the eigenvalues of Φ and the idiosyncratic errors, $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$, are postulated.

ASSUMPTION 1 *All eigenvalues of Φ , denoted by λ_i , for $i = 1, 2, \dots, N$, are distinct and have the following invariant conditional moments*

$$\left. \begin{aligned} E(\lambda_i^s | \Upsilon_t, \mathbf{P}, \mathbf{u}_{t-s}) &= a_s, \\ E(\lambda_i^s | \Upsilon_t, \mathbf{P}, \mathbf{B}) &= b_s(\boldsymbol{\beta}), \\ E(\lambda_i^s | \Upsilon_t, \mathbf{P}, \mathbf{\Gamma}) &= c_s(\mathbf{\Gamma}), \end{aligned} \right\} \quad (3)$$

for all $s = 1, 2, \dots$, and $i = 1, 2, \dots, N$, where $\Upsilon_t = (\mathbf{y}_{t-M}; \mathbf{x}_t, \mathbf{x}_{t-1}, \dots; \mathbf{f}_t, \mathbf{f}_{t-1}, \dots; \bar{y}_{w,t-1}, \bar{y}_{w,t-2}, \dots)$, \mathbf{P} is $N \times N$ matrix containing the eigenvectors of Φ as column vectors (in any order), and the coefficient vector $\boldsymbol{\beta}$ and the coefficient matrices \mathbf{B} and $\mathbf{\Gamma}$ are defined in model (1).

ASSUMPTION 2 *The idiosyncratic shocks, $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$, in (1) are serially uncorrelated and weakly cross sectionally dependent with zero means and finite variances.*

The above assumptions allow dependence between λ_i and the loadings β_i and γ_{ij} . Also since by assumption \mathbf{u}_t is serially uncorrelated then

$$E(\mathbf{w}'\mathbf{u}_t | \Upsilon_t, \Phi, \mathbf{B}, \mathbf{\Gamma}) = 0, \text{ and hence } E(\mathbf{w}'\mathbf{u}_t | \Upsilon_t, \mathbf{P}) = 0. \quad (4)$$

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

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

where by construction $E(v_{wt} | \Omega_t) = 0$, and v_{wt} , $t = 1, 2, \dots$ are serially uncorrelated, although they could be conditionally heteroskedastic.

Solving (1) recursively forward from the initial state, \mathbf{y}_{-M} , we have

$$\mathbf{y}_t = \Phi^{t+M} \mathbf{y}_{-M} + \sum_{s=0}^{t+M-1} \Phi^s (\mathbf{B}\mathbf{x}_{t-s} + \mathbf{\Gamma}\mathbf{f}_{t-s} + \mathbf{u}_{t-s}). \quad (5)$$

Hence, using the spectral decomposition of $\Phi = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ is a diagonal matrix with eigenvalues of Φ on its diagonal and the columns of \mathbf{P} are the associated eigenvectors of Φ , we obtain

$$\bar{y}_{wt} = \mathbf{w}'\mathbf{P}\mathbf{\Lambda}^{t+M}\mathbf{P}^{-1}\mathbf{y}_{-M} + \sum_{s=0}^{t+M-1} \mathbf{w}'\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}(\mathbf{B}\mathbf{x}_{t-s} + \mathbf{\Gamma}\mathbf{f}_{t-s} + \mathbf{u}_{t-s}). \quad (6)$$

Let $\Psi_{bt} = (\mathbf{P}, \Upsilon_t, \mathbf{B})$, $\Psi_{\Gamma t} = (\mathbf{P}, \Upsilon_t, \mathbf{\Gamma})$, $\Psi_{ut} = (\mathbf{P}, \Upsilon_t, \mathbf{u}_{t-s})$ and $\mathbf{S}_t = (\mathbf{P}, \Upsilon_t)$. It is clear that $\mathbf{S}_t \subset \Psi_{bt}$, $\mathbf{S}_t \subset \Psi_{\Gamma t}$, and $\mathbf{S}_t \subset \Psi_{ut}$, and by the chain rule of expectations we obtain

$$E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{B} | \mathbf{S}_t) = E[E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{B} | \Psi_{bt}) | \mathbf{S}_t] = E[\mathbf{P}E(\mathbf{\Lambda}^s | \Psi) \mathbf{P}^{-1}\mathbf{B} | \mathbf{S}_t].$$

Similarly,

$$E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma} | \mathbf{S}_t) = E[E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma} | \Psi_{\Gamma t}) | \mathbf{S}_t] = E[\mathbf{P}E(\mathbf{\Lambda}^s | \Psi_{\Gamma t}) \mathbf{P}^{-1}\mathbf{\Gamma} | \mathbf{S}_t].$$

But under (3) we have, $E(\mathbf{\Lambda}^s | \Psi_{ut}) = a_s \mathbf{I}_N$, $E(\mathbf{\Lambda}^s | \Psi_{bt}) = b_s(\beta) \mathbf{I}_N$, and $E(\mathbf{\Lambda}^s | \Psi_{\Gamma t}) = c_s(\mathbf{\Gamma}) \mathbf{I}_N$.

Hence

$$\begin{aligned} E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{B} | \mathbf{S}_t) &= E[\mathbf{P}b_s(\beta)\mathbf{P}^{-1}\mathbf{B} | \mathbf{S}_t] \\ &= E[b_s(\beta)\mathbf{B} | \mathbf{S}_t]. \end{aligned}$$

Similarly,

$$\begin{aligned} E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma} | \mathbf{S}_t) &= E[E(\mathbf{P}\mathbf{\Lambda}^s\mathbf{P}^{-1}\mathbf{\Gamma} | \Psi_{\Gamma t}) | \mathbf{S}_t] = E[\mathbf{P}c_s(\mathbf{\Gamma})\mathbf{P}^{-1}\mathbf{\Gamma} | \mathbf{S}_t] \\ &= E[c_s(\mathbf{\Gamma})\mathbf{\Gamma} | \mathbf{S}_t]. \end{aligned}$$

Finally,

$$E(\mathbf{P}\Lambda^s\mathbf{P}^{-1}\mathbf{u}_{t-s}|\mathbf{S}_t) = E[E(\mathbf{P}\Lambda^s\mathbf{P}^{-1}\mathbf{u}_{t-s}|\Psi_{ut})|\mathbf{S}_t] = E[\mathbf{P}a_s\mathbf{I}_N\mathbf{P}^{-1}\mathbf{u}_{t-s}|\mathbf{S}_t] = a_sE(\mathbf{u}_{t-s}|\mathbf{S}_t).$$

Taking expectations of both sides of (6) conditional on \mathbf{S}_t , we now have

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

Using the results derived above together with (4) we obtain

$$\begin{aligned} E(\bar{y}_{wt}|\mathbf{S}_t) &= (\mathbf{w}'\mathbf{y}_{-M})a_{t+M} + \sum_{s=0}^{t+M-1} \mathbf{w}'E[b_s(\boldsymbol{\beta})\mathbf{B}|\mathbf{S}_t]\mathbf{x}_{t-s} + \\ &\quad \sum_{s=0}^{t+M-1} \mathbf{w}'E[c_s(\boldsymbol{\Gamma})\mathbf{\Gamma}|\mathbf{S}_t]\mathbf{f}_{t-s} + \sum_{s=1}^{t+M-1} a_sE(\mathbf{w}'\mathbf{u}_{t-s}|\mathbf{S}_t), \end{aligned}$$

and finally taking expectations conditional on the aggregate information set (and noting that $\Omega_t \subset \mathbf{S}_t$)

$$\begin{aligned} E(\bar{y}_{wt}|\Omega_t) &= (\mathbf{w}'\mathbf{y}_{-M})E(a_{t+M}|\Omega_t) + \sum_{s=0}^{t+M-1} \mathbf{w}'E[b_s(\boldsymbol{\beta})\mathbf{B}|\Omega_t]\mathbf{x}_{t-s} \\ &\quad + \sum_{s=0}^{t+M-1} \mathbf{w}'E[c_s(\boldsymbol{\Gamma})\mathbf{\Gamma}|\Omega_t]\mathbf{f}_{t-s} + \sum_{s=1}^{t+M-1} a_sE(\bar{u}_{w,t-s}|\Omega_t). \end{aligned}$$

In the case where β_i and γ_i are identically distributed across i ,

$$E[b_s(\boldsymbol{\beta})\mathbf{B}|\Omega_t] = b_s\mathbf{I}_N, \tag{7}$$

$$E[c_s(\boldsymbol{\Gamma})\mathbf{\Gamma}|\Omega_t] = \boldsymbol{\tau}_N\mathbf{c}'_s, \tag{8}$$

where $b_s = E[b_s(\boldsymbol{\beta})\beta_i]$, $\boldsymbol{\tau}_N$ is $N \times 1$ vector of ones, and $\mathbf{c}_s = E[c_s(\boldsymbol{\Gamma})\boldsymbol{\gamma}_i]$; and we have

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

The above result holds for any finite M and t . It also holds for $M \rightarrow \infty$, if $\{a_s\}, \{b_s\}$ and $\{c_s\}$ decay sufficiently fast. If $|\lambda_i| < 1 - \epsilon$, for some strictly positive constant $\epsilon > 0$, $|\beta_i| < K$ and $\|\boldsymbol{\gamma}_i\| < K$, for some finite constant $K < \infty$ and for all i , then the distributed lag coefficients in the aggregate function decay exponentially. But if λ_i are draws from a uniform distribution, for example, 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 would exhibit long memory. Under both of these situations and for M sufficiently large, and a finite initial value, $\mathbf{w}'\mathbf{y}_{-M}$, we have

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

where $\eta_{t-s} = E(\bar{u}_{w,t-s} | \Omega_t)$. Note that $\sum_{s=1}^{\infty} a_s \eta_{t-s} = E[\sum_{s=1}^{\infty} a_s \bar{u}_{w,t-s} | \Omega_t]$. The optimal aggregate function will then be

$$\bar{y}_{wt} = \sum_{s=0}^{\infty} b_s \bar{x}_{w,t-s} + \sum_{s=0}^{\infty} \mathbf{c}'_s \mathbf{f}_{t-s} + \sum_{s=1}^{\infty} a_s \eta_{t-s} + v_{wt}. \quad (9)$$

It is important to note that the above result holds for any finite N .

3.1 Limiting behavior of the aggregate function

The persistence of the aggregate variable, \bar{y}_{wt} , to shocks depends on the decay rates of the distributed lag coefficients, $\{a_s\}, \{b_s\}$ and $\{c_s\}$. Suppose that the distribution of eigenvalues complies with the following assumption.

ASSUMPTION 3 *The conditional moments, $E(\lambda_i^s | \boldsymbol{\Upsilon}_t, \mathbf{P}, \mathbf{u}_{t-s}) = a_s$, in Assumption 1 are absolute summable, namely $\sum_{s=0}^{\infty} |a_s| < K < \infty$.*

As noted above, in the case where λ_i , the eigenvalues of Φ , are independent draws from a distribution with support that excludes unity, then a_s will be absolute summable.

Since \mathbf{u}_t is serially uncorrelated

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

But since the errors, $(u_{1t}, u_{2t}, \dots, u_{Nt})$ are cross sectionally weakly dependent under Assumption 2, $\sup_t [\text{Var}(\bar{u}_{wt})] \rightarrow 0$, as $N \rightarrow \infty$, (See Chudik, Pesaran, and Tosetti (2010)), and under Assumption 3, $\sum_{s=1}^{\infty} a_s^2 < K$. Hence, for each t , $\sum_{s=1}^{\infty} a_s \bar{u}_{w,t-s} \xrightarrow{q.m} 0$. Also since $\sum_{s=1}^{\infty} a_s \eta_{t-s} = E(\sum_{s=1}^{\infty} a_s \bar{u}_{w,t-s} | \Omega_t)$, it follows that

$$\sum_{s=1}^{\infty} a_s \eta_{t-s} \xrightarrow{q.m} 0, \quad (10)$$

and hence for each t we have

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

The behavior of the aggregation error, v_{wt} , as $N \rightarrow \infty$ depends on the nature of the processes generating x_{it} , \mathbf{f}_t , and u_{it} , as well as the degree of cross section dependence of the coefficients in Φ . To this end we postulate the following assumptions.

ASSUMPTION 4 *The micro regressors, x_{it} , for $i = 1, 2, \dots, N$, in (1) are generated accordingly to the following common factor model,*

$$x_{it} = \sum_{k=1}^{m_x} \alpha_{ik} g_{kt} + v_{xit}, \quad (11)$$

where the m_x common factors, $\mathbf{g}_t = (g_{1t}, g_{2t}, \dots, g_{m_x t})'$, are covariance stationary with absolute summable autocovariances, and $\|E(\mathbf{g}_t \mathbf{g}_t')\| < K$. The factor loadings $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im_x})'$ are independently and identically distributed across i with common mean $E(\boldsymbol{\alpha}_i) = \boldsymbol{\alpha}$, and satisfy $\|E(\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i')\| < K$. The individual specific components v_{xit} have zero mean, are uncorrelated, cross sectionally weakly dependent, and independently distributed from the remaining random variables in (1).

ASSUMPTION 5 *The common factors, \mathbf{f}_t , in (1) are stationary with absolute summable auto-covariances, and $\|E(\mathbf{f}_t\mathbf{f}_t')\| < K$.*

ASSUMPTION 6 *$\|\Phi\| \leq \rho < 1$ for all realizations of random variables in Φ , and the coefficients in ϕ_{ij} , for $i, j = 1, 2, \dots, N$, are independently distributed from \mathbf{u}_t , β and Γ . In addition, the coefficients β_i and γ_i are identically and independently distributed across i , with means $E(\beta_i) = \beta$ and $E(\gamma_i) = \gamma$, and $\|E(\gamma_i\gamma_i')\| < K$.*

ASSUMPTION 7 *The column and row norms of the matrix of eigenvectors, \mathbf{P} , are bounded.*

Assumption 4 essentially decomposes x_{it} into a cross-sectionally strongly dependent component, $\sum_{k=1}^{m_x} \alpha_{ik}g_{kt}$, and a cross sectionally weakly dependent component, v_{xit} . See, also Chudik, Pesaran, and Tosetti (2010). The independence of Λ from β and Γ implies $b_s = a_s\beta$, and $\mathbf{c}_s = a_s\gamma$. Therefore, the sequences $\{b_s\}$ and $\{\mathbf{c}_s\}$ will be absolutely summable under Assumption 3.

The following proposition establishes sufficient conditions for aggregation error to vanish as $N \rightarrow \infty$.

Proposition 1 *Under Assumptions 1-6, and for any weights vector \mathbf{w} , satisfying the granularity conditions in (2), we have*

$$\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \xrightarrow{q.m.} 0, \text{ as } N \rightarrow \infty. \quad (12)$$

If in addition Assumption 7 holds, then

$$\sum_{s=0}^{\infty} (\mathbf{w}'\Phi^s \Gamma \mathbf{f}_{t-s} - \mathbf{c}'_s \mathbf{f}_{t-s}) \xrightarrow{q.m.} 0, \quad (13)$$

$$\sum_{s=0}^{\infty} (\mathbf{w}'\Phi^s \mathbf{B} \mathbf{x}_{t-s} - b_s \bar{x}_{w,t-s}) \xrightarrow{q.m.} 0, \quad (14)$$

and the aggregation error $v_{wt} \xrightarrow{q.m.} 0$, as $N \rightarrow \infty$. See Appendix A for a proof.

Under the assumptions in Proposition 1, the aggregation error v_{wt} tends to zero in quadratic mean and we have

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

Crucial assumption for $v_{wt} \xrightarrow{q.m.} 0$ is weak dependence of innovations, which implies an appropriate bound on $\|\Sigma_u\| = \|E(\mathbf{u}_t \mathbf{u}_t')\|$, and a sufficient bound on the norm of Φ , which implies $\sum_{s=1}^{\infty} E \|\Phi^s\| \leq \sum_{s=1}^{\infty} E \|\Phi\|^s < K$. If on the other hand $\sum_{s=1}^{\infty} E \|\Phi^s\|$ is not bounded as $N \rightarrow \infty$, or \mathbf{u}_t is strongly cross sectionally dependent, then the aggregation error v_{wt} does not necessarily converge to zero and could be sizeable.

4 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 (9). 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,

$$\bar{\theta}_N = \frac{1}{N} \boldsymbol{\tau}'_N \boldsymbol{\theta} = \frac{1}{N} \boldsymbol{\tau}'_N (\mathbf{I}_N - \Phi)^{-1} \boldsymbol{\beta}, \quad (15)$$

where $\boldsymbol{\theta} = (\mathbf{I}_N - \Phi)^{-1} \boldsymbol{\beta} = (\boldsymbol{\beta} + \Phi \boldsymbol{\beta} + \Phi^2 \boldsymbol{\beta} + \dots)$ is the $N \times 1$ vector of individual long-run coefficients, and $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones. Suppose that β_i are identically distributed across i with mean β and the conditions set out in (3) are satisfied. Under these assumptions equation (7) holds and $E(\Phi^s \boldsymbol{\beta}) = E\{E[b_s(\boldsymbol{\beta}) \mathbf{B} | \Omega_t]\} = b_s \mathbf{I}_N$ for any $s = 0, 1, \dots$. Hence, the elements of $\boldsymbol{\theta}$ have a common mean, $E(\theta_i) = \theta = \sum_{\ell=0}^{\infty} b_s$, which does not depend on elements of \mathbf{P} . If, in addition, the sequence of random variables θ_i is ergodic in mean, then for sufficiently large N , $\bar{\theta}_N$ 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 λ_i are not independently distributed, and irrespective of whether micro shocks contain a common factor.

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

$$\|Var(\boldsymbol{\theta})\| = O(N^{1-\epsilon}), \text{ for some } \epsilon > 0, \quad (16)$$

in which case $\|Var(\bar{\theta}_N)\| \leq N^{-1} \|Var(\boldsymbol{\theta})\| = O(N^{-\epsilon}) \rightarrow 0$ as $N \rightarrow \infty$ and $\bar{\theta}_N \xrightarrow{q.m.} \theta$. Condition (16) need not always hold. The condition (16) can be violated if there is a high degree of dependence

of micro coefficients β_i across i , and or if there is a dominant unit in the underlying model in which case the column norm of Φ becomes unbounded in N .

The mean of β_i is straightforward to identify from the aggregate relation since $E(\beta_i) = b_0$. But further restrictions are needed for identification of $E(\lambda_i)$ from the aggregate model. Similarly to Pesaran (2003) and Lewbel (1994), independence of β_i and λ_i would be sufficient for the identification of the moments of the micro parameters, λ_i . Under the assumption that β_i and λ_i are independently distributed, all moments of λ_i can be identified as

$$E(\lambda_i^s) = \frac{b_s}{b_0}. \quad (17)$$

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 λ_i and λ_i has a beta distribution over $(0, 1)$,

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

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

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

and $\theta = b_0(p + q - 1) / (q - 1)$. Another example is uniform distribution for λ_i on interval $[a_1, a_2]$, $a_1 > -1$, $a_2 < 1$. Equation (17) can be solved to obtain (see Robinson, 1978),

$$a_1 = \frac{b_1 - \sqrt{3(b_0 b_2 - b_1^2)}}{b_0}, \quad \text{and} \quad a_2 = \frac{b_1 + \sqrt{3(b_0 b_2 - b_1^2)}}{b_0}.$$

5 Impulse Responses of Micro and Macro Shocks

This section considers the effects of micro and macro shocks in the factor augmented VAR model given by (1), but without the exogenous regressors. Including them is of little consequence for the

analysis that follows. We set $\beta = \mathbf{0}$ and write (1) as

$$\mathbf{y}_t = \Phi \mathbf{y}_{t-1} + \Gamma \mathbf{f}_t + \mathbf{u}_t. \quad (18)$$

The common factors are assumed to follow the VAR model,

$$\mathbf{f}_t = \Psi \mathbf{f}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (19)$$

where Ψ is an $m \times m$ matrix of coefficients, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{mt})'$ is the $m \times 1$ vector of macro shocks. Both, the reduced form micro innovations in \mathbf{u}_t , and the macro shocks in $\boldsymbol{\varepsilon}_t$, are assumed to be serially uncorrelated with zero means. Also without loss of generality it is assumed that \mathbf{u}_t and $\boldsymbol{\varepsilon}_{t'}$ are independently distributed for all t and t' . In addition, it is assumed that $Var(\mathbf{u}_t) = \boldsymbol{\Sigma}_u$, where $\boldsymbol{\Sigma}_u$ is a positive definite matrix such that its column sum does not expand at the rate N , namely $\|\boldsymbol{\Sigma}_u\|_1 = O(N^{1-\epsilon})$, for some positive constant $\epsilon > 0$. This condition is sufficient and necessary for weak cross section dependence of micro innovations. See Chudik, Pesaran, and Tosetti (2010). Finally, without loss of generality, the variance matrix of macro shocks $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_\varepsilon$ is assumed to be diagonal matrix. The number of lags in VAR models (18) and (19) is restricted to one only for expositional convenience.

Combining (18) and (19) we have

$$\mathbf{G} \mathbf{z}_t = \mathbf{H} \mathbf{z}_{t-1} + \mathbf{v}_t, \quad (20)$$

where $\mathbf{z}_t = (\mathbf{y}'_t, \mathbf{f}'_t)'$, $\mathbf{v}_t = (\mathbf{u}'_t, \boldsymbol{\varepsilon}'_t)'$,

$$\mathbf{G} = \begin{pmatrix} \mathbf{I}_N & -\Gamma \\ \mathbf{0}_{m \times N} & \mathbf{I}_m \end{pmatrix}, \text{ and } \mathbf{H} = \begin{pmatrix} \Phi & \mathbf{0}_{N \times m} \\ \mathbf{0}_{m \times N} & \Psi \end{pmatrix}.$$

The vector \mathbf{v}_t consists of both the micro shocks, $\{u_{jt} = v_{jt}, \text{ for } j = 1, 2, \dots, N\}$, and the macro shocks $\{\varepsilon_{jt} = v_{N+j,t}, \text{ for } j = 1, 2, \dots, m\}$. The matrix \mathbf{G} is upper triangular with ones on its main

diagonal with the inverse given by

$$\mathbf{G}^{-1} = \begin{pmatrix} \mathbf{I}_N & \mathbf{\Gamma} \\ \mathbf{0}_{m \times N} & \mathbf{I}_m \end{pmatrix}.$$

Pre-multiplying both sides of equation (20) by \mathbf{G}^{-1} yields the following reduced form VAR model for \mathbf{z}_t ,

$$\mathbf{z}_t = \mathbf{C}\mathbf{z}_{t-1} + \mathbf{G}^{-1}\mathbf{v}_t, \quad (21)$$

where $\mathbf{C} = \mathbf{G}^{-1}\mathbf{H}$. Assuming that all eigenvalues of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are within the unit circle,

$$\mathbf{z}_t = \sum_{s=0}^{\infty} \mathbf{C}^s \mathbf{G}^{-1} \mathbf{v}_{t-s},$$

where

$$\mathbf{C}^s = \begin{pmatrix} \mathbf{\Phi}^s & \sum_{\ell=0}^{s-1} \mathbf{\Phi}^{\ell} \mathbf{\Gamma} \mathbf{\Psi}^{s-\ell} \\ \mathbf{0}_{m \times N} & \mathbf{\Psi}^s \end{pmatrix}, \text{ for } s = 0, 1, 2, \dots .$$

5.1 Generalized impulse response functions

In order to analyze the effects of macro and micro shocks on the individual units or on the aggregate variable, we use the generalized impulse response functions (GIRF) proposed by Koop, Pesaran, and Potter (1996) and further developed by Pesaran and Shin (1998). The generalized impulse responses have the property of being invariant to the ordering of the variables, which is of particular importance in a large system. The GIRF of y_{it} for a unit shock to the j -th innovation, v_{jt} , is defined by

$$\mathcal{GIRF}_{ij}(s) = E(y_{i,t+s} | v_{jt} = \sqrt{\sigma_{jj}}, \mathcal{I}_{t-1}) - E(y_{i,t+s} | \mathcal{I}_{t-1}) = \sigma_{jj}^{-1/2} \mathbf{e}'_{i,v} \mathbf{C}^s \mathbf{G}^{-1} \mathbf{\Sigma} \mathbf{e}_{j,v}, \quad (22)$$

where \mathcal{I}_{t-1} is the information set at time $t-1$, $\mathbf{e}_{i,v}$ is an $(N+m) \times 1$ selection vector that selects the i -th element of $\mathbf{v}_t = (\mathbf{u}'_t, \boldsymbol{\varepsilon}'_t)'$,

$$\mathbf{\Sigma}_{N+m \times N+m} = E(\mathbf{v}_t \mathbf{v}'_t) = \begin{pmatrix} \mathbf{\Sigma}_u & \mathbf{0}_{N \times m} \\ \mathbf{0}_{m \times N} & \mathbf{\Sigma}_{\varepsilon} \end{pmatrix},$$

and σ_{jj} is the j -th diagonal element of Σ . The sub-matrix $\Sigma_\varepsilon = E(\varepsilon_t \varepsilon_t')$ was assumed to be diagonal matrix without any loss of generality. However, the off-diagonal elements of the matrix Σ_u are in general non-zero and no restrictions, besides the weak cross section dependence assumption, are placed on Σ_u .

In addition to a shock to an individual variable, we also consider the GIRF of y_{it} for a composite shock, $\mathbf{a}'_v \mathbf{v}_t$, which is given by

$$\mathcal{GIRF}_i(s, \mathbf{a}_v) = \sigma_a (\mathbf{a}'_v \Sigma \mathbf{a}_v)^{-1} \mathbf{e}'_{i,v} \mathbf{C}^s \mathbf{G}^{-1} \Sigma \mathbf{a}_v. \quad (23)$$

Normally we set $\sigma_a = (\mathbf{a}'_v \Sigma \mathbf{a}_v)^{1/2}$, in which case GIRF function (23) shows the effects of one unit composite shock. We also find it useful in the empirical application below to set the size of the composite shock to match the standard error of the innovations to the optimal aggregate function.

The analysis below distinguishes between the effects of micro and macro shocks, noting that the former are identified by $j = 1, 2, \dots, N$, and the latter by $j = N+1, N+2, \dots, N+m$. Also the effects of macro shocks on micro and macro variables can be obtained by setting $j = N+1, \dots, N+m$, and $i = 1, 2, \dots, N, N+1, \dots, N+m$ in the above expressions. But to simplify the notations we denote the effects of micro shocks by $\mathfrak{h}_{ij}(s) \equiv \mathcal{GIRF}_{ij}(s)$ for $j = 1, 2, \dots, N$, and the effects of macro shocks by $\mathfrak{g}_{ij}(s) \equiv \mathcal{GIRF}_{ij}(s)$, for $j = N+1, N+2, \dots, N+m$. Similarly, the GIRF of y_{it} for a composite micro shock, $\mathbf{a}'_u \mathbf{u}_t$, will be denoted by $\mathfrak{h}_i(s, \mathbf{a}_u) \equiv \mathcal{GIRF}_i(s, \mathbf{a}_u)$, where $\mathbf{a}_u = (\mathbf{a}'_u, \mathbf{0}'_m)'$, and the GIRF of y_{it} for a composite macro shock, $\mathbf{a}'_\varepsilon \varepsilon_t$, by $\mathfrak{g}_i(s, \mathbf{a}_\varepsilon) \equiv \mathcal{GIRF}_i(s, \mathbf{a}_\varepsilon)$, where $\mathbf{a}_\varepsilon = (\mathbf{0}'_N, \mathbf{a}'_\varepsilon)'$.

5.2 Effects of macro shocks

Using (22), it is easily seen that the effects of a composite macro shock are given by

$$\mathfrak{g}_i(s, \mathbf{a}_\varepsilon) = (\mathbf{a}'_\varepsilon \Sigma_\varepsilon \mathbf{a}_\varepsilon)^{-\frac{1}{2}} \mathbf{e}'_{i,u} \left(\sum_{\ell=0}^s \Phi^\ell \Gamma \Psi^{s-\ell} \right) \Sigma_\varepsilon \mathbf{a}_\varepsilon, \text{ for } i = 1, 2, \dots, N, \quad (24)$$

where $\mathbf{e}_{i,u}$ is an $N \times 1$ selection vector that selects the i^{th} element. If we set $\mathbf{a}_\varepsilon = \mathbf{e}_{j,\varepsilon}$, where $\mathbf{e}_{j,\varepsilon}$ is an $m \times 1$ selection vector that selects the j^{th} element of ε_t , we obtain the effects of a unit macro shock to the j -th factor on the individual units, $i = 1, 2, \dots, N$. These effects can be aggregated

across units to get the following results for the aggregate effect of the (composite) macro shock:

$$\mathbf{g}_w(s, \mathbf{a}_\varepsilon) = \sum_{i=1}^N w_i \mathbf{g}_i(s, \mathbf{a}_\varepsilon) = (\mathbf{a}'_\varepsilon \Sigma_\varepsilon \mathbf{a}_\varepsilon)^{-\frac{1}{2}} \mathbf{w}' \left(\sum_{\ell=0}^s \Phi^\ell \Gamma \Psi^{s-\ell} \right) \Sigma_\varepsilon \mathbf{a}_\varepsilon, \quad (25)$$

where as before, w_i , for $i = 1, 2, \dots, N$, are the aggregation weights.

Now consider the associated GIRF based on the optimal aggregate function, defined by

$$\mathbf{g}(s, \mathbf{a}_\varepsilon) = E[\mathbf{g}_w(s, \mathbf{a}_\varepsilon) | \Omega_t],$$

where as before $\Omega_t = (\bar{y}_{w,t-1}, \bar{y}_{w,t-2}, \dots, \mathbf{f}_t, \mathbf{f}_{t-1}, \dots)$ is the aggregate information set. Assuming N is large, Γ is independently distributed of Φ , $E(\gamma_i) = \gamma$ for all i , and that the eigenvalues of Φ satisfy Assumption 1 we have

$$\mathbf{g}(s, \mathbf{a}_\varepsilon) = (\mathbf{a}'_\varepsilon \Sigma_\varepsilon \mathbf{a}_\varepsilon)^{-1/2} \left(\sum_{\ell=0}^s a_\ell \gamma^\ell \Psi^{s-\ell} \right) \Sigma_\varepsilon \mathbf{a}_\varepsilon, \quad (26)$$

for $s = 0, 1, 2, \dots$. Note that $\mathbf{g}(s, \mathbf{a}_\varepsilon)$ does not depend on the individual micro parameters or the weights; and shows clearly how the persistence effects of the common factors (given by the powers of Ψ) compound with the heterogeneity in eigenvalues of the micro coefficient matrix, Φ , as reflected in the decay rate of a_s . The resulting aggregate effects of macro shocks can be long lasting and very sluggish, as can be seen from the Monte Carlo and the empirical results that follow.

5.3 Effects of micro shocks

Using equation (22), the effects of a unit (composite) micro shock are given by

$$\mathbf{h}_i(s, \mathbf{a}_u) = (\mathbf{a}'_u \Sigma_u \mathbf{a}_u)^{-\frac{1}{2}} \mathbf{e}'_{i,u} \Phi^s \Sigma_u \mathbf{a}_u, \text{ for } i = 1, 2, \dots, N. \quad (27)$$

where as before $\mathbf{e}_{i,u}$ is an $N \times 1$ selection vector that selects the i^{th} element of \mathbf{u}_t . For $\mathbf{a}_u = \mathbf{e}_{j,u}$, $j = 1, 2, \dots, N$, we obtain the effects of a unit shock to the j -th unit on the i -th variable. The effects of a micro shock on the aggregate variable is given by the cross section average of the individual effects in (27),

$$\mathbf{h}_w(s, \mathbf{a}_u) = \sum_{i=1}^N w_i \mathbf{h}_i(s, \mathbf{a}_u) = (\mathbf{a}'_u \Sigma_u \mathbf{a}_u)^{-\frac{1}{2}} \mathbf{w}' \Phi^s \Sigma_u \mathbf{a}_u. \quad (28)$$

By setting \mathbf{a}_u to \mathbf{w} in the above expression we obtain the impulse responses of the effects of an average micro shock, $\mathbf{w}'\mathbf{u}_t$, namely $\mathfrak{h}_w(s, \mathbf{w}) = (\mathbf{w}'\boldsymbol{\Sigma}_u\mathbf{w})^{-\frac{1}{2}} \mathbf{w}'\boldsymbol{\Phi}^s\boldsymbol{\Sigma}_u\mathbf{w}$. Note that in general $\mathfrak{h}_w(s, \mathbf{w})$ differs from the distributed lag coefficients, $a_s = E(\mathbf{w}'\boldsymbol{\Phi}^s\boldsymbol{\tau})$, in the optimal aggregate function. However, $\mathfrak{h}_w(s, \mathbf{w})$ and a_s coincide (apart from a scaling constant) if the micro shocks are uncorrelated. The advantage of using $\mathfrak{h}_w(s, \mathbf{w})$ over a_s lies in the fact that the former allows for possible weak cross section dependence in the micro errors, whilst the latter does not.

If the cross section dependence of micro units is weak conditionally on the macro shocks and their lags, then the effects of a unit micro shock on the aggregate variable become negligible, as established in the following proposition.

Proposition 2 *Suppose that the aggregation weights satisfy the granularity conditions in (2), $\|\boldsymbol{\Sigma}_u\| = O(N^{1-\epsilon})$, for some $\epsilon > 0$, and $E\|\boldsymbol{\Phi}^s\| < K$ for some constant $K < \infty$. Then for any given $j \in \{1, 2, 3, \dots\}$, and $s \in \{0, 1, 2, \dots\}$, and for any nonzero vector, $\mathbf{a}_u \neq \mathbf{0}$, the aggregate effects of a unit micro shock, given by equation (28), satisfy*

$$\lim_{N \rightarrow \infty} E|\mathfrak{h}_w(s, \mathbf{a}_u)| = 0. \quad (29)$$

Above proposition implies that effects of a shock to the j -th micro unit on the aggregate variable become negligible, namely $\lim_{N \rightarrow \infty} E|\mathfrak{h}_w(s, \mathbf{e}_{j,u})| = 0$.

6 Excessive Persistence: A Monte Carlo Investigation

The question of persistence in macro variables, such as consumer price inflation or real exchange rates, is an important issue in economics. In this section, using Monte Carlo techniques, we investigate the extent to which aggregation can generate excessive persistence in macro variables. In particular, we examine the possible sources of such persistence focussing on heterogeneity of micro parameters, persistence of unobserved common factor, or a combination of both. We employ the impulse response functions of micro and macro shocks developed in the previous section to compare the persistence of shocks under alternative scenarios, using the disaggregate and aggregate specifications.

Three different Monte Carlo designs are considered. We start with a benchmark case where the

panel data are generated from weakly dependent AR(1) micro relations. The objective here is to see how well the true persistence of the aggregate variable can be estimated from the distributed lag coefficients of the optimal aggregate equation, and to illustrate Proposition 2 by showing how fast the aggregate effects of micro shocks decay with N . The Monte Carlo design is then extended by introducing common factors into the panel, which allows us to examine the relative importance of macro shocks and the aggregation process for the persistence of the aggregates. Finally, we consider a more general setting where we also allow for neighborhood effects and examine their importance in generating persistence at the aggregate level.

6.1 Experiments based on micro relations without a common factor or neighborhood effects

Initially we begin with AR(1) micro relations considered by Granger (1980) and Robinson (1978), but we allow for weak cross section dependence of innovations,

$$y_{it} = \lambda_i y_{i,t-1} + u_{it}, \text{ for } i = 1, 2, \dots, N. \quad (30)$$

The autoregressive micro coefficients are generated as $\lambda_i \sim IIDU(0, \lambda_{\max})$, for $i = 1, 2, \dots, N$. Two options for λ_{\max} are considered, (a) $\lambda_{\max} = 0.9$ and (b) $\lambda_{\max} = 1$, that yield different rates of delay in the lag coefficients in the optimal aggregate equation. These choices are motivated by the theoretical considerations and the empirical application that follows.

Although, the original analyses of Robinson and Granger assume the idiosyncratic innovations, u_{it} , to be cross sectionally independent, as noted earlier our theoretical results continue to hold even if these innovations are cross sectionally correlated so long as the dependence is weak in the sense discussed in Chudik, Pesaran, and Tosetti (2010). To allow for weak cross section dependence in the errors, we generate $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$ according to the following spatial autoregressive process,

$$\mathbf{u}_t = \rho_u \mathbf{S}_u \mathbf{u}_t + \boldsymbol{\vartheta}_t, \quad 0 < \rho_u < 1, \quad (31)$$

where $\boldsymbol{\vartheta}_t = (\vartheta_{1t}, \vartheta_{2t}, \dots, \vartheta_{Nt})'$, $\vartheta_{it} \sim IIDN(0, \sigma_\vartheta^2)$, for $i = 1, 2, \dots, N$, and the $N \times N$ dimensional spatial weights matrix \mathbf{S}_u is

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

The parameter ρ_u is set equal to 0.4, which ensures that the errors are cross sectionally weakly dependent, and the variance σ_y^2 is set equal to $N/\text{tr}(\mathbf{\Sigma}_u)$ so that on average $\text{Var}(u_{it}) = 1$, where $\mathbf{\Sigma}_u = \mathbf{R}_u \mathbf{R}'_u$, and $\mathbf{R}_u = (\mathbf{I} - \rho_u \mathbf{S}_u)^{-1}$. Micro observations are generated using model (30) with 50 burn-in data points and the initial values, $y_{i,-50} = 0$ for all i .

We refer to relations in (30) and (31) as the disaggregate relations without a common factor or neighborhood effects, although the innovations are spatially dependent. As to the aggregate variable we use the simple cross section average, $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}$.

The optimal aggregate equation, (9), simplifies considerably under this Monte Carlo set up. There are no micro regressors and no common factors in these experiments, which implies that $b_s = \gamma_s = 0$, and, under the assumption of uniformly distributed micro AR parameters on the interval $[0, \lambda_{\max}]$, the distributed lag coefficients in the optimal aggregate equation are given by $a_s = (1 + s)^{-1} \lambda_{\max}^s$, for $s = 0, 1, 2, \dots$

The objective of Monte Carlo experiments is to estimate the distributed lag coefficients a_s for $s = 1, 2, \dots$ from disaggregate and the associated aggregate models. These coefficients in the aggregate equation are the moments of the autoregressive micro parameters.

There are several ways that a_s can be estimated, and we consider the following two options. One option is to estimate an AR(p) process for the aggregate variable, \bar{y}_t , and then compute the corresponding moving average representation. We denote this estimator by \tilde{a}_s and refer to it as the "aggregate estimator". A potential disadvantage of this approach is that it relies on the selection of the truncation lag order, p . Coefficients a_s need not be absolute summable, in which case the aggregate variable will be a long memory process and a rather large value of p might be needed to capture the slowly decaying tail effects of the lag distribution of the aggregate function. Accordingly, when selecting the lag order for the aggregate specification we set the maximum lag

order to \sqrt{T} and use the Akaike information criterion (AIC) to choose the lag order. Advantage of this macro approach is that it does not rely on correct specification of the micro relations.

An alternative estimator of $\{a_s\}$ can be based on disaggregate estimates, $\hat{\lambda}_i$. Recall that under our assumptions, $\{\lambda_i, i = 1, 2, \dots, N\}$, are random draws from the distribution of λ , and $a_s = E(\mathbf{w}'\Phi^s\boldsymbol{\tau}) = E(\lambda^s)$. Hence, for sufficiently large N and T , the distributed lag coefficients can be consistently estimated by $\hat{a}_s = \frac{1}{N} \sum_{i=1}^N \hat{\lambda}_i^s$, where $\hat{\lambda}_i$ is the least square estimator of λ_i in the autoregressions (30). The distribution F_λ is not the same as the distribution of $\hat{\lambda}_i$ when T is finite, but \hat{a}_s is consistent at least when $T \rightarrow \infty$ followed by $N \rightarrow \infty$. Both the aggregate and the disaggregate estimators are biased when T is small.

Initially, we also report the estimates of a_s based on a "representative agent" specification where the aggregate model is assumed to follow the AR(1) process with the coefficient, $E(\lambda)$. Under this representation the distributed lag coefficients of the aggregate model are given by $\check{a}_s = [E(\lambda)]^s$ and can be consistently estimated by $\left(N^{-1} \sum_{i=1}^N \hat{\lambda}_i\right)^s$. In the case where $\lambda_i \sim IIDU(0, \lambda_{\max})$, we have $\check{a}_s = (\lambda_{\max}/2)^s$, where $\lambda_{\max}/2$ is the population mean of λ_i . It is clear that for any given value of λ_{\max} , \check{a}_s will decay much faster than the true value of a_s given by $a_s = E(\lambda^s) = (s+1)^{-1} \lambda_{\max}^s$.

6.1.1 Aggregate and disaggregate estimators of a_s

We first report average (across 2000 replications) estimates of a_s using the aggregate and the disaggregate estimators, denoted by \hat{a}_s and \check{a}_s , respectively. These estimates are compared to the true values $a_s = (s+1)^{-1} \lambda_{\max}^s$, and the distributed lag coefficients for a representative unit which is given by $\check{a}_s = (\lambda_{\max}/2)^s$. These experiments were carried out for different combinations of N and T . But initially we present a graphic representation of the results for $(N, T) = (200, 100)$ in Figure 1. This sample combination was selected since it is close to the dimensions of the data sample used in the empirical section. Panel A of the figure shows the estimates computed using the true lag orders for the underlying disaggregated models, and panel B the estimates based on lagged orders selected by AIC (see below for more details). The chart on the left of Figure 1 presents the results for the experiments with $\lambda_{\max} = 0.9$, and the ones on the right relate to the case where $\lambda_{\max} = 1$. In the former case, a_s declines at a geometric rate and one would expect the aggregate and disaggregate estimators to perform reasonably well in large samples. Although in small samples both estimators are biased downward, which is the well know small sample bias in

estimating autoregressive coefficients. In line with Kendall's (1954) approximation formula,⁴ the bias is larger in experiments with $\lambda_{\max} = 1$, where there is also a long memory effect since in this case a_s decays rather slowly. As to be expected, the results in Figure 1 also show a much faster rate of decay in the estimates based on the representative unit, as compared to the other two estimators.

The plots in part B of Figure 1 are computed based on the unit-specific AR(p_i) regressions

$$y_{it} = \sum_{\ell=1}^{p_i} \phi_{i\ell} y_{i,t-\ell} + u_{it}, \text{ for } i = 1, 2, \dots, N, \quad (32)$$

where p_i is chosen by AIC with the maximum lag set to 4. The disaggregate estimator, \hat{a}_s , is then computed as

$$\hat{a}_s = \frac{1}{N} \sum_{i=1}^N \hat{\alpha}_{is}, \quad (33)$$

where $\hat{\alpha}_i(L) = \hat{\phi}_i^{-1}(L, p_i) = 1 + \sum_{s=1}^{\infty} \hat{\alpha}_{is} L^s$, $\hat{\phi}_i(L, p_i) = 1 - \hat{\phi}_{i1}L - \hat{\phi}_{i2}L^2 - \dots - \hat{\phi}_{ip_i}L^{p_i}$, and $\hat{\phi}_{i\ell}$, $\ell = 1, 2, \dots, p_i$ represent the least squares estimators of the autoregressive coefficients in (32). For the aggregate estimator, the lag order is also selected by AIC, with the maximum lag order set to \sqrt{T} . The estimates in part B of Figure 1 are very similar to those in part A, suggesting that not knowing the lag orders is not a problem for the sample size combinations under consideration.⁵

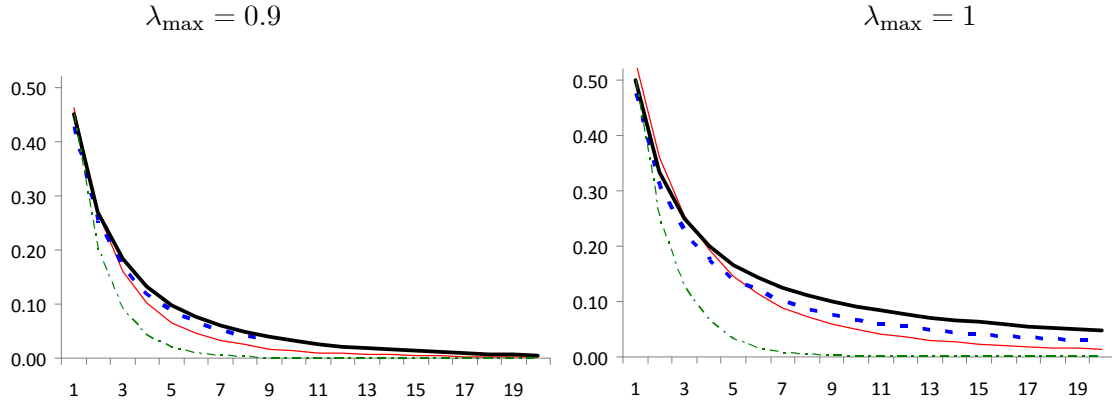
The bias and the root mean square error (RMSE) of the estimators for different values of $N \in \{10, 50, 100, 200\}$ and $T \in \{50, 100, 200\}$ are summarized in Table 1. The left part of Table 1 presents bias and RMSE for estimates of a_s averaged over horizons 1 - 12, and the right panel presents the results for the estimates averaged over horizons 13 - 24. The aggregate estimator, \tilde{a}_s , in all cases has substantially larger RMSE (up to 4-7 times) compared to the disaggregate estimator, \hat{a}_s . Such a large difference in RMSE is generally observed in both experiments, with $\lambda_{\max} = 0.9$ and 1. Overall, the disaggregate estimator, \hat{a}_s , performs much better than the aggregate estimator, \tilde{a}_s .

⁴ $E(\hat{\lambda}_i) - \lambda_i = -1/(1 + 3\lambda_i) + O(T^{-3/2})$

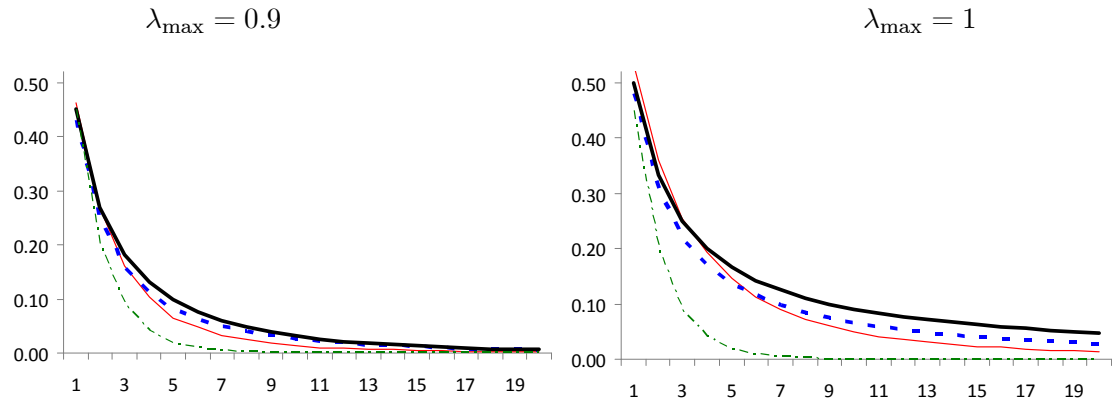
⁵The results based on the Schwarz lag orderation criterion (not reported here) were slightly better compared to the results based on the Akaike criterion.

Figure 1: Distributed lag coefficients $\{a_s\}$ of the optimal aggregate function for the experiments without a common factor and neighborhood effects; $T = 100$ and $N = 200$

Panel A: Known lag orders used for the underlying disaggregated models



Panel B: Lag orders of the underlying disaggregate models are selected by AIC



The true value is represented by solid thick black line (a_s), the disaggregate estimator is represented by thick dashed blue line (\hat{a}_s), the aggregate estimator is represented by thin red line (\tilde{a}_s), and the distributed lag coefficients of the representative agent (\check{a}_s) are represented by dashed green line.

Table 1: Bias and RMSE of aggregate and disaggregate estimators (averaged over horizons $s = 1$ to 12 and $s = 13$ to 24) of the distributed lag coefficients $\{a_s\}$ in the optimal aggregate equation. (Experiments without a common factor or neighborhood effects, for different values of N and T).

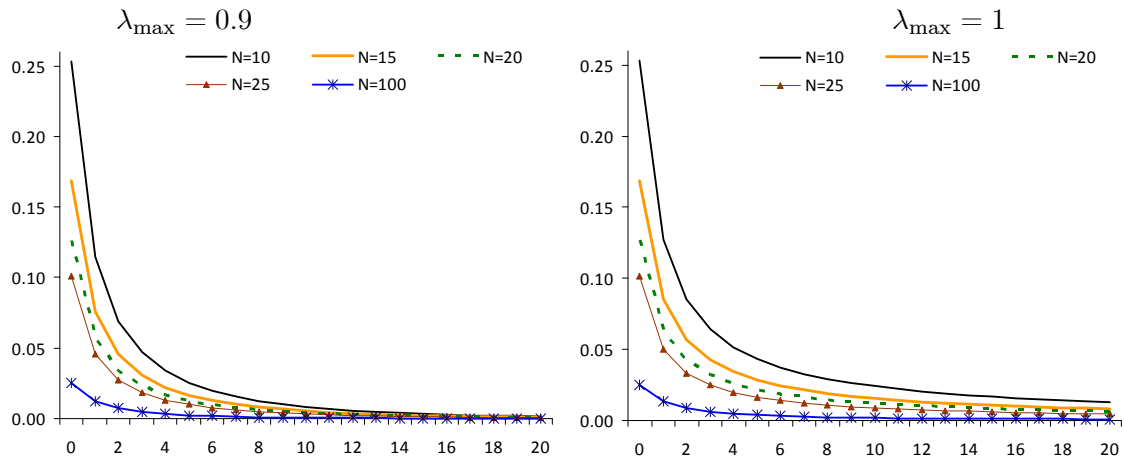
		Bias ($\times 100$)											
		Estimates averaged over horizons from $s = 1$ to 12						Estimates averaged over horizons from $s = 13$ to 24					
N \ T		50		100		200		50		100		200	
		\tilde{a}	\hat{a}	\tilde{a}	\hat{a}	\tilde{a}	\hat{a}	\tilde{a}	\hat{a}	\tilde{a}	\hat{a}	\tilde{a}	\hat{a}
(a) Experiments with $\lambda_{\max} = 0.9$													
10		-3.25	-2.45	-2.05	-1.34	-0.93	-0.63	-0.41	-0.15	-0.43	-0.12	-0.45	-0.04
50		-3.23	-2.52	-1.86	-1.27	-0.82	-0.69	-0.42	-0.17	-0.48	-0.10	-0.49	-0.07
100		-3.31	-2.54	-1.98	-1.30	-0.80	-0.71	-0.47	-0.17	-0.49	-0.11	-0.51	-0.08
200		-3.33	-2.54	-1.82	-1.31	-0.75	-0.69	-0.49	-0.17	-0.52	-0.11	-0.53	-0.07
(b) Experiments with $\lambda_{\max} = 1$ (long memory in aggregate variable)													
10		-5.91	-5.00	-2.36	-2.90	0.53	-1.33	-4.16	-3.08	-3.22	-2.23	-2.03	-1.29
50		-5.75	-5.07	-1.90	-2.72	1.79	-1.39	-4.01	-3.07	-3.55	-2.19	-1.94	-1.31
100		-5.72	-4.98	-2.06	-2.80	1.87	-1.31	-4.13	-3.06	-3.55	-2.23	-2.12	-1.25
200		-5.72	-4.99	-1.89	-2.72	2.06	-1.45	-4.08	-3.06	-3.55	-2.18	-1.88	-1.35
RMSE ($\times 100$)													
(a) Experiments with $\lambda_{\max} = 0.9$													
10		10.89	5.77	8.64	5.15	7.30	4.91	3.26	1.38	2.17	1.06	1.63	1.00
50		10.20	3.27	7.82	2.37	6.31	2.01	2.92	0.56	1.78	0.45	1.52	0.39
100		10.25	2.92	7.68	1.92	6.07	1.54	3.09	0.42	1.77	0.32	1.44	0.28
200		10.15	2.74	7.48	1.63	5.97	1.18	2.67	0.34	1.72	0.24	1.49	0.21
(b) Experiments with $\lambda_{\max} = 1$ (long memory in aggregate variable)													
10		14.08	8.69	12.62	7.89	11.83	8.05	6.44	4.96	6.49	4.25	6.37	4.48
50		13.45	5.88	10.78	4.15	10.02	3.53	7.55	3.39	5.56	2.66	5.52	2.21
100		13.33	5.40	10.73	3.55	9.47	2.61	6.34	3.25	5.54	2.47	5.19	1.77
200		13.39	5.21	10.54	3.13	9.48	2.18	6.59	3.14	5.38	2.30	5.39	1.61

Notes: The aggregate estimator (\tilde{a}_s) is computed as the moving average representation of the estimated aggregate AR(p) process. The disaggregate estimator (\hat{a}_s) is computed by aggregation of the estimated disaggregate moving average representations, and is defined by (33). Lags in the aggregate AR and the micro AR regressions are chosen by AIC criterion. The maximum lag in the aggregate regression is set equal to the integer part of \sqrt{T} , and the maximum lag in micro regressions is set to 4. Details of Monte Carlo design are described in Subsection 6.1.

6.1.2 Effects of micro shocks

Effects of a unit shock to micro innovations on the individual units could be large, but as established in Proposition 2, the aggregate effects tend towards zero as the number of cross section units is increased. This is illustrated in Figure 2, which plots the aggregate effects of a unit shock to u_{1t} for different values of N . The aggregate effects of the micro shock are largest when $N = 10$, but decrease linearly in N , and are very close to zero for $N = 100$.

Figure 2: Effects of one unit micro shock on the aggregate variable in the experiments without neighborhood effects



This figure plots function $h_w(s, 1)$ given by (28) for different values of N in experiments without neighborhood effects.

6.2 Experiments based on micro relations with a common factor but without neighborhood effects

The model in this case allows for strong cross section dependence in the residuals by considering the AR(1) models with unobserved common factor

$$y_{it} = \lambda_i y_{i,t-1} + \gamma_i f_t + u_{it}, \text{ for } i = 1, 2, \dots, N. \quad (34)$$

The factor loadings are generated as $\gamma_i \sim IIDN(0.5, 0.1)$, for $i = 1, 2, \dots, N$, and the common factor, f_t , is generated according to the following relatively persistent AR(1) process,

$$f_t = \psi f_{t-1} + \varepsilon_t, \varepsilon_t \sim IIDN(0, 1 - \psi^2), \psi = 0.9, \quad (35)$$

with 50 burn-in data points and the initialization $f_{-50} = 0$. Innovations $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$ and the autoregressive micro parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_N)'$ are generated in the same way as before.

The main objective of the present set of experiments is to consider the small sample properties of the alternative estimates of the aggregate effects of a macro shock. The GIRF function for the effects of a macro shock is given by (26). Since we have only one macro shock, ε_t , we denote the effects of the macro shock by $\mathbf{g}(s)$ in this section. These effects combine both sources of persistence, the AR coefficient, ψ , in the process for the common factor, and the heterogeneity in the micro AR coefficients, λ_i . We also examine how these two sources of persistence combine to produce the overall observed persistence of the aggregate.

6.2.1 Effects of macro shocks

Before proceeding further with the results for the aggregate and the disaggregate estimators, it is useful to illustrate how the persistence in the common factor combines with the heterogeneity in the micro AR parameters. The effects of a unit shock to ε_t in the disaggregate model with heterogeneous parameters for the case of uniformly distributed eigenvalues on the interval $[0, \lambda_{\max}]$, is given by

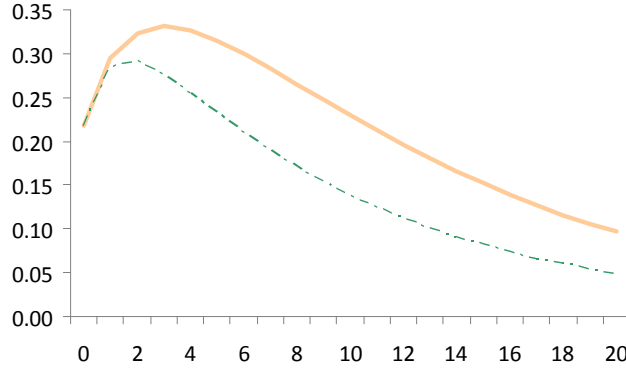
$$\mathbf{g}(s) = \kappa \sum_{\ell=0}^s \frac{\lambda_{\max}^{\ell}}{1+\ell} \psi^{s-\ell}. \quad (36)$$

where $\kappa = \sigma_{\varepsilon}^{1/2} \gamma$. The result for the representative agent model is given by

$$\check{\mathbf{g}}(s) = \kappa \sum_{\ell=0}^s \left(\frac{\lambda_{\max}}{2} \right)^{\ell} \psi^{s-\ell}, \quad (37)$$

where $\lambda_{\max}/2$ is the population mean of λ . It is clear that $\mathbf{g}(s)$ decays slower than $\check{\mathbf{g}}(s)$, particularly as s rises. Figure 3 plots the two distributed lag functions for values of $\psi = \lambda_{\max} = 0.9$. The two impulse responses are identical on impact (at $s = 0$), but begin to deviate from each other quite substantially for large values of s . It takes 19 time periods for $\mathbf{g}(s)$ to decline below half of the initial impact whereas it takes only 12 quarters for the function $\check{\mathbf{g}}(s)$ based on the representative agent model. The differences between the two functions become even larger if we consider the long memory case where $\lambda_{\max} = 1$.

Figure 3: Effects of a macro shock on the aggregate variable under dynamic heterogeneity and the associated representative agent model



The thick solid orange line represents $\mathbf{g}(s)$ for the heterogeneous dynamic model, whilst the thin dashed green line represents $\tilde{\mathbf{g}}(s)$ for the associated representative agent model, defined by (36) and (37), respectively. Both functions are evaluated at $\lambda_{\max} = \psi = 0.9$.

6.2.2 Estimation of the aggregate effects of a macro shock

We consider the aggregate estimator, denoted by $\tilde{\mathbf{g}}(s)$, which is computed using estimates based on an $\text{AR}(p)$ process fitted to the aggregate observations. The lag order is selected by AIC and the aggregate AR process is estimated by least squares. Thus apart from a scaling constant, $\tilde{\mathbf{g}}(s)$ is the same as \tilde{a}_s .

The disaggregate estimator, $\hat{\mathbf{g}}(s)$, is computed by cross section aggregation of the individual impulse responses, $\hat{\mathbf{g}}_i(s)$, namely

$$\hat{\mathbf{g}}(s) = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{g}}_i(s), \quad s = 0, 1, 2, \dots \quad (38)$$

where $\hat{\mathbf{g}}_i(s)$, $i = 1, 2, \dots, N$, provide estimates of the effects of macro shocks on individual units in the underlying disaggregate model (21). Micro AR coefficients are consistently estimated by least squares using the following cross sectionally augmented disaggregated regressions

$$y_{it} = \sum_{\ell=1}^{p_i} \phi_{i\ell} y_{i,t-\ell} + \delta_{i0} \bar{y}_t + \sum_{\ell=1}^{q_i} \delta_{i\ell} \bar{y}_{t-\ell} + \eta_{it}, \quad \text{for } i = 1, 2, \dots, N, \quad (39)$$

where p_i and q_i are chosen by AIC, with $p_{\max} = q_{\max} = 4$. The use of cross section averages, \bar{y}_t , and its lags to proxy for the unobserved common factor is justified by Chudik and Pesaran (2010). In

the present context the use of cross section averages tend to perform better than using the principal components of y_{it} . The aggregate model was approximated by the following AR(p) specification model

$$\bar{y}_t = \sum_{\ell=1}^p \pi_{\ell} \bar{y}_{t-\ell} + \varepsilon_{\bar{y}t}. \quad (40)$$

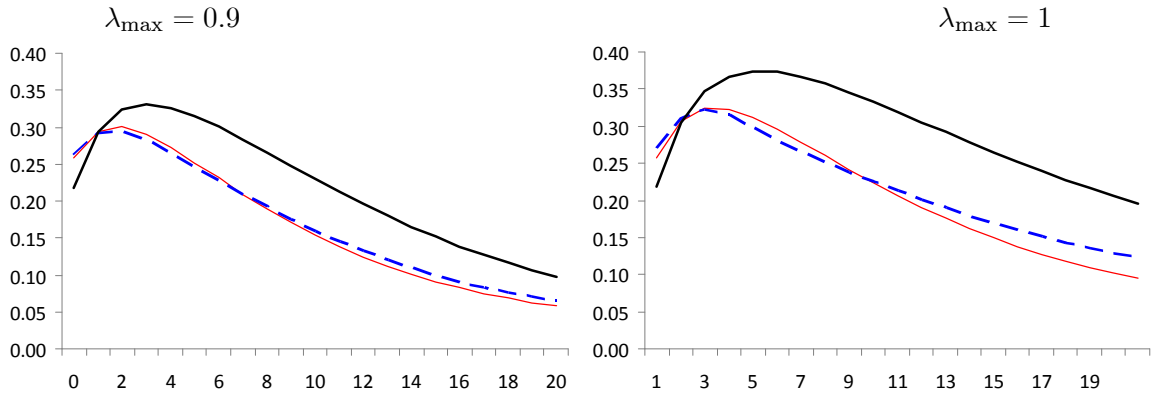
The lag order, p , is chosen by AIC with p_{\max} set to the integer part of \sqrt{T} .

6.2.3 Results for experiments with a common factor but without neighborhood effects

First we show a graphical representation of the results for the average (across 2000 replications) values of the disaggregate estimator, $\hat{\mathbf{g}}(s)$, the aggregate estimator, $\tilde{\mathbf{g}}(s)$, together with the true values, $\mathbf{g}(s)$, given by equation (36). Figure 4 reports findings for the sample size $(N, T) = (200, 100)$. The performance of both estimators in terms of bias is very similar. The disaggregate estimator slightly outperforms the aggregate estimator at longer horizons, similarly to the experiments without the common factor. The bias of both estimators, however, is quite substantial in both sets of experiments with $\lambda_{\max} = 0.9$ or 1.

RMSE together with bias for various values of $N \in \{10, 50, 100, 200\}$ and $T \in \{50, 100, 200\}$ are reported in Table 2. Unlike the results in Table 1, the difference in RMSE of the two estimators is not very large. The disaggregate estimator marginally outperforms the aggregate estimator in terms of RMSE when $\lambda_{\max} = 0.9$, but not when $\lambda_{\max} = 1$.

Figure 4: The effects of a macro shock on the aggregate variable for the experiments with a common factor but without neighborhood effects, $T = 100$, and $N = 200$



The solid thick black line represents the true values, $\mathbf{g}(s)$, the disaggregate estimator is represented by the thick dashed blue line, $\hat{\mathbf{g}}(s)$, and the aggregate estimator, $\tilde{\mathbf{g}}(s)$, is shown by the thin solid red line.

Table 2: Bias and RMSE of aggregate and disaggregate estimators (averaged over horizons $s = 1$ to 12 and $s = 13$ to 24) of the effects of macro shocks on the aggregate variable (Experiments with common factor and without neighborhood effects, for different values of N and T)

		Bias ($\times 100$)											
		Estimates averaged over horizons from $s = 1$ to 12						Estimates averaged over horizons from $s = 13$ to 24					
N \ T		50		100		200		50		100		200	
		\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}
(a) Experiments with $\lambda_{\max} = 0.9$													
10		-11.13	-11.07	-5.59	-5.67	-2.96	-3.19	-8.27	-7.94	-6.26	-5.66	-4.73	-4.00
50		-10.12	-9.96	-6.11	-6.17	-3.37	-3.66	-6.13	-5.61	-4.53	-3.78	-2.55	-1.66
100		-10.55	-10.37	-6.07	-6.12	-3.23	-3.50	-6.17	-5.59	-4.43	-3.69	-2.73	-1.82
200		-10.21	-10.01	-5.80	-5.83	-3.20	-3.44	-6.23	-5.70	-4.56	-3.81	-3.00	-2.08
(b) Experiments with $\lambda_{\max} = 1$ (long memory in aggregate variable)													
10		-13.23	-13.31	-6.37	-7.30	-1.70	-3.71	-15.46	-14.17	-11.78	-10.11	-7.51	-5.99
50		-13.67	-13.25	-8.20	-8.65	-4.24	-5.87	-13.51	-11.55	-10.29	-7.92	-6.14	-4.09
100		-14.29	-13.88	-7.98	-8.30	-4.50	-5.81	-14.08	-12.13	-10.37	-7.88	-6.88	-4.64
200		-13.88	-13.30	-7.89	-8.14	-4.33	-5.58	-14.41	-12.26	-10.40	-7.88	-6.60	-4.36
RMSE ($\times 100$)													
(a) Experiments with $\lambda_{\max} = 0.9$													
10		17.84	17.63	14.13	13.55	11.33	10.65	11.94	11.88	9.41	9.18	7.81	7.57
50		14.37	14.21	9.93	9.72	7.11	6.93	11.39	11.31	8.19	7.87	6.35	6.06
100		14.07	13.91	9.77	9.63	6.73	6.63	13.06	12.96	8.64	8.33	6.44	6.06
200		14.34	14.17	9.42	9.27	6.40	6.34	18.54	18.28	8.60	8.26	6.52	6.11
(b) Experiments with $\lambda_{\max} = 1$ (long memory in aggregate variable)													
10		21.41	21.07	17.02	16.19	13.78	12.80	19.27	19.22	16.37	15.47	13.47	12.51
50		17.86	17.59	12.74	12.81	8.91	9.54	19.73	19.22	14.33	12.85	10.16	8.66
100		18.04	17.83	12.20	12.29	8.39	8.98	19.54	18.84	14.44	12.83	10.49	8.75
200		17.75	17.45	11.82	11.94	7.83	8.53	20.36	19.93	14.22	12.54	10.23	8.48

Notes: The aggregate estimator is based on the aggregate AR(p) process. The disaggregate estimator is obtained by aggregating of the effects of the macro shock on individual units, as in (38). Lags in the aggregate and the underlying disaggregate regressions are chosen by AIC criterion with the maximum lag set equal to the integer part of \sqrt{T} and 4, respectively. Details of Monte Carlo design are described in Subsection 6.2.

6.3 Experiments with neighborhood effects

To allow for neighborhood effects we used the following data generating process

$$y_{1t} = \lambda_1 y_{1,t-1} + \gamma_1 f_t + u_{1t}, \quad (41)$$

and

$$y_{it} = \theta_i y_{i-1,t-1} + \lambda_i y_{i,t-1} + \gamma_i f_t + u_{it}, \text{ for } i = 2, 3, 4, \dots, N, \quad (42)$$

where each unit, except the first, has one left neighbor ($y_{i-1,t-1}$). The lagged coefficients, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_N)'$, the factor loadings, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_N)'$, the unobserved common factor, f_t , and the micro innovations, $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$, are generated as before. The neighborhood coefficients, θ_i , are generated as *IIDU* $(0, 1 - \lambda_i)$, for $i = 2, 3, \dots, N$, to ensure bounded variances as $N \rightarrow \infty$. Specifically, $\|\boldsymbol{\Phi}\|_\infty \leq |\lambda_i| + |\theta_i| < 1$, where (see Pesaran and Chudik (2010))

$$\boldsymbol{\Phi} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ \theta_2 & \lambda_2 & 0 & \cdots & 0 \\ 0 & \theta_3 & \lambda_3 & & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & & \theta_N & \lambda_N \end{pmatrix}.$$

The focus of the experiments is on the estimation of the aggregate effects of a macro shock. The aggregate estimator, $\tilde{\mathbf{g}}(s)$, is computed as before. Own lags and the coefficients corresponding to the neighboring unit in the disaggregate estimator $\hat{\mathbf{g}}(s)$, are consistently estimated using regressions similar to (39), but augmented with neighboring units,

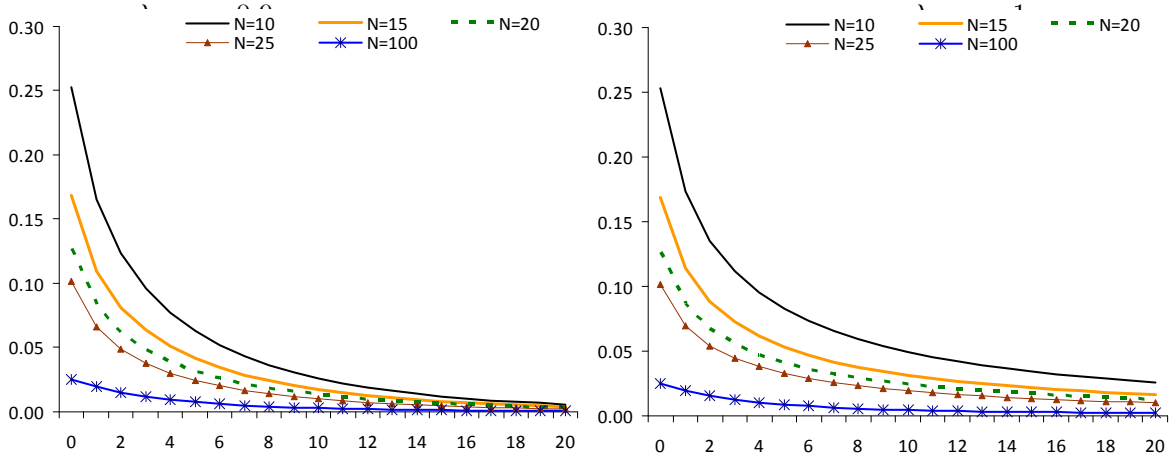
$$y_{it} = \sum_{\ell=1}^{p_i} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=1}^{p_i} \theta_{i\ell} y_{i-1,t-\ell} + \delta_{i0} \bar{y}_t + \sum_{\ell=1}^{q_i} \delta_{i\ell} \bar{y}_{t-\ell} + \eta_{it}, \text{ for } i = 2, 3, \dots, N. \quad (43)$$

Initially we used AIC criterion to select the lag orders, p_i and q_i (with the maximum lag set to 4). But we encountered unstable roots in the dynamics of the disaggregate specification. We switched to the Schwarz Bayesian Criterion (SBC), and obtained stable roots. The results below are based on SBC .

6.3.1 Effects of micro shocks in the experiments with neighborhood effects

Before presenting the results of the macro shocks, it is interesting to see if the inclusion of neighborhood effects in the disaggregated model has had any significant impact on the importance of micro shocks for the aggregate variable. According to Proposition 2 the inclusion of neighborhood effects should not affect the outcomes if N is sufficiently large and the cross section dependence induced by the neighborhood effects is weak. The impulse responses of the effects of a unit shock to u_{1t} on \bar{y}_{wt} , given by $h_w(s, 1)$ in (28), for different values of N are displayed in Figure 5. As compared to the estimates without neighborhood effects (in Figure 2), the inclusion of neighborhood effects generates more persistence, but as expected the effects of the micro shock on the aggregate variable become negligible as N increases.

Figure 5: Effects of a micro shock on the aggregate variable in experiments with neighborhood effects



This figure plots function $h_w(s, 1)$ defined by (28) for different values of N in the experiments with neighborhood effects.

6.3.2 Estimation results for experiments with neighborhood effects

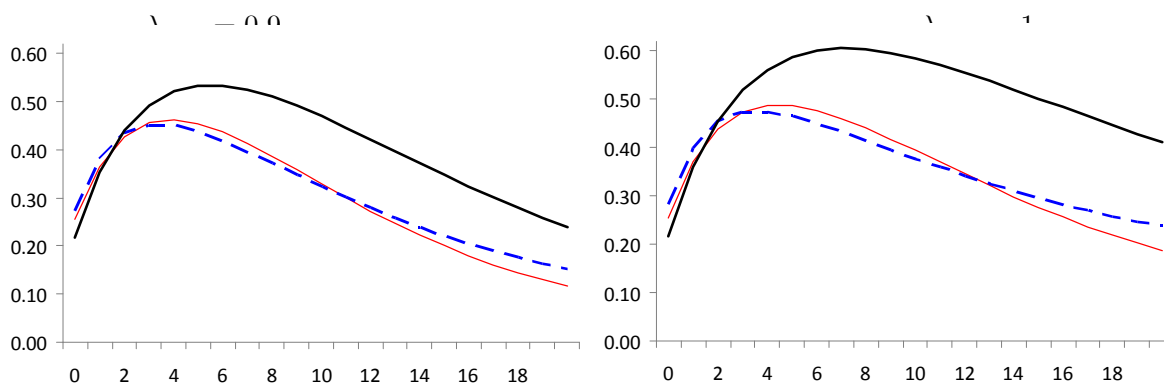
The consequence of the left neighbor in this set-up is that it is not straightforward to analytically calculate the optimal aggregate function $\mathbf{g}(s)$. Nevertheless, it is straightforward to numerically compute the effects of a macro shock, given by $\mathbf{g}_w(s)$ and defined by equation (25). Figure 6 displays the estimates of the aggregate estimator, $\tilde{\mathbf{g}}(s)$, the disaggregate estimator, $\hat{\mathbf{g}}(s)$, and $\mathbf{g}_w(s)$ computed using $\mathbf{w} = N^{-1}(1, 1, \dots, 1)'$, for $N = 200$ and $T = 100$. Similarly to the experiments without neighborhood effects, there is a significant bias in the estimation of the aggregate effects of

macro shocks. These effects are underestimated, particularly at longer horizons. The disaggregate estimator performs only marginally better in terms of bias as compared to the aggregate estimators at longer horizons.

Table 3 provides summary statistics (bias and RMSE) for different choices of N and T . These results suggest that estimation of aggregate effects of macro shocks is subject to greater sampling uncertainty when neighborhood effects are present, as compared to the results reported in Table 2. The performance of aggregate and disaggregate estimators is very similar, with the disaggregate estimator doing marginally better at longer horizons.

We also computed the disaggregate estimator, $\hat{\mathbf{g}}(s)$, based on the regressions without neighborhood units, to evaluate how the misspecification of neighbors could affect the results. The omission of neighborhood units from the disaggregate regressions seems to have little adverse effects on the performance of the disaggregate estimator.

Figure 6: The effects of a macro shock on the aggregate variable for the experiments with a common factor and neighborhood effects, $T = 100$, and $N = 200$



The solid thick black line represents the true value, $\mathbf{g}_w(s)$, the disaggregate estimator is represented by the thick dashed blue line, $\hat{\mathbf{g}}(s)$, and the aggregate estimator is represented by the thin solid red line, $\tilde{\mathbf{g}}(s)$.

Table 3: Bias and RMSE of aggregate and disaggregate estimators (averaged over horizons $s = 1$ to 12 and $s = 13$ to 24) of the effects of a unit macro shock on the aggregate variable. (Experiments with common factors and with neighborhood effects, for different values of N and T).

		Bias ($\times 100$)											
		Estimates averaged over horizons from $s = 1$ to 12						Estimates averaged over horizons from $s = 13$ to 24					
N \ T		50		100		200		50		100		200	
		\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}	\tilde{g}	\hat{g}
(a) Experiments with $\lambda_{\max} = 0.9$													
10		-13.68	-13.63	-6.17	-7.31	-1.68	-3.58	-14.05	-11.70	-10.07	-7.83	-6.93	-4.47
50		-17.89	-17.99	-9.31	-11.10	-4.70	-7.14	-18.01	-11.08	-12.70	-7.38	-7.99	-4.10
100		-17.94	-17.80	-9.15	-10.10	-4.36	-6.28	-17.89	-10.46	-13.03	-6.94	-7.99	-3.74
200		-18.22	-17.67	-8.96	-9.60	-4.25	-5.82	-18.84	-9.85	-12.81	-6.50	-7.25	-3.38
(b) Experiments with $\lambda_{\max} = 1$ (long memory in aggregate variable)													
10		-16.51	-16.96	-6.41	-9.34	-1.10	-5.82	-24.04	-21.33	-17.61	-14.57	-12.13	-10.05
50		-21.31	-21.63	-11.45	-14.13	-5.02	-9.80	-28.92	-22.78	-21.83	-17.18	-13.59	-11.48
100		-22.15	-21.67	-11.57	-13.22	-5.49	-9.13	-30.37	-24.83	-22.27	-17.86	-13.53	-11.53
200		-22.29	-21.43	-12.02	-12.99	-5.62	-8.57	-31.59	-26.66	-22.07	-18.17	-13.26	-11.21
RMSE ($\times 100$)													
(a) Experiments with $\lambda_{\max} = 0.9$													
10		24.66	23.36	17.93	16.52	13.83	12.43	24.51	22.80	16.17	13.65	12.22	9.76
50		24.63	24.19	16.34	16.97	11.36	12.39	25.64	23.56	19.11	15.42	13.73	10.82
100		24.26	24.01	15.84	16.15	10.32	11.21	28.20	25.79	19.21	16.10	13.75	11.07
200		23.98	23.55	15.05	15.30	9.63	10.37	26.36	24.53	19.13	16.28	13.48	10.99
(b) Experiments with $\lambda_{\max} = 1$ (long memory in aggregate variable)													
10		28.46	27.65	20.59	19.56	15.65	15.36	35.22	34.28	26.10	22.60	19.19	16.90
50		28.31	28.45	19.06	20.83	12.67	15.67	38.46	35.15	28.82	24.05	19.91	16.85
100		28.04	28.04	18.23	19.68	11.87	14.55	38.74	34.91	28.58	24.21	20.15	16.81
200		27.98	27.86	18.07	19.21	11.44	13.80	39.38	35.71	28.53	24.48	19.59	16.41

Notes: The aggregate estimator, $\tilde{g}(s)$, is computed as impulse response function based on the estimated aggregate AR(p) process. The disaggregate estimator, $\hat{g}(s)$, is given by aggregation of the effects of a macro shock on individual units, and is defined in (38). Own lags and coefficients corresponding to the neighboring unit are estimated using the regression given by (43). Lags in the aggregate regression are chosen by AIC criterion with the maximum lag set equal to the integer part of \sqrt{T} , and lags in the underlying disaggregate regressions are chosen by SBC criterion with the maximum lag of 4. Details of Monte Carlo design are described in Subsection 6.3.

7 Inflation Persistence: Aggregation or Common Factor Persistence

Proper understanding of aggregate inflation behavior and how it relates to individual prices at the micro level is crucial for the conduct of monetary policy. 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. so that our respective conclusions can be compared more readily.⁶ We show that dynamic heterogeneity as well as persistent common factors are needed for explaining the observed persistence of the aggregate inflation. Dynamic heterogeneity alone can not explain the persistence of the aggregate inflation, rather it is the combination of factor persistence and cross section heterogeneity that seem to be responsible for the high persistence of aggregate inflation as compared to the persistence of the underlying individual inflation series.

7.1 Data

The inflation series for the i -th price category is computed as $y_{it} = 400 \cdot [\ln(p_{it}) - \ln(p_{i,t-1})]$, where p_{it} is the seasonally adjusted consumer price index of unit i at time t . Units are individual categories of the consumer price index (e.g. bread, wine, repairs, medical services,...) and the time dimension is quarterly covering the period 1985Q1 to 2004Q2, altogether 78 quarterly observations per price category. We have data on 85 categories in Germany, 145 in France and 168 in Italy. The aggregate inflation measure is (annualized quarterly) computed as $\bar{y}_{wt} = \sum_{i=1}^N w_i y_{it}$, 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,

⁶We are grateful to Altissimo et al. for providing us with their data set.

and all measures of persistence reported below are therefore unconditional.

7.2 Disaggregate price relations without a common factor or neighborhood effects

We begin our empirical investigation by considering benchmark autoregressive processes without allowing for any strong cross section dependence amongst the different price categories. We consider separate $AR(p_i)$ processes and allow the lag order, p_i , to differ across i ,⁷

$$y_{it} = \sum_{\ell=1}^{p_i} \phi_{i\ell} y_{i,t-\ell} + u_{it}. \quad (44)$$

The optimal aggregate function (9), under the assumption that micro models are generated by (44) with weakly cross sectionally dependent innovations u_{it} , reduces to

$$\bar{y}_{wt} = \sum_{s=1}^{\infty} a_s \eta_{t-s} + v_{wt}. \quad (45)$$

The objective is to see if the persistence of the aggregate inflation as measured by a_s , can be explained in terms of the heterogeneity of $\phi_{i\ell}$ across i . It is clear that the persistence of aggregate inflation must match that of the underlying series if $\phi_{i\ell} = \phi_{\ell}$, for all i . To this end, we estimate GIRFs for the aggregate inflation using the above disaggregate and aggregate specifications. The aggregate macro shock in the case of the disaggregate inflation equations is defined as the composite shock, $\mathbf{w}'\mathbf{u}_t$, where \mathbf{w} is the vector of weights in the CPI basket, and for comparability with the aggregate model it is calibrated on impact to be equal to the standard error of the innovations in the aggregate model. For the derivations of the GIRFs see Section 5.

7.2.1 Estimation results

The fit of the estimated disaggregate relations is in most cases relatively high, and the average values for the adjusted R^2 over the different product categories are 49%, 36% and 43%, for Germany, France and Italy, respectively. The number of lags in the aggregate AR model is chosen by AIC with maximum lag set equal to the integer part of \sqrt{T} , and the number of lags in the disaggregate price relations is chosen by SBC criterion with maximum lag 4.

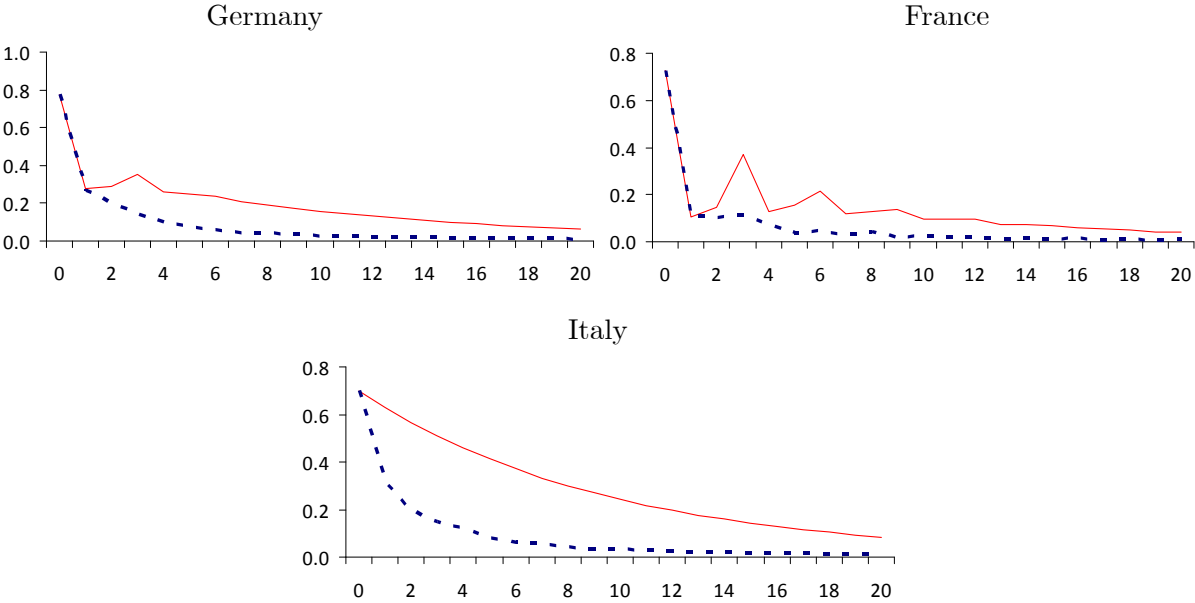
⁷An intercept is also included in the AR regressions, but omitted for expositional simplicity.

Figure 7 displays the effects of a unit shock in the aggregate model and the average aggregate effects of micro shocks using the disaggregate model. For all the three countries considered, the latter estimates are much less persistent as compared to the persistence of shocks using the aggregate model. This is in contrast to the Monte Carlo results obtained in the case of the experiments without a common factor. This could be due to the fact that the errors of the different micro equations are assumed to be weakly dependent, whilst in reality there might exist strong error cross section dependencies that impact the aggregate model and renders the individual micro relations misspecified as a result of possible missing unobserved common factors. The possibility that there are missing factors in the micro relations will be investigated below.

A similar conclusion is also reached if we consider the distributed lag coefficients of the optimal aggregate equation, (45), estimated using aggregate and disaggregate specifications. These estimates are displayed in Figure 8, and show similar patterns as in Figure 7, with the aggregate estimators being much more persistence than the disaggregate estimators.⁸

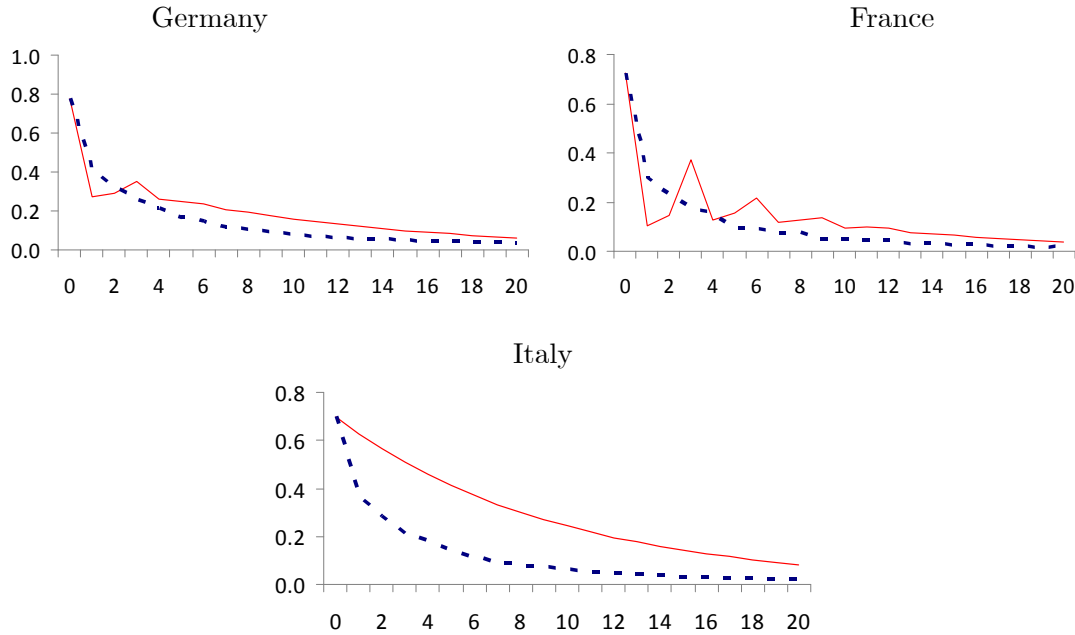
⁸To allow easier comparisons, in Figure 8 the variance of the shocks are normalized to unity in the MA representations so that a_0 and its estimates have the same value as the impulse response functions on impact.

Figure 7: IRF of the aggregate model and IRF of the average micro and macro shocks in the disaggregate model based on the price relations without neighborhood effects and without common factors.



IRF of the aggregate model is represented by thin red line and IRF of an average micro shock is shown by thick dashed blue line.

Figure 8: Distributed lag coefficients $\{a_s\}$ of the optimal aggregate equation based on the price relations without neighborhood effects and without common factors



The disaggregated estimates \hat{a}_s are represented by thick dashed blue line, and the aggregate estimates \tilde{a}_s are represented by thin solid red line.

7.3 Micro relations with common factors and neighborhood effects

Following Chudik and Pesaran (2010), we now investigate the possibility that there are missing 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 the Appendix B. An alternative possibility would be define neighbors in terms of their proximity as measured by flows of transactions between different commodity categories using input-output tables. But based on the Monte Carlo experiments reported earlier, 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 \mathcal{N}_i be the index set neighbors for unit i , and define the following local averages

$$\eta_{it} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} y_{jt} = \mathbf{s}'_i \mathbf{y}_t, \quad i = 1, 2, \dots, N, \quad (46)$$

where $|\mathcal{N}_i|$ is the number of neighbors of unit i , assumed to be small and fixed as $N \rightarrow \infty$, \mathbf{s}_i is the corresponding $N \times 1$ sparse weights vector with $|\mathcal{N}_i|$ nonzero elements. η_{it} represents the local average of unit i . No unit is assumed to be dominant in the sense discussed by Pesaran and Chudik (2010).

We follow Pesaran (2006) and its extension to dynamic panels in Pesaran and Chudik (2010), and model the effects of unobserved common factors by mean of cross section averages, at the national and sectoral levels. Accordingly, we use an economy wide average, $\bar{y}_t = N^{-1} \sum_{j=1}^N y_{jt}$, and the sectoral averages

$$\bar{y}_{\alpha t} = \frac{1}{|\mathcal{S}_\alpha|} \sum_{j \in \mathcal{S}_\alpha} y_{jt} = \mathbf{w}'_\alpha \mathbf{y}_t, \quad \text{for } \alpha \in \{f, g, s\}, \quad (47)$$

where \mathcal{S}_α is the index set of units belonging to sector $\alpha =$ food and beverages sector (f), goods sector (g) and services sector (s), $|\mathcal{S}_\alpha|$ is the number of units in sector α , and \mathbf{w}_α is the corresponding sectoral weights vector. This set up allows us to accommodate up to four unobserved common factors.

The following regressions are estimated by least squares for the price category i belonging to sector α , (intercepts are included but not shown)⁹

$$y_{it} = \sum_{\ell=1}^{p_i} \phi_{ii,\ell} y_{i,t-\ell} + \sum_{\ell=1}^{p_i^n} \delta_{i\ell} \eta_{i,t-\ell} + \sum_{\ell=0}^{\bar{p}_i} h_{i\ell} \bar{y}_{t-\ell} + \sum_{\ell=0}^{p_i^\alpha} h_{\alpha i \ell} \bar{y}_{\alpha,t-\ell} + u_{it}, \quad \text{for } i \in \mathcal{S}_\alpha. \quad (48)$$

Similar equations are also estimated for energy price categories, but without sectoral averages. It is useful to re-write equation (48) in the following way,

$$y_{it} = \phi_{ii}(L) y_{i,t-1} + \delta_i(L) \eta_{i,t-1} + \mathbf{h}'_i(L) \boldsymbol{\zeta}_t + u_{it}, \quad (49)$$

⁹We also estimated the individual price equations, (48), without the sectoral effects and obtained similar results. The inclusion of the sectoral effects introduces additional persistence at the aggregate level, which seems to be important particularly in the case of Italy.

where $\zeta_t = (\bar{y}_t, \bar{y}_{ft}, \bar{y}_{gt}, \bar{y}_{st})'$ is the 4×1 vector of national and sectoral cross section averages, and $\mathbf{h}'_i(L) = [h_i(L), h_{fi}(L), h_{gi}(L), h_{si}(L)]$. Stacking equations (49) for $i = 1, 2, \dots, N$ gives the following VAR model,

$$\mathbf{y}_t = \Phi(L)\mathbf{y}_{t-1} + \mathbf{H}(L)\zeta_t + \mathbf{u}_t, \quad (50)$$

where

$$\Phi(L) = \begin{pmatrix} \phi_{11}(L) & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \phi_{NN}(L) \end{pmatrix} + \begin{pmatrix} \delta_1(L)\mathbf{e}'_{N1} \\ \vdots \\ \delta_N(L)\mathbf{e}'_{NN} \end{pmatrix}. \quad (51)$$

$\mathbf{H}(L)\zeta_t$ is a proxy for the effects of unobserved common factor(s). Lags (p_i, p_i^n, \bar{p}_i , and p_i^α) are chosen by SBC with the maximum lag set equal to 2.

7.3.1 Estimation results

Table 4 summarizes the statistical significance of the various coefficients in the price equations (48), for Germany, France and Italy. The parameters are grouped into those characterizing own lagged effects (ϕ_{iil}), lagged neighborhood effects ($\delta_{i\ell}$), country effects (h_{ail}), and sectoral effects ($h_{\alpha i\ell}$, for $\alpha = f, g, s$). All four types of effects are statistically important, although perhaps not surprisingly own lag effects are more important statistically as compared to the other effects. At the 5% significance level, own lag 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 section 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 benchmark AR 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 section averages, compared to the own lagged coefficients. As before, the fit is relatively high in most cases. The average \bar{R}^2 is 56% in Germany, 48% in France, and 51% in Italy (median values are 61%, 52%, and 54%, respectively).

Table 4: Summary statistics for individual price relations, (48), with common factor(s) and neighborhood effects.

	No. of estimated coef.	No. of significant coef (<i>at the 5% nominal level</i>)	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%

Using the estimates of $\phi_{ii}(L)$ and $\delta_i(L)$, for $i = 1, 2, \dots, N$, we compute eigenvalues of the companion matrix corresponding to the VAR polynomial matrix $\Phi(L)$ defined in (51). The modulus of eigenvalues for Germany and France is bounded by 0.93 and all eigenvalues with the exception of one are bounded by 0.86 in Italy. If the support of the eigenvalues distribution does not cover the unit circle then regardless of the functional form of the cross section distribution of the eigenvalues, heterogeneity in eigenvalues can not generate long memory in the aggregates.

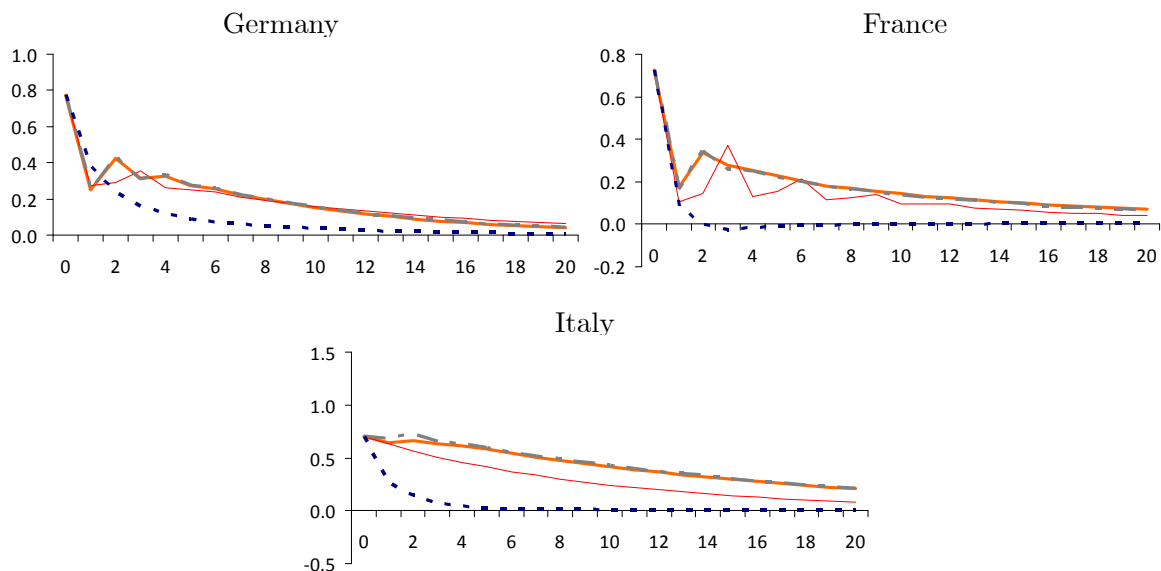
We compute the aggregate effects of micro and macro shocks along the lines explained in Section 5. To this end we estimate a VAR model in $\zeta_t = (\bar{y}_{at}, \bar{y}_{ft}, \bar{y}_{gt}, \bar{y}_{st})'$, and use it in conjunction with (50) to compute the impulse responses of the shocks to the factors and to the individual inflation equations. The aggregate effects of micro shocks are computed as before by aggregating the individual effects of shocks across units using the CPI aggregation weights. On impact the effects of aggregate and macro shocks are set equal to one standard error of the AR model for the aggregate inflation.

The estimated aggregate effects of macro shocks are reported in Figure 9 and are found to be highly persistent. The effects of shocks to factors are very similar, irrespective of whether the national or a composite of the four sectoral factors is shocked. Figure 9 also shows that the degree of persistence of the aggregate inflation is in fact underestimated by the aggregate model, as compared

to the effects of the factor shocks in the disaggregate model. In contrast, the effects of a composite micro shock decays much faster than the factor shocks, or the effects of the shocks in the aggregate model. It is the combination of factor persistence and the dynamics of individual inflation series that yield the highly persistence effects of macro shocks shown in Figure 9.

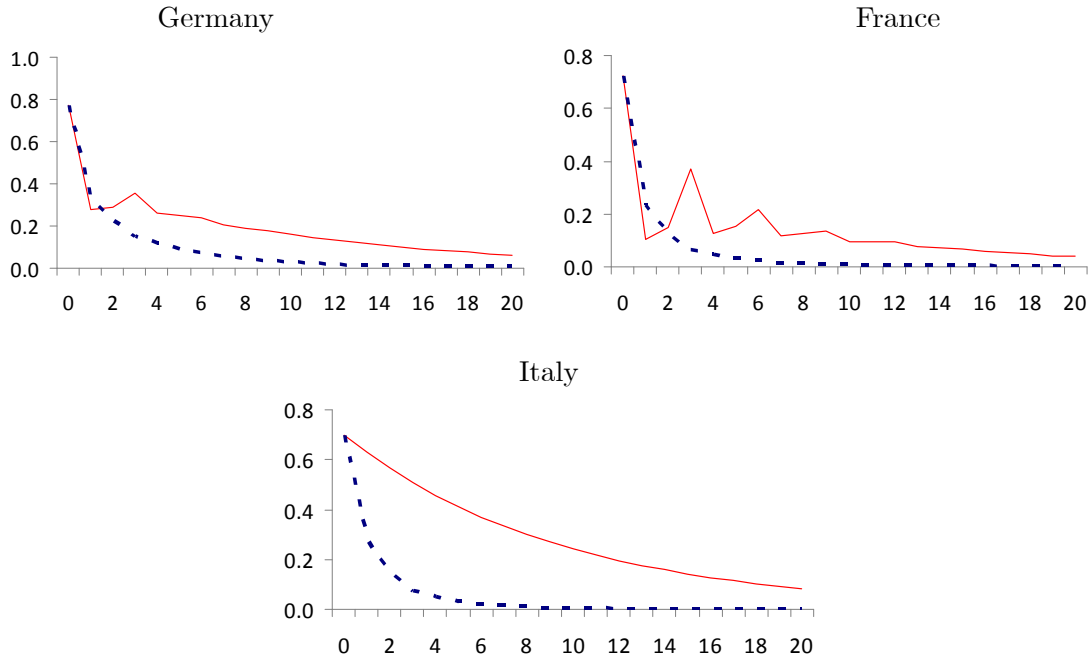
Dynamic heterogeneity alone does not seem to be sufficient for explaining the observed persistence of the aggregate inflation. This point is also apparent if we consider the distributed lag coefficients of the optimal aggregate equation. Figure 10 presents the estimates \hat{a}_s based on the underlying micro model with neighbors and common factors and compares it with the aggregate MA representation \tilde{a}_s . Coefficients are again scaled to match the initial impacts of the impulse responses in Figure 9, to allow for easier comparisons. It can be observed that the disaggregate estimates, \hat{a}_s , decay much faster than the aggregate estimates, \tilde{a}_s . The disaggregate estimates of the distributed lag coefficients decay even faster than the corresponding estimates in the benchmark AR micro model, which is not surprising and could be due to the omission of common factors in the benchmark model.

Figure 9: Impulse response functions of micro and macro shocks on aggregate inflation



The thin red line represents the IRF of the aggregate model. The other lines are based on the disaggregated model, specified by (48). The thick orange line gives the IRF of the shock to the national factor, the dashed thick gray line represents the IRF of a composite of four sectoral shocks, and the thick dashed blue line gives the IRF of an average micro shock in the disaggregate model.

Figure 10. Distributed lag coefficients $\{a_s\}$ of the optimal aggregate equation based on the individual price relations given by (48)



The aggregate estimates, \tilde{a}_s , are represented by the thin red line, and the disaggregated estimates, \hat{a}_s , by the thick dashed blue line.

7.4 Further discussion of the empirical results: aggregation or factor persistence

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.

$$y_{it} = \psi_i(L) \varepsilon_t + \varphi_i(L) u_{it},$$

where $\psi_i(L)$ and $\varphi_i(L)$ are unit-specific polynomials, ε_t is a serially *uncorrelated* unobserved common factor innovation orthogonal to u_{it} , and u_{it} is $IID(0, \sigma_i^2)$. Altissimo et al. find that the persistence of aggregate inflation originates from the unobserved common component, $\psi_i(L) \varepsilon_t$, and that the persistence of the aggregate idiosyncratic component, $\sum_{i=1}^N w_i \psi_i(L) u_{it}$, is relatively small. The latter finding is in line with our results, which shows that \hat{a}_s seems to decline at a

geometric rate in Figure 10. Their analysis focuses on the roots of $\psi_i(L)$, but does not study whether one could decompose $\psi_i(L)$ into the products $\gamma_i(L)\theta(L)$, in which case one could write $\psi_i(L)\varepsilon_t = \gamma_i(L)\theta(L)\varepsilon_t = \gamma_i(L)f_t$ where $f_t = \theta(L)\varepsilon_t$ could be viewed as a serially correlated unobserved common factor. As a result they attribute any possible persistence that might be due to the common factor(s) 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 amount 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. Results in this paper suggests that the interaction of the persistence in common factors and the eigenvalues heterogeneity is 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 and for the conduct of monetary policy.

8 Conclusion

This paper extends the literature on aggregation of linear dynamic models in a number of directions. After a brief review of Granger's contribution to the aggregation literature, 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 micro and macro shocks, allowing for weak cross section 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 heterogeneous panels which are worthy of further investigations. The present paper should be seen as a small step in this direction.

References

- Altissimo, F., B. Mojon, and P. Zaffaroni (2009). Can aggregation explain the persistence of inflation? *Journal of Monetary Economics* 56, 231–241.
- Chudik, A. and M. H. Pesaran (2010). Infinite dimensional VARs and factor models. *Forthcoming in Journal of Econometrics*.
- Chudik, A., M. H. Pesaran, and E. Tosetti (2010). Weak and strong cross section dependence and estimation of large panels. *Forthcoming in the Econometrics Journal*.
- 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(2), 206–210.
- Giacomini, R. and C. W. 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(2), 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. Routledge, 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), 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.
- Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika* 41(3), 403–404.

- Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119–147.
- Lewbel, A. (1994). Aggregation and simple dynamics. *The American Economic Review* 84(4), 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.
- Mandelbrot, B. B. and J. W. van Ness (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review* 10(4), 422–437.
- Pesaran, M. and A. Chudik (2010). Econometric analysis of high dimensional VARs featuring a dominant unit. ECB Working Paper No. 1194.
- 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., 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(4), 861–888.
- Pesaran, M. H. and Y. Shin (1998). Generalised impulse response analysis in linear multivariate models. *Economics Letters* 58, 17–29.
- 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(3), 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), 1–18.
- Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics* 120, 75–102.

A Mathematical Appendix

Proof of Proposition 1. Using (5) with $M \rightarrow \infty$, finite initial value $\mathbf{w}'\mathbf{y}_{-M}$, and under Assumption 6, which is sufficient for $\max_i |\lambda_i| < 1$, we have

$$\bar{y}_{wt} = \sum_{s=0}^{\infty} \mathbf{w}'\Phi^s (\mathbf{B}\mathbf{x}_{t-s} + \Gamma\mathbf{f}_{t-s}) + \sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s}, \quad (\text{A.1})$$

Therefore, in view of (9), (10), and (A.1) we must also have

$$v_{wt} - \sum_{s=0}^{\infty} (\mathbf{w}'\Phi^s \mathbf{B}\mathbf{x}_{t-s} - b_s \bar{x}_{w,t-s}) - \sum_{s=0}^{\infty} (\mathbf{w}'\Phi^s \Gamma\mathbf{f}_{t-s} - \mathbf{c}'_s \mathbf{f}_{t-s}) - \sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \xrightarrow{q.m} 0. \quad (\text{A.2})$$

It is now easy to see from (A.2) that $v_{wt} \xrightarrow{q.m} 0$ when results (12)-(14) hold. We establish results (12)-(14) below.

To prove (12), consider

$$\text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \right) = \text{Var} \left[E \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \mid \Phi \right) \right] + E \left[\text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \mid \Phi \right) \right].$$

Since $E(\mathbf{u}_t \mid \Phi) = 0$, for all t ,

$$\text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \right) = E \left[\text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \mid \Phi \right) \right].$$

Also since \mathbf{u}_t is serially uncorrelated (by Assumption 2),

$$\left\| \text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \mid \Phi \right) \right\| = \left\| \mathbf{w}' \left(\sum_{s=0}^{\infty} \Phi^s \Sigma_u \Phi'^s \right) \mathbf{w} \right\| \leq \|\mathbf{w}\|^2 \|\Sigma_u\|^2 \sum_{s=0}^{\infty} \|\Phi^s\|^2.$$

Hence

$$E \left[\text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \right) \right] \leq \|\mathbf{w}\|^2 \|\Sigma_u\|^2 \sum_{s=0}^{\infty} E \|\Phi^s\|^2.$$

But $\|\mathbf{w}\|^2 = O(N^{-1})$, and $\lim_{N \rightarrow \infty} N^{-1} \|\Sigma_u\|^2 = 0$, by the weak cross section dependence of innovations \mathbf{u}_t postulated in Assumption 2. Also $\sum_{s=0}^{\infty} E \|\Phi^s\|^2 < K$ under Assumption 6. Hence, it follows that $\lim_{N \rightarrow \infty} \text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \mathbf{u}_{t-s} \right) = 0$, and noting that $E(\mathbf{u}_t \mid \Phi) = 0$, for all t , completes the proof of (12).

To establish (13), consider

$$\begin{aligned} \text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \Gamma\mathbf{f}_{t-s} - \sum_{s=0}^{\infty} \mathbf{c}'_s \mathbf{f}_{t-s} \right) &= \text{Var} \left[E \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \Gamma\mathbf{f}_{t-s} - \sum_{s=0}^{\infty} \mathbf{c}'_s \mathbf{f}_{t-s} \mid \Omega_t \right) \right] + \\ &+ E \left[\text{Var} \left(\sum_{s=0}^{\infty} \mathbf{w}'\Phi^s \Gamma\mathbf{f}_{t-s} - \sum_{s=0}^{\infty} \mathbf{c}'_s \mathbf{f}_{t-s} \mid \Omega_t \right) \right]. \quad (\text{A.3}) \end{aligned}$$

But Assumption 1 implies that $E(\Phi^s \Gamma \mid \Omega_t) = E[c_s(\Gamma) \Gamma \mid \Omega_t] = \tau_N \mathbf{c}'_s$ (see the arguments used to derive equation (8)). Therefore

$$E(\mathbf{w}'\Phi^s \Gamma\mathbf{f}_{t-s} - \mathbf{c}'_s \mathbf{f}_{t-s} \mid \Omega_t) = 0, \quad (\text{A.4})$$

and the first summand on the right side of (A.3) is zero. Consider now the second term on the right side of (A.3). Let $\mathbf{V}_\Gamma = \mathbf{\Gamma} - E(\mathbf{\Gamma}) = \mathbf{\Gamma} - \tau_N \boldsymbol{\gamma}'$ and note that under Assumption 6, \mathbf{V}_Γ is distributed independently of Φ . Hence $Cov(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \boldsymbol{\gamma}' \mathbf{f}_{t-s}, \sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t) = 0$, and

$$Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{\Gamma} \mathbf{f}_{t-s} | \Omega_t\right) = Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t\right) + Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \boldsymbol{\gamma}' \mathbf{f}_{t-s} | \Omega_t\right). \quad (\text{A.5})$$

But

$$\begin{aligned} Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t\right) &= Var\left[E\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t, \Phi\right) | \Omega_t\right] \\ &\quad + E\left[Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t, \Phi\right) | \Omega_t\right] \end{aligned} \quad (\text{A.6})$$

$E(\mathbf{V}_\Gamma) = \mathbf{0}$ by construction, and

$$E\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t, \Phi\right) = 0. \quad (\text{A.7})$$

In addition, using the matrix norm inequality and noting that $\|\mathbf{V}_\Gamma \mathbf{f}_{t-s} \mathbf{f}'_{t-s} \mathbf{V}'_\Gamma\| \leq \|\mathbf{V}_\Gamma\|^2 \|\mathbf{f}_{t-s} \mathbf{f}'_{t-s}\|$ for any realization of random variables in \mathbf{V}_Γ , we have

$$\begin{aligned} Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t, \Phi\right) &= \sum_{s=0}^{\infty} \mathbf{w}' \Phi^s E\left(\mathbf{V}_\Gamma \mathbf{f}_{t-s} \mathbf{f}'_{t-s} \mathbf{V}'_\Gamma | \Omega_t, \Phi\right) \Phi^{s'} \mathbf{w} \\ &\leq \|\mathbf{w}\|^2 E\left(\|\mathbf{V}_\Gamma\|^2 | \Omega_t, \Phi\right) \sum_{s=0}^{\infty} \|\Phi\|^{2s} \|\mathbf{f}_{t-s} \mathbf{f}'_{t-s}\| \end{aligned} \quad (\text{A.8})$$

Taking expectation of (A.8) and noting that $\|\mathbf{w}\|^2 = O(N^{-1})$, $E\left(\|\mathbf{V}_\Gamma\|^2 | \Omega_t, \Phi\right) = O(1)$ (by Assumption 6, which postulates that γ_i are independently and identically distributed across i , and γ_i is independently distributed of Φ), $\sup_t E(\|\mathbf{f}_t \mathbf{f}'_t\|) = E(\|\mathbf{f}_t \mathbf{f}'_t\|) < K$ by Assumption 5, and $\sum_{s=0}^{\infty} E(\|\Phi\|^{2s}) < K$ by Assumption 6, we have

$$E\left[Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t, \Phi\right)\right] = O(N^{-1}),$$

which together with results (A.6) and (A.7) implies

$$E\left[Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \mathbf{V}_\Gamma \mathbf{f}_{t-s} | \Omega_t\right)\right] = O(N^{-1}). \quad (\text{A.9})$$

Next we consider the second summand on the right side of (A.5). We have

$$\begin{aligned} Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \boldsymbol{\gamma}' \mathbf{f}_{t-s} | \Omega_t\right) &= Var\left[E\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \boldsymbol{\gamma}' \mathbf{f}_{t-s} | \Omega_t, \mathbf{P}\right) | \Omega_t\right] \\ &\quad + E\left[Var\left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \boldsymbol{\gamma}' \mathbf{f}_{t-s} | \Omega_t, \mathbf{P}\right) | \Omega_t\right]. \end{aligned}$$

But

$$\begin{aligned} E \left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \gamma' \mathbf{f}_{t-s} \mid \Omega_t, \mathbf{P} \right) &= \sum_{s=0}^{\infty} \mathbf{w}' \mathbf{P} E(\Lambda^s \mid \Omega_t, \mathbf{P}) \mathbf{P}^{-1} \tau_N \gamma' \mathbf{f}_{t-s} \\ &= \sum_{s=0}^{\infty} a_s \mathbf{w}' \tau_N \gamma' \mathbf{f}_{t-s} = \sum_{s=0}^{\infty} a_s \gamma' \mathbf{f}_{t-s}, \end{aligned}$$

and since $\mathbf{f}_{t-s} \subset \Omega_t$ for any $s = 0, 1, 2, \dots$,

$$Var \left[E \left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \gamma' \mathbf{f}_{t-s} \mid \Omega_t, \mathbf{P} \right) \mid \Omega_t \right] = Var \left[\sum_{s=0}^{\infty} a_s \gamma' \mathbf{f}_{t-s} \mid \Omega_t \right] = 0. \quad (\text{A.10})$$

Furthermore, using the independence of λ_i and λ_j for $i \neq j$, we have

$$\begin{aligned} Var \left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \gamma' \mathbf{f}_{t-s} \mid \Omega_t, \mathbf{P} \right) &= Var \left(\sum_{s=0}^{\infty} \mathbf{w}' \mathbf{P} \Lambda^s \mathbf{P}^{-1} \tau_N \gamma' \mathbf{f}_{t-s} \mid \Omega_t, \mathbf{P} \right) \\ &= \sum_{\ell=0}^{\infty} \sum_{s=0}^{\infty} \sum_{i=1}^N \vartheta_i^2 \delta_i^2 \left[E \left(\lambda_i^{\ell+s} \mid \Omega_t, \mathbf{P} \right) - E \left(\lambda_i^s \mid \Omega_t, \mathbf{P} \right) E \left(\lambda_i^{\ell} \mid \Omega_t, \mathbf{P} \right) \right] \gamma' \mathbf{f}_{t-s} \mathbf{f}_{t-\ell} \gamma, \\ &= \sum_{i=1}^N \vartheta_i^2 \delta_i^2 \cdot \sum_{\ell=0}^{\infty} \sum_{s=0}^{\infty} (a_{s+\ell} - a_s a_{\ell}) \gamma' \mathbf{f}_{t-s} \mathbf{f}_{t-\ell} \gamma, \end{aligned} \quad (\text{A.11})$$

where ϑ_i is the i -th element $\boldsymbol{\vartheta} = \mathbf{w}' \mathbf{P}$, and δ_i is the i -th element of $\boldsymbol{\delta} = \mathbf{P}^{-1} \tau_N$. Note that $|a_s| \leq |a_{\ell}|$ for $s > \ell$ and $|a_{s+\ell}| \geq |a_s a_{\ell}|$. Therefore $|a_{s+\ell} - a_s a_{\ell}| \leq 2|a_{s+\ell}| \leq 2|a_s|$. Taking expectations of (A.11) and also noting that Assumptions 5 and 6 imply existence of a positive constant $K < \infty$, such that $\sum_{\ell=0}^{\infty} |E(\gamma' \mathbf{f}_{t-s} \mathbf{f}_{t-\ell} \gamma)| < K$, where the constant K does not depend on s, t or ℓ , we have

$$\begin{aligned} E \left[Var \left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \tau_N \gamma' \mathbf{f}_{t-s} \mid \Omega_t, \mathbf{P} \right) \right] &\leq E \left(\sum_{i=1}^N \alpha_i^2 \delta_i^2 \right) \sum_{\ell=0}^{\infty} \sum_{s=0}^{\infty} 2|a_s| |E(\gamma' \mathbf{f}_{t-s} \mathbf{f}_{t-\ell} \gamma)| \\ &\leq E \left(\sum_{i=1}^N \alpha_i^2 \delta_i^2 \right) \sum_{s=0}^{\infty} \left(2|a_s| \sum_{\ell=0}^{\infty} |E(\gamma' \mathbf{f}_{t-s} \mathbf{f}_{t-\ell} \gamma)| \right) \\ &\leq 2K \cdot E \left(\sum_{i=1}^N \alpha_i^2 \delta_i^2 \right) \cdot \sum_{s=0}^{\infty} |a_s| \rightarrow 0, \text{ as } N \rightarrow \infty, \end{aligned} \quad (\text{A.12})$$

where $\sum_{s=0}^{\infty} |a_s| < K$, under Assumption 3, and $\sum_{i=1}^N \alpha_i^2 \delta_i^2 = O(N^{-1})$ under Assumption 7 and by the granularity conditions on \mathbf{w} . It follows from (A.10) and (A.12) that

$$E \left[Var \left(\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \Gamma \mathbf{f}_{t-s} \mid \Omega_t \right) \right] \rightarrow 0, \text{ as } N \rightarrow \infty,$$

which, together with result (A.4) imply that $\sum_{s=0}^{\infty} \mathbf{w}' \Phi^s \Gamma \mathbf{f}_{t-s} - \sum_{s=0}^{\infty} \mathbf{c}'_s \mathbf{f}_{t-s} \xrightarrow{q.m.} 0$, as $N \rightarrow \infty$. This completes the proof of result (13).

Result (14) is established in a similar way as results (12)-(13). Under Assumption 4, the micro regressors are given by the factor model (11). Assumptions about the factors \mathbf{g}_t and the loadings $\boldsymbol{\alpha}_i$ are the same as the assumptions about the factors \mathbf{f}_t and the loadings $\boldsymbol{\gamma}_i$. Therefore, using similar arguments as in the proof

of result (13), it can be shown that

$$\sum_{s=0}^{\infty} \left(\mathbf{w}' \Phi^s \mathbf{B} \sum_{k=1}^{m_x} \boldsymbol{\alpha}_{.k} g_{kt} - b_s \sum_{k=1}^{m_x} E(\boldsymbol{\alpha}_k) g_{kt} \right) \xrightarrow{q.m.} 0, \text{ as } N \rightarrow \infty, \quad (\text{A.13})$$

where $\boldsymbol{\alpha}_{.k} = (\alpha_{1k}, \alpha_{2k}, \dots, \alpha_{Nk})'$. The same assumptions are postulated for innovations v_{xit} and u_{it} . In particular, the innovations v_{xit} are cross sectionally weakly dependent (and uncorrelated). Hence, using similar arguments as in the proof of result (12), it can be shown that

$$\sum_{s=0}^{\infty} (\mathbf{w}' \Phi^s \mathbf{B} \mathbf{v}_{xt}) \xrightarrow{q.m.} 0, \text{ as } N \rightarrow \infty. \quad (\text{A.14})$$

Using results (A.13) and (A.14), and noting that $\bar{x}_{wt} \xrightarrow{q.m.} E(x_{it})$ completes the proof of (14). ■

Proof of Proposition 2. Taking the absolute values of (28) and applying the matrix norm inequality yield

$$|\mathfrak{h}_w(s, \mathbf{a}_N)| \leq \|\mathbf{w}\| \|\Phi^s\| \left\| \frac{\boldsymbol{\Sigma}_u \mathbf{a}_N}{(\mathbf{a}'_N \boldsymbol{\Sigma}_u \mathbf{a}_N)^{1/2}} \right\|, \text{ for } s = 0, 1, 2, \dots, \quad (\text{A.15})$$

and for every possible realization of the random elements in the matrix Φ . $\boldsymbol{\Sigma}_u = \text{Var}(\mathbf{u}_t)$ is symmetric and nonnegative definite and therefore there exists a matrix \mathbf{C} such that $\boldsymbol{\Sigma}_u = \mathbf{C}\mathbf{C}'$.

$$\left\| \frac{\boldsymbol{\Sigma}_u \mathbf{a}_N}{(\mathbf{a}'_N \boldsymbol{\Sigma}_u \mathbf{a}_N)^{1/2}} \right\| = \left\| \mathbf{C} \frac{\mathbf{C}' \mathbf{a}_N}{\|\mathbf{C}' \mathbf{a}_N\|} \right\| \leq \|\mathbf{C}\| \frac{\|\mathbf{C}' \mathbf{a}_N\|}{\|\mathbf{C}' \mathbf{a}_N\|} \leq \|\mathbf{C}\| \quad (\text{A.16})$$

The assumption of weak cross section dependence of micro innovations implies existence of a positive constant $\epsilon > 0$ such that $\|\mathbf{C}\| = [\lambda_1(\boldsymbol{\Sigma}_u)]^{1/2} = O(N^{(1-\epsilon)/2})$. Using inequality (A.16) in (A.15), taking expectations, and using the condition $E\|\Phi^s\| < K$, yields

$$E|\mathfrak{h}_w(s, j)| \leq K \|\mathbf{w}\| \|\mathbf{C}\| = O(N^{-\frac{\epsilon}{2}}),$$

for any $j = 1, 2, 3, \dots$, and any $s = 0, 1, \dots$, where $\|\mathbf{w}\| = O(N^{-1/2})$ by granularity conditions in (2). Result (29) now easily follows. ■

B Groups defining neighboring product categories

Group	List of product categories
1	Meat; Fish and seafood; Milk, cheese and eggs
2	Fruit; Vegetables
3	Sugar, jam, honey, chocolate and confectionery; Food products n.e.c.
4	Coffee, tea and cocoa; Mineral waters, soft drinks, fruit and vegetable juices
5	Spirits; Wine; Beer
6	Clothing materials; Garments; Other articles of clothing and clothing accessories; Footwear including repair
7	Services for the maintenance and repair of the dwelling; Water supply; Refuse collection; Sewerage collection; Other services relating to the dwelling n.e.c.
8	Electricity; Gas; Liquid fuels; Solid fuels; Heat energy
9	Repair of furniture, furnishings and floor coverings; Repair of household appliances; Repair of audio-visual, photographic and information processing equipment; Maintenance and repair of other major durables for recreation and culture
10	Major household appliances whether electric or not and small electric household appliances; Glassware, tableware and household utensils; Tools and equipment for house and garden; Non-durable household goods.
11	Pharmaceutical products; Other medical products; therapeutic appliances and equipment
12	Medical services; paramedical services; Dental services; Hospital services; Out-patient services
13	Motor cars; Motor cycle; Spares parts and accessories for personal transport equipment; Bicycles
14	Passenger transport by railway; Passenger transport by road; Passenger transport by air; Passenger transport by sea and inland waterway; Combined passenger transport; Other purchased transport services
15	Equipment for the reception, recording and reproduction of sound and pictures; Photographic and cinematographic equipment and optical instruments; Information processing equipment; Recording media
16	Recreational and sporting services; Cultural services
17	Books; Newspapers and periodicals; Miscellaneous printed matter; stationery and drawing materials
18	Restaurants, cafés and the like; Canteens
19	Package holidays; Accommodation services
20	Jewelry, clocks and watches; Other personal effects
21	Insurance connected with the dwelling; Insurance connected with health; Insurance connected with transport Other insurance