National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 579

Aggregating Rational Expectations Models in the Presence of Unobserved Micro Heterogeneity

Eric Jondeau **Florian Pelgrin**

First version: August 2009
Current version: August 2009

This research has been carried out within the NCCR FINRISK project on
"Macro Risk, Systemic Risks and International Finance"

Aggregating Rational Expectations Models in the Presence of Unobserved Micro Heterogeneity[1]

Eric Jondeau

Faculty of Business and Economics (HEC)

Swiss Finance Institute

University of Lausanne, Switzerland

Florian Pelgrin

Faculty of Business and Economics (HEC)

Department of Economics and Econometrics

University of Lausanne, Switzerland

First draft: August 2009

Current draft: August 2009

[1]We are grateful to audiences at the Society for Computational Economics (Montréal, 2007), the Workshop on Computational and Financial Econometrics (Geneva, 2007), the seminar of the European Central Bank (October 2008), and the Macro Brownbag of the University of Lausanne (April, 2009), Macro Workshop in Toulouse (June 2009). All errors are our own. Corresponding author: Eric Jondeau, eric.jondeau@unil.ch. This research has been carried out within the NCCR FINRISK project on "Macro Risk, Systemic Risks and International Finance".

Abstract

Our paper addresses the correction of the aggregation bias in linear rational expectations models when there is some unobserved micro-parameter heterogeneity and only macro data are available. Starting from Lewbel (1994), we propose two new consistent estimators, which rely on a flexible parametric specification of the cross-sectional parameter distributions and account for the dependence across coefficients inherent in such models. A Monte-Carlo study reveals that the finite-sample and asymptotic properties of the proposed estimators correct the aggregation bias found with the maximum-likelihood and generalized-method-of-moments approaches. As a by-product, we can also infer the cross-sectional distribution of the parameters. Finally, we reassess the empirical evidence about the New Keynesian Phillips curve and explain the apparent discrepancy between micro- and macro-based estimates of the average persistence of inflation.

JEL Classification Code: C13, C20, E20.

Keywords: Aggregation, Rational expectations models, Heterogeneity.

1 Introduction

From the literature on aggregation, it is well known that ignoring heterogeneity in dynamic models results in inconsistent maximum-likelihood (ML) and generalized-method-of-moments (GMM) estimators. Imposing homogeneity on the aggregate dynamics forces heterogeneity into the error term. The error term is then unavoidably correlated with (some of) the regressors, resulting in inconsistent parameter estimates (Pesaran and Smith, 1995).[1]

When micro data are available, such an inconsistency can generally be avoided by using an appropriate estimation technique that accounts for parameter heterogeneity (e.g., the mean group estimator or the random coefficient estimator).[2] However, the problem is much more acute when only aggregate data are available, as we cannot control for the heterogeneity at the micro level. This issue is particularly relevant because in many situations disaggregate data are simply unavailable or only available over short periods of time. For instance, measures of consumption expenditure or investment reported in National Accounts are, in reality, the aggregation of the consumption or investment decisions made by a large number of micro-units (households or firms). Since such data are observed down to a certain level of disaggregation only, the extent of the heterogeneity of decision rules across micro-units is often only partially observed or measured.[3] Therefore it is critical to have an estimation technique, based on macro data only, that provides consistent estimates of the aggregate dynamics.

Only a few papers have concentrated on the reliability of macro information in circum-

[1] For a review of the aggregation literature, see Barker and Pesaran (1990), Stoker (1993), and Blundell and Stoker (2005).

[2] A flurry of papers has investigated how estimates of micro-parameters can be used to improve the estimation of macro-parameters and the prediction of aggregate macro-economic time series (Pesaran, Pierse, and Lee, 1994; Pesaran and Smith, 1995; Pesaran, Smith, and Im, 1996; Van Garderen, Lee, and Pesaran, 2000; Pesaran, 2003).

[3] The availability of surveys may partially overcome this problem. However, they are far from covering all of the economic fields.

venting the aggregation bias in the presence of unobserved micro heterogeneity (Lewbel, 1992, 1994; Pesaran, 2003). Our paper proposes some solutions to this problem in the general class of linear dynamic rational expectations models, when both the time series dimension (T) and the cross-sectional dimension (N) are large. More specifically, we provide two new consistent estimators that fully exploit both the heterogeneity-based specification of the aggregate dynamics and the cross-sectional dependence of the coefficients of interest. They both allow for the recovery of reliable information regarding the distribution of micro-parameters.

We consider the so-called aggregate "hybrid" (or second-order) linear rational expectations model, in which the only source of misspecification results from the aggregation of heterogeneous individual dynamics (e.g., households or firms):

$$Y_t = \bar{\omega}\, Y_{t-1} + (1-\bar{\omega})\, E_t\, Y_{t+1} + \bar{\lambda}\, X_t + \bar{\varepsilon}_t, \tag{1}$$

$$X_t = \bar{\rho}\, X_{t-1} + \bar{v}_t, \tag{2}$$

where Y_t and X_t denote respectively the cross-section mean of the dependent and forcing variables at date t ($t = 1, \cdots, T$) and E_t denotes the expectation conditional on the information available at date t. This model provides a convenient framework for representing many macroeconomic behaviors, such as the dynamics of price or wage inflation, consumption, investment, or inventories (see, among others, Smets and Wouters, 2007).

We first show that the estimation of Eqs. (1) and (2) by standard econometric techniques, such as ML and GMM, leads to substantial biases in the parameters of interest in the presence of unobserved micro heterogeneity. This arises from the correlation of the error terms $\bar{\varepsilon}_t$ and $\bar{v}_t$ with the regressors due to some unobserved heterogeneity in the individual realizations of the random parameters ω, λ, and ρ. Then we argue that, once the genuine structure of the error terms is properly taken into account, the aggregate variables, Y_t and X_t, do not share the same estimable hybrid dynamics as the individual variables. In particular, Y_t has an infinite autoregressive distributed lags (ARDL) representation and X_t has an infinite autoregressive representation. The intuition is that the addition of lags removes any correlation between the error terms and the regressors, so that the error terms are whitened as $N \to \infty$.

Armed with these aggregate specifications, we describe two consistent estimation techniques based on a flexible parameterization of the cross-sectional parameter distributions, following a long tradition in the aggregation literature (Granger, 1980). The first method, called the parametric estimator, is based on a regression approach, while the second method exploits the auto- and cross-covariance properties of the aggregate processes through a minimum distance (MD) estimation. As a by-product, both proposed estimators also provide reliable information on the cross-sectional distributions of the micro-parameters, even when only macro data are available. The intuition is that the moments of the cross-sectional distribution of the individual parameters can be backed out from the parameter estimates.

Our paper is related to a set of contributions that investigate the dynamics of aggregate processes resulting from heterogenous micro-units. Robinson (1978), Granger (1980), and Gonçalves and Gouriéroux (1988), among others, explored the contemporaneous aggregation of autoregressive processes. Zaffaroni (2004) provided additional asymptotic results in the case of a general ARMA process when the number of micro-units is infinite. Forni and Lippi (1997) and Canova (2007) reviewed the relevance of this issue across a wide range of empirical applications in macroeconomics. These contributions do not consider the estimation issue when only macro data are available. In contrast, Lewbel (1994) aimed at correcting the aggregation bias in a linear dynamic model using only macro data. He showed that aggregation results in an infinite autoregressive form when random parameters are independent from each other. Pesaran (2003) used a forecast-based approach and provided some extensions to Lewbel's result when the random parameters are not independent of each other and the forcing variable dynamics are exogenously given. While these two papers are complementary to ours, they also differ in a number of important ways. First, contrary to Lewbel's (1994) and Pesaran's (2003) approaches, the dynamics of the forcing variable is specified in our framework, so that the model can be solved recursively forward. The micro-parameters in the resulting reduced form are not distributed independently from each other and the forcing variable coefficient is correlated with the forcing variable. Second, we provide general conditions that ensure an infinite order autoregressive distributed lag representation and show that a truncated dynamics can be consistently estimated. Moreover, our consistent estimation methods fully account for the set of non-linear restrictions that drives the aggregate parameters.

Using Monte Carlo simulations, we show that the bias implied by unobserved heterogeneity in standard ML and GMM estimators is substantial. For realistic calibrations, we show that the bias in the persistence parameter estimate of Y_t can be as high as 0.2. In addition, the proposed estimation procedures perform very well in two dimensions. First, they result in essentially unbiased parameter estimates. Second, they also provide consistent estimates of the moments of the cross-sectional distribution of the random parameters.

Finally, we use our estimation techniques for the estimation of a New Keynesian Phillips curve. In this context, we show that the aggregation of heterogeneous behaviors likely explains some puzzles encountered in the literature, such as the discrepancy between micro- and macro-based evidence of the average inflation persistence. The aggregate specifications given here may be used to enhance the efficiency of aggregate models by incorporating relevant information that may be known about the distribution of simple dynamic behavior across individuals. Alternatively, restrictions on the distribution of individual parameters may be tested with aggregate data. The proposed methods might be used to assess whether a forecast of the aggregate variable using heterogeneity-correcting estimates may produce an improved prediction mean squared error over a forecast of the aggregate variable using ML or GMM estimation of structural or time-series models.

The rest of the paper is structured as follows. In Section 2, we describe our framework and discuss the main assumptions as well as their implications. In Section 3, we show how to derive the *exact* aggregate dynamics that account for the presence of unobserved micro heterogeneity. Then, we describe the two proposed estimators, starting from an extension of the regression-based approach developed by Lewbel (1994). Section 4 provides Monte Carlo simulations of the finite-sample properties of the proposed estimators in the context of rational expectations models. We compare them to the aggregate ML and the GMM estimators. Section 5 presents an empirical application based on the French New Keynesian Phillips curve. Some concluding remarks are provided in Section 6.

2 The Dynamics of Micro-units

In this section, we introduce a general class of linear dynamic rational expectations models to describe the individual dynamics.[4] We consider the following structural model for the variable of interest, $y_{i,t}$:

$$y_{i,t} = \omega_i \; y_{i,t-1} + (1 - \omega_i) \, E_t \; y_{i,t+1} + \lambda_i \; x_{i,t} + \varepsilon_{i,t}, \tag{3}$$

where i indexes micro-units ($i = 1, \cdots, N$) and t indexes time periods ($t = 1, \cdots, T$). The functional form is the same across micro-units, but parameters ω_i and λ_i may vary across micro-units. For the sake of simplicity, the forcing variable, $x_{i,t}$, is defined as an

[4] Linearity is a key assumption for implementing the proposed methods. As stressed by Kelajian (1980) and Stoker (1993), individual linear (dynamic) models are the only models that allow for recoverability of the distribution of the individual parameters, so that estimation of the behavioral parameters can be based on aggregate data. An implication is that we can infer individual reduced-form parameter behavior from aggregate reduced-form parameters, while this is almost unfeasible for structural parameters. Our results also apply to the estimation of *ad hoc* reduced-form time series models (i.e., which are not derived from a structural model).

AR(1) process:

$$x_{i,t} = \rho_i \; x_{i,t-1} + v_{i,t}, \tag{4}$$

where parameter ρ_i can also vary across micro-units.[5]

As shown in Appendix 1, solving Eqs. (3) and (4) recursively forward yields the following reduced form:

$$y_{i,t} = \phi_i \; y_{i,t-1} + \beta_i \; x_{i,t} + u_{i,t}, \tag{5}$$

with $u_{i,t} = \varepsilon_{i,t} / (1 - \omega_i) = \varepsilon_{i,t} (1 + \phi_i)$, $\phi_i = \omega_i / (1 - \omega_i)$, and $\beta_i = \lambda_i (1 + \phi_i) / (1 - \rho_i)$. Consequently, the aggregation of linear dynamic rational expectations models is only feasible at the expense of relaxing the assumption of independence of parameters, common in the aggregation literature. In our case, (1) micro-parameters are not distributed independently from each other since ϕ_i and β_i are correlated with each other, and (2) β_i and $x_{i,t}$ are correlated since β_i depends on ρ_i. More specifically, we make the following assumption regarding the micro-parameters.

Assumption 1 (Micro-parameters) *(1) Micro-parameters ϕ_i and ϕ_j are i.i.d. realizations of the random parameter ϕ for any $i \neq j$, with $0 \leq \phi < 1$. (2) Micro-parameters ρ_i and ρ_j are i.i.d. realizations of the random parameter ρ for any $i \neq j$, with $|\rho| < 1$. (3) Micro-parameters λ_i satisfy $|\lambda_i| < +\infty$ for any i. (4) ϕ, λ, and ρ are mutually independent. (5) $(\phi_i, \lambda_i, \rho_i)$ are distributed independently from $\varepsilon_{i,t}$ and $v_{i,t}$.*

The stationarity conditions in (A1.1) and (A1.2) ensure that central and noncentral individual moments exist so that the cross-sectional distribution of each random coefficient is well defined.[6] Given the bijective relation between ω and ϕ, we mainly focus on ϕ, since it drives the intrinsic persistence properties of $y_{i,t}$. Assumptions (A1.4) and (A1.5) are standard in the aggregation literature.

[5] In the technical report available upon request, we discuss the analytical expression of the aggregate process when $x_{i,t}$ is an AR(p) process or when the complete model is driven by a VAR(p) process.

[6] Assumption (1.2) also guarantees that there is no individual unit root parameter, which would dominate at the aggregate level and would generate long-memory features (see Abadir and Talmain, 2002; Zaffaroni, 2004).

We now describe our assumptions regarding the innovations $\varepsilon_{i,t}$ and $v_{i,t}$. Following a standard specification in the aggregation and panel literature (Forni and Lippi, 1997; Pesaran, 2003), we allow the innovations to be decomposed into a common component and an idiosyncratic (individual-specific) component:

$$\varepsilon_{i,t} = \kappa_i \, \epsilon_t + \eta_{i,t}, \tag{6}$$

$$v_{i,t} = \tilde{\kappa}_i \, \tilde{\epsilon}_t + \tilde{\eta}_{i,t}, \tag{7}$$

where ϵ_t and $\tilde{\epsilon}_t$ denote the common innovations, $\eta_{i,t}$ and $\tilde{\eta}_{i,t}$ the idiosyncratic innovations, and κ_i and $\tilde{\kappa}_i$ some micro-parameters. Equations (6) and (7) satisfy Assumption 2.

Assumption 2 (Error terms) *(1) ϵ_t and $\tilde{\epsilon}_t$ are white noise processes with mean zero and variance σ_ϵ^2 and $\sigma_{\tilde{\epsilon}}^2$ respectively. (2) $\eta_{i,t}$ and $\tilde{\eta}_{i,t}$ are white noise processes with mean zero and variance σ_η^2 and $\sigma_{\tilde{\eta}}^2$ respectively. (3) ϵ_t, $\tilde{\epsilon}_t$, $\eta_{i,t}$, and $\tilde{\eta}_{j,t}$ are orthogonal at any lag and lead. (4) Micro-parameters $(\kappa_i, \tilde{\kappa}_i)$ and $(\kappa_j, \tilde{\kappa}_j)$ are independent draws from the distribution of the random parameters $(\kappa, \tilde{\kappa})$ for any $i \neq j$. (5) κ and $\tilde{\kappa}$ are mutually independent. (6) $(\kappa, \tilde{\kappa})$ and (ϕ, λ, ρ) are mutually independent. (7) $E(\kappa) = E(\tilde{\kappa}) = 1$.*

The white noise assumption in (A2.1) and (A2.2) may be relaxed at the cost of specifying the corresponding dynamics. This is also the case for the independence assumption in (A2.5) and (A2.6). Assumption (A2.7) is an identification condition. Assumptions 1 and 2 imply that the $x_{i,s}$ terms have finite second-order moments and are distributed independently of $\varepsilon_{i,t}$ for all i, t and $s \leq t$. This implication is required for consistent estimation of the aggregate dynamics.

When the number of micro-units, N, is infinite (e.g., households or firms), idiosyncratic components cancel out in the aggregate due to the law of large numbers, while the common component, even when negligible at the micro level, completely predominates at the macro level (Granger, 1980, 1987).[7] In contrast, when the population is naturally segmented into a finite number of groups (Blundell and Stoker, 2005), some intermediate-level data may be available (e.g., sectors for firms or classes for households). In this case, the idiosyncratic components do not necessarily cancel out at the aggregate level and thus may coexist with a common shock.

3 Aggregation and Correction for Heterogeneity

In this section, we first describe the aggregate dynamics of both structural and reduced-form models, which imply that GMM and ML estimators are unavoidably biased in this context. Then, we argue that the aggregation bias can be corrected once we acknowledge that the aggregate dynamics do not share the same estimable hybrid representation of the individual models. Lastly, we propose consistent estimation methods.

3.1 Aggregate dynamics

Throughout the paper, we adopt the following notation. Let w_i denote the individual weight such that $\sum_{i=1}^{N} w_i = 1$ and $w_i > 0$.[8] The weights are independent of the parameters of interest. The cross-sectional mean of the individual variable $z_{i,t}$ is denoted $Z_{w,t} = \sum_{i=1}^{N} w_i \; z_{i,t}$. The cross-sectional mean of the individual random parameter ϑ_i is denoted $E(\vartheta) = \sum_{i=1}^{N} w_i \; \vartheta_i$, with $\tilde{\vartheta}_i \equiv \vartheta_i - E(\vartheta)$.

[7] For an interesting discussion, see Forni and Lippi (1997, Chapter 4, pp. 53–63), and Barker and Pesaran (1990), especially the contribution of Granger (1990) therein. In fact, in the case of an infinite autoregressive distributed lag model, Pesaran (2003) showed that the contribution of the idiosyncratic shocks to the aggregate function will depend on the rate at which the distributed lagged coefficients of the forcing variables decay as $N \to \infty$. Zaffaroni (2004) pointed out that, in the case of autoregressive processes, when the autoregressive coefficients are heterogenous and close to non-stationarity, the average of the idiosyncratic components might not disappear, so that the limit of the aggregate of purely idiosyncratic units can have a non-degenerate distribution. In the sequel, we will exclude the possibility of non-stationary individual processes.

[8] These weights may be determined in different ways. In the case of firms, the weights may be based on the relative value added. For countries or goods, one may follow Pesaran, Schuermann, and Weiner (2004) and use output shares or trade weights. In the case of a large N, one may assume equal weights, $w_i = 1/N$.

The aggregate dynamics of the structural equation (3) is obtained by averaging over the micro-units:

$$\sum_{i=1}^{N} w_i \ y_{i,t} = \sum_{i=1}^{N} w_i \ \omega_i \ y_{i,t-1} + \sum_{i=1}^{N} w_i \ (1-\omega_i) \ E_t \ y_{i,t+1} + \sum_{i=1}^{N} w_i \ \lambda_i \ x_{i,t} + \sum_{i=1}^{N} w_i \ \varepsilon_{i,t}.$$

This expression can be rewritten as:

$$Y_{w,t} = E(\omega) Y_{w,t-1} + (1 - E(\omega)) E_t \ Y_{w,t+1} + E(\lambda) X_{w,t} + \bar{\varepsilon}_{w,t}, \tag{8}$$

where the error term,

$$\bar{\varepsilon}_{w,t} = \varepsilon_{w,t} + \sum_{i=1}^{N} w_i \ \tilde{\omega}_i \left(y_{i,t-1} - E_t \ y_{i,t+1}\right) + \sum_{i=1}^{N} w_i \ \tilde{\lambda}_i \ x_{i,t},$$

combines the aggregate shock, $\varepsilon_{w,t}$, and the effect of parameter heterogeneity on the aggregate variables.

Accordingly, the reduced-form equation is given by:

$$Y_{w,t} = E(\phi) Y_{w,t-1} + E(\beta) X_{w,t} + \bar{u}_{w,t}, \tag{9}$$

with:

$$\bar{u}_{w,t} = u_{w,t} + \sum_{i=1}^{N} w_i \ \tilde{\phi}_i \ y_{i,t-1} + \sum_{i=1}^{N} w_i \ \tilde{\beta}_i \ x_{i,t}.$$

From Eq. (4), the cross-aggregation of the forcing variable dynamics yields:

$$X_{w,t} = E(\rho) X_{w,t-1} + \bar{v}_{w,t}, \tag{10}$$

with:

$$\bar{v}_{w,t} = v_{w,t} + \sum_{i=1}^{N} w_i \ \tilde{\rho}_i \ x_{i,t-1}.$$

We observe that imposing homogeneity on the aggregate parameters forces heterogeneity into the residual terms, so that the error terms in Eqs. (8), (9), and (10) are unavoidably correlated with the dependent variables. In a technical appendix available upon request, we show that the standard GMM and ML estimation techniques fail to account for the residual properties in this framework and therefore yield biased estimates (see Pesaran and Smith, 1995; Imbs, Jondeau, and Pelgrin, 2007). In particular, the persistence parameters driving the dynamics of $Y_{w,t}$ and $X_{w,t}$ are typically overestimated. For instance, this result is consistent with empirical evidence regarding inflation dynamics, which reveals a discrepancy between micro- and macro-based estimates of the average persistence of the inflation rate (Imbs, Jondeau, and Pelgrin, 2007; Altissimo, Mojon, and Zaffaroni, 2009).

3.2 Rewriting the aggregate dynamics

We now rewrite Eqs. (9) and (10) in order to obtain consistent estimators of the aggregate linear dynamic model. For this purpose, we need to remove any correlation between the micro-variables and the corresponding random reduced-form parameter in the error term. This is done by introducing additional lags in Eqs. (9) and (10), where the parameters reflect the cross-sectional moments of the random parameters. This methodology, which was originally described by Lewbel (1994), leads to specify the dynamics of $Y_{w,t}$ as an infinite ARDL model in $Y_{w,t}$ and $X_{w,t}$, and the dynamics of $X_{w,t}$ as an infinite AR process. These results are stated in Proposition 1.

Proposition 1 *Assume that each micro-unit is described by the linear rational expectations model (3) and (4) and that the individual processes have been initialized to zero at time $t \to -\infty$. Then, under Assumptions 1 and 2, as $N \to \infty$, the aggregate variables $Y_{w,t}$ and $X_{w,t}$ are specified as:*

$$Y_{w,t} = \sum_{s=1}^{\infty} A_s \; Y_{w,t-s} + \sum_{s=0}^{\infty} B_s \; X_{w,t-s} + U_{w,t}, \tag{11}$$

$$X_{w,t} = \sum_{s=1}^{\infty} C_s \; X_{w,t-s} + V_{w,t}, \tag{12}$$

where $U_{w,t} = (1 + A_1)\epsilon_t$ *and* $V_{w,t} = \tilde{\epsilon}_t$. *The parameters are given by* $A_s = E(a_s)$, $C_s = E(c_s)$, $\forall s \geq 1$, $B_s = E(b_s)$, $\forall s \geq 0$, *and:*

$$
\begin{aligned}
a_1 &= \phi & a_{s+1} &= (a_s - A_s)\,\phi, \\
b_0 &= \beta & b_{s+1} &= (a_{s+1} - A_{s+1})\,\beta + (b_s - B_s)\,\rho, \\
c_1 &= \rho & c_{s+1} &= (c_s - C_s)\,\rho,
\end{aligned}
$$

with $\sum_{s=0}^{\infty} |A_s| < +\infty$, $\sum_{s=0}^{\infty} |C_s| < +\infty$, $A_0 = 1$, *and* $C_0 = 1$.

Proof: See Appendix 2.

Several points are worth commenting. First, if we fully account for the unobserved parameter heterogeneity of the individual data generating processes, an exact finite order autoregressive distributed lag representation for $Y_{w,t}$ and $X_{w,t}$ does not exist (Lewbel, 1994; Pesaran, 2003). This statement remains true even if we specify some parametric distributions for the random parameters (ϕ, ρ, λ). This result has an important implication regarding the *correct* estimable aggregate dynamics of $Y_{w,t}$ and $X_{w,t}$. If the individual dynamics are well described by the heterogeneous equations that we postulate, then the estimation of the aggregate dynamics with the same form as the underlying individual dynamics yields inconsistent parameter estimates. On the one hand, if we acknowledge the presence of parameter heterogeneity and if micro data are available, the standard mean group and random coefficient estimators are consistent, since they correspond to the average of the cross-sectional parameters (see the expressions for A_1, B_0, and C_1). However, they are plugged into a mis-specified dynamics and therefore do not allow for a recovery of the genuine aggregate dynamics. On the other hand, the GMM and ML estimators are based on the empirical counterparts of Eqs. (8), (9), and (10), i.e.,

$$Y_{w,t} = \bar{\omega}\, Y_{w,t-1} + (1-\bar{\omega}) E_t\, Y_{w,t+1} + \bar{\lambda}\, X_{w,t} + \bar{\varepsilon}_{w,t}, \tag{13}$$

$$Y_{w,t} = \bar{\phi}\, Y_{w,t-1} + \bar{\beta}\, X_{w,t} + \bar{u}_{w,t}, \tag{14}$$

$$X_{w,t} = \bar{\rho}\, X_{w,t-1} + \bar{v}_{w,t}. \tag{15}$$

The resulting parameter estimates therefore correspond to the aggregate dynamics in the hypothetical model with homogenous micro-units. This result explains why GMM and ML estimates of the aggregate dynamics based on Eqs. (13)–(15) are difficult to align with disaggregate evidence (as for instance, the inflation dynamics). This difficulty comes from trying to generate plausible aggregate dynamics with models that do not feature heterogeneity.

Second, Proposition 1 clearly shows that the aggregate dynamics depend on the distribution of the random parameters and that some of the non-central cross-sectional moments can be recovered from the aggregate parameters. Indeed, the parameters A_s and C_s are nonlinear transformations of the non-central moments of ϕ and ρ. It is straightforward to show that:

$$A_{s+1} = E(a_{s+1}) = E\left(\phi^{s+1}\right) - \sum_{r=1}^{s} A_r\, E\left(\phi^{s-r+1}\right).$$

In particular, the first four moments of the distribution of ϕ are given by:[9]

$$\begin{aligned} E(\phi) &= A_1, \\ V(\phi) &= A_2, \\ sk(\phi) &= (A_3 - A_1 A_2) \,/\, (A_2)^{3/2}, \\ ku(\phi) &= \left(A_4 - 2A_1A_3 + A_1^2 A_2 + A_2^2\right) / \, (A_2)^2. \end{aligned}$$

[9] If $0 \leq \phi_i < 1$, then one has $0 \leq A_1 < 1$ and $0 \leq A_s \leq (1 - A_1)\, A_1 \leq 1/4$ for $s = 2, 3$. The same result applies for ρ_i, provided that the distribution of ρ is defined over $[0, 1[$. See Appendix 2.

Since parameters A_s and C_s are functions of the moments of the distributions of ϕ and ρ, respectively, some necessary conditions must hold for the aggregate behavior to be derived from disaggregate behavior. In particular, it is well known that the non-central moments $m_s(\vartheta) = E(\vartheta^s)$ of any (non-degenerate) random variables ϑ, defined on $[0, 1[$, satisfy

$$1 > m_1(\vartheta) \geq \cdots \geq m_s(\vartheta) \geq 0, \qquad \forall s \geq 1,$$

and that $m_s(\vartheta) \rightarrow 0$ as $s \rightarrow \infty$ (see Stuart and Ord, 1994). Moreover, since $\{A_s\}$ and $\{C_s\}$ are absolute summable, it implies that A_s and C_s tend toward zero as s goes to infinity. In other words, for a sufficiently large number of lags (say $s > K_0$), the infinite sums in Eqs. (11) and (12) can be truncated.

Third, no attribute of the univariate distribution of β other than the cross-sectional mean can be inferred from aggregate data. Given the expression of β, the following lemma shows that the B_s terms can be expressed as a function of the non-central moments of ρ and ϕ, as well as the cross-sectional mean of λ.

Lemma 1 *Assume that each micro-unit is described by the linear rational expectations model (3) and (4). Then, under Assumption 1, as* $N \rightarrow \infty$*, the aggregate parameters* B_s*,* $\forall s \geq 0$*, are defined by:*

$$\begin{aligned} B_0 &= E(\beta) = \lambda_0 \left(1 + E(\phi)\right) E\left(\frac{1}{1-\rho}\right), \\ B_s &= \lambda_0 \sum_{j=0}^{s} \sum_{k=0}^{j} \tilde{A}_k \left[E(\phi^{j-k}) + E(\phi^{j-k+1})\right] E\left(\frac{\rho^{s-j}}{1-\rho}\right) - \sum_{j=0}^{s-1} B_j \; E(\rho^{s-j}) \quad \textit{for } s \geq 1, \end{aligned}$$

with $\lambda_0 = E(\lambda)$*,* $\tilde{A}_0 = 1$*, and* $\tilde{A}_s = -A_s$*,* $\forall s \geq 1$*. If* $\{E(\phi^s)\}$ *and* $\{E(\rho^s)\}$ *are absolute summable, then* $\{B_s\}$ *is absolute summable and thus* $B_s \underset{s \rightarrow \infty}{\rightarrow} 0$*.*

Proof: See Appendix 3.

Since only the first moment of λ intervenes in the aggregate coefficients B_s, we cannot recover all of the characteristics of the distribution of λ or β. The intuition is that, once the aggregate dynamics of $Y_{w,t}$ is correctly specified, the dispersion of λ (as well as other higher moments of λ) across micro-units does not play any specific role.

Fourth, another useful representation of the aggregate processes $Y_{w,t}$ and $X_{w,t}$ is the infinite moving average representation. Indeed, it is straightforward to show that:

$$\begin{aligned} y_{i,t} &= \beta_i \sum_{s=0}^{\infty} \Theta_{i,s}\ v_{i,t-s} + \sum_{s=0}^{\infty} \phi_i^s\ u_{i,t-s}, \\ x_{i,t} &= \sum_{s=0}^{\infty} \rho_i^s\ v_{i,t-s}, \end{aligned}$$

where $\Theta_{i,s} = \sum_{r=0}^{s} \rho_i^r\ \phi_i^{s-r}$. Since $y_{i,t}$ and $x_{i,t}$ are stationary processes (Assumption 1), we assume an infinite past for mathematical convenience. The corresponding aggregate processes are (as $T \to \infty$):

$$Y_{w,t} = \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i\ \beta_i\ \Theta_{i,s}\ v_{i,t-s} + \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i\ \phi_i^s\ u_{i,t-s}, \tag{16}$$

$$X_{w,t} = \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i\ \rho_i^s\ v_{i,t-s}. \tag{17}$$

Therefore, Eqs. (16) and (17) can be rewritten asymptotically as follows.

Proposition 2 *Assume that each micro-unit is described by the linear rational expectations model (3) and (4) and that the individual processes have been initialized to zero at the infinite past. If $\{E(\phi^s)\}$ and $\{E(\rho^s)\}$ are absolute summable and Assumptions 1 and 2 hold, then the aggregate variables $Y_{w,t}$ and $X_{w,t}$ have the following representation as*

$N \to \infty$:

$$Y_{w,t} = \sum_{s=0}^{\infty} \delta_s \ \tilde{\epsilon}_{w,t-s} + \sum_{s=0}^{\infty} \psi_s \ \epsilon_{w,t-s}, \tag{18}$$

$$X_{w,t} = \sum_{s=0}^{\infty} \gamma_s \ \tilde{\epsilon}_{w,t-s}, \tag{19}$$

where $\delta_0 = E(\beta)$, $\psi_0 = 1 + E(\phi)$, $\gamma_0 = 1$, and $\delta_s = E(\beta \ \Theta_s)$, $\psi_s = E(\phi^s) + E(\phi^{s+1})$, $\gamma_s = E(\rho^s)$, with $\sum_{s=0}^{\infty} |\delta_s| < \infty$, $\sum_{s=0}^{\infty} |\psi_s| < \infty$, and $\sum_{s=0}^{\infty} |\gamma_s| < \infty$.

Proof: See Appendix 4.

In this set-up, we can also recover the non-central cross-sectional moments of ρ and ϕ by using the ψ_s and γ_s terms.[10] In contrast, only the first moment of β (or λ) can be backed out from the aggregate parameter δ_0. For $s > 0$, the δ_s terms provide information regarding the co-moments of β with ϕ and ρ. When the random coefficients ϕ, λ, and ρ are i.i.d. and mutually uncorrelated, the computation of these co-moments is given in the following lemma.

Lemma 2 *Assume that each micro-unit is described by the linear rational expectations model (3) and (4). Then, under Assumption 1, as $N \to \infty$, the aggregate parameters δ_s, $\forall s \geq 0$, are defined by:*

$$\delta_s = \lambda_0 \sum_{k=0}^{s} \left[E(\phi^{s-k}) + E(\phi^{s-k+1})\right] E\left(\frac{\rho^k}{1-\rho}\right) \qquad s \geq 0,$$

with $\lambda_0 = E(\lambda)$.

Proof: See Appendix 5.

In summary, the aggregate processes are rewritten in order to exploit the genuine structure of the error terms. Moreover, consistent estimation of the aggregate parameters may

[10] Conditions required on the parameters γ_s for an infinite moving average representation to be derived from the aggregation of an AR(1) have been given by Gonçalves and Gouriéroux (1988).

yield substantial information regarding the cross-sectional distribution of the random parameters. We are now in a position to propose two consistent estimators. Our starting point is a straightforward extension of the regression-based approach developed by Lewbel (1994). We then rely on a parametric specification for the distributions of ρ and ϕ. The first method is based on estimating the ARDL specification given in Proposition 1. The second method is based on a minimum distance estimation, which exploits the theoretical covariance properties of $Y_{w,t}$ and $X_{w,t}$ given in Proposition 2.

3.3 The unrestricted approach

The approach originally proposed by Lewbel (1994) consists of truncating the infinite sums in Proposition 1 and estimating the resulting dynamic.[11] In the context of linear dynamic rational expectations models, we generalize Lewbel's estimator as follows.

Definition 1 *The unrestricted estimator of the aggregate linear dynamic rational expectations model (11) and (12) with truncation lags (K_ϕ, K_β, K_ρ) is the ML estimator of the parameter set $\zeta_{Unr} = (A_1, \cdots, A_{K_\phi}, B_0, \cdots, B_{K_\beta}, C_1, \cdots, C_{K_\rho})'$ in the model:*

$$\begin{aligned} Y_{w,t} &= \sum_{s=1}^{K_\phi} A_s \, Y_{w,t-s} + \sum_{s=0}^{K_\beta} B_s \, X_{w,t-s} + (1 + A_1)\epsilon_t, && (20) \\ X_{w,t} &= \sum_{s=1}^{K_\rho} C_s \, X_{w,t-s} + \tilde{\epsilon}_t. && (21) \end{aligned}$$

In doing so, the first K_ϕ moments (respectively, K_ρ moments) of the distribution function of ϕ (respectively, ρ) are obtained from estimates of $A_1, \cdots, A_{K_\phi}$ (respectively, $C_1, \cdots, C_{K_\rho}$) after estimating the unconstrained truncated model defined above.

From Definition 1, three issues are of particular concern in the unrestricted estimation

[11] An alternative approach consists of truncating the infinite sums of the infinite moving average representation (Proposition 2) and estimating the corresponding dynamic. To some extent, this is the approach proposed by Pesaran (2003). However, it turns out that the aggregate dynamics is much more difficult to estimate with this approach and our simulation results tend to favor the regression-based approach.

of the aggregate model (20) and (21). First, the estimates of $\{A_s\}$, $\{B_s\}$, and $\{C_s\}$ are directly related to the characteristics of the cross-sectional distributions of ϕ, β, and ρ. In the unrestricted estimation, these relationships are not explicitly taken into account. In particular, all non-central moments of ρ and ϕ must be nonnegative. This involves a set of non-linear inequality restrictions on the parameter set, which are essentially intractable for large K_ϕ or K_ρ when the parameter distributions are not specified. At best, this implies a loss of efficiency in the estimation of the aggregate parameters. Second, the problem of the truncation remainder may be critical in small samples.[12] The unrestricted approach does not allow for the introduction of a large number of lags because any additional lag requires the estimation of one more parameter. This, in turn, means that the remainder of the truncated sums may contribute to the aggregation bias.[13] Last, one may argue that Eqs. (20) and (21) are observationally equivalent to a model in which homogenous individual (reduced-form) dynamics are driven by an $ARDL(K_\phi, K_\beta)$ for $y_{i,t}$ and an $AR(K_\rho)$ for $x_{i,t}$. Since individual data are not observable, we cannot disentangle the two equivalent specifications from a statistical point of view.

3.4 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 distributions, so that the parameters defining these distributions can be estimated from the aggregate equations. Following Granger (1980) and Gonçalves and Gouriéroux (1988), we assume that random parameters (say ϑ) are independently drawn from a Beta distribution:

$$f(\vartheta) = \frac{\vartheta^{p-1}\,(1-\vartheta)^{q-1}}{B(p,q)},$$

[12]In principle, the truncation remainder may significantly contribute to the aggregation bias. The problem of the truncation remainder was addressed long ago in the case of geometric distributed lag models. It can be resolved, for example, along the lines suggested by Dhrymes (1971) or Pesaran (1973).

[13]In his empirical illustration, Lewbel (1994) estimated a model with three lags corresponding to the first three moments of the distribution. The estimation of the cross-sectional skewness of ϕ turns out to be very sensitive to the specification of the model. This result is confirmed by the Monte Carlo experiments reported in Section 4.

where $p, q \in (0, \infty)$ and $B(p, q)$ is the beta function.

This parametric distribution is of great interest for several reasons. First, it naturally covers the range of values $[0, 1]$, in which empirical investigators are typically interested. Second, the non-central moments of a Beta distribution can be very easily computed (given p and q) with:

$$m_j(\vartheta) \equiv E\left(\vartheta^j\right) = \frac{B\left(p+j, q\right)}{B(p, q)} = \frac{\Gamma\left(p+q\right)\Gamma\left(p+j\right)}{\Gamma\left(p+q+j\right)\Gamma\left(p\right)}.$$

Third, the Beta distribution is able to reproduce a wide range of distribution shapes. For $p = q = 1$, we have a uniform distribution. For $0 < p < 1 < q$, the distribution is continuously decreasing, while for $0 < q < 1 < p$, it is continuously increasing. For $0 < p, q < 1$, the distribution is U-shaped. For $1 < p < q$, the distribution is rightward asymmetric bell-shaped, while for $1 < q < p$, the asymmetry is leftward. However, it is worth noting that $\{m_j\}$ is absolute summable provided that $q > 1$. In the sequel, we impose that $q_\rho > 1$ and $q_\phi > 1$. These conditions insure that $\{B_s\}$ and $\{\delta_s\}$ are absolute summable, i.e., the corresponding sums can be truncated. These conditions are not necessary to ensure the convergence of $\{A_s\}$ and $\{C_s\}$.

As shown by Lemma 1 and 2, the sequence of parameters in $\{B_s\}$ and $\{\delta_s\}$ are affected by the properties of λ only through the cross-sectional mean $\lambda_0 = E(\lambda)$. Accordingly, specifying the distribution of λ does not convey more information. Therefore the unknown parameters are those of the distribution of ρ, (p_ρ and q_ρ), and ϕ, (p_ϕ and q_ϕ), as well as $\lambda_0 = E(\lambda)$. These parameters can then be estimated as follows.

Definition 2 *The restricted parametric estimator of the aggregate linear rational expectations model (11) and (12) is the ML estimator of the parameter set* $\zeta_{Par} = (p_\phi, q_\phi, p_\rho, q_\rho, \lambda_0)'$

in the model:

$$Y_{w,t} = \sum_{s=1}^{K_\phi} A_s \; Y_{w,t-s} + \sum_{s=0}^{K_\beta} B_s \; X_{w,t-s} + (1 + A_1)\epsilon_{w,t}, \tag{22}$$

$$X_{w,t} = \sum_{s=1}^{K_\rho} C_s \; X_{w,t-s} + \tilde{\epsilon}_{w,t}, \tag{23}$$

where:

$$A_s = E\left(\phi^s\right) - \sum_{r=1}^{s-1} A_r \; E\left(\phi^{s-r}\right) \qquad and \qquad C_s = E\left(\rho^s\right) - \sum_{r=1}^{s-1} C_r \; E\left(\rho^{s-r}\right),$$

with:

$$E\left(\phi^s\right) = \frac{B\left(p_\phi + s, q_\phi\right)}{B\left(p_\phi, q_\phi\right)} \qquad and \qquad E\left(\rho^s\right) = \frac{B\left(p_\rho + s, q_\rho\right)}{B\left(p_\rho, q_\rho\right)}.$$

The B_s terms are given in Lemma 1.

An obvious advantage of the parametric estimator is that it greatly reduces the number of unknown parameters that must be estimated. In particular, the parameters of the Beta distributions are consistently estimated if we can capture the shape of each distribution with a finite number of lags.

3.5 The minimum distance estimator

Using the infinite moving average representation of the aggregate processes (Proposition 2), we also define a minimum distance estimator that aims at minimizing the distance between the theoretical moments and co-moments of $Y_{w,t}$ and $X_{w,t}$ and their empirical counterparts. Provided that ϕ, ρ, and λ are mutually independent (Assumption A1.4), one can define the non-central moments of ϕ and ρ from the Beta distribution and obtain the implied theoretical variances and covariances of $Y_{w,t}$ and $X_{w,t}$. For this purpose, we first need the following result that gives the theoretical autocovariance and cross-covariance functions of the aggregate processes.

Lemma 3 *Under Assumptions 1 and 2, the autocovariance and cross-covariance functions of the aggregate processes $Y_{w,t}$ and $X_{w,t}$ are, as $N \rightarrow \infty$:*

$$
\begin{aligned}
Cov\left(Y_{w,t}, Y_{w,t-h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \delta_s \; \delta_{s+h} + \sigma_{\epsilon}^2 \sum_{s=0}^{\infty} \psi_s \; \psi_{s+h}, \qquad \forall h \geq 0, \\
Cov\left(X_{w,t}, X_{w,t-h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \gamma_s \; \gamma_{s+h}, \qquad \forall h \geq 0, \\
Cov\left(Y_{w,t}, X_{w,t-h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \delta_{s+h} \; \gamma_s, \qquad \forall h \geq 0, \\
Cov\left(Y_{w,t}, X_{w,t+h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \delta_s \; \gamma_{s+h}, \qquad \forall h \geq 0,
\end{aligned}
$$

where $\delta_s = E\left(\beta \; \Theta_s\right)$ is defined in Lemma 2, $\psi_s = E(\phi^s) + E(\phi^{s+1})$, and $\gamma_s = E(\rho^s)$.

Proof: See Appendix 6.

In this respect, the MD estimator, denoted ζ_{MD}, is given by:

$$
\zeta_{MD} \in \arg\min_{\zeta} \left(C\left(\zeta\right) - \hat{C}\right)' \hat{\Omega} \left(C\left(\zeta\right) - \hat{C}\right),
$$

where $C\left(\zeta\right)$ and $\hat{C}$ denote the set of k functions of the parameters of interest and the empirical counterparts, and $\hat{\Omega}$ is a weighting matrix. Under some standard regularity conditions, ζ_{MD} is consistent and asymptotically normally distributed.[14] Our MD estimator is defined as follows.

Definition 3 *The MD estimator of the aggregate linear rational expectations model (11) and (12) is the parameter set $\zeta_{MD} = (p_\phi, q_\phi, p_\rho, q_\rho, \lambda_0, \sigma_\epsilon^2, \sigma_{\tilde{\epsilon}}^2)'$ that minimizes the distance $(C\left(\zeta\right) - \hat{C})'\hat{\Omega}(C\left(\zeta\right) - \hat{C})$ with:*

$$
C\left(\zeta\right) = \left\{Cov\left(Y_t, Y_{t-h}\right), Cov\left(X_t, X_{t-h}\right), Cov\left(Y_t, X_{t-h}\right), Cov\left(Y_{t-h+1}, X_t\right)\right\}_{h \geq 0}.
$$

[14] A usual choice for the weighting matrix is an estimate of the inverse of the covariance matrix of $\hat{C}$. The asymptotic distribution of $\sqrt{T}(\hat{\zeta}_{MD} - \zeta_0)$ does not depend upon the form of $\hat{\Omega}$, but only upon its limiting value Ω, with ζ_0 being the true value of the parameter set. The construction of the weighting matrix is discussed in Section 4 below.

Moreover, the MD estimator of ζ satisfies the constraints that the sequences $\{\delta_s\}$, $\{\psi_s\}$, and $\{\gamma_s\}$ are absolute summable.

The restricted parametric and MD estimators share some common features. First, the dependence between the structural parameters is explicitly taken into account. Second, the flexible parametric representation of the random coefficient distribution reduces the number of unknown parameters and ensures that the autoregressive parameters in the aggregate dynamics are consistent with the theoretical constraints. Therefore, they are less likely to suffer from finite-sample biases and over-fitting. Yet, these estimators are based on different time series representations. The parametric estimator relies on an ARDL representation, while the MD estimator is based on a moving-average representation. Moreover, the parametric estimator is likelihood-based, whereas the MD estimator is an M-estimator and may depend, at least in finite samples, on the choice of the weighting matrix $\hat{\Omega}$.

4 Monte Carlo Simulations

In this section, we perform Monte Carlo simulations to evaluate the finite and large sample properties of our proposed estimators relative to the (inefficient) unrestricted approach and the (inconsistent) GMM and ML estimators. More specifically, we focus on the ability of these estimators to provide an unbiased estimate of the aggregate parameters, as well as the ability of a given approach to reproduce the distribution of each micro-parameter. Indeed, the moments of the cross-sectional distribution of ϕ and ρ can be recovered, provided that the forcing variable is driven by an AR(1) dynamics. We assume that the data-generating process is given by Eqs. (3) and (4). Consequently, even when ϕ_i, λ_i, and ρ_i are assumed to be mutually uncorrelated for all i, the reduced-form parameters are correlated with each other through the relation $\beta_i = \lambda_i \left(1 + \phi_i\right) / \left(1 - \rho_i\right)$.

The simulation design is as follows. All results reported below are based on $2,000$ simulations. We experimented with different values of the time and individual dimensions. In the sequel, our benchmark is given by $T = 200$ and $N = 100$. For each parameter ϕ, λ, and ρ, we specify a Beta distribution, $B(p,q)$, where p and q vary across parameters.[15] Table 1 reports the distributional assumptions of the cases studied here.

[Insert Table 1 around here]

The parameter values for our benchmark are initially set to reproduce standard estimates encountered in the dynamic stochastic general equilibrium model literature (e.g., inflation or real wage dynamics). In particular, it is assumed that ρ is generated by a leftward asymmetric bell-shaped Beta distribution, with $p_\rho = 36$ and $q_\rho = 4$. The forcing variable is thus rather persistent, with an average persistence of $E(\rho) = 0.9$. The distribution of ϕ is symmetric ($p_\phi = q_\phi = 2$), with a mean value of $E(\phi) = 0.5$. Finally, λ is generated by a rightward asymmetric bell-shaped distribution ($p_\lambda = 10$, $q_\lambda = 90$) with a mean value of $E(\lambda) = 0.1$ (corresponding to $E(\beta) = 1.65$). In subsequent cases, we change the distribution of one parameter in turn and test the robustness of the results obtained in our baseline scenario. Finally, the remaining parameters (including the variances and covariances of the error terms) are calibrated to fit the characteristics of the French data described in Section 5.[16]

We also need to set the lag order in the unrestricted and parametric estimators as well as the number of autocovariances and the optimal weighting matrix in the MD approach. The unrestricted regression is carried out by setting $(K_\phi, K_\beta, K_\rho) = (4, 0, 4)$ in Eqs. (20)

[15] For a given value of p and q, the first four moments of ϑ can then be deduced as

$$\begin{aligned} E(\vartheta) &= p/(p+q) \\ V(\vartheta) &= pq/\left((p+q)^2(p+q+1)\right) \\ sk(\vartheta) &= 2(q-p)\sqrt{1+p+q}/(\sqrt{pq}(p+q+2)) \\ ku(\vartheta) &= 3+\frac{6\left[p^3+p^2(1-2q)+q^2(1+q)-2pq(2+q)\right]}{pq(p+q+2)(p+q+3)}. \end{aligned}$$

[16] Several additional simulation experiments are available upon request in a technical report. All in all, our main conclusions remain robust.

and (21), such that the first four moments of ϕ and ρ can be identified as well as the mean value of β or λ.[17] In the case of the parametric estimator, we include $T/20$ lags for the two autoregressive parameters ϕ and ρ. We also estimate the first four co-moments of β with ρ and ϕ. Last, the MD estimator is evaluated to provide the best fit of the variance and the first ten autocovariances of $Y_{w,t}$ and $X_{w,t}$, using the expressions in Definition 3 and Lemma 3. The weighting matrix is based on linearly decreasing weights, as in Newey and West (1987). The infinite sums in Lemma 3 are truncated to 100 terms. The results presented below are robust to higher-order lag truncations.

For each sample, the parameters are estimated using the techniques described in Section 3. We then compute the median value of the first four moments of ρ, ϕ, and the mean of λ over the $2,000$ samples. For comparative purposes, we also report ML and GMM estimates of the aggregate hybrid model where heterogeneity is ignored. These estimators are arguably biased (see Pesaran and Smith, 1995; Imbs, Jondeau, and Pelgrin, 2007). However, they provide a benchmark from which it is possible to evaluate the ability of the proposed estimators to overcome the aggregation bias.[18]

We first discuss the relative performance of the estimators in capturing the persistence of the forcing variable. Since the conclusions are robust across the different cases and remain almost unchanged when the time and individual dimensions vary, Table 2 reports the results for three different values of p_ρ and q_ρ, corresponding to medium, low, and high persistence of the forcing variable.[19] The first column corresponds to the theoretical value of the first four moments of the true distribution of ρ. The second and third columns

[17] This lag specification corresponds to the estimator defined by Lewbel (1994), i.e., without accounting for either the dependence between the coefficients or the correlation between β and the forcing variable. An increase in K_β has only minimal end effects.

[18] GMM estimation of Eq. (13) is obtained using four lags of $Y_{w,t}$ and $X_{w,t}$ as instruments. ML estimation is obtained by the joint OLS estimation of Eqs. (14) and (15) with $\bar{u}_{w,t}$ and $\bar{v}_{w,t}$ assumed to be uncorrelated.

[19] Aggregation of AR(1) processes is well known in the aggregation literature (see Robinson, 1978; Granger, 1980). The proposed estimators are designed to correct the aggregation bias of the dependent variable $Y_{w,t}$ and do not convey new results in the case of an $AR(1)$ process, with the exception of the cross-sectional moments of each parameters.

report the ML and GMM estimates. The subsequent columns describe the results for the unrestricted, parametric, and MD approaches, for which we provide estimates of the first four moments of ρ.

[Insert Table 2 around here]

For the three cases, the ML and GMM estimates do not display any sizeable finite-sample bias. Although theoretical analysis shows that both estimators asymptotically overestimate the true value of the parameter $E(\rho)$, this result is not surprising, since the autoregressive parameter is known to be downward biased in finite samples, even in a correctly specified model (Sawa, 1978). The simulation experiment shows that both effects compensate for each other. Regarding the procedures that correct for heterogeneity, it is worth noting that the unrestricted approach performs reasonably well for the first two moments but provides excessively large estimates of the skewness and kurtosis of the distribution of ρ. In contrast, the parametric and MD approaches yield almost unbiased estimates of all of the distribution moments in our benchmark (Panel A) and in the case of low persistence (Panel B). In the case of a higher persistence (Panel C), the correcting estimators do not perform as well as in the previous cases because estimation is affected by the near-unit root behavior of some simulated series.

We now discuss the results of the Monte Carlo simulations for the dynamics of the endogenous variable, reported in Table 3 ($N = 100$, $T = 200$). The estimation of the cross-sectional distribution of ϕ is clearly the most challenging, because it is very likely to reflect all the biases due to heterogeneity. We begin with the benchmark (Panel A).As already discussed in Section 3, the commonly used ML and GMM procedures asymptotically overestimate the true intrinsic persistence, $E(\phi)$, in the presence of unaccounted heterogeneity. As expected, these two estimators are severely biased. The biases are as large as 0.2 for the ML and GMM estimators. In contrast, the proposed heterogeneity-correcting estimators are almost unbiased. At the same time, these estimators sharply contrast in terms of cross-sectional higher moment estimates. The unrestricted approach performs rather poorly, especially for the kurtosis. In contrast, the parametric and MD approaches yield unbiased estimates of these moments.

[Insert Table 3 around here]

Regarding the parameter of the forcing variable, two points are worth emphasizing. First, the unrestricted regression estimator does not perform well, since it does not impose the restrictions implied by the rational expectations model. Second, the parametric and MD estimators display no sizeable bias. All in all, these results reveal that, in finite samples, the parametric and MD approaches significantly reduce the heterogeneity biases incurred by the ML and GMM estimators. In addition, both heterogeneity-correcting techniques provide accurate estimates of the moments of the cross-sectional distribution of ρ and ϕ, and outperform the unrestricted approach.

We now assess the robustness of our benchmark results by varying the characteristics of the cross-section distribution of the parameters. First, we reduce the persistence of the forcing variable, with a mean value $E(\rho)$ decreasing from 0.9 to 0.85 (Case 2). The results, reported in Panel B, clearly show that the biases in the aggregate estimates are not significantly altered when the forcing variable is less persistent.[20] Second, we reduce the dispersion of the intrinsic persistence of the endogenous variable, measured by ϕ, from 0.224 to 0.167 (Case 3). As reported in Panel C, this results in smaller biases in the estimation of $E(\phi)$. The two proposed heterogeneity-correcting techniques yield almost unbiased estimate of the cross-sectional moments of ϕ, while the over-estimation of the mean value of ϕ with ML and GMM is still as large as 0.1 and 0.17, respectively. All the estimation techniques perform better in this case (irrespective of the persistence of the forcing variable). The intuition is that the reduction in the dispersion of ϕ reinforces the role of the extrinsic persistence. Since the first moment of ρ is well captured

[20] Reducing (respectively, increasing) the dispersion of the persistence parameter of the forcing variable (ρ) implies a decrease (respectively, an increase) in the bias of the intrinsic persistence of the endogenous variable (ϕ) for all methods.

by all approaches, this tends to significantly reduce the bias of the lagged endogenous variable. In contrast, an increase in the dispersion in the intrinsic persistence yields the opposite effect. Third, we consider the case of a decrease in the intrinsic persistence, with the mean value of ϕ decreased to 0.4 (Case 4). As reported in Panel D, there is no particular change in the bias of the various estimators (irrespective of the forcing variable persistence). The two proposed heterogeneity-correcting estimators again outperform the standard ML and GMM techniques as well as the unrestricted regression estimator. An increase in the intrinsic persistence yields the same conclusion. As a final change in the value of the parameter set, we increase the mean value of λ from 0.1 to 0.15 (Case 5, Panel E). This leads to slightly more biased estimates of the intrinsic persistence.

Unsurprisingly, these results suggest that the sources of persistence, intrinsic *versus* extrinsic, and their distributions are fundamental in explaining the heterogeneity bias in the estimation of the parameter of the lagged endogenous variable. The level of extrinsic persistence of the forcing variable (ρ), as well as the level of intrinsic persistence of the variable of interest (ϕ), directly affect the magnitude of the ML and GMM biases in $E(\phi)$. The higher the persistence in the dynamics of the micro-units is, the larger the bias in the standard estimation techniques. It is worth emphasizing that heterogeneity-correcting estimators are much less sensitive to this problem.

Finally, we assess the robustness of our simulation results for different values of N and T. Table 4 presents the results obtained with $T = 200$ and $N = 1000$. Three points are worth commenting. First, the two proposed heterogeneity-correcting estimators still outperform the unrestricted estimator and the standard ML and GMM estimators. In particular, both estimators reproduce the first four cross-sectional moments of ϕ more accurately than the unrestricted estimator. Second, as in Table 3, the parametric and MD estimators exhibit similar performances: the relative efficiency of both estimators, measured by the standard errors reported in brackets, is case-dependent. Third, comparing with Table 3, we notice that the median bias of the proposed estimators on the average value of each parameter reduces with the number of cross-sectional units (N). In contrast, the median bias of the ML and GMM estimators is not reduced when N is increased. Finally, Table 5 shows that the conclusions above are robust when both T and N are large ($T = N = 1000$).

[Insert Tables 4 and 5 around here]

In summary, these results clearly state that the parametric and MD estimators perform very well for a wide range of parameters. In all cases, these estimators outperform the unrestricted estimator as well as the ML and GMM estimators in correcting the aggregation bias in the presence of unobserved parameter heterogeneity. Moreover, the parametric and MD estimators provide reliable information regarding the moments of the cross-sectional distribution. Finally, we also tested the robustness of our results with respect to the decomposition of the shock (idiosyncratic *versus* common factor), the number of lags, the number of moment conditions, and the choice of the weighting matrices. The results, which are available upon request, show that our conclusions remain unchanged.

5 Application

In this section, we empirically address the following questions: What can be inferred about the aggregation bias of the inflation dynamics, especially the New Keynesian Phillips Curve (NKPC)? What can be inferred about the cross-sectional distribution of the parameters?

After bursting onto the scene of mainstream monetary economics, the New Keynesian Phillips curve has been the focus of two important empirical debates. First, to what extent can a purely forward-looking pricing behavior be reconciled with observed inflation persistence? Second, to what extent do properly measured marginal costs affect inflation dynamics? Both issues are crucial and have recently been hotly debated, with good reason. In that respect, one conclusion drawn from the vast body of evidence is that price dynamics are heterogeneous and that reported inflation persistence may be an artifact of aggregation. More specifically, macroeconomic estimates have been criticized on the ground that the average macro duration of price stickiness and the aggregate intrinsic persistence of inflation are inconsistent with micro evidence (see Bils and Klenow, 2005; Angeloni et al., 2006; Dhyne et al., 2007). This apparent contradiction between micro and macro evidence may be explained by the fact that the aggregate Phillips curve results from the aggregation of individual price dynamics and that heterogeneity is key in explaining the discrepancy (Leith and Maley, 2007; Imbs, Jondeau, and Pelgrin, 2007).

In this respect, we analyze the aggregation bias of the NKPC by assuming that it results from firms or sectoral heterogeneity. Following Galí and Gertler (1999), the cross-sectional model is given by:

$$\begin{aligned} \pi_{i,t} &= \omega_i \; \pi_{i,t-1} + (1-\omega_i) E_t \; \pi_{i,t+1} + \lambda_i \; mc_{i,t} + \varepsilon_{i,t}, \\ mc_{i,t} &= \rho_i \; mc_{i,t-1} + v_{i,t}, \end{aligned}$$

where $\pi_{i,t}$ and $mc_{i,t}$ are the inflation rate and the real marginal cost, respectively, and the parameters $(\omega_i, \lambda_i, \rho_i)$ are i.i.d. realizations of the random variables (ω, λ, ρ) (Assumption 1). The corresponding reduced form is written as:

$$\pi_{i,t} = \phi_i \; \pi_{i,t-1} + \beta_i \; mc_{i,t} + u_{i,t},$$

where $\phi_i = 1/(1-\omega_i)$, $\beta_i = \lambda_i(1+\phi_i)/(1-\rho_i)$, and $u_{i,t} = (1+\phi_i)\varepsilon_{i,t}$. Ignoring sectoral heterogeneity leads to the following macro model:

$$\begin{aligned} \pi_{w,t} &= \bar{\phi} \; \pi_{w,t-1} + \bar{\beta} \; mc_{w,t} + \bar{u}_{w,t}, \\ mc_{w,t} &= \bar{\rho} \; mc_{w,t-1} + \bar{v}_{w,t}, \end{aligned}$$

where $\pi_{w,t}$ is the aggregate inflation and $mc_{w,t}$ the aggregate marginal costs, both variables being defined as the weighted average of the corresponding sectoral variable. The error terms are defined as $\bar{u}_{w,t} = u_{w,t} + \sum_{i=1}^{N} w_i \; \tilde{\phi}_i \; \pi_{i,t-1} + \sum_{i=1}^{N} w_i \; \tilde{\beta}_i \; mc_{i,t}$, and $\bar{v}_{w,t} = v_{w,t} + \sum_{i=1}^{N} w_i \; \tilde{\rho}_i \; mc_{i,t-1}$ respectively. The heterogeneity-correcting estimators detailed in Section 3 rely on the following truncated dynamics:

$$\begin{aligned} \pi_{w,t} &= \sum_{s=1}^{K_\phi} A_s \; \pi_{w,t-s} + \sum_{s=0}^{K_\beta} B_s \; mc_{w,t-s} + U_{w,t}, \\ mc_{w,t} &= \sum_{s=1}^{K_\rho} C_s \; mc_{w,t-s} + V_{w,t}, \end{aligned}$$

where $\{A_s\}$, $\{B_s\}$, and $\{C_s\}$ are defined in Proposition 1.

To estimate these models, we use French data on prices and marginal costs. Our data are quarterly from 1978:1 to 2005:3, for a total of 111 observations. Interestingly, the French Statistical Institute (INSEE) also provides quarterly sectoral information on these quantities covering all French economic activity, using the same sector definitions.[21] Consequently, we can compare our estimates with those of the ML and GMM approaches, as well as with those of the standard panel estimators, in particular the Mean Group (MG) and Random Coefficient (RC) estimators, based on sectoral data.[22] In addition, we assess the reliability of the cross-sectional moment estimates obtained using heterogeneity-correcting approaches.

Before going through the details of the results, it is worth describing how we imple-

[21]The inflation rate is computed as the quarter-on-quarter growth rate of the value-added deflator. In computing a measure of the real marginal cost, we follow Sbordone (2001) or Galí, Gertler, and Lopez-Salido (2001) and define $mc_{i,t}$ as the (logarithm) deviation of the share labor income in value added from its sample mean. We have data on output price, wages, and employment for sixteen sectors of the French economy, comprising all industries. Coverage includes agriculture, manufacturing (six sectors), and services (nine sectors). It is quite difficult to obtain similar information of homogenous quality for any country, including the United States.

[22]The MG estimator, introduced in Swamy (1970), consists in an arithmetic average of sector-specific parameter estimates. The RC estimator is defined as a weighted average of the sectoral estimates, with weights inversely proportional to their covariance matrix. For a discussion, see Hsiao and Pesaran (2004).

ment the heterogeneity-correcting estimators. As explained before, implementation of the unrestricted and parametric estimators requires some truncation lags, K_ϕ, K_ρ, and K_β, while the MD estimator requires assumptions about the number of auto- and cross-covariances and the weighting matrix. On the one hand, we investigated several values for the number of lags and did not find any sizeable difference. We report estimates with $(K_\phi, K_\beta, K_\rho) = (4, 0, 4)$ for the unrestricted estimator and $(K_\phi, K_\beta, K_\rho) = (8, 0, 8)$ for the parametric estimator. On the other hand, the MD estimator is evaluated to provide the best fit of the variance and the first ten auto- and cross-covariances of $\pi_{w,t}$ and $mc_{w,t}$. Finally, the weighting matrix is based on linearly decreasing weights, as in Newey and West (1987).[23]

Results are reported in Table 6. The first column reports the cross-sector moments of the parameters. Although they may be rather poor proxies of the moments of the actual distribution of parameters across micro-units, they do provide a benchmark for evaluating the ability of the competing estimators to reproduce the shape of the cross-sectional distribution of the parameters. The next four columns report the estimation of the standard aggregate estimation techniques (ML and GMM estimators) and panel estimation techniques (MG and RC estimators). The last three columns report the unrestricted, parametric, and MD estimates. They include the estimate of the mean parameter as well as the next three moments.

[Insert Table 6 around here]

Several points are worth commenting. First, the ML and GMM estimates of the intrinsic persistence ($\bar{\phi}$) sharply contrast with those of the proposed estimators and the panel estimators. Indeed, the heterogeneity-correcting (and panel) estimators tend to favor less persistent inflation dynamics and thus a more forward-looking specification of the

[23] We also experimented with more lags (for the unrestricted and parametric approaches), and with more moment conditions, more lags of the autocovariance functions, and alternative weighing matrices (for the MD approach), with minimal end effects. These results are not reported to save space, but are available upon request.

NKPC. These results are consistent with the simulations reported in Section 4, which show that standard aggregate estimates of $E(\phi)$ are asymptotically upward biased. In addition, with the exception of the MD estimator, the proposed heterogeneity-correcting methods reinforce the role of the forcing variable as a key determinant of the inflation dynamics. In particular, the parametric estimate of $E(\lambda)$ is substantially larger than those of the GMM and ML, and broadly consistent with the panel estimators. Therefore, the inflation dynamics inherit more extrinsic persistence through the real marginal cost when unobserved sectoral heterogeneity is taken into account.

Second, the heterogeneity-correcting estimates are close to those obtained from panel techniques on sectoral data.[24] The parametric and MD estimators provide parameter estimates of the three parameters that are very close to the MG and RC estimates (based on sectoral averages). Our results suggest that these estimation techniques are able to correct the heterogeneity bias.

Third, if we compare the heterogeneity-correcting estimation techniques, we notice that the unrestricted estimation approach only partially offsets the aggregation bias. It tends to overestimate the mean value of ϕ and underestimate the mean value of λ. In addition, it provides very imprecise estimates of the other moments of the cross-sectional distribution of the parameters. If one considers the distribution of ρ, the table provides evidence that both the standard deviation and the kurtosis of the distribution are too large, whereas the skewness estimate is too low. Such extreme estimates of the higher moments of the distribution have been already documented by Lewbel (1994) for the parameters driving U.S. consumption expenditure dynamics. In contrast, the parametric and MD estimators provide similar estimates of the higher moments, which turn out to closely reproduce the moments implied by sectoral estimates.

[24]The MG and RC estimates are based on sectoral estimates of $(\rho_i, \phi_i, \lambda_i)$. More disaggregate data are simply not available. Therefore, the proposed estimators can be compared only to the sectoral characteristics of the inflation dynamics.

6 Conclusion

In this paper, we addressed the correction of the aggregation bias in rational expectations models when only macro data are available. We considered general linear (reduced-form) dynamic models, which encompass linear aggregate rational expectations models, such as, for instance, the New Keynesian Phillips curve or any Euler equation. Starting from Lewbel (1994), we proposed two new consistent estimators that account for the dependence across parameters encountered in such models. The first estimator is based on a flexible parametric specification of the cross-sectional parameter distributions. The second estimator exploits the autocovariance properties of the aggregate processes through an MD estimation. A Monte Carlo study revealed that the finite-sample properties of the proposed estimators outperform those of the unrestricted approach and that we can correct the bias of the standard ML and GMM approaches. As a by-product, we can also infer the cross-sectional distribution of the parameters under certain restrictions.

Finally, we studied the aggregation bias regarding the New Keynesian Phillips curve based on aggregate French data. Empirical evidence suggests that standard estimation techniques overestimate the intrinsic persistence (e.g., the contribution of lagged inflation) and underestimate the role of the forcing variable, the real marginal cost. In contrast, the proposed estimators suggest that the forward-looking component is predominant and that the real marginal cost is key in determining the inflation dynamics.

Appendices

6.1 Appendix 1: Derivation of the reduced form of a linear rational expectations model

The full model rests on the following system:

$$\begin{aligned} y_{i,t} &= \omega_i \; y_{i,t-1} + (1-\omega_i) \;\; E_t \; y_{i,t+1} + \lambda_i \; x_{i,t} + \varepsilon_{i,t}, & (24) \\ x_{i,t} &= \rho_i \; x_{i,t-1} + v_{i,t}. & (25) \end{aligned}$$

The characteristic equation with respect to $y_{i,t}$ writes:

$$\left(1 - \omega_i \; L - (1-\omega_i) \;\; L^{-1}\right) y_{i,t} = \lambda_i \; x_{i,t} + \varepsilon_{i,t},$$

where L denotes the lag operator. Determining the roots of the characteristic equation, $\delta_{1i} = \omega_i/(1-\omega_i) = \phi_i$ and $\delta_{2i} = 1$, one obtains:

$$y_{i,t} = \phi_i \; y_{i,t-1} + \lambda_i \, (1+\phi_i) \, E_t \sum_{k=0}^{\infty} x_{i,t+k} + u_{i,t},$$

where $u_{i,t} = (1+\phi_i) \, \varepsilon_{i,t}$. Using the dynamics of the forcing variable, one gets:

$$\begin{aligned} y_{i,t} &= \phi_i \; y_{i,t-1} + \lambda_i \, (1+\phi_i) \, E_t \sum_{k=0}^{\infty} \rho_i^k \; x_{i,t} + u_{i,t} \\ &= \phi_i \; y_{i,t-1} + \frac{\lambda_i \, (1+\phi_i)}{(1-\rho_i)} x_{i,t} + u_{i,t}. \end{aligned}$$

This can be re-written as:

$$\begin{aligned} y_{i,t} &= \phi_i \; y_{i,t-1} + \beta_i \; x_{i,t} + u_{i,t}, \\ x_{i,t} &= \rho_i \; x_{i,t-1} + v_{i,t}, \end{aligned}$$

where $\beta_i \equiv \lambda_i \, (1+\phi_i) \, / \, (1-\rho_i)$.

6.2 Appendix 2: Proof of Proposition 1.

6.2.1 Determination of the aggregate dynamics

The individual dynamic model is given by:

$$\begin{aligned} y_{i,t} &= \phi_i \; y_{i,t-1} + \beta_i \; x_{i,t} + u_{i,t} \\ x_{i,t} &= \rho_i \; x_{i,t-1} + v_{i,t}. \end{aligned}$$

The cross-section aggregate model is defined by:

$$\begin{aligned} Y_{w,t} &= E(\phi_i \; y_{i,t-1}) + E(\beta_i \; x_{i,t}) + U_{w,t} \\ X_{w,t} &= E(\rho_i \; x_{i,t-1}) + V_{w,t}. \end{aligned}$$

The first equation can be rewritten as follows:

$$Y_{w,t} = A_1 \; Y_{w,t-1} + B_0 \; X_{w,t} + \sum_{i=1}^{N} w_i(\phi_i - A_1) y_{i,t-1} + \sum_{i=1}^{N} w_i(\beta_i - B_0) x_{i,t} + \sum_{i=1}^{N} w_i \; u_{i,t},$$

where $A_1 = E(\phi)$ and $B_0 = E(\beta)$.

Starting from this expression, we can iterate the procedure and obtain:

$$\begin{aligned} Y_{w,t} &= A_1 \; Y_{w,t-1} + B_0 \; X_{w,t} + \sum_{i=1}^{N} w_i(\phi_i - A_1)\,(\phi_i \; y_{i,t-2} + \beta_i \; x_{i,t-1} + u_{i,t-1}) \\ &\quad + \sum_{i=1}^{N} w_i(\beta_i - B_0)\,(\rho_i \; x_{i,t-1} + v_{i,t}) + \sum_{i=1}^{N} w_i \; d_{i,0} \; \varepsilon_{i,t} \\ &= A_1 \; Y_{w,t-1} + B_0 \; X_{w,t} + \sum_{i=1}^{N} w_i\,(\phi_i - A_1)\,\phi_i \; y_{i,t-2} + \sum_{i=1}^{N} w_i\,(\phi_i - A_1)\,\beta_i \; x_{i,t-1} \\ &\quad + \sum_{i=1}^{N} w_i\,(\beta_i - B_0)\,\rho_i \; x_{i,t-1} + \sum_{i=1}^{N} w_i \; d_{i,0} \; \varepsilon_{i,t} + \sum_{i=1}^{N} w_i\,(\phi_i - A_1)\,d_{i,1} \; \varepsilon_{i,t-1} \\ &\quad + \sum_{i=1}^{N} w_i \; e_{i,1} \; v_{i,t}, \end{aligned}$$

where $d_{i,0} = 1 + \phi_i$, $d_{i,1} = (\phi_i - A_1)\,(1 + \phi_i)$ and $e_{i,1} = \beta_i - B_0$.

Then, replacing $y_{i,t-2}$ by $\phi_i \; y_{i,t-3} + \beta_i \; x_{i,t-2} + u_{i,t-2}$, and $x_{i,t-1}$ by $\rho_i \; x_{i,t-2} + v_{i,t-1}$, we get:

$$\begin{aligned} Y_{w,t} &= A_1 \; Y_{w,t-1} + A_2 \; Y_{w,t-2} + B_0 \; X_{w,t} + B_1 \; X_{w,t} \\ &\quad + \sum_{i=1}^{N} w_i \left\{ \left[(\phi_i - A_1)\,\phi_i - A_2 \right] \phi_i - A_3 \right\} y_{i,t-3} \\ &\quad + \sum_{i=1}^{N} w_i \left\{ \left[(\phi_i - A_1)\,\phi_i - A_2 \right] \beta_i + \left[(\phi_i - A_1) + (\beta_i - B_0)\,\rho_i - B_1 \right] \rho_i - B_2 \right\} x_{i,t-2} \\ &\quad + \sum_{i=1}^{N} w_i \; d_{i,0} \; \varepsilon_{i,t} + \sum_{i=1}^{N} w_i \; d_{i,0}\,(\phi_i - A_1)\,\varepsilon_{i,t-1} + \sum_{i=1}^{N} w_i \; d_{i,0} \left[(\phi_i - A_1)\,\phi_i - A_2 \right] \varepsilon_{i,t-2} \\ &\quad + \sum_{i=1}^{N} w_i \; e_{i,1} \; v_{i,t} + \sum_{i=1}^{N} w_i \left[(\phi_i - A_1)\,\beta_i + (\beta_i - B_0)\,\rho_i - B_1 \right] v_{i,t-1}. \end{aligned}$$

Iterating again yields:

$$Y_{w,t} = \sum_{s=1}^{\infty} A_s \; Y_{w,t-s} + \sum_{s=0}^{\infty} B_s \; X_{w,t-s} + \sum_{s=0}^{\infty}\sum_{i=1}^{N} w_i \; d_{i,s} \; \varepsilon_{i,t-s} + \sum_{s=1}^{\infty}\sum_{i=1}^{N} w_i \; e_{i,s} \; v_{i,t-s},$$

where the sequences of parameters $A_s = E(a_s)$ and $B_s = E(b_s)$ are defined as:

$$\begin{array}{lll} a_1 = \phi, & a_{s+1} = (a_s - A_s)\,\phi, & s = 1, \cdots, \infty, \\ b_0 = \beta, & b_{s+1} = (a_{s+1} - A_{s+1})\,\beta + (b_s - B_s)\,\rho, & s = 0, \cdots, \infty, \end{array}$$

and

$$\begin{array}{lll} d_{i,0} = 1 + \phi_i & d_{i,s} = \phi_i \; d_{i,s-1} - A_s(1 + \phi_i) & s = 0, \cdots, \infty, \\ e_{i,1} = \beta_i - B_0 & e_{i,s} = \frac{d_{i,s}}{1+\phi_i}\beta_i + \rho_i \; e_{i,s-1} - B_s & s = 1, \cdots, \infty. \end{array}$$

Similarly, it is straightforward to show that the aggregation of the forcing variable dynamics:

$$X_{w,t} = \sum_{s=1}^{\infty} C_s \; X_{w,t-s} + \sum_{s=0}^{\infty}\sum_{i=1}^{N} w_i \; f_{i,s} \; v_{i,t-s},$$

where the sequence of parameters $C_s = E(c_s)$ is related to the moments of the distribution of ρ as follows:

$$c_1 = \rho, \qquad c_{s+1} = (c_s - C_s)\,\rho, \qquad s = 1, \cdots, \infty,$$

and

$$f_{i,0} = 1 \qquad f_{i,s} = \rho_i \; f_{i,s-1} - C_s \qquad s = 0, \cdots, \infty.$$

Now, as $N \to \infty$, we have:

$$\begin{array}{l} \sum_{s=0}^{\infty}\sum_{i=1}^{N} w_i \; d_{i,s} \; \varepsilon_{i,t-s} \xrightarrow{p} (1 + E(\phi))\,\epsilon_t, \\ \sum_{s=1}^{\infty}\sum_{i=1}^{N} w_i \; e_{i,s} \; v_{i,t-s} \xrightarrow{p} 0, \\ \sum_{s=0}^{\infty}\sum_{i=1}^{N} w_i \; f_{i,s} \; v_{i,t-s} \xrightarrow{p} \tilde{\epsilon}_t. \end{array}$$

Therefore, the dynamics of the aggregate variables $Y_{w,t}$ and $X_{w,t}$ when $N \to \infty$ are given by:

$$\begin{array}{rcl} Y_{w,t} & = & \sum_{s=1}^{\infty} A_s \; Y_{w,t-s} + \sum_{s=0}^{\infty} B_s \; X_{w,t-s} + U_{w,t}, \\ X_{w,t} & = & \sum_{s=1}^{\infty} C_s \; X_{w,t-s} + V_{w,t}, \end{array}$$

with $U_{w,t} = (1 + E(\phi))\epsilon_t$ and $V_{w,t} = \tilde{\epsilon}_t$.

In particular, it can be shown without further restrictions that the first three coefficients of

A_s and C_s are positive. Indeed, $A_1 = m_1 = E(\phi) > 0$, $A_2 = m_2 - m_1^2 = V(\phi) > 0$, with $m_2 = E(\phi^2)$. In the case of A_3, we have:

$$A_3 = m_3 - A_1 m_2 - A_2 m_1 = m_3 - 2m_1 m_2 + m_1^3,$$

with $m_3 = E(\phi^3)$. Using a Gauss decomposition, we obtain:

$$A_3 = m_1 \left(m_1 - \frac{m_2}{m_1} \right)^2 + m_3 - \frac{m_2^2}{m_1}.$$

Using the Cauchy-Schwarz inequality, it is straightforward to show that $m_2^2 \leq m_1 m_3$. Therefore, A_3 (and C_3) is positive. Finally, A_2 and A_3 are bounded above by 1/4. The proof is given in Lewbel (1994). Note that the positiveness of all coefficients A_s and C_s can be stated for specific distributions, as for instance the Beta or the uniform distribution.
□

6.2.2 Convergence of $\{A_s\}$ and $\{C_s\}$

We show the result for the A_s terms. The proof is the same for the C_s terms. Let $m_s = E(\phi^s)$ denote the non-central moment of order s of the cross-section distribution of ϕ. As shown, for instance, in Stuart and Ord (1994), the non-central moments m_s of a random variable defined over $(0, 1)$ satisfy

$$1 \geq m_1 \geq \cdots \geq m_s \geq 0, \forall s \geq 1,$$

and $m_s \underset{s\to\infty}{\to} 0$. Then, we have:

$$A_{s+1} = E(a_{s+1}) = E\left(\phi^{s+1}\right) - \sum_{r=1}^{s} A_r\ E\left(\phi^{s-r+1}\right) = m_{s+1} - \sum_{r=1}^{s} A_r\ m_{s-r+1},$$

so that:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ m_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ m_s & \cdots & m_1 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_{s+1} \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_{s+1} \end{pmatrix}.$$

It is worth noting that the solution $A = (A_1, A_2, \cdots)'$ of the previous linear system is given by the sequence of coefficients of the series $A(z) = \sum_{i=1}^{\infty} A_{i+1} z^i$ that satisfies:

$$(1 + zM(z))A(z) = M(z), \tag{26}$$

where $M(z)$ denotes the series $\sum_{j=0}^{\infty} m_{j+1}\ z^j$, with the m_j terms are the non-central moments $E(X^j)$ of a non degenerate real random variable defined on $[0, 1]$, and $\varphi(z) = 1 + zM(z)$ is the moment generating function of X. In order to prove that assertion, we must show that the A_s

terms are the solution of the initial recurrent system. After expanding Eq. (26), we get:

$$(1 + zM(z))A(z) = \sum_{k=0}^{\infty} \left(\sum_{i=1}^{k+1} A_i \; m_{k+1-i} \right) z^k = \sum_{k=0}^{\infty} m_{k+1} \; z^k,$$

with $m_0 = 1$ (by convention). Using Eq. (26), we obtain:

$$m_{k+1} = \sum_{i=1}^{k+1} A_i \; m_{k+1-i} = A_1 \; m_k + A_2 \; m_{k-1} + \cdots + A_k \; m_1 + A_{k+1},$$

for all $k \geq 0$, and:

$$\begin{aligned} A_{k+1} &= m_{k+1} - \sum_{i=1}^{k} A_i \; m_{k+1-i} \\ &= m_{k+1} - (A_1 \; m_k + A_2 \; m_{k-1} + \cdots + A_k \; m_1), \end{aligned}$$

for all $k \geq 0$. Finally, the radius of convergence of $M(z)$ is at least one, since the sequence $\{m_j\}$ is bounded. Therefore, the series $M(z)$ is an analytical function of the real variable $z \in]-1;1[$ (and of the complex variable z in the open disk $\mathcal{D}(0,1)$).

Starting from this rewriting, we can now use a Tauberian theorem in order to show that the series $A(1) = \sum_{i=0}^{\infty} A_{i+1}$ is absolute summable (i.e., $A_s \to 0$ as $s \to \infty$). For instance, we can use the following result. If a function $z \mapsto u(z)$, $z \in [0,1[$, defined by a series $\sum_{k \geq 0} u_k \; z_k$ with a radius of convergence larger than or equal to one, has positive coefficients and if it admits a limit ℓ as $z \to 1^-$, then the series $u(1) = \sum_{k \geq 0} u_k$ is absolute summable and it converges to ℓ.[25] Assuming that the coefficients A_s are positive, which is the case if we specify a Beta distribution (among others), we must show that (i) $A(z)$ has a limit when $z \to 1^-$ and (ii) the series $A(z)$ has a radius of convergence larger than or equal to one. First, we write:

$$A(z) = \frac{M(z)}{1 + zM(z)}.$$

Since the coefficients of the series $M(z)$ are positive, it turns out that the function $z \mapsto M(z)$ is increasing over $[0,1[$. Therefore, there exists a limit, which can be infinite, $\lambda \in \mathbb{R} \cup \{\infty\}$. This implies that $A(z)$ converges to $\ell = \lambda/(\lambda + 1) \in \mathbb{R}$ or $\ell = 1$ as $z \to 1^-$. Second, regarding the radius of convergence, it is worth noting that the series $A(z)$ is the ratio of two series with a radius of convergence at least equal to one. Therefore, we only need to show that the denominator $\varphi(z) = 1 + zM(z)$ has no zero on the open disk $\mathcal{D}(0,1)$. In doing so, we can use the following lemma.

[25]The condition regarding the positiveness of the A_s terms can be relaxed by showing that $k \; u_k$ is bounded among others.

Lemma 4 *Assume that X is a non degenerate random variable with support $[0,1]$ and that $m_j = E(X^j)$ is the non-central moment of order j. The integral form of $\varphi(z)$ is given by:*

$$\varphi(z) = \int_0^1 \frac{d\mu(x)}{1 - z\ x},$$

where $\mu(x)$ represents the corresponding probability measure associated to X. Moreover, if X is characterized by a probability density function, then:

$$\varphi(z) = \int_0^1 \frac{f(x)}{1 - z\ x} dx.$$

□

Using the previous lemma, we obtain:

$$\varphi(z) = \int_0^1 \frac{1 - \bar{z}\ x}{|1 - z\ x|^2} d\mu(x).$$

Therefore $\varphi(z) = 0$ is equivalent to:

$$\bar{z} = \frac{\int_0^1 \frac{d\mu(x)}{|1-x\ z|^2}}{\int_0^1 \frac{x\ d\mu(x)}{|1-x\ z|^2}} > 1.$$

This implies that $\varphi(z) = 0$ if and only if $z > 1$ and $\varphi(z) = 1 + zM(z)$ has no zero on the open disk $\mathcal{D}(0,1)$. The result follows. □

Proof of Lemma 4: Using the definition of the moment generating function, we have:

$$\varphi(z) = \sum_{k=0}^{\infty} z^k\ E(X^k) = E\left(\sum_{k=0}^{\infty} z^k\ X^k\right),$$

where the second right-hand side equality is a consequence of the theorem of dominated convergence. Indeed, for a given z in the open disk $\mathcal{D}(0,1)$, one has $0 \leq |z|\ X(\omega) < 1$, which implies that the sequence $\left(\sum_{k=0}^{n} |z|^k\ X(\omega)^k\right)_{n\geq 0}$ converges for all $\omega \in \Omega$ and has a limit $g(\omega) = \frac{1}{1-|z|\ X(\omega)}$. Consequently, the sequence of partial sums $s_n(\omega) = \sum_{k=0}^{n} z^k\ X(\omega)^k$ converges for all ω and is majored (in modulus) by the function $g(\omega)$. In addition,

$$E(g) = \int_\Omega \frac{1}{1 - |z|\ X(\omega)} dP(\omega) = \int_0^1 \frac{d\mu(x)}{1 - |z|\ x} < \infty.$$

Therefore, using the theorem of dominated convergence, we get:

$$\varphi(z) = \lim_{n\to\infty} E(s_n) = E\left(\lim_{n\to\infty} s_n\right),$$

which implies:

$$\varphi(z) = E\left(\sum_{k=0}^{\infty} z^k\ X^k\right) = E\left(\frac{1}{1 - z\ X}\right),$$

and:

$$\varphi(z) = \int_0^1 \frac{d\mu(x)}{1 - z\ x}.$$

□

6.3 Appendix 3: Proof of Lemma 1

From Proposition 1, we get:

$$\begin{aligned} B_0 &= E(b_0) = E(\beta) = E\left(\frac{\lambda(1+\phi)}{1-\rho}\right) \\ &= E(\lambda)\, E(1+\phi)\, E\left(\frac{1}{1-\rho}\right), \end{aligned}$$

using Assumption 1.
More generally, we have $B_s = E(b_s)$. Using again Proposition 1, we have:

$$\begin{aligned} B_s &= E(b_s) = E\left[(a_s - A_s)\,\beta + (b_{s-1} - B_{s-1})\,\rho\right] \\ &= E\left((a_s - A_s)\,\beta\right) + E\left((b_{s-1} - B_{s-1})\,\rho\right). \end{aligned}$$

We now give the recursion formulae for both right hand side expressions. For the first term, we obtain:

$$\begin{aligned} E\left((a_s - A_s)\,\beta\right) &= E\left(a_s\beta\right) - A_s E\left(\beta\right) = E\left(a_{s-1}\phi\beta\right) - A_{s-1}E\left(\phi\beta\right) - A_s E\left(\beta\right) \\ &= E\left((a_{s-2} - A_{s-2})\,\phi^2\beta\right) - A_{s-1}E\left(\phi\beta\right) - A_s E\left(\beta\right) \\ &= E\left(\phi^s\beta\right) - \sum_{k=1}^{s} A_k E\left(\phi^{s-k}\beta\right) \\ &= E(\lambda)\left[E\left(\phi^s\,(1+\phi)\right) - \sum_{k=1}^{s} A_k E\left(\phi^{s-k}\,(1+\phi)\right)\right] E\left(\frac{1}{1-\rho}\right) \\ &= E(\lambda)\sum_{k=0}^{s} \tilde{A}_k E\left(\phi^{s-k} + \phi^{s+1-k}\right) E\left(\frac{1}{1-\rho}\right), \end{aligned}$$

with $\tilde{A}_0 = 1$ and $\tilde{A}_s = -A_s$ for $s \geq 1$.

For the second term, the proof is similar. We have:

$$\begin{aligned} E\left(b_{s-1}\rho\right) &= E\left(\left[\left(a_{s-1}-A_{s-1}\right)\beta+\left(b_{s-2}-B_{s-2}\right)\rho\right]\rho\right) \\ &= E\left(b_{s-2}\rho^{2}\right)+E\left(\left(a_{s-1}-A_{s-1}\right)\beta\rho\right)-E\left(B_{s-2}\rho^{2}\right) \\ &= E\left(\lambda\right)\sum_{j=0}^{s-1}\sum_{k=0}^{j}\tilde{A}_{k}E\left(\phi^{j-k}+\phi^{j+1-k}\right)E\left(\frac{\rho^{s-j}}{1-\rho}\right)-\sum_{j=0}^{s-2}B_{j}E\left(\rho^{s-j}\right). \end{aligned}$$

Finally, we obtain:

$$\begin{aligned} B_{s} &= E\left(\left(a_{s}-A_{s}\right)\beta\right)+E\left(b_{s-1}\rho\right)-E\left(B_{s-1}\rho\right) \\ &= E\left(\lambda\right)\sum_{j=0}^{s}\sum_{k=0}^{j}\tilde{A}_{k}E\left(\phi^{j-k}+\phi^{j+1-k}\right)E\left(\frac{\rho^{s-j}}{1-\rho}\right)-\sum_{j=0}^{s-1}B_{j}E\left(\rho^{s-j}\right). \end{aligned}$$

□

6.4 Appendix 4: Proof of Proposition 2

Each micro-unit satisfies Eq. (4) and (5). By virtue of Assumptions 1 and 2, we can write:

$$y_{i,t} = \frac{\beta_i}{\left(1-\phi_i L\right)\left(1-\rho_i L\right)}v_{i,t}+\frac{u_{i,t}}{1-\phi_i L}, \tag{27}$$

$$x_{i,t} = \frac{v_{i,t}}{1-\rho_i L}, \tag{28}$$

or equivalently:

$$\begin{aligned} y_{i,t} &= \sum_{s=0}^{\infty}\beta_i\ \Theta_{i,s}\ v_{i,t-s}+\sum_{s=0}^{\infty}\phi_i^{s}\ u_{i,t-s}, \\ x_{i,t} &= \sum_{s=0}^{\infty}\rho_i^{s}\ v_{i,t-s}, \end{aligned}$$

since:

$$\begin{aligned} \frac{1}{\left(1-\phi_i L\right)\left(1-\rho_i L\right)} &= \frac{1}{\rho_i-\phi_i}\left[\frac{\rho_i}{1-\rho_i L}-\frac{\phi_i}{1-\phi_i L}\right]=\frac{1}{\rho_i-\phi_i}\left\{\sum_{s=0}^{\infty}\left(\rho_i^{s+1}-\phi_i^{s+1}\right)L^{s}\right\} \\ &= \frac{1}{\rho_i-\phi_i}\left\{\sum_{s=0}^{\infty}\left(\rho_i-\phi_i\right)\left[\sum_{r=0}^{s}\rho_i^{s}\ \phi_i^{s-r}\right]L^{s}\right\} \\ &= \sum_{s=0}^{\infty}\sum_{r=0}^{s}\rho_i^{r}\ \phi_i^{s-r}\ L^{s}=\sum_{s=0}^{\infty}\Theta_{i,s}\ L^{s}, \end{aligned}$$

where $\Theta_{i,s}=\sum_{r=0}^{s}\rho_i^{r}\ \phi_i^{s-r}$.

After aggregating, we obtain:

$$
\begin{aligned}
Y_{w,t} &= \sum_{i=1}^{N} w_i \sum_{s=0}^{\infty} \beta_i \ \Theta_{is} \ v_{i,t-s} + \sum_{i=1}^{N} w_i \sum_{s=0}^{\infty} \phi_i^s \ u_{i,t-s} \\
&= \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i \ E\left(\beta \ \Theta_s\right) v_{i,t-s} + \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i \left(\beta_i \ \Theta_{is} - E\left(\beta \ \Theta_s\right)\right) v_{i,t-s} \\
&\quad + \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i \ E\left(\phi^s(1+\phi_i)\right) \varepsilon_{i,t-s} + \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i \left(\phi_i^s(1+\phi_i) - E\left(\phi^s(1+\phi)\right)\right) \varepsilon_{i,t-s} \\
&= \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i \ E\left(\beta \ \Theta_s\right) v_{i,t-s} + \sum_{s=0}^{\infty} \sum_{i=1}^{N} w_i \ E\left(\phi^s(1+\phi_i)\right) \varepsilon_{i,t-s},
\end{aligned}
$$

under Assumption 2. As $N \to \infty$,

$$
\begin{aligned}
Y_{w,t} &= \sum_{s=0}^{\infty} E\left(\beta \ \Theta_s\right) \sum_{i=1}^{N} w_i \ v_{i,t-s} + \sum_{s=0}^{\infty} E\left(\phi^s(1+\phi)\right) \sum_{i=1}^{N} w_i \ \varepsilon_{i,t-s} \\
&= \sum_{s=0}^{\infty} E\left(\beta \ \Theta_s\right) \tilde{\epsilon}_{t-s} + \sum_{s=0}^{\infty} E\left(\phi^s + \phi^{s+1}\right) \epsilon_{t-s}, \\
&= \sum_{s=0}^{\infty} \delta_s \ \tilde{\epsilon}_{t-s} + \sum_{s=0}^{\infty} \psi_s \ \epsilon_{t-s}.
\end{aligned}
$$

where $\delta_s = E\left(\beta \ \Theta_s\right)$ and $\psi_s = E\left(\phi^s\right) + E(\phi^{s+1})$. Similarly, we have:

$$
X_{w,t} = \sum_{i=1}^{N} \sum_{s=0}^{\infty} w_i \ \rho_i^s \ v_{i,t-s} = \sum_{s=0}^{\infty} \gamma_s \ \tilde{\epsilon}_{t-s},
$$

where $\gamma_s = E\left(\rho^s\right)$.
In addition, it is straightforward to show that $\{\psi_s\}$ (respectively, $\{\gamma_s\}$) is absolute summable if $\{\phi^s\}$ (respectively, $\{\rho^s\}$) is absolute summable. This in turn implies that $\psi_s \to 0$ (respectively, $\gamma_s \to 0$) as $s \to \infty$. Finally, $\{\delta_s\}$ is absolute summable if both $\{\rho^s\}$ and $\{\phi^s\}$ are absolute summable.

□

6.5 Appendix 5: Proof of Lemma 2

In order to show Lemma 2, we need to determine the general expression of $\delta_s = E(\beta\ \Theta_s)$. We have:

$$\begin{aligned}
\delta_s &= E(\beta\ \Theta_s) = \sum_{i=1}^{N} w_i\ \beta_i \left(\sum_{r=0}^{s} \rho_i^r\ \phi_i^{s-r}\right) = E\left[\beta\left(\sum_{r=0}^{s} \rho^r\ \phi^{s-r}\right)\right] \\
&= E(\beta) \sum_{r=0}^{s} E(\rho^r)\, E\left(\phi^{s-r}\right) = E(\beta) \sum_{r=0}^{s} E(\rho^r)\, E\left(\phi^{s-r}\right).
\end{aligned}$$

We know from the expression of the reduced form of the model that:

$$\beta_i = \frac{\lambda_i (1+\phi_i)}{1-\rho_i} = \lambda_i (1+\phi_i) \sum_{k=0}^{\infty} \rho_i^s.$$

After replacing this expression in δ_s, we obtain:

$$\begin{aligned}
\delta_s &= E(\beta\ \Theta_s) = \sum_{i=1}^{N} w_i\ \beta_i\ \Theta_{is} \\
&= \sum_{i=1}^{N} w_i \left\{\left[\lambda_i (1+\phi_i) \sum_{k=0}^{\infty} \rho_i^k\right]\left[\sum_{r=0}^{s} \rho_i^r\ \phi_i^{s-r}\right]\right\} \\
&= \sum_{i=1}^{N} w_i\ \lambda_i \left\{\sum_{r=0}^{s} \left(\phi_i^{s-r} + \phi_i^{s-r+1}\right) \sum_{k=0}^{\infty} \rho_i^{k+r}\right\} \\
&= E(\lambda) \sum_{r=0}^{s} \left(E\left(\phi^{s-r}\right) + E\left(\phi^{s-r+1}\right)\right) \sum_{k=0}^{\infty} E\left(\rho^{k+r}\right) \\
&= E(\lambda) \sum_{r=0}^{s} E\left(\frac{\rho^r}{1-\rho}\right) \left(E\left(\phi^{s-r}\right) + E\left(\phi^{s-r+1}\right)\right).
\end{aligned}$$

Since $E(\lambda)$ is bounded, $E(\phi^s) \to 0$, and $E(\rho^s) \to 0$ for s sufficiently large, we have $\delta_s \to 0$ as $s \to \infty$.

□

6.6 Appendix 6: Proof of Lemma 3

From Proposition 2, we have:

$$\begin{aligned}
Y_{w,t} &= \sum_{s=0}^{\infty} \delta_s\ \tilde{\epsilon}_{t-s} + \sum_{s=0}^{\infty} \psi_s\ \epsilon_{t-s}, \\
X_{w,t} &= \sum_{s=0}^{\infty} \gamma_s\ \tilde{\epsilon}_{t-s}.
\end{aligned}$$

Since ϵ_{t-s} and $\tilde{\epsilon}_{t-s}$ are independent (under Assumption 2), the autocovariance function of $Y_{w,t}$ can be characterized as follows, as $N \to \infty$:

$$
\begin{aligned}
cov\left(Y_{w,t}, Y_{w,t-h}\right) &= Cov\left(\sum_{s=0}^{\infty} \delta_s \ \tilde{\epsilon}_{t-s} + \sum_{s=0}^{\infty} \psi_s \ \epsilon_{t-s}, \sum_{s=0}^{\infty} \delta_{s+h} \ \tilde{\epsilon}_{t-s} + \sum_{s=0}^{\infty} \psi_{s+h} \ \epsilon_{t-s}\right) \\
&= \sum_{s=0}^{\infty} \delta_s \ \delta_{s+h} V\left(\tilde{\epsilon}_{t-s}\right) + \sum_{s=0}^{\infty} \psi_s \ \psi_{s+h} V\left(\epsilon_{t-s}\right) \\
&= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \delta_s \ \delta_{s+h} + \sigma_{\epsilon}^2 \sum_{s=0}^{\infty} \psi_s \ \psi_{s+h}, \qquad h \geq 0.
\end{aligned}
$$

Following the same line of reasoning, for $h \geq 0$:

$$
\begin{aligned}
Cov\left(X_{w,t}, X_{w,t-h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \gamma_s \ \gamma_{s+h}, \\
Cov\left(Y_{w,t}, X_{w,t-h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \delta_{s+h} \ \gamma_s, \\
Cov\left(Y_{w,t}, X_{w,t+h}\right) &= \sigma_{\tilde{\epsilon}}^2 \sum_{s=0}^{\infty} \delta_s \ \gamma_{s+h}.
\end{aligned}
$$

□

References

[1] Abadir, K., and G. Talmain, 2002, Aggregation, Persistence and Volatility in a Macro Model, Review of Economic Studies 69, 749–779.

[2] Angeloni, I., L. Aucremanne, M. Ehrmann, J. Galí, A. Levin, and F. Smets, 2006, New Evidence on Inflation Persistence and Price Stickiness in the Euro Area: Implications for Macro Modelling, Journal of the European Economic Association 4, 562–574.

[3] Altissimo, F., B. Mojon, and P. Zaffaroni, 2009, Can Aggregation Explain the Persistence of Inflation? Journal of Monetary Economics 56, 231–241.

[4] Barker, T., and H. Pesaran, 1990, Disaggregation in Econometric Modelling, Routledge, London.

[5] Bils, M., and P. Klenow, 2005, Some Evidence on the Importance of Sticky Prices, Journal of Political Economy 112, 947–985.

[6] Blundell, R., and T.M. Stoker, 2005, Heterogeneity and Aggregation, Journal of Economic Literature XLIII, 347–391.

[7] Canova, F., 2007, Applied Macroeconomic Research, Princeton University Press.

[8] Dhrymes, P.J., 1971, Distributed Lags: Problems of Estimation and Formulation, San Francisco, CA: Holden-Day.

[9] Dhyne, E., L. Alvarez, H. Le Bihan, G. Veronese, D. Dias, J. Hoffman, N. Jonker, P. Lünnemann, F. Rumler, and J. Vilmunen, 2004, Price Setting in the Euro Area: Some Stylized Facts From Individual Consumer Price Data, ECB Working Paper Series, No. 524.

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

[11] Galí, J., and M. Gertler, 1999, Inflation Dynamics: A Structural Econometrics Analysis, Journal of Monetary Economics 44, 195–222.

[12] Galí, J., M. Gertler, and D. López-Salido, 2001, European Inflation Dynamics, European Economic Review 45, 1237–1270.

[13] Gonçalves, E., and C. Gouriéroux, 1988, Agrégation de Processus Autorégressifs d'ordre 1, Annales d'Économie et de Statistique 12, 127–149.

[14] Granger, C.W.J., 1990, Aggregation of Time-Series Variables – A Survey, in Barket, T., and M.H. Pesaran (eds), Disaggregation in Econometric Modelling, Routledge, London, 17–34.

[15] Granger, C.W.J., 1987, Implications of Aggregation with Common Factors, Econometric Theory 3, 209–222.

[16] Granger, C.W.J., 1980, Long Memory Relationships and the Aggregation of Dynamic Models, Journal of Econometrics 14, 227–238.

[17] Hsiao, C., and M.H. Pesaran, 2004, Random Coefficient Panel Data Models, CESifo Working Paper Series, No. 1233.

[18] Imbs, J., E. Jondeau, and F. Pelgrin, 2007, Aggregating Phillips Curves, Swiss Finance Institute Research Paper No 07-06.

[19] Kelejian, H.H., 1980, Aggregation and Disaggregation of Non-Linear Equations, in Evaluation of Econometric Models, by Kmenta, J., and J.B. Ramsey (eds), Academic Press, New York.

[20] Leith, C., and J. Malley, 2007, A Sectoral Analysis of Price-Setting Behaviour in U.S. Manufacturing Industries, Review of Economics and Statistics 89, 335–342.

[21] Lewbel, A., 1992, Aggregation with Log-Linear Models, Review of Economic Studies 59, 635–642.

[22] Lewbel, A., 1994, Aggregation and Simple Dynamics, American Economic Review 84, 905–918.

[23] Newey, W.K., and K.D. West, 1987, A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703–708.

[24] Pesaran, M.H., 1973, The Small Sample Problem of Truncation Remainders in the Estimation of Distributed Lag Models with Auto-correlated Errors, International Economic Review 14, 120–131.

[25] Pesaran, M.H., 2003, Aggregation of Linear Dynamic Models: An Application to Life-Cycle Consumption Models under Habit Formation, Economic Modelling 20, 383–415.

[26] Pesaran, M.H., R.G. Pierse, and M.S. Kumar, 1989, Econometric Analysis of Aggregation in the Context of Linear Prediction Models, Econometrica 57, 861–888.

[27] Pesaran, M.H., R.G. Pierse, and K.C. Lee, 1994, Choice Between Disaggregate and Aggregate Specifications Estimated by Instrumental Variables Methods, Journal of Business and Economic Statistics 12, 11–21.

[28] Pesaran, M.H., and R. Smith, 1995, Estimation of Long-run Relationships from Dynamic Heterogenous Panels, Journal of Econometrics 68, 79–114.

[29] Pesaran, M.H., T. Schuermann, and S.M. Weiner, 2004, Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model, Journal of Business and Economic Statistics 22, 129–162.

[30] Pesaran, M.H., R. Smith, and K.S. Im, 1996, Dynamic Linear Models For Heterogeneous Panels, Chapter 8 in L. Mátyás and P. Sevestre (eds.), The Econometrics of Panel Data: A Handbook of the Theory With Applications, Kluwer Academic Publishers, 145–195.

[31] Robinson, P.M., 1978, Statistical Inference for a Random Coefficient Autoregressive Model, Scandinavian Journal of Statistics 5, 163–168.

[32] Sawa, T., 1978, The Exact Moment of the Least Square Estimator for the Autoregressive Model, Journal of Econometrics 8, 159–172.

[33] Sbordone, A., 2001, Prices and Unit Labor Costs: A New Test of Price Stickiness, Journal of Monetary Economics 49, 265–292.

[34] Smets, F., and R. Wouters, 2007, Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, American Economic Review 97, 586–606.

[35] Stoker, T.J., 1993, Empirical Approaches to the Problem of Aggregation over Individuals, Journal of Economic Literature 31, 1827–1874.

[36] Stuart, A., and K. Ord, 1994, Kendall's Advanced Theory of Statistics, vol. I, sixth edition, Hodder Arnold.

[37] Swamy, P.A.V.B., 1970, Efficient Inference in a Random Coefficient Regression Model, Econometrica 38, 311–323.

[38] Van Garderen, K.J., K. Lee, and M.H. Pesaran, 2000, Cross-sectional Aggregation of Non-Linear Models, Journal of Econometrics 95, 285–331.

[39] Zaffaroni, P., 2004, Contemporaneous Aggregation of Linear Dynamic Models in Large Economies, Journal of Econometrics 120, 75–102.

Captions

Table 1: This table presents the Monte Carlo simulation experiments performed in this paper. (p_ρ, q_ρ), (p_ϕ, q_ϕ), and (p_λ, q_λ) denote the parameters p and q of the Beta distribution corresponding to ρ, ϕ, and λ, respectively.

Table 2: This table reports the results of the Monte Carlo simulation experiments for the persistence parameter of the forcing variable. For each case, the table displays the estimates of the mean value of ρ obtained using the ML and GMM techniques and the first four moments of ρ estimated using the heterogeneity-correcting techniques. The table also reports, in square brackets, the standard error of the estimate of $E(\rho)$. The experiments are based on $2,000$ samples of $N = 100$ and $T = 200$. Truncation lags are $(K_\phi, K_\beta, K_\rho) = (4, 0, 4)$ for the unrestricted estimator and $(T/20, T/20, T/20)$ for the parametric estimator. The MD estimator is based on ten moment conditions.

Table 3: This table reports results of the Monte Carlo simulation experiments described in Table 1 for $N = 100$ and $T = 200$. For each case, the table displays the estimates of the mean values of ϕ and λ obtained using the ML and GMM techniques and the first four moments of ϕ as well as the mean of λ estimated using the heterogeneity-correcting techniques. The table also reports, in square brackets, the standard error of the estimates of $E(\phi)$ and $E(\lambda)$. The experiments are based on $2,000$ samples. Truncation lags are $(K_\phi, K_\beta, K_\rho) = (4, 0, 4)$ for the unrestricted estimator and $(T/20, T/20, T/20)$ for the parametric estimator. The MD estimator is based on ten moment conditions.

Table 4: This table reports results of the Monte Carlo simulation experiments described in Table 1 for $N = 1000$ and $T = 200$. See Table 3 for details.

Table 5: This table reports results of the Monte Carlo simulation experiments described in Table 1 for $N = 1000$ and $T = 1000$. See Table 3 for details.

Table 6: This table reports parameter estimates of the New Keynesian Phillips curve, using French aggregate data. The estimators are the ML and GMM techniques based on aggregate data, the MG and RC techniques based on sectoral data, and the heterogeneity-correcting techniques. The estimations are based on the following assumptions. Truncation lags are $(K_\phi, K_\beta, K_\rho) = (4, 4, 4)$ for the unrestricted estimator and $(10, 4, 10)$ for the parametric estimator. The MD estimator is based on ten moment conditions.

Table 1: **Characteristics of the Monte Carlo simulation experiments**

Simulations	ρ		ϕ		λ	
	p_ρ	q_ρ	p_ϕ	q_ϕ	p_λ	q_λ
Benchmark	36	4	2	2	10	90
Case 2	34	6	2	2	10	90
Case 3	36	4	4	4	10	90
Case 4	36	4	2	3	10	90
Case 5	36	4	2	2	15	85

Table 2: **Estimation of the forcing variable dynamics (simulated data)**

Param.	Moments	True value	ML	GMM	Unrestricted regression	Parametric regression	MDE
Panel A: Benchmark $((p_{rho}, q_{rho}) = (36, 4))$							
ρ	mean	0.900	0.904	0.905	0.888	0.897	0.893
			[0.014]	[0.015]	[0.027]	[0.017]	[0.018]
	std dev.	0.047			0.025	0.026	0.048
	skweness	-0.813			-4.573	-0.457	-0.769
	kurtosis	3.829			23.912	3.264	3.729
Panel B: Low persistence in the forcing variable $((p_{rho}, q_{rho}) = (34, 6))$							
ρ	mean	0.850	0.853	0.854	0.837	0.843	0.849
			[0.016]	[0.017]	[0.026]	[0.019]	[0.003]
	std dev.	0.056			0.054	0.045	0.056
	skweness	-0.598			-4.323	-0.496	-0.593
	kurtosis	3.384			20.866	3.223	3.375
Panel C: High persistence in the forcing variable $((p_{rho}, q_{rho}) = (38, 2))$							
ρ	mean	0.950	0.956	0.953	0.930	0.970	0.934
			[0.011]	[0.011]	[0.024]	[0.005]	[0.014]
	std dev.	0.034			0.010	0.003	0.034
	skweness	-1.259			-4.693	-0.159	-0.998
	kurtosis	5.183			18.846	3.037	4.343

Table 3: Estimation of the endogenous variable dynamics (simulated data)

Param.	Moments	True value	ML	GMM	Unrestricted regression	Parametric regression	MDE
Panel A: Benchmark							
ϕ	mean	0.500	0.697	0.721	0.544	0.519	0.520
			[0.054]	[0.155]	[0.047]	[0.051]	[0.067]
	std dev.	0.224			0.241	0.246	0.213
	skweness	0.000			0.030	-0.049	-0.052
	kurtosis	2.143			6.639	2.071	2.267
λ	mean	0.100	0.111	0.074	0.155	0.107	0.110
			[0.031]	[0.051]	[0.054]	[0.036]	[0.038]
Panel B: Low persistence in the forcing variable (case 2)							
ϕ	mean	0.500	0.604	0.707	0.528	0.515	0.499
			[0.049]	[0.180]	[0.056]	[0.050]	[0.035]
	std dev.	0.224			0.222	0.211	0.221
	skweness	0.000			0.335	-0.013	0.003
	kurtosis	2.143			3.868	2.277	2.178
λ	mean	0.100	0.125	0.077	0.138	0.098	0.098
			[0.029]	[0.068]	[0.035]	[0.035]	[0.017]
Panel C: Low dispersion in ϕ (case 3)							
ϕ	mean	0.500	0.605	0.673	0.518	0.502	0.516
			[0.044]	[0.161]	[0.047]	[0.050]	[0.045]
	std dev.	0.167			0.176	0.183	0.171
	skweness	0.000			-0.260	-0.001	-0.060
	kurtosis	2.455			9.449	2.405	2.513
λ	mean	0.100	0.130	0.093	0.164	0.110	0.122
			[0.034]	[0.052]	[0.056]	[0.042]	[0.048]
Panel D: Small value of ϕ (case 4)							
ϕ	mean	0.400	0.532	0.685	0.433	0.420	0.424
			[0.055]	[0.199]	[0.047]	[0.049]	[0.056]
	std dev.	0.200			0.223	0.214	0.188
	skweness	0.286			0.402	0.177	0.202
	kurtosis	2.357			7.612	2.355	2.454
λ	mean	0.100	0.130	0.080	0.161	0.110	0.132
			[0.032]	[0.057]	[0.052]	[0.040]	[0.051]
Panel E: Increase in λ (case 5)							
ϕ	mean	0.500	0.746	0.734	0.556	0.531	0.534
			[0.047]	[0.146]	[0.047]	[0.055]	[0.112]
	std dev.	0.224			0.232	0.252	0.177
	skweness	0.000			0.103	-0.090	-0.075
	kurtosis	2.143			9.898	2.048	2.509
λ	mean	0.150	0.138	0.121	0.226	0.163	0.163
			[0.038]	[0.057]	[0.075]	[0.048]	[0.073]

Table 4: **Estimation of the endogenous variable dynamics (simulated data)**

Param.	Moments	True value	ML	GMM	Unrestricted regression	Parametric regression	MDE
Panel A: Benchmark							
ϕ	mean	0.500	0.708	0.718	0.548	0.525	0.514
			[0.038]	[0.154]	[0.047]	[0.047]	[0.060]
	std dev.	0.224			0.235	0.253	0.184
	skweness	0.000			0.319	-0.069	-0.037
	kurtosis	2.143			6.679	2.031	2.397
λ	mean	0.100	0.112	0.079	0.161	0.108	0.109
			[0.029]	[0.048]	[0.051]	[0.034]	[0.066]
Panel B: Low persistence in the forcing variable (case 2)							
ϕ	mean	0.500	0.610	0.721	0.525	0.513	0.498
			[0.041]	[0.171]	[0.056]	[0.051]	[0.057]
	std dev.	0.224			0.223	0.203	0.213
	skweness	0.000			0.333	-0.011	0.005
	kurtosis	2.143			3.987	2.297	2.240
λ	mean	0.100	0.121	0.076	0.142	0.093	0.104
			[0.025]	[0.067]	[0.039]	[0.036]	[0.034]
Panel C: Low dispersion in ϕ (case 3)							
ϕ	mean	0.500	0.612	0.672	0.516	0.495	0.511
			[0.034]	[0.159]	[0.047]	[0.049]	[0.039]
	std dev.	0.167			0.183	0.193	0.162
	skweness	0.000			0.278	0.006	-0.048
	kurtosis	2.455			8.285	2.361	2.508
λ	mean	0.100	0.131	0.096	0.166	0.110	0.111
			[0.032]	[0.054]	[0.052]	[0.041]	[0.048]
Panel D: Small value of ϕ (case 4)							
ϕ	mean	0.400	0.545	0.684	0.434	0.415	0.411
			[0.040]	[0.189]	[0.047]	[0.048]	[0.051]
	std dev.	0.200			0.227	0.221	0.188
	skweness	0.286			0.235	0.211	0.242
	kurtosis	2.357			7.180	2.325	2.456
λ	mean	0.100	0.132	0.080	0.159	0.109	0.122
			[0.032]	[0.053]	[0.059]	[0.037]	[0.052]
Panel E: Increase in λ (case 5)							
ϕ	mean	0.500	0.735	0.733	0.556	0.536	0.548
			[0.028]	[0.144]	[0.047]	[0.047]	[0.086]
	std dev.	0.224			0.234	0.258	0.176
	skweness	0.000			-0.023	-0.113	-0.173
	kurtosis	2.143			10.549	1.991	2.485
λ	mean	0.150	0.155	0.124	0.234	0.173	0.177
			[0.043]	[0.054]	[0.073]	[0.051]	[0.067]

Table 5: **Estimation of the endogenous variable dynamics (simulated data)**

Param.	Moments	True value	ML	GMM	Unrestricted regression	Parametric regression	MDE
Panel A: Benchmark							
ϕ	mean	0.500	0.724	0.628	0.563	0.529	0.505
			[0.020]	[0.066]	[0.047]	[0.020]	[0.024]
	std dev.	0.224			0.248	0.230	0.223
	skweness	0.000			0.253	-0.086	-0.035
	kurtosis	2.143			12.306	2.123	2.157
λ	mean	0.100	0.097	0.069	0.131	0.102	0.100
			[0.013]	[0.016]	[0.022]	[0.016]	[0.018]
Panel B: Low persistence in the forcing variable (case 2)							
ϕ	mean	0.500	0.632	0.607	0.542	0.532	0.500
			[0.019]	[0.073]	[0.056]	[0.020]	[0.017]
	std dev.	0.224			0.234	0.209	0.224
	skweness	0.000			0.188	-0.089	-0.001
	kurtosis	2.143			6.459	2.237	2.143
λ	mean	0.100	0.109	0.074	0.124	0.097	0.099
			[0.011]	[0.022]	[0.016]	[0.014]	[0.008]
Panel C: Low dispersion in ϕ (case 3)							
ϕ	mean	0.500	0.636	0.584	0.537	0.518	0.504
			[0.017]	[0.063]	[0.047]	[0.021]	[0.012]
	std dev.	0.167			0.194	0.156	0.167
	skweness	0.000			0.908	-0.030	-0.009
	kurtosis	2.455			16.371	2.516	2.456
λ	mean	0.100	0.110	0.085	0.139	0.101	0.103
			[0.013]	[0.018]	[0.025]	[0.013]	[0.021]
Panel D: Small value of ϕ (case 4)							
ϕ	mean	0.400	0.567	0.548	0.450	0.427	0.403
			[0.020]	[0.082]	[0.047]	[0.022]	[0.020]
	std dev.	0.200			0.235	0.190	0.200
	skweness	0.286			0.777	0.192	0.275
	kurtosis	2.357			10.865	2.377	2.352
λ	mean	0.100	0.110	0.076	0.137	0.097	0.105
			[0.013]	[0.019]	[0.023]	[0.013]	[0.021]
Panel E: Increase in λ (case 5)							
ϕ	mean	0.500	0.756	0.641	0.570	0.524	0.539
			[0.018]	[0.053]	[0.047]	[0.021]	[0.030]
	std dev.	0.224			0.235	0.241	0.219
	skweness	0.000			0.148	-0.073	-0.114
	kurtosis	2.143			19.636	2.066	2.189
λ	mean	0.150	0.131	0.105	0.189	0.157	0.149
			[0.018]	[0.020]	[0.032]	[0.021]	[0.024]

Table 6: Estimation of the New Keynesian Phillips curve using French aggregate data

Moments	Sectoral moments	ML	GMM	MG	RC	Unrestricted regression	Parametric regression	MDE
ρ								
mean	0.937	0.973	0.970	0.937	0.931	0.964	0.922	0.955
		[0.020]	[0.019]	[0.003]	[0.019]	[0.094]	[0.027]	[0.027]
std	0.051					0.450	0.026	0.020
skew	-1.117					-4.782	-0.596	-0.839
kurt	2.840					17.946	3.474	3.990
ϕ								
mean	0.467	0.673	0.770	0.467	0.404	0.587	0.469	0.411
		[0.064]	[0.131]	[0.024]	[0.080]	[0.094]	[0.084]	[0.288]
std	0.308					0.361	0.291	0.305
skew	0.175					-3.508	0.108	0.323
kurt	1.758					16.716	1.798	1.821
λ								
mean	0.058	0.025	0.000	0.058	0.044	0.026	0.043	0.023
		[0.020]	[0.007]	[0.013]	[0.045]	[0.017]	[0.026]	[0.049]