Economics Letters 124 (2014) 341–347

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Estimating aggregate autoregressive processes when only macro data are available

Eric Jondeau [a,*], Florian Pelgrin [b]

[a] Swiss Finance Institute and Faculty of Business and Economics (HEC Lausanne), University of Lausanne, CH 1015 Lausanne, Switzerland

[b] EDHEC Business School, France

HIGHLIGHTS

- The aggregation of individual AR(1) models is an infinite AR process.
- We estimate the aggregate process when only macro data is available.
- A parametric and a minimum distance estimator for the aggregate dynamics are proposed.
- The estimators recover the moments of the distribution of the AR parameters.
- The estimators perform very well, even with finite samples.

ARTICLE INFO

Article history:
Received 22 March 2014
Received in revised form
11 June 2014
Accepted 13 June 2014
Available online 20 June 2014

JEL classification:
C2
C13

Keywords:
Autoregressive process
Aggregation
Heterogeneity

ABSTRACT

The aggregation of individual random AR(1) models generally leads to an AR(∞) process. We provide two consistent estimators of aggregate dynamics based on either a parametric regression or a minimum distance approach for use when only macro data are available. Notably, both estimators allow us to recover some moments of the cross-sectional distribution of the autoregressive parameter. Both estimators perform very well in our Monte-Carlo experiment, even with finite samples.

1. Introduction

Aggregation is a critical and widely acknowledged issue in theoretical and empirical economics research. As noted by Pesaran and Chudik (2014), among the different aspects of the aggregation problem, the identification and estimation of certain distributional features of the micro-parameters from aggregate relations are important issues, especially when only macro data are available (Robinson, 1978; Granger, 1980; Forni and Lippi, 1997). Notably, identifying such features requires the researchers to derive the optimal aggregate function and to make explicit the back out between "macro" and "micro" parameters. Yet, only a few papers have examined the reliability of macro information in circumventing the aggregation bias in the presence of unobserved micro heterogeneity (Lewbel, 1994; Pesaran, 2003; Carvalho and Dam, 2010; Mayoral, 2013). Our paper contributes to this stream of the literature by providing a solution to this problem for autoregressive models when the time-series and cross-sectional dimensions are both large.

We propose two consistent estimation techniques that rely on a flexible parametric specification of the distribution of the micro-parameters and on the estimation of the hyper-parameters of this cross-sectional distribution. The first method is based on maximum likelihood estimation, while the second method is based on minimum distance estimation. Both methods explicitly account for the set of non-linear restrictions that drive the aggregate parameters and allow us to recover reliable information on the distribution of the micro-parameters. Using Monte Carlo simulation, we show that both methods perform very well, even with relatively small samples.

* Corresponding author. Tel.: +41 21 692 33 49.
E-mail addresses: eric.jondeau@unil.ch (E. Jondeau), florian.pelgrin@edhec.edu (F. Pelgrin).

http://dx.doi.org/10.1016/j.econlet.2014.06.012
0165-1765/

2. The model

Consider the random AR(1) model for $i = 1, \ldots, N$:

$$x_{i,t} = \rho_i\, x_{i,t-1} + \upsilon_{i,t}, \tag{1}$$

where ρ_i denotes an individual-specific random parameter and $\upsilon_{i,t}$ is an error term. For instance, such dynamics may represent consumption expenditures across households (Lewbel, 1994), consumer price inflation across subindices (Altissimo et al., 2009), real exchange rates across sectors (Imbs et al., 2005), or real marginal cost across industries (Imbs et al., 2011). Innovation $\upsilon_{i,t}$ is decomposed into a common component (ϵ_t) and an idiosyncratic (individual-specific) component ($\eta_{i,t}$):

$$\upsilon_{i,t} = \kappa_i\, \epsilon_t + \eta_{i,t}, \tag{2}$$

where κ_i denotes a scaling parameter. The macro variable results from the aggregation of micro-units, with the use of time-invariant nonrandom weights $W_N = (w_1, \ldots, w_N)'$, with $\sum_{i=1}^N w_i = 1$, so that $X_{N,t} = \sum_{i=1}^N w_i x_{i,t}$. The cross-sectional moments of ρ are $\tilde{E}_N(\rho^s) = \sum_{i=1}^N w_i \rho_i^s$, for all $s = 1, 2, \ldots$. Moreover, the following assumptions hold:

Assumption 1. $|\rho| \le c < 1$ almost surely for some constant c. Random parameters have finite variance and higher moments.

Assumption 2. ϵ_t and $\eta_{i,t}$ are white noise processes with mean zero and variance σ_ϵ^2 and σ_η^2, respectively; ϵ_t and $\eta_{i,t}$ are mutually orthogonal at any lag and lead; $\{\epsilon, \eta_i\}$ and $\{\rho_i, \kappa_i\}$ are mutually independent for all i; ρ_i and κ_i are mutually independent; $E(\kappa) = 1$.

Assumption 3. As $N \to \infty$, $\|W_N\| = O\left(N^{-1/2}\right)$ and $w_i / \|W_N\| = O\left(N^{-1/2}\right)$ for all $i \in \mathbb{N}$.

Assumption 1 guarantees that there are no individual unit root parameters that would dominate at the aggregate level (Zaffaroni, 2004). This assumption implies that the limit aggregate (as $N \to \infty$) has a short memory with an exponentially decaying autocorrelation function.[1] Eqs. (1) and (2) together with Assumption 2 provide a parsimonious form of (statistical) cross-sectional dependence, which is common in the aggregation literature (Forni and Lippi, 1997; Zaffaroni, 2004). The aggregation mechanism depends solely on the characteristics of the common component of the error term, i.e., our specification and assumptions rule out the presence of an idiosyncratic component at the aggregate level.[2] Assumption 3 is a granularity condition, which insures that the weights are not dominated by a few of the cross-sectional units (Gabaix, 2011; Pesaran and Chudik, 2014).[3]

3. Aggregate dynamics

Using the moving average (MA) representation in Eqs. (1)–(2), we can straightforwardly show that the aggregate process, $X_{N,t}$, has the following dynamics[4]:

$$X_{N,t} = \sum_{k=0}^{\infty}\left(\sum_{i=1}^{N} w_i \rho_i^k \kappa_i\right)\epsilon_{t-k} + \sum_{k=0}^{\infty}\left(\sum_{i=1}^{N} w_i \rho_i^k \eta_{i,t-k}\right). \tag{3}$$

When N becomes large, by virtue of the strong law of large moments, the limit aggregate dynamics is obtained.

Proposition 1. *Suppose that Assumptions 1–3 hold. Given the disaggregate model defined in Eqs. (1)–(2), the limit aggregate process as $N \to \infty$ has the following dynamics:*

$$X_t = \sum_{s=0}^{\infty} \gamma_s\, \epsilon_{t-s} \quad (MA\ form), \tag{4}$$

$$X_t = \sum_{s=1}^{\infty} C_s\, X_{t-s} + \epsilon_t \quad (AR\ form), \tag{5}$$

where $X_{N,t} \xrightarrow{L^2} X_t$ and $\tilde{E}_N(\rho^s) \xrightarrow{a.s.} \tilde{E}(\rho^s)$ as $N \to \infty$. Parameters γ_s are defined as $\gamma_s = \tilde{E}(\rho^s)$, with $\sum_{s=0}^\infty |\gamma_s| < \infty$. Parameters C_s are defined by $C_0 = 1$, $C_s = \tilde{E}(c_s)$, $\forall s \ge 1$ with $c_1 = \rho$ and $c_{s+1} = (c_s - C_s)\,\rho$, with $\sum_{s=0}^\infty |C_s| < \infty$.

Proof. Gonçalves and Gouriéroux (1988) and Lewbel (1994). See Appendix A.

The absence of an idiosyncratic component in Eqs. (4) and (5) is a direct consequence of Assumptions 1–3. Eq. (4) shows that the impulse-response coefficients γ_s are the noncentral moments of the random parameter ρ.[5] Eq. (5) shows that aggregation leads to an infinite autoregressive model for X_t (see Robinson, 1978 and Lewbel, 1994). The autoregressive parameters C_s are nonlinear transformations of the noncentral moments of ρ and satisfy the following nonhomogeneous difference equations (for $s \ge 1$), which turn out to be useful in the estimation with only macro data:

$$C_{s+1} = \tilde{E}(c_{s+1}) = \tilde{E}\left(\rho^{s+1}\right) - \sum_{r=1}^{s} C_r \tilde{E}\left(\rho^{s-r+1}\right), \tag{6}$$

and $\sum_{s=1}^\infty C_s = \sum_{s=1}^\infty \tilde{E}(\rho^s)/(1 + \sum_{s=1}^\infty \tilde{E}(\rho^s)) < 1$ almost surely. In addition, the long-run multiplier is given by $1/(1 - \sum_{s=1}^\infty C_s) = \sum_{s=0}^\infty \tilde{E}(\rho^s)$. With the exception of a degenerate distribution for ρ (Dirac distribution), the aggregate dynamics is richer than the individual dynamics because of the nonergodicity of the random AR(1) process. Conversely, when parameters C_s are known or estimated, the cross-sectional moments can be easily deduced. For instance, the cross-sectional mean and variance are $\tilde{E}(\rho) = C_1$ and $\tilde{V}(\rho) = C_2$, respectively, and the standardized skewness and kurtosis are $\tilde{S}(\rho) = (C_3 - C_1 C_2)/(C_2)^{3/2}$ and $\tilde{K}(\rho) = (C_4 - 2C_1C_3 + C_1^2 C_2 + C_2^2)/(C_2)^2$, respectively.

4. Estimation

The estimation approach that was originally proposed by Lewbel (1994) consists in truncating the infinite sums in Proposition 1 and estimating the resulting dynamics:

$$X_{N,t} = \sum_{s=1}^{K} C_s\, X_{N,t-s} + V_{N,t}, \tag{7}$$

[1] Assumption 1 can be relaxed to allow for long-memory effects. This point is further discussed in Section 5.

[2] The contribution of idiosyncratic shocks through network effects or nongranularity has been discussed in recent papers (e.g., Gabaix, 2011 and Acemoglu et al., 2012).

[3] Our results extend to the case of (time-varying) stochastic weights. Such an extension requires at least that the weights be distributed independently from the stochastic process defining the random variable.

[4] Put differently, it is an ARMA(N, $N-1$) in the absence of common roots in the individual processes (Granger and Morris, 1976).

[5] Noncentral moments $\gamma_s = \tilde{E}(\rho^s)$ of any (nondegenerate) random variable ρ, defined on [0, 1), satisfy: $1 > \gamma_1 \ge \cdots \ge \gamma_s \ge 0$, $\forall s \ge 1$, and $\gamma_s \to 0$ as $s \to \infty$. See Appendix B for additional properties of noncentral moments.

where K is a truncation lag. In doing so, the first K moments of the distribution function of ρ can be backed out from the estimates of $C_1, \ldots, C_K$.

Two issues are of particular concern in the unrestricted estimation of the aggregate model in Eq. (7). First, the problem of the truncation remainder may be critical in small samples, as any additional lag requires the estimation of one more parameter. Second, the implications of the theoretical relations between aggregate parameters C_s and cross-sectional moments γ_s (Eq. (6)) are not explicitly taken into account. In fact, two sets of restrictions should be imposed on the estimation of the aggregate parameters in Eq. (7). On the one hand, taking the functional dependence between C_s and γ_s, one needs to impose autoregressive parameters to be consistent with the moments γ_s defining a well-behaved cross-sectional distribution. This restriction results in the positivity of the infinite Hankel matrix associated with Choquet's representation (Choquet, 1969; Gonçalves and Gouriéroux, 1988) and thus the nonnegativity of the principal minors associated to the Hankel matrix. On the other hand, as micro-parameters ρ_i are assumed to be defined over [0, 1), the moments γ_s have a set of restrictions, described by Hausdorff's moment conditions. In Appendix B, we provide details on these restrictions. We also show that such restrictions and the so-called "restricted estimator" help at improving the estimation of cross-sectional moments, although the bias remains large especially for the cross-sectional skewness and kurtosis. To some extent, this result is consistent with the view that the identification of the micro-parameters can often be obtained only by imposing more structure on the micro-processes as for instance on the distribution of the parameter that drives heterogeneity (see Forni and Lippi, 1997).

4.1. The parametric estimator

A natural way to circumvent the issues raised by the unrestricted approach is to adopt a parametric representation of the cross-sectional distribution, so that the parameters defining the distribution can be estimated from the aggregate equation. Several densities have been proposed to describe the distribution of an autoregressive parameter, such as a uniform distribution (Linden, 1999), a Beta distribution (Granger, 1980; Gonçalves and Gouriéroux, 1988), and a polynomial distribution (Chang, 2006). We assume that the random parameter ρ is i.i.d. and drawn from a parametric distribution, $f(\rho;\theta)$, parameterized by θ.

Definition 1. Suppose that Assumptions 1–3 hold and that $\rho \sim f(\rho;\theta)$. The parametric estimator of the aggregate autoregressive model in Eq. (5) is the ML estimator of the parameter set $\xi_P = \left(\theta, \sigma_\epsilon^2\right)'$ in the model:

$$X_{N,t} = \sum_{s=1}^{K} C_s X_{N,t-s} + \epsilon_t, \tag{8}$$

where $C_s = \tilde{E}\left(\rho^s\right) - \sum_{r=1}^{s-1} C_r \tilde{E}\left(\rho^{s-r}\right)$ and $\tilde{E}(\rho^s) = \int \rho^r f(\rho;\theta)d\rho$.

The asymptotic distribution of the estimator $\hat{\xi}_P$ is $\sqrt{T}(\hat{\xi}_P - \xi_0) \to N(0, J_0^{-1} I_0 J_0^{-1})$, where ξ_0 is the true value of the parameter, $I_0 = \lim_{T\to\infty} V\left(\sqrt{T}\frac{\partial}{\partial\xi}\log L_T(\xi_0)\right)$, $J_0 = \lim_{T\to\infty}\left(\frac{\partial^2}{\partial\xi_i\partial\xi_j}\log L_T(\xi_0)\right)$, and $L_T(\xi_0)$ is the likelihood of the model evaluated at ξ_0.

An obvious advantage of the parametric estimator is that it greatly reduces the number of unknown parameters to be estimated. Moreover, the Beta or polynomial distributions are sufficiently flexible to capture most of the possible shapes of the distribution of the micro-parameters.

The limit aggregate model (Eq. (5)) involves an infinite number of parameters. Berk (1974) has shown that standard $\sqrt{T}$-consistent and asymptotically normally distributed estimates are obtained when an infinite autoregressive process is approximated by a finite autoregressive process, such as an AR(K), as long as the truncation lag K does not increase too much or too slowly with respect to the sample size. While standard selection criteria generally produce severe finite-sample biases in this situation, a general-to-specific approach can be implemented in order to provide a data-dependent rule (Kuersteiner, 2005).

4.2. The minimum distance estimator

Using the infinite MA representation of the limit aggregate process (Eq. (4)), we also define a minimum distance (MD) estimator. To circumvent the dimensionality problem, we proceed as before and rely on a parametric distribution for ρ. This approach allows us to compute all of the terms involving $\tilde{E}(\rho^s)$ in a straightforward manner. The MD estimator aims to minimize the distance between the theoretical moments of $X_{N,t}$ and their empirical counterparts.

Definition 2. Suppose that Assumptions 1–3 hold and that $\rho \sim f(\rho;\theta)$. The MD estimator of the moving average model in Eq. (4) is the parameter set $\xi_{MD} = \left(\theta, \sigma_\epsilon^2\right)'$ that minimizes the distance $(\hat{\Gamma} - \Gamma(\xi))'\hat{\Omega}(\hat{\Gamma} - \Gamma(\xi))$, where $\Gamma(\xi)$ and $\hat{\Gamma}$ denote the set of the first k auto-covariances of $X_{N,t}$ and their empirical counterparts, respectively, and $\hat{\Omega}$ is a weighting matrix. $\Gamma(\xi) = \{\mathrm{Cov}(X_{N,t}, X_{N,t-h})\}_{h=1,\ldots,k}$, with $\mathrm{Cov}\left(X_{N,t}, X_{N,t-h}\right) = \sigma_\epsilon^2 \sum_{s=0}^{\infty} \gamma_s\ \gamma_{s+h}$ and $\gamma_s = \tilde{E}(\rho^s)$.

Given the standard optimal weighting matrix $\Omega = \Sigma^{-1}$, where Σ is the asymptotic covariance matrix of the sample auto-covariances, the asymptotic distribution of $\hat{\xi}_{MD}$ is $\sqrt{T}(\hat{\xi}_{MD} - \xi_0) \to N(0, (D_0'\Sigma^{-1}D_0)^{-1})$, where $D_0 = \partial\Gamma(\xi)/\partial\xi$ is evaluated at ξ_0.

For parametric and MD estimators, the flexible parametric representation of the parameter distribution reduces the number of unknown parameters and ensures that the autoregressive parameters in the aggregate dynamics are consistent with the theoretical constraints. Therefore, these estimators are less likely to suffer from finite-sample biases and overfitting.

5. Simulation experiment

We perform a Monte Carlo simulation to evaluate the finite-sample and large-sample properties of our proposed estimators. The simulation is designed as follows. We simulate the data generating process given by Eqs. (1) and (2). We consider different values for the number of micro-units ($N = 25$ and 100) and for the number of observations ($T = 250$ and 1000). We assume that ρ is generated by a bell-shaped Beta distribution, with different sets of parameters: $(p, q) = (5, 5)$, (7.5, 2.5), and (8.5, 1.5).[6] These sets of parameters correspond to an average autoregressive parameter of $\tilde{E}(\rho) = 0.5$, 0.75, and 0.85, respectively.[7] All experiments are based on 1000 samples. We then compute the average value of the first four moments of ρ over the simulated samples.

Table 1 reports the results for the three different values of (p, q), which correspond to medium, low, and high persistence. The first

[6] The Beta distribution of a random variable Y is defined as $f(y;\theta) = \frac{1}{B(p,q)}\frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}}$, where $a \le y \le b$. The noncentral moments of the standardized variable $X = (Y-a)/(b-a)$ are given by $m_s(X) \equiv \tilde{E}(X^s) = B(p+s, q)/B(p, q)$. The Beta distribution generates long memory when $q \le 1$.

[7] We also relax Assumption 1 and allow for long-memory effects (e.g., $p = 9$ and $q = 1$). Results provide evidence that the cross-sectional skewness and kurtosis might be biased in the presence of long-memory. This result can be explained by the facts that (i) we cannot impose the relevant constraint for long-range dependence, namely that the (infinite) sum of the autoregressive parameters equals 1 in the presence of long-memory, and that (ii) observing long-range dependence in an aggregate series consistent with micro evidence proves to be difficult (Beran, 1994). Results are not reported here but are available upon request.

Table 1
Estimation of the aggregate dynamics.

Moments	True value	ML	Unrestricted regression	Parametric regression	MDE	ML	Unrestricted regression	Parametric regression	MDE
		Case 1: $N = 25, T = 250$				**Case 2:** $N = 25, T = 1000$			
$(p, q) = (\mathbf{5}, \mathbf{5})$									
Mean	0.500	0.505	0.484	0.487	0.476	0.518	0.499	0.502	0.498
Variance	0.023	–	0.040	0.022	0.022	–	0.028	0.020	0.021
Skewness	0.000	–	9.290	0.040	0.056	–	11.489	0.006	0.012
Kurtosis	2.539	–	−15.259	2.629	2.646	–	−101.769	2.618	2.615
$(p, q) = (\mathbf{7.5}, \mathbf{2.5})$									
Mean	0.750	0.771	0.734	0.739	0.735	0.785	0.750	0.754	0.752
Variance	0.017	–	0.038	0.018	0.016	–	0.027	0.015	0.016
Skewness	−0.638	–	10.275	−0.436	−0.376	–	8.488	−0.562	−0.548
Kurtosis	3.103	–	24.882	2.968	2.961	–	103.144	3.095	3.066
$(p, q) = (\mathbf{8.5}, \mathbf{1.5})$									
Mean	0.850	0.868	0.831	0.835	0.836	0.889	0.850	0.852	0.852
Variance	0.012	–	0.038	0.015	0.011	–	0.024	0.011	0.011
Skewness	−1.084	–	6.632	−0.758	−0.606	–	4.909	−0.963	−0.937
Kurtosis	4.164	–	23.508	3.648	3.458	–	238.678	4.000	3.956
		Case 3: $N = 100, T = 250$				**Case 4:** $N = 100, T = 1000$			
$(p, q) = (\mathbf{5}, \mathbf{5})$									
Mean	0.500	0.504	0.483	0.487	0.476	0.514	0.496	0.498	0.494
Variance	0.023	–	0.039	0.021	0.021	–	0.028	0.021	0.022
Skewness	0.000	–	14.237	0.043	0.054	–	12.898	0.013	0.018
Kurtosis	2.539	–	−191.714	2.640	2.647	–	−108.330	2.607	2.589
$(p, q) = (\mathbf{7.5}, \mathbf{2.5})$									
Mean	0.750	0.766	0.729	0.735	0.735	0.781	0.746	0.750	0.748
Variance	0.017	–	0.039	0.018	0.014	–	0.025	0.016	0.017
Skewness	−0.638	–	15.523	−0.425	−0.356	–	15.002	−0.562	−0.541
Kurtosis	3.103	–	−87.568	2.945	2.951	–	−29.148	3.063	3.033
$(p, q) = (\mathbf{8.5}, \mathbf{1.5})$									
Mean	0.850	0.862	0.828	0.831	0.839	0.886	0.847	0.850	0.851
Variance	0.012	–	0.036	0.015	0.009	–	0.025	0.011	0.011
Skewness	−1.084	–	1.836	−0.732	−0.580	–	5.015	−0.970	−0.930
Kurtosis	4.164	–	240.777	3.562	3.432	–	230.569	3.978	3.904

The table reports the results of the Monte Carlo simulation experiments for the estimation of the persistence parameters in the aggregate dynamics. For each case, the table displays the estimates of the mean value of the ML estimator $\bar{\rho}$ and the first four moments of ρ estimated using the heterogeneity-correcting techniques. The experiments are based on 1000 samples of $N = (25; 100)$ and $T = (250; 1000)$. Truncation lag is $K = 4$ for the unrestricted estimator (as in Lewbel, 1994) and $T/20$ for the parametric estimator. The MD estimator is based on the first $T/10$ auto-covariances of $X_{N,t}$ and the infinite sum in Definition 2 is truncated to 100 terms.

column corresponds to the theoretical value of the first four moments of the true distribution of ρ. For comparative purposes, in the second column, we report the ML estimates of the aggregate hybrid model when heterogeneity is ignored, i.e., $X_{N,t} = \bar{\rho} X_{N,t-1} + \epsilon_t$. This estimator is arguably inconsistent (see Pesaran and Smith, 1995; Imbs et al., 2011).[8] However, it provides a benchmark from which we can evaluate the ability of the proposed estimators to overcome the aggregation bias. For the three levels of persistence, the ML estimator exhibits limited finite-sample bias ($T = 250$, Cases 1 and 3). Although the theoretical analysis shows that this estimator asymptotically overestimates the true value of the parameter, $\tilde{E}(\rho)$, the well-known downward bias of the autoregressive parameter in finite sample compensates for this effect, even in a correctly specified model (Sawa, 1978). For large sample ($T = 1000$, Cases 2 and 4), as expected, the ML estimator clearly overestimates the true value of $\tilde{E}(\rho)$.

The table also reports the results for the heterogeneity-correcting approaches. To evaluate these approaches, we report the first four moments of the cross-sectional distribution of ρ ($\tilde{E}(\rho), \tilde{V}(\rho), \tilde{S}(\rho), \tilde{K}(\rho)$). In the unrestricted approach, the first four moments are directly implied by the first four autoregressive parameters $\hat{C}_1, \ldots, \hat{C}_4$. As the table shows, this approach performs reasonably well for the first two moments but provides excessively large estimates (in absolute values) of the skewness and kurtosis.

In contrast, the parametric and MD approaches yield almost unbiased estimates for all of the distribution moments. These unbiased estimates are obtained irrespective of the persistence. In the case with high persistence ($\tilde{E}(\rho) = 0.85$), the heterogeneity-correcting estimators slightly underestimate the skewness (in absolute value) and the kurtosis of the cross-sectional distribution even with large sample. Finally, we observe that the finite-sample and large-sample properties of these estimators are not affected by the use of a relatively small number of micro-units N.

6. Conclusion

This paper opens several avenues for future research. First, the relation between micro and macro parameters can be exploited in a different manner. When disaggregate data are available, the dynamics of the aggregate process can be determined completely (Jondeau and Pelgrin, 2014). Second, the proposed methods could be used to assess whether a forecast of the aggregate variable using heterogeneity-correcting estimates produces an improved prediction mean squared error over a forecast of the aggregate variable using ML or GMM estimation of structural or time-series models.

Appendix A. Proof of Proposition 1

We consider the infinite MA representation of the random AR(1) model, for $i = 1, \ldots, N$:

$$x_{i,t} = \sum_{s=0}^{\infty} \rho_i^s v_{i,t-s} = \sum_{s=0}^{\infty} \rho_i^s \kappa_i \epsilon_{t-s} + \sum_{s=0}^{\infty} \rho_i^s \eta_{i,t-s}.$$

[8] It can be shown that when T and N tend jointly or sequentially to infinity, the limit in probability of the ML estimator of $\bar{\rho}$ is $\tilde{E}(\rho/(1-\rho))/\tilde{E}(1/(1-\rho)) > \tilde{E}(\rho)$. See Appendix C.

Then the aggregate process, $X_{N,t}$ is defined as:

$$X_{N,t} = \sum_{s=0}^{\infty}\sum_{i=1}^{N} w_i \rho_i^s v_{i,t-s} = \sum_{s=0}^{\infty}\left(\sum_{i=1}^{N} w_i \rho_i^s \kappa_i\right)\epsilon_{t-s} + \sum_{s=0}^{\infty}\left(\sum_{i=1}^{N} w_i \rho_i^s \eta_{i,t-s}\right).$$

As $N \to \infty$, by virtue of the strong law of large numbers and Assumptions 1–3,

$$\left(\sum_{i=1}^{N} w_i \rho_i^s \kappa_i\right) \overset{a.s}{\to} \tilde{E}\left(\rho^s \kappa\right) = \tilde{E}\left(\rho^s\right)\tilde{E}\left(\kappa\right) = \tilde{E}\left(\rho^s\right),$$

$$\left(\sum_{i=1}^{N} w_i \rho_i^s \eta_{i,t-s}\right) \overset{a.s}{\to} \tilde{E}\left(\rho^s \eta\right) = 0.$$

Therefore, we find Eq. (4) in Proposition 1:

$$X_t = \sum_{s=0}^{\infty} \gamma_s\, \epsilon_{t-s},$$

with $\gamma_s = \tilde{E}(\rho^s)$ and $\sum_{s=0}^{\infty}\gamma_s^2 < \infty$. Moreover, Assumption 1 insures that $\sum_{s=0}^{\infty}|\gamma_s| < \infty$.

Remark 1. Following the representation of Choquet (1969), $X_{N,t}$ results from the aggregation of random AR(1) processes if and only if:

$$\forall n,\ \forall \alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{R}, \quad \sum_{i=0}^{n}\sum_{j=0}^{n}\alpha_i \alpha_0 \gamma_{i+j} \geq 0.$$

It turns out that this condition is equivalent to the positivity of the infinite Hankel matrix (see Gonçalves and Gouriéroux, 1988). Therefore, the principal minors of the infinite Hankel matrix must be nonnegative for all $s \geq 0$:

$$\det\left(\mathcal{H}_s\right) \geq 0,$$

where

$$\mathcal{H}_s = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_s \\ \gamma_1 & \gamma_2 & & \gamma_{s+1} \\ \vdots & & \ddots & \vdots \\ \gamma_s & \gamma_{s+1} & \cdots & \gamma_{2s} \end{pmatrix},$$

with $\gamma_0 = 1$.

From Brockwell and Davis (1991), the infinite AR representation (5) in Proposition 1 follows:

$$X_t = \sum_{s=1}^{\infty} C_s\, X_{t-s} + \epsilon_t,$$

where $C_0 = 1$, $C_s = \tilde{E}(c_s)$, $\forall s \geq 1$ with $c_1 = \rho$, $c_{s+1} = (c_s - C_s)\,\rho$, and $\sum_{s=0}^{\infty}|C_s| < \infty$.

Remark 2. Taking the first K autoregressive coefficients, we obtain the first K moments of the cross-sectional distribution of ρ, by writing relation (6) as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -C_1 & 1 & 0 & \cdots & 0 \\ -C_2 & -C_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -C_K & -C_{K-1} & \cdots & -C_1 & 1 \end{pmatrix}\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_K \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_K \end{pmatrix},$$

or, equivalently, $\Sigma\Gamma = \delta$. Therefore, we have $\Gamma = \Sigma^{-1}\delta$ and the conditions on the principal minors of the Hankel matrix can be imposed (see Remark 1).

Remark 3. The proof of the convergence in mean square of the aggregate process is the same as in Pesaran and Chudik (2014).

Appendix B. Restricted estimator

In this appendix, we briefly describe our strategy to estimate the "restricted estimator" briefly presented in Section 4. In the regression

$$X_t = \sum_{s=1}^{K} C_s\, X_{t-s} + \epsilon_t,$$

the sequence $\{C_s\}_{s=1}^{K}$ should satisfy the condition that the implied moments are consistent with a well-defined distribution. To do this, we deduce the sequence of moments $\gamma_s = \tilde{E}(\rho^s)$ from the estimate of C_s for $s = 1, \ldots, K$, by solving the relation:

$$C_{s+1} = \gamma_{s+1} - \sum_{r=1}^{s} C_r\, \gamma_{s-r+1},$$

with $C_1 = \gamma_1 = \tilde{E}(\rho)$. Then, we compute the moment conditions defined as the nonnegativity of the determinant of the corresponding Hankel matrices, $|\mathcal{H}_s| \geq 0$, for all $s = 1, \ldots, \lfloor K/2 \rfloor$ (see Gonçalves and Gouriéroux, 1988, and Appendix A). In other words, for a given number of lags K, the parameters $\{C_s\}_{s=1}^{K}$ will be consistent with a well-defined cross-section distribution for ρ_i if the resulting Hankel matrices all have a nonnegative determinant $|\mathcal{H}_s| \geq 0$, for $s = 1, \ldots, \lfloor K/2 \rfloor$.

In addition, as parameters ρ_i are drawn from a distribution over $[0, 1)$, there are restrictions that moments γ_s must satisfy to be consistent with such a distribution, a problem known as Hausdorff's moment problem. These additional restrictions are given by $(-1)^n \Delta^n \gamma_s \geq 0$, for $s, n = 0, 1, \ldots, K$, where Δ is the first difference operator applied to the γ_s terms (i.e., $\Delta\gamma_s = \gamma_s - \gamma_{s-1}$) (see Shohat and Tamarkin, 1943).[9]

Using these inequalities, the "restricted estimator" is defined as:

$$\hat{\gamma} = \underset{\gamma \in \Theta^\star}{\texttt{argmin}} \sum_{t=K+1}^{T} \epsilon_t^2(\gamma),$$

where $\gamma = (\gamma_1, \ldots, \gamma_K)'$ and $\epsilon_t(\gamma) = X_t - \sum_{s=1}^{K} C_s(\gamma) X_{t-s}$. The $C_s(\gamma)$ terms are defined from the mapping between the autoregressive and moving average coefficients. The constrained set $\Theta^\star$ is given by:

$$\Theta^\star = \{\gamma : |\mathcal{H}_s| \geq 0 \text{ for } s = 1, \ldots, \lfloor K/2 \rfloor \text{ and } (-1)^n \Delta^n \gamma_s \geq 0 \text{ for } s, n = 0, 1, \ldots, K\}.$$

Taking this definition, we proceed with Monte Carlo simulations and estimate the so-called "restricted estimator". The results reported in Table A.1 show that this estimator performs much better than the unrestricted estimator. In particular, it clearly reduces the range of the skewness estimators. It also produces kurtosis estimates that are restricted to be positive. However, the range of values for the kurtosis is still extremely large and unrealistic compared to the expected values. The estimator still compares unfavorably with the parametric and MD estimators.

All in all, Monte Carlo simulations suggest that imposing restrictions on coefficients implied by the aggregation process is not enough to significantly improve the efficiency and especially the fit of the cross-sectional moments. To some extent, this result is consistent with the view that the identification of micro-parameters can often be obtained only by imposing more structure on the micro-processes as for instance on the distribution of the parameter that drives heterogeneity (see also Forni and Lippi, 1997).

[9] For $n = 0$, we have $\gamma_s \geq 0$, for all $s = 1, 2, \ldots$. For $n = 1$, we have $\gamma_s \leq \gamma_{s-1}$, for all $s = 2, 3, \ldots$. For $n = 2$, we have $\gamma_s - \gamma_{s-1} \leq \gamma_{s-1} - \gamma_{s-2}$, for all $s = 3, 4, \ldots$.

Table A.1
Estimation of the aggregate dynamics.

Moments	True value	Unrestricted regression	Restricted regression	Unrestricted regression	Restricted regression
		Case 1: $N = 25, T = 250$		**Case 2:** $N = 25, T = 1000$	
$(p, q) = (\mathbf{5}, \mathbf{5})$					
Mean	0.500	0.484	0.482	0.499	0.499
Variance	0.023	0.040	0.033	0.028	0.029
Skewness	0.000	9.290	−3.513	11.489	−3.065
Kurtosis	2.539	−15.259	318.489	−101.769	211.098
$(p, q) = (\mathbf{7.5}, \mathbf{2.5})$					
Mean	0.750	0.734	0.729	0.750	0.749
Variance	0.017	0.038	0.031	0.027	0.025
Skewness	−0.638	10.275	−1.112	8.488	0.876
Kurtosis	3.103	24.882	155.583	103.144	125.184
$(p, q) = (\mathbf{8.5}, \mathbf{1.5})$					
Mean	0.850	0.831	0.825	0.850	0.847
Std dev.	0.012	0.038	0.029	0.024	0.022
Variance	−1.084	6.632	−2.622	4.909	0.059
Kurtosis	4.164	23.508	77.123	238.678	47.027
		Case 3: $N = 100, T = 250$		**Case 4:** $N = 100, T = 1000$	
$(p, q) = (\mathbf{5}, \mathbf{5})$					
Mean	0.500	0.483	0.482	0.496	0.496
Variance	0.023	0.039	0.033	0.028	0.028
Skewness	0.000	14.237	−3.229	12.898	−2.444
Kurtosis	2.539	−191.714	297.319	−108.330	219.626
$(p, q) = (\mathbf{7.5}, \mathbf{2.5})$					
Mean	0.750	0.729	0.725	0.746	0.745
Variance	0.017	0.039	0.032	0.025	0.024
Skewness	−0.638	15.523	−0.938	15.002	2.287
Kurtosis	3.103	−87.568	153.801	−29.148	130.639
$(p, q) = (\mathbf{8.5}, \mathbf{1.5})$					
Mean	0.850	0.828	0.822	0.847	0.844
Std dev.	0.012	0.036	0.027	0.025	0.024
Variance	−1.084	1.836	−2.292	5.015	0.420
Kurtosis	4.164	240.777	86.224	230.569	39.298

The table reports the results of the Monte Carlo simulation experiments for the estimation of the persistence parameters in the aggregate dynamics. For each case, the table displays the first four moments of ρ estimated using the unrestricted and restricted Lewbel approaches. The experiments are based on 1000 samples of $N = (25; 100)$ with $T = 250$. Truncation lag is $K = 4$ (as in Lewbel, 1994).

Appendix C. Maximum likelihood estimator

Suppose that the true model results from the aggregation of random AR(1) model and thus the (limit) aggregate process is (Proposition 1):

$$X_t = \sum_{s=1}^{\infty} C_s X_{t-s} + \epsilon_t,$$

where $C_0 = 1, C_s = \tilde{E}(c_s),\ \forall s \geq 1$ and $c_1 = \rho, c_{s+1} = (c_s - C_s)\rho$. Ignoring heterogeneity leads to the following AR(1) model on aggregate data:

$$X_t = \bar{\rho} X_{t-1} + \epsilon_t.$$

The (conditional) maximum likelihood estimator of $\bar{\rho}$,

$$\hat{\rho}_{\text{ML}} = \left(\sum_{t=2}^{T} X_{t-1}^2\right)^{-1} \left(\sum_{t=2}^{T} X_t X_{t-1}\right),$$

is no longer a consistent estimator of the expected value $\tilde{E}(\rho)$ (cross-sectional mean) as $T \to \infty$ given the nonergodicity of the random coefficient AR(1) model (Robinson, 1978). This result can be shown by studying the joint or sequential limit of the conditional maximum likelihood estimator of $\bar{\rho}$. In doing so, we define:

$$S_{T,N} = \left(\sum_{t=2}^{T} X_{N,t-1}^2\right)^{-1} \left(\sum_{t=2}^{T} X_{N,t} X_{N,t-1}\right),$$

where $X_{N,t}$ is the sum of the individual processes (without loss of generality), $X_{N,t} = \sum_{i=1}^{N} x_{i,t}$ with $x_{i,t} = \rho_i x_{i,t-1} + v_{i,t}$. Then, if N is fixed and $T \to \infty$, $\hat{\rho}_{\text{ML}}$ is not a consistent estimator of $\tilde{E}(\rho)$.

Proposition 1 (*Limit when N is Fixed and* $T \to \infty$). *Suppose that* $\tilde{E}\left((1 - |\rho|)^{-3}\right) < \infty$. *Then, for any N:*

$$S_{N,T} \underset{T\to\infty}{\overset{p}{\to}} S_N = \left(\sum_{i=1}^{N} \frac{1}{1-\rho_i^2}\right)^{-1} \left(\sum_{i=1}^{N} \frac{\rho_i}{1-\rho_i^2}\right).$$

Proof. See Jondeau and Pelgrin (2014).

Moreover, using this proposition, the sequential limit of the conditional likelihood estimator can be characterized.

Proposition 2 (*Sequential Limit as* $N \to \infty$). *Suppose that* $\tilde{E}\left((1 - |\rho|)^{-3}\right) < \infty$. *Then:*

$$S_N \underset{N\to\infty}{\overset{p}{\to}} \left(\tilde{E}\left(\frac{1}{1-\rho^2}\right)\right)^{-1} \tilde{E}\left(\frac{\rho_i}{1-\rho^2}\right).$$

Proof. See Jondeau and Pelgrin (2014).

Remark 4. The convergence of $S_{N,T}$ when T is fixed and $N \to \infty$ leads to:

$$S_{N,T} \underset{T\to\infty}{\overset{d}{\to}} \tilde{S}_T,$$

where $\tilde{S}_T = \left(\sum_{t=2}^{T} X_{t-1}^2\right)^{-1} \left(\sum_{t=2}^{T} X_t X_{t-1}\right)$, $X = (X_1, \ldots, X_T)' \sim \mathcal{N}_T(0, \Omega)$ with $E(X_t) = 0$ and $E(X_s X_h) = \tilde{E}\left(\rho^{|s-h|}(1-\rho^2)^{-1}\right)$ for

$s, h = 1, \ldots, T$, and:

$$\tilde{S}_T \underset{T\to\infty}{\overset{p}{\to}} \left(\tilde{E}\left(\frac{1}{1-\rho^2}\right)\right)^{-1} \tilde{E}\left(\frac{\rho_i}{1-\rho^2}\right).$$

The same results hold when N and T tend to infinity simultaneously.

Proposition 3 (*Joint Limit*)**.** *Suppose that Assumptions 1–2 hold true,* $\tilde{E}\left((1-|\rho|)^{-3}\right) < \infty$, *and* $\min(N,T) \to \infty$. *Then,*

$$S_{N,T} \underset{(N,T)\to\infty}{\overset{p}{\to}} \left(\tilde{E}\left(\frac{1}{1-\rho^2}\right)\right)^{-1} \tilde{E}\left(\frac{\rho_i}{1-\rho^2}\right).$$

Proof. See Jondeau and Pelgrin (2014).

Remark 5. Irrespective of the limit theory (joint or sequential), we have:

$$\left(\tilde{E}\left(\frac{1}{1-\rho^2}\right)\right)^{-1} \tilde{E}\left(\frac{\rho_i}{1-\rho^2}\right) > \tilde{E}(\rho).$$

Proof. Since $\Pr[0 \le \rho < 1] = 1$ and using Lehmann (1966), there exists a set with positive measure such that $\tilde{\text{Cov}}\left(\rho, \frac{1}{1-\rho^2}\right) > 0$. A similar result holds when $\Pr[-1 < \rho \le 0] = 1$ (See Jondeau and Pelgrin, 2014).

Remark 6. To get the limit distribution, some rate conditions must be imposed (see Phillips and Moon, 1999). For instance, the limit distribution can be derived under the pertinent assumption that $N/T(N) \to 0$ (i.e., T is a function of the number of micro-units and grows faster to infinity than the number of micro-units) (See Jondeau and Pelgrin, 2014).

Remark 7. Following Robinson (1978), a consistent estimator of $\tilde{E}(\rho)$ can be defined as:

$$\hat{\rho} = \frac{\gamma_X(1) - \gamma_X(3)}{\gamma_X(0) - \gamma_X(2)} = \frac{\sum_{t=2}^{T} X_{N,t}X_{N,t-1} - \sum_{t=4}^{T} X_{N,t}X_{N,t-3}}{\sum_{t=1}^{T} X_{N,t}^2 - \sum_{t=3}^{T} X_{N,t}X_{N,t-2}},$$

and its asymptotic distribution can be characterized under mild conditions.

References

Acemoglu, D., Carvalho, V.M., Ozdaglar, A., Tahbaz-Salehi, A., 2012. The network origins of aggregate fluctuations. Econometrica 80, 1977–2016.

Altissimo, F., Mojon, B., Zaffaroni, P., 2009. Can aggregation explain the persistence of inflation? J. Monetary Econ. 56, 231–241.

Beran, J., 1994. Statistics for Long-Memory Processes. In: Monograph on Statistics and Applied Probability, vol. 61. Chapman and Hall.

Berk, K.N., 1974. Consistent autoregressive spectral estimates. Ann. Statist. 2, 489–502.

Brockwell, P.J., Davis, R.A., 1991. Times Series: Theory and Methods, second ed. In: Springer Series in Statistics, Springer Verlag.

Carvalho, C., Dam, N.A., 2010. The cross-sectional distribution of price stickiness implied by aggregate data. Federal Reserve Bank of New York Staff Report 419.

Chang, T.T.L., 2006. The polynomial aggregated AR(1) model. Econom. J. 9, 98–122.

Choquet, G., 1969. Lectures on Analysis, vol. II, Representation Theory. W.A. Benjamin, New York.

Forni, M., Lippi, M., 1997. Aggregation and the Microfoundations of Dynamic Macroeconomics. Clarendon Press, Oxford.

Gabaix, X., 2011. The granular origins of aggregate fluctuations. Econometrica 79, 733–772.

Gonçalves, E., Gouriéroux, C., 1988. Agrégation de processus autorégressifs d'ordre 1. Ann. Écon. Statist. 12, 127–149.

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

Granger, C.W.J., Morris, M.J., 1976. Time series modelling and interpretation. J. R. Stat. Soc. 139, 246–257.

Imbs, J., Jondeau, E., Pelgrin, F., 2011. Sectoral Phillips curves and the aggregate Phillips curve. J. Monetary Econ. 58, 328–344.

Imbs, J., Muntaz, A., Ravn, M.O., Rey, H., 2005. PPP strikes back: aggregation and the real exchange rate. Quart. J. Econ. 120, 1–43.

Jondeau, E., Pelgrin, F., 2014. Heterogeneity in panel data models: an extended mean group estimator. Mimeo.

Kuersteiner, G.M., 2005. Automatic inference for infinite order vector autoregressions. Econometric Theory 21, 85–115.

Lehmann, E.L., 1966. Some concepts of dependence. Ann. Math. Stat. 37, 1137–1153.

Lewbel, A., 1994. Aggregation and simple dynamics. Amer. Econ. Rev. 84, 905–918.

Linden, M., 1999. Time series properties of aggregated AR(1) processes with uniformly distributed coefficients. Econom. Lett. 64, 31–36.

Mayoral, L., 2013. Heterogeneous dynamics, aggregation, and the persistence of economic shocks. Internat. Econom. Rev. 54, 1295–1307.

Pesaran, M.H., 2003. Aggregation of linear dynamic models: an application to life-cycle consumption models under habit formation. Ecol. Modell. 20, 383–415.

Pesaran, M.H., Chudik, A., 2014. Aggregation in large dynamic panels. J. Econometrics 178, 273–285.

Pesaran, M.H., Smith, R., 1995. Estimation of long-run relationships from dynamic heterogenous panels. J. Econometrics 68, 79–114.

Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67, 1057–1111.

Robinson, P.M., 1978. Statistical inference for a random coefficient autoregressive model. Scand. J. Statist. 5, 163–168.

Sawa, T., 1978. The exact moment of the least squares estimator for the autoregressive model. J. Econometrics 8, 159–172.

Shohat, J.A., Tamarkin, J.D., 1943. The Problem of Moments. In: Mathematical Surveys, vol. 1. American Mathematical Society, New York, NY.

Zaffaroni, P., 2004. Contemporaneous aggregation of linear dynamic models in large economies. J. Econometrics 120, 75–102.