

# Heterogeneous dynamics, aggregation and the persistence of economic shocks.

Laura Mayoral\*

Institute for Economic Analysis, (CSIC)

## Abstract

It has been recently emphasized that, if individuals have heterogeneous dynamics, estimates of shock persistence based on aggregate data are significantly higher than those derived from its disaggregate counterpart. However, a careful examination of the implications of this statement on the various tools routinely employed to measure persistence is missing in the literature.

This paper formally examines this issue. We consider a disaggregate linear model with heterogeneous dynamics and compare the values of several measures of persistence across aggregation levels. Interestingly, we show that the average persistence of aggregate shocks, as measured by the impulse response function (IRF) of the aggregate model or by the average of the individual IRFs, is identical on all horizons. Interestingly, similar conclusions hold when the basic framework is extended in several directions, including non-linear and VAR models.

In contrast, other popular persistence measures, such as the sum of the autoregressive coefficients or the largest autoregressive root, tend to be higher the higher the aggregation level. We argue, however, that this should be seen more as an undesirable property of these measures than as evidence of different average persistence across aggregation levels. The results are illustrated in an application using U.S. inflation data.

Key words: Heterogeneous dynamics, aggregation, persistence, impulse response function, sum of the autoregressive coefficients, non-linear models, U.S. inflation persistence.

---

\*I am very grateful to Manuel Arellano, Lola Gadea, Albert Marcet, Thies van Rens, Enrique Sentana, Paolo Zaffaroni and participants in seminars at the Universities of Harvard, MIT, Glasgow, Southampton, Pompeu Fabra, CEMFI and IAE. Financial support from the Spanish Ministry of Education through grant SEJ2006-00369 and from the Fundación Ramón Areces is gratefully acknowledged. The usual disclaimer applies.

## 1. INTRODUCTION

In this paper we are interested in the measurement of the persistence of economic shocks across different aggregation levels in a context where the underlying processes are allowed to have heterogeneous dynamics. By persistence, we refer to the speed and pattern of adjustment of the process (or processes) of interest to economic shocks of different natures.

Heterogeneous dynamics at the individual level have been found to be important in a wide variety of contexts, such as in the speed of reversion of real exchange rates (Imbs et al., 2005 and Crucini and Shintani, 2008) and of income shocks (Hu and Ng, 2004), in the dynamics of saving behavior (Haque et al., 2000), in inflation dynamics (Altissimo et al., 2006a and the references therein), in labor demand across firms (Zhang and Small, 2006), etc. In fact, one could argue that the existence of some degree of heterogeneity across individuals is likely to be the rule rather than the exception in most contexts.

A recent strand of literature has pointed out that, if the underlying DGP is a dynamic model with heterogeneous coefficients, estimates of shock persistence based on aggregate data are significantly higher than those derived from averaging the corresponding persistence measures computed with disaggregate data. Two explanations have been provided to account for this phenomenon. On the one hand, building on the results of Pesaran and Smith (1995), it has been argued that estimates computed with aggregate data are biased and that the sign of the bias is positive (see Imbs et al., 2005). On the other, influenced by the results in Robinson (1978) and Granger (1980), some authors have suggested that the aggregation of heterogeneous dynamic processes is not an innocuous operation and that it may tend to increase overall shock response (see Altissimo et al., 2009 among others).

Several empirical studies corroborate these arguments by finding estimates of persistence that vary considerably across aggregation levels and are, in general, higher, the higher the level of aggregation (see Imbs et al., 2005, Altissimo et al., 2009, 2006b, Clark, 2006, Lünnemann and Mathä, 2004, etc). In contrast, some authors have reported very similar persistence values across aggregation levels in highly heterogeneous datasets (Crucini and Shintani, 2008, Gadea and Mayoral, 2009). The conclusions of these articles are typically drawn by comparing averages of individual persistence measures with their corresponding

values computed with aggregate data. However, different papers employ different tools to measure persistence and therefore, it is not clear whether these findings are related to specific properties of the considered data or whether the use of a particular persistence measure can systematically bias the conclusions towards a specific direction.

This paper examines the relationship among measures of persistence of aggregate shocks computed at different levels of aggregation. We start by considering a linear model with a large number of individual processes that may have heterogeneous dynamics, the dynamic random coefficients model, and we use the most widely employed measures to establish the comparisons, namely, the impulse response function and other popular scalar measures such as the sum of the autoregressive coefficients, the largest autoregressive root, the half-life, etc.

Our results demonstrate that not all the measures routinely employed in applications to establish persistence comparisons fair equally well. We show that the response to an aggregate shock, as measured by the impulse response function (IRF) of the aggregate model or by the average of the individual IRFs, is the same on all horizons with or without individual dynamic heterogeneity. This implies that, according to the IRF, the aggregation of heterogeneous units does not magnify the *average* response to a shock.

In contrast, other popular persistence measures, such as the sum of the autoregressive coefficients (SAC), are not invariant to aggregation when individual heterogeneity is allowed for, being typically larger, the higher the level of aggregation. However, we argue that this should be interpreted more as an undesirable property of these measures, rather than as a sign of different average persistence across aggregation levels. The reason is the following: the SAC was introduced as a summary of the IRF because it has a direct relation with the cumulative impulse response (CIR), namely,  $SAC=1-CIR^{-1}$  (Andrews and Chen, 1994). Typically, when applied to disaggregate data, the SAC is computed for each of the micro-units and its average is used as a measure of micro-persistence. Then, this value is compared to the SAC obtained with aggregate data. However, since the relation between the CIR and the SAC is not linear, the average SAC has not a direct connection with the average CIR. We show that, while the average CIR remains constant across aggregation levels, a straightforward application of Jensen's inequality ensures that the average SAC tends

to increase systematically with the level of aggregation. Thus, by relying on the latter measure, one could conclude that the average response to a shock increases with the level of aggregation when, in fact, a more thorough analysis of the IRF would suggest the opposite. Similar problems appear when other tools, such as the largest autoregressive root (LAR), are employed for these purposes.

Macroeconomic models typically involve several variables and often imply non-linear heterogeneous dynamics. We have extended the basic framework by allowing for non-linear behaviour and VAR models at the disaggregate level and it is shown that similar conclusions still hold in these cases. These results have some important implications on macroeconomic modelling. Some authors have warned against the practice of using averages of microparameters to calibrate macro models (Altissimo et al., 2009). Our findings qualify this statement by showing that the adequacy of this procedure entirely depends on the type of microparameter employed. Thus, averages of individual IRFs can be used in macro calibration independently of whether there is heterogeneity or not at the individual level in a wide variety of situations, including both linear and non-linear individual behavior under heterogeneity. However, as mentioned above, this conclusion cannot be extended to other popular parameters.

Finally, to illustrate the theoretical results, U.S. inflation data at different levels of aggregation has been analyzed and it has been shown that the IRF remains fairly constant across them. Furthermore, we find that other measures, such as the SAR, systematically increase with the level of aggregation, as reported in previous studies. Thus, by focusing only on the latter set of measures, one can easily conclude that the effect of aggregate shocks changes across aggregation levels when, in fact, a careful examination of the IRF would suggest the opposite.

The structure of this paper is as follows. Section 2 presents a simple model for the individual data, the dynamic linear random coefficients model (RCM), and defines some tools for measuring the average impact of an aggregate shock on the economy. Section 3 compares the measures of average micro-persistence derived in Section 2 with those obtained from the corresponding aggregate model. Section 4 extends the results obtained in previous sections in several directions. In particular, it allows for non-linear dynamics, VAR models,

correlation among the parameters of the disaggregate models and richer linear dynamics. Section 5 investigates, by using Monte Carlo simulation, whether it is possible to obtain good approximations to the population relationships established in the previous sections when the corresponding sample counterparts are employed. Section 6 presents an empirical illustration that analyzes the persistence of U.S. inflation data at different levels of aggregation and Section 7 concludes. The Appendix summarizes the conditions for the holding of a law of large numbers relating the sample average of the individual processes with the corresponding expectation.

## 2. HETEROGENEITY AND PERSISTENCE AT THE MICRO-LEVEL

Consider the problem of assessing the impact of an aggregate shock when micro data is available and the different units can be heterogeneous. This section presents a standard linear model with heterogeneous dynamics and reviews some measures of shock persistence that can be employed, or have been employed in the past, for analyzing this issue. Section 3 applies similar measures to the aggregate model and compares the two sets of results, while Section 4 extends the analysis in several directions.

A simple but commonly postulated model for microeconomic behavior that allows for heterogeneous dynamics is the Random Coefficients Model.<sup>1</sup> It describes an economy in which each unit  $i$  satisfies a linear regression with a Koyck lag given by:

$$y_{it} = a_i y_{it-1} + b_i' x_{it} + \nu_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

$$\nu_{it} = \rho_i u_t + \varepsilon_{it}, \quad (2)$$

where  $t$  denotes time,  $y_{it}$  and  $x_{it}$  are observable variables, the innovation  $\nu_{it}$  is the sum of two orthogonal, zero-mean martingale difference sequences, one common to all agents and one idiosyncratic, with variances  $\sigma_u^2 > 0$  and  $\sigma_\varepsilon^2$ , respectively;  $a_i$ ,  $b_i$  and  $\rho_i$  are unknown coefficients that are assumed to be particular draws from the variables  $a = \bar{a} + \eta^a$ ,  $b = \bar{b} + \eta^b$  and  $\rho = 1 + \eta^\rho$ , where  $\eta^k$ , for  $k \in \{a, b, \rho\}$ , are zero-mean random variables with variance  $\sigma_k^2$ . In this section, we further assume that the latter variables are mutually independent. The

---

<sup>1</sup>See Hsiao and Pesaran (2007) for a recent survey and Pesaran et al. (2002) and Pesaran (2006) for details on the estimation of this model.

distribution of  $a$  has bounded support in the interval  $(-1,1]$  and it is assumed that  $E_I(a^h)$  exists for all  $h$ , where  $E_I(\cdot)$  denotes expectation across the distribution of individuals.<sup>2</sup> The process  $x_{it}$  is assumed to be independent of  $u_t$  and the expectation  $E_I(b'x_t)$  is assumed to exist. If  $x_{it}$  is just a constant, then (1) is simply the first-order autoregressive model. The distribution of agents can be discrete or continuous and the number of agents is assumed to be countably or uncountably infinite.

Suppose now that, at time  $t$ , a unitary aggregate shock occurs and one is interested in measuring its *average* impact over time on a system like (1), populated by heterogeneous individuals. For each unit  $i$ , the impact of this shock  $h$  periods ahead can be evaluated through the IRF, defined as the difference between two forecasts (see Koops et al., 1996)

$$IRF_i(t, h) = E(y_{it+h}|u_t = 1; z_{it-1}) - E(y_{it+h}|u_t = 0; z_{it-1}) \quad (3)$$

where the operator  $E(\cdot|.)$  denotes the best mean squared error predictor and  $z_{it-1} = \left( y_{it-1}, y_{it-2}, \dots, x_{it-1}, x_{it-2}, \dots \right)'$ . Application of this definition to (1) yields

$$IRF_i(t, h) = \rho_i a_i^h, \text{ for } h \geq 0. \quad (4)$$

The average response to this aggregate shock can be computed as the expected value of (4) over the distribution of units, that we denote as *average disaggregate IRF* ( $IRF_{dis}$ ), given by

$$IRF_{dis}(t, h) = E_I(IRF(t, h)) = E_I(a^h), \text{ for } h \geq 0, \quad (5)$$

since in this section we assume that  $a$  and  $\rho$  are independent and  $E_I(\rho)$  has been normalized to 1 (the former assumption will be relaxed in Section 4). Thus, for the simple DGP considered here,  $IRF_{dis}$  is given by the  $h$ -th moment of the distribution of  $a$ .

Since the IRF is an infinite vector of numbers, it is a rather unwieldy measure of persistence. For this reason, scalar measures are frequently preferred, such as the Cumulated Impulse Response (*CIR*), which evaluates the total cumulative effect of a shock over time, and the half life (*HL*), defined as the number of periods it takes until half the effect of a shock dissipates. Using expression (5), it is straightforward to define the corresponding

---

<sup>2</sup>A sufficient, although not necessary, condition for this assumption to hold for all  $h$  is that the support of  $a$  is strictly contained in the interval  $[-1,1]$ .

versions of these measures at the micro level. The average cumulative response (denoted as  $CIR_{dis}$ ) can be computed as

$$CIR_{dis} = \sum_{h=0}^{\infty} IRF_{dis}(t, h), \quad (6)$$

which, for the simple model considered in this section, yields  $CIR_{dis} = E_I \left( \frac{1}{1-a} \right)$ . As for the *average disaggregate* HL ( $HL_{dis}$ ), it can be defined as the value of  $h$  that verifies<sup>3</sup>

$$IRF_{dis}(t, h)|_{h=HL_{dis}} = 0.5. \quad (7)$$

In many applications average shock persistence is evaluated using the *sum of the autoregressive coefficients* (SAC) and the *largest autoregressive root*. Typically, they are computed for each individual time series and, then, averages (or distribution quantiles) are reported as measures of average *micro* persistence; See Altissimo et al. (2006a), Bilke (2005), Clark (2006), Lünnemann and Mathä (2004), etc.

Finally, other authors have employed an alternative definition of ‘disaggregate’ IRF to the one presented in this section to draw their conclusions about persistence. This is the case of Imbs et al. (2005), who first construct an ‘artificial’ representative agent model,  $y_t^*$ , that has the same autoregressive structure as the units  $y_{it}$  but whose AR coefficients are given by the average of the individual-specific coefficients. Then, they compute the ‘disaggregate’ IRF by applying the standard definition of the IRF in (3) on  $y_t^*$ .

In the following section we assess the performance of the above-defined measures of persistence when used to compare persistence across aggregation levels.

### 3. INDIVIDUAL HETEROGENEITY AND PERSISTENCE AT THE MACRO LEVEL

This section compares the response to the same aggregate shock over time when it is computed from the micro model (1) or from its corresponding aggregate counterpart. To establish the relation, we first focus on the IRF and, then, the behavior of other popular

---

<sup>3</sup>Notice that, since the HL is a nonlinear function of the IRF, the average of the individual HLs does not coincide, in general, with the HL associated with  $IRF_{dis}$  defined in (7).

persistence measures is also examined. For the sake of clarity, this section deals with a very simple model and more general results are presented in Section 4.

How is the IRF computed from aggregate data related to the average of the individual impulse responses defined in Section 2? In a representative agent economy, the answer is trivial. This is because the individual and the aggregate models share the same dynamics so the IRFs derived from each model are also the same. When the individuals are heterogeneous, however, the dynamics of the aggregate process are generally different from those of the individual units. If all the micro-units are stationary and there is a finite number of them, it is still easy to derive the relationship between the two sets of IRFs.

Consider the simplest case: let  $y_{1t}$  and  $y_{2t}$  be defined as in (1) and (2), and let  $Y_t = (y_{1t} + y_{2t})/2$  be the aggregate process. If  $y_{it}$ , for  $i = \{1, 2\}$ , is weakly stationary (assuming that  $b_i = 0$  for simplicity), then

$$y_{1t} = \rho_1 u_t + a_1 \rho_1 u_{t-1} + a_1^2 \rho_1 u_{t-2} \dots + \epsilon_{1t}, \quad (8)$$

$$y_{2t} = \rho_2 u_t + a_2 \rho_2 u_{t-1} + a_2^2 \rho_2 u_{t-2} \dots + \epsilon_{2t}, \quad (9)$$

and

$$Y_t = \frac{((\rho_1 + \rho_2)u_t + (a_1 \rho_1 + a_2 \rho_2)u_{t-1} + (a_1^2 \rho_1 + a_2^2 \rho_2)u_{t-2} \dots + \epsilon_{1t} + \epsilon_{2t})}{2}, \quad (10)$$

where  $\epsilon_{it} = \sum_{j=0}^{\infty} a_i^j \varepsilon_{it-j}$ ,  $i = \{1, 2\}$ . Since the IRF to a unitary change of  $u_t$  is given by the coefficients of  $u$  in the expressions above, it is clear that the IRF of  $Y_t$  is given by the average of the individual impulse responses. Thus, if consistent estimators are employed, the limit of the IRF estimated with aggregate data and that of the average of the micro IRFs should be the same since they are estimates of the same population quantity.

The former derivation has important limitations since it relies heavily on the existence of the MA( $\infty$ ) representation of the micro-units and of the aggregate process. Thus, if some of the processes at the disaggregate level contain a unit root, it could not be applied. Furthermore, if the number of individual processes is allowed to go to infinity, the existence of the MA( $\infty$ ) representation of the aggregate process is not guaranteed, even when all the micro units are weakly stationary (see Granger, 1980). For the sake of generality, in the following we consider the AR representation instead, since it does not present these problems. The calculations, however, are not so straight forward as when the MA one is

employed.

We now analyze the relation between the aggregate and individual IRFs in the more general case where there is an infinite number of individual processes that can be stationary or not. Before computing the aggregate IRF, it is first necessary to consider the aggregation of model (1) in these circumstances. This problem has been addressed by Lewbel (1994), who followed the “stochastic” approach to aggregation introduced by Kelejian (1980) and formalized by Stoker (1984). The latter author defines an aggregate function as the expected value over the distribution of agents of the micro relations (see Stoker, 1984, Definition 2)

$$Y_t = E_I(y_t) = \int y_t P(y_t|\theta_t) dy_t, \quad (11)$$

where  $\theta(t)$  is a parameter vector that could vary over time but not across individuals and  $E_I(\cdot)$  gives the time path of the dependent variable mean.

To apply the results in Lewbel (1994), we further assume that  $B = E_I(b)$ ,  $X_t = E_I(x_t)$ , and  $E_I(\nu_t) = u_t$  exist. For simplicity, we also consider that the variables  $b$  and  $x$  are uncorrelated for all  $t$ , and that  $a$  is independent of the distribution of  $(b'x + \nu)$ . Some of these conditions can be easily relaxed without any change in the main result, as will be shown in Section 4. So,

$$Y_t = E_I(ay_{t-1}) + B'X_t + u_t, \quad (12)$$

where  $Y_t = E_I(y_t)$ . Lewbel (1994) showed that, under the above-mentioned assumptions, expression (12) can be written as

$$Y_t = \sum_{s=1}^{\infty} A_s Y_{t-s} + B'X_t + u_t, \quad (13)$$

for constants  $A_1, A_2, \dots$ , defined as  $A_s = E(\alpha_s)$ , where  $\alpha_1 = a$  and  $\alpha_s = (\alpha_{s-1} - A_{s-1})a$  for  $s > 1$ .

These constants can be easily shown to satisfy the equation

$$A_s = m_s - \sum_{r=1}^{s-1} m_{s-r} A_r, \quad (14)$$

where  $m_s = E_I(a^s)$ . Thus, model (13) can be employed to evaluate how the aggregate variable  $Y_t$  responds to changes in the aggregate shock,  $u_t$ .

### 3.1 Comparing aggregate versus disaggregate response

As in Section 2, we consider a unitary common shock occurring at time  $t$  and now we are interested in its effect on the aggregate variable  $Y_{t+h}$ , for  $h \geq 1$ . The (aggregate) IRF (denoted henceforth as  $IRF_{AG}$ ) can be easily computed as

$$IRF_{AG}(t, h) = E(Y_{t+h}|u_t = 1; Z_{t-1}) - E(Y_{t+h}|u_t = 0; Z_{t-1}), \quad (15)$$

where  $Z_{t-1} = (Y_{t-1}, Y_{t-2}, \dots, X_{t-1}, X_{t-2}, \dots)$ . Application of this definition to (13) yields

$$IRF_{AG}(t, h) = \begin{cases} IRF_{AG}(t, 0) = 1, \\ \sum_{j=1}^h A_j IRF_{AG}(h-j), \text{ if } h \geq 1. \end{cases} \quad (16)$$

Although this expression seems complex, it can, in fact, be notably simplified. Notice that, from equation (14),  $m_s = \sum_{r=0}^{s-1} m_r A_{s-r}$ . Iterating (16) from  $h = 1$ , it is straightforward to see that<sup>4</sup>

$$\begin{aligned} IRF_{AG}(t, 1) &= m_1 = E_I(a) = IRF_{dis}(t, 1) \\ IRF_{AG}(t, 2) &= m_2 = E_I(a^2) = IRF_{dis}(t, 2) \\ &\dots \\ IRF_{AG}(t, h) &= m_h = E_I(a^h) = IRF_{dis}(t, h). \end{aligned} \quad (17)$$

Thus, there is a direct link between the (average) micro and the aggregate response to a common shock: the IRF computed in the aggregate model is just the expected value of the individual IRFs and, as will be shown in the next section, this is also true under less stringent assumptions than the ones imposed above. Thus, the aggregation of heterogeneous processes like (1) does not amplify the *average* response to a given shock; if the aggregate process is highly persistent, it is not so only because the micro-units are heterogeneous but because, on average, they are highly persistent too.

An interesting consequence of the discussion above is that estimates of the aggregate IRF can also be obtained with either aggregate or disaggregate data. As will be shown in Section 5, using disaggregate data can bring about important efficiency gains.

---

<sup>4</sup>A related and independently developed result has recently been put forward by Caballero and Engel (2007). They consider a model of infrequent price adjustment at the disaggregate level and show that the response to a monetary shock is the same at the micro and at the macro level.

The result above does not imply that other properties of the micro processes are invariant to aggregation. As is well known, the aggregation of short-memory stationary processes may induce long memory or nonstationary behavior in the aggregate variable (see the Appendix for details). Similarly, if some of the individual processes are I(1) while the others are strictly stationary, the aggregate will also be an I(1) process. Nevertheless, even in this case, the above-described relationship among IRFs across different aggregation levels still applies.

To illustrate this argument, consider again model (1) under the assumptions of Sections 2 and 3, with  $a$  having support in the interval  $(0, 1)$ . Clearly, since  $a_i < 1$  for all  $i$ , all individual processes are (short-memory) stationary.<sup>5</sup> Thus, they admit a Wold representation (for simplicity, we assume that  $b_i = 0$  for all  $i$ )

$$y_{it} = \nu_{it} + a_i \nu_{it-1} + a_i^2 \nu_{it-2} + \dots, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (18)$$

The aggregate process  $Y_t = E_I(y_t)$  could be computed by taking expectations in (18), provided the expected value of the right-hand side yields a well-defined MA representation. This will happen as long as the resulting MA coefficients are square-summable, that is, if  $\sum_{j=0}^{\infty} E_I(a^j)^2 < \infty$ . As shown by Zaffaroni (2004), the holding of this condition crucially depends on the behavior of the distribution of  $a$  in a neighborhood of 1, as will be explained in more detail in Section 4. Consider, for instance, that  $a$  follows a Beta( $p, q$ ) distribution in the interval  $(0,1)$ . The moments of  $a$  are given by

$$m_j = E(a^j) = \frac{B(p + j/2, q)}{B(p, q)},$$

where  $p, q > 0$  are the parameters of the distribution and  $B(p, q)$  is the Beta function. It holds, for large  $j$ , that

$$m_j \approx (p + j/2)^{-q}. \quad (19)$$

Clearly, for values of  $q$  smaller than 0.5, the sequence  $\{m_j\}_{j=0}^{\infty}$  is not square-summable, implying that the MA representation of  $Y_t$  is not mean-square convergent and, so,  $Y_t$  is not covariance-stationary. Notice, however, that the AR expansion described in (12) exists and is well defined, since  $A(L)Y_t$  is a mean-square convergent random variable. If  $0.5 < q < 1$ ,

---

<sup>5</sup>An stationary process is short (long)-memory if its autocorrelation function is (is not) summable.

$Y_t$  is stationary and the  $k$ -th autocorrelation can be approximated by  $\rho_k \approx ck^{1-2q}$ . Then, if  $q \in (0.5, 1)$  the correlations are not summable and  $Y_t$  is a long memory process.<sup>6</sup> Notice, however, that the aggregate IRF derived from  $Y_t$  would equal the average of the micro IRFs derived from the stationary processes, since all the conditions needed for obtaining (17) are fulfilled.

### 3.2 Other measures of persistence

We now examine the performance of the remaining measures of persistence mentioned in Section 2.

The aggregate versions of the CIR and the HL are defined as

$$CIR_{AG} = \sum_{h=0}^{\infty} IRF_{AG}(t, h),$$

and

$$IRF_{AG}(t, h)|_{h=HL_{AG}} = 0.5,$$

respectively. Clearly, the equality between the  $IRF_{AG}$  and the  $IRF_{dis}$ , established in (17), implies that the CIR and the HL also present the same population values across aggregation levels.

Many empirical papers focus on the sum of autoregressive coefficients for comparing shock response across aggregation levels. In these articles, the SAC computed with aggregate data (denoted henceforth as  $SAC_{AG}$ ) is compared with the average of the individual SACs ( $SAC_{dis}$ ). This measure was originally introduced by Andrews and Chen (1994) because it has a direct relation with the CIR through the expression

$$SAC = 1 - CIR^{-1}, \tag{20}$$

and so, “*different values of the SAC can be interpreted easily in terms of persistence because they correspond straightforwardly to different values of the CIR*” (Andrews and Chen, 1994).

However, although this relation holds when a single series is employed, it is clear that relation (20) no longer holds when the SAC and the CIR are replaced by the corresponding

---

<sup>6</sup>More specifically,  $Y_t$  could be approximated by a fractionally-integrated process with order of integration  $d = (1 - q)$ . See Granger (1980) and Zaffaroni (2004) for details.

averages. In fact, although the CIR remains constant across aggregation levels, the SAC increases systematically with the level of aggregation. To see this, notice that, for the simple model described in (1),  $CIR_{dis}$  is given by  $CIR_{dis} = 1 + E_I(a) + E_I(a^2) + \dots = E_I\left(\frac{1}{1-a}\right)$  and recall that  $CIR_{dis} = CIR_{AG}$ . Since  $\frac{1}{1-a}$  is a convex function, Jensen's inequality implies that

$$SAC_{AG} = 1 - \left(E_I\left(\frac{1}{1-a}\right)\right)^{-1} > 1 - \left(\frac{1}{1-E_I(a)}\right)^{-1} = E_I(a) = SAC_{dis}, \quad (21)$$

so the aggregate SAC is bigger than the average of the individual SACs unless there is no heterogeneity, in which case both measures are equal. Notice, however, that this result does not imply that average persistence increases with the level of aggregation: it only implies that under individual heterogeneity,  $SAC_{dis}$  is a poor summary of  $CIR_{dis}$ , and therefore, is not a suitable tool for establishing persistence comparisons.

The inequality in (21) can be illustrated with a simple example. Consider, again, a collection of individuals behaving as in (1) (with  $b_i = 0$ , for all  $i$  for simplicity) with a heterogenous autoregressive parameter,  $a$ , that follows a  $U(0,1)$  distribution. Thus, the (population) value of the average of the individual SACs,  $SAC_{dis}$ , is equal to 0.5. With respect to  $CIR_{dis}$ , the moments of the uniform distribution, given by  $E(a^h) = (h+1)^{-1}$ , are not summable and so  $CIR_{dis} = \sum_{h=0}^{\infty} E(a^h) = \infty$ .

As for the corresponding aggregate values, notice that, in this case,  $Y_t$  can be obtained by taking expectations in (18) since the aggregate  $MA(\infty)$  representation is well defined.<sup>7</sup> Bearing in mind that  $A(z) = M(z)^{-1}$  for all  $z$ , where  $M(z) = \sum_{j=0}^{\infty} E(a^j) z^j$ , it follows that, in particular,  $A(1) = M(1)^{-1} = \left(\sum_{h=0}^{\infty} E(a^h)\right)^{-1} = 0$ . Thus,  $SAC_{AG} = 1 - A(1) = 1$ . Finally, it is easy to check that  $CIR_{AG} = CIR_{dis} = \infty$ . So, although the CIR remains constant across aggregation levels, the SAC increases considerably (from 0.5 to 1).

The LAR suffers from similar problems to the SAC. For instance, for the example above, it is easy to see that  $LAR_{dis} = SAR_{dis} = 0.5$  and  $LAR_{AG} = SAR_{AG} = 1$ , where  $LAR_{dis}$  and  $LAR_{AG}$  are the average of the individual LARs and the LAR associated with  $Y_t$ , respectively. Thus, the LAR should not be employed to compare persistence across aggregation levels either.

---

<sup>7</sup> $Y_t$  is stationary with variance  $\sigma_u^2 \sum_{j=0}^{\infty} (j+1)^{-2} < \infty$ .

Finally, in some applications, an alternative definition of disaggregate IRF to the one presented in this paper has been employed (see Imbs et al., 2005). As mentioned in Section 2, the strategy consists of fitting autoregressive models to each of the individual processes and to compute the average of the AR coefficient estimates. Next, an “averaged” model is constructed, having the same AR structure as the individuals but whose coefficients are given by these averaged estimates. Then, the IRF of this artificially generated process is computed as in a model with homogeneous coefficients. Thus, if the units are AR(1) processes, estimates of the function

$$\overline{IRF}(t, h) = E_I(a)^h, \text{ for } h \geq 0, \quad (22)$$

will be provided as measures of micro-persistence. But clearly, since the IRF is not a linear function, (22) does not capture the average response of the micro units. Furthermore, whenever the support of  $a$  is positive,  $a^h$  is strictly convex and the application of Jensen’s inequality yields

$$E_I(IRF(t, h)) = E_I(a^h) > \overline{IRF}(t, h) = E_I(a)^h, \text{ for all } h > 1. \quad (23)$$

It follows that, if estimates of (22) are compared with those of  $IRF_{AG}$ , differences might arise but mainly due to the underestimation of the average shock response at the micro level rather than to an overestimation of the response when aggregate data is employed.<sup>8</sup>

#### 4. EXTENSIONS

A more realistic approach than the one analyzed in Sections 2 and 3 would entail considering fewer assumptions of independence and uncorrelatedness between the variables of the model, vector autoregressive processes, as well as allowing for more general individual dynamics at the disaggregate level. In addition, it would be interesting to examine whether the relationship established in Section 3 holds when non-linear models are considered. This section extends the results of Section 3 in these directions.

---

<sup>8</sup>Gadea and Mayoral (2009) have shown that the reduction in real exchange rate persistence found by Imbs et al. (2005) using sectoral real exchange rates is due to their definition of average micro-persistence. When the disaggregate IRF defined in (5) is employed, standard estimates of persistence are recovered.

#### 4.1. Relaxing independence and uncorrelatedness assumptions

The random coefficient model considered in the previous section can be thought of as the reduced form representation of an underlying economic model. Thus, it is reasonable to think that the forces that drive the heterogeneity in the coefficients  $a$ ,  $b$  and  $\rho$  are related to some fundamentals like preferences, technology, borrowing constraints, etc. As a result, correlation among these coefficients is likely to arise. We now analyze how the existence of these correlations affects the results in the previous section.

Consider the model defined in (1) and (2) where now  $a$  and  $\rho$  are correlated. For simplicity, we assume that  $b_i = 0$  for all  $i$ . The average of the individual IRFs is given by the expected value of (4), that is

$$IRF_i(t, h) = E_I(a^h \rho), \text{ for } h \geq 0. \quad (24)$$

To compute  $IRF_{AG}$ , we first need to obtain the aggregate process. This case has not been considered in the original work by Lewbel (1994) so we derive it here. Define

$$\begin{aligned} \gamma_1 &= a; \quad \delta_1 = E_I(a\rho); & m_s^* &= E_I(a^s \rho) \\ \delta_s &= m_s^* - \sum_{l=1}^{s-1} \delta_l m_{s-l}^*; & \gamma_{s+1} &= (\gamma_s - \delta_s) a; \end{aligned}$$

and

$$\phi_s = (\gamma_s - \delta_s)\rho.$$

Notice that  $\delta_s = E_I(\gamma_s \rho)$  for all  $s$ , which implies that  $E_I(\phi_s) = E_I(\gamma_s \rho - \delta_s \rho) = 0$ , since  $E_I(\rho)$  is normalized to 1. For any positive integer  $s$ , one can obtain

$$\begin{aligned} E_I(\gamma_s y_{t-s}) &= \delta_s Y_{t-s} + E_I((\gamma_s - \delta_s) y_{t-s}) \\ &= \delta_s Y_{t-s} + E_I((\gamma_s - \delta_s) (a y_{t-s-1} + \rho u_{t-s})) \\ &= \delta_s Y_{t-s} + E_I(\gamma_{s+1} y_{t-s-1} + \phi_s u_{t-s}) \\ &= \delta_s Y_{t-s} + E_I(\gamma_{s+1} y_{t-s-1}), \end{aligned} \quad (25)$$

since  $E_I(\phi_s u_{t-s}) = E_I(\phi_s) u_{t-s} = 0$ . Starting with  $s = 1$  and applying (25) repeatedly to  $Y_t = E_I(a y_{t-1}) + u_t$ , it follows that

$$Y_t = \sum_{j=1}^{\infty} \delta_j Y_{t-j} + u_t.$$

From (15), the aggregate IRF is given by

$$IRF_{AG}(t, h) = \begin{cases} IRF_{AG}(t, 0) = 1, \\ \sum_{j=1}^h \delta_j IRF_{AG}(h - j), \text{ if } h \geq 1. \end{cases} \quad (26)$$

Finally, to compare (24) with (26), iterate expression  $m_s^* = \sum_{r=0}^{s-1} m_r^* \delta_{s-r}$  from  $s = 1$ . One obtains that  $IRF_{AG}(t, h) = m_h^* = E_I(a^h \rho) = IRF_{dis}(t, h)$ , as before.

The previous results unveil a potential source of discrepancies between disaggregate and aggregate persistence estimates. In applications it is customary to compute disaggregate persistence measures without paying attention to the potential correlation between the autoregressive coefficients,  $a$ , and the individual effect of the common shock,  $\rho$ . Thus, suppose that  $a$  and  $\rho$  are in fact autocorrelated but that this correlation is not taken into account to compute the average of the individual responses to the common shock, that is, they are calculated as  $E_I(a^h)$  for the simple DGP considered here. If the correlation between  $a$  and  $\rho$  is positive, then  $E_I(a^h)$  is strictly smaller than  $E_I(a^h \rho)$  (and viceversa). In these cases, *naïve* estimates of the average individual responses, calculated as the sample counterpart of  $E_I(a^h)$ , will tend to be systematically smaller than estimates of the IRF computed with aggregate data, since the latter would be consistent estimates of  $E_I(a^h \rho)$ .

Figure 1 illustrates this point. IRF1 to IRF3 are computed as in (24) and correspond to three different levels of correlation between  $a$  and  $\rho$ , more specifically,  $\pi_{a\rho} = \{0.3, 0.6, 0.9\}$ , respectively, where  $\pi$  denotes correlation. IRF0 corresponds to the *naïve* IRF and has been constructed as if  $a$  and  $\rho$  were independent, that is, as  $E(a^h)$ .<sup>9</sup>

For small values of  $\pi_{a\rho}$ , the difference between the *naïve* IRF and the true one is likely to be small. However, arbitrarily large differences between these two functions can arise as  $\pi_{a\rho}$  increases, as illustrated in Figure I. In addition, discrepancies may also be important if the variances of  $\rho$  or  $a$  are large.

**(Figure 1 about here)**

---

<sup>9</sup>The coefficients  $a$  have been drawn from a uniform distribution,  $U(0.67, 1)$ , that corresponds to Level of Persistence III (LP III) and  $\sigma_a = 0.1$ , see Section 5 for details. As for  $\rho$ , it is generated as  $\rho = \beta(a - \bar{a}) + 1 + e$ , where  $e$  is  $N(0, 1)$ ,  $\bar{a}$  is the mean of  $a$  and  $\beta = \pi_{a\rho} / (\sqrt{(1 - \pi_{a\rho})} \sigma_a)$ . Defined in this way,  $\rho$  has mean equal to 1 and is able to display the desired correlation with  $a$ .

;

If  $a$  and  $b$  are correlated, it is obtained that (see Lewbel, 1994)

$$Y_t = \sum_{s=1}^{\infty} A_s Y_{t-s} + B' X_t + \sum_{s=1}^{\infty} Cov(\alpha_s, b) X_{t-s} + u_t, \quad (27)$$

where  $\alpha_1 = a$  and  $\alpha_{s+1} = (\alpha_s - E(\alpha_s)) a$ , for  $s \geq 1$ . However, it is straightforward to check that the expression of the aggregate IRF in (16) is not altered by the addition of this new term and, hence, the relation between the micro and macro IRFs established in Section 3 is preserved. If  $X_t$  is simply a vector of ones, then, provided  $\sum_{s=1}^{\infty} Cov(\alpha_s, b) < \infty$ , the new term is just an addition to the aggregate intercept. Otherwise, in order to have consistent estimates of the aggregate responses, an increasing number of lags of  $X_t$  should be included in the aggregate equation.

If the distributions of  $b$  and  $x_t$  are correlated, the aggregate model becomes

$$Y_t = \sum_{s=1}^{\infty} A_s Y_{t-s} + B' X_t + Cov(b, x_t) + u_t. \quad (28)$$

If the  $Cov(b, x_t)$  is finite and constant over time, it is simply an additional term in the intercept of the aggregate model and it does not have any impact on the expression of the IRF. Thus, none of the main conclusions change.

## 4.2. Non-linear models

Analyzing non-linear models presents two obvious problems: firstly, there exists an unlimited amount of them and, secondly, it has been difficult to determine their properties, in particular, those concerning aggregation. To circumvent these issues, in this section we make use of White's Theorem (Granger 2008), which states that any arbitrary time series,  $y_t$ , such that  $E(y_t) < \infty$  and  $Prob(y_t = 0) = 0$ , can be written as a time varying parameter (TVP) AR(1) process

$$y_t = a_t y_{t-1} + \varpi_t, \quad (29)$$

where the sequences  $\{a_t\}$  and  $\{\varpi_t\}$  exist and  $a_t$  is  $F_{t-1}$ -measurable, where  $F_{t-1}$  is the  $\sigma$ -field  $\sigma(\dots, y_{t-2}, y_{t-1})$  and  $(\varpi_t, F_t)$  is a martingale difference sequence.<sup>10</sup>

---

<sup>10</sup>Some popular non-linear models can be seen as parametric TVP linear models, such as the smooth regression transition, the logistic STR (LSTR), the exponential STR (ESTR) and various forms of the threshold autoregressive (STAR) models, as discussed by Teräsvirta (2006).

Thus, we consider a group of arbitrary heterogeneous processes,  $y_{it}$ , verifying the conditions above. Using White's theorem and assuming further that they share a common factor, they can be written as a TVP linear model

$$\begin{aligned} y_{it} &= a_{it}y_{it-1} + \varpi_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \\ \varpi_{it} &= \rho_i u_t + \varepsilon_{it}. \end{aligned} \quad (30)$$

This linear representation of non-linear processes is very convenient since it allows us to employ an analogous strategy as that used in Section 3 to obtain a similar set of results. In the following, we assume that all the relevant expectations exist and that  $\rho$  and the variables in the sequence  $\{a_t\}$  are independent.

Using (3), it is easy to check that the impulse response for individual  $i$  to a shock happening at  $t$ ,  $h$  periods ahead is given by

$$IRF_i(t, h) = a_{it+h}a_{it+h-1}\dots a_{i1}\rho_i,$$

whereas the average response over the cross-sectional distribution is

$$IRF_{dis}(t, h) = E_I(a_{it+h}a_{it+h-1}\dots a_{i1}).$$

To compare this expression to the IRF of the aggregate process, we first need to obtain the latter by aggregating the individual processes in (30). We follow again Lewbel (1994) and define

$$\begin{aligned} \alpha_{1t}^* &= a_t; \quad A_{st}^* = E_I(\alpha_{st}^*), \\ \alpha_{s+1t}^* &= (\alpha_{st}^* - A_{st}^*)a_{t-s}. \end{aligned}$$

Following a similar strategy as before, it is obtained that

$$Y_t = E_I(y_t) = \sum_{k=1}^{\infty} A_{kt}^* Y_{t-k} + u_t. \quad (31)$$

and the aggregate IRF can be obtained by applying (15) to (31),

$$\begin{aligned} IRF_{AG}(t, 1) &= A_{1,t+1}, \\ IRF_{AG}(t, h) &= \sum_{j=1}^h A_{j,t+h} IRF_{AG}(t, h-j). \end{aligned} \quad (32)$$

Iterating (32) is easy to check that

$$IRF_{AG}(t, 1) = E(a_{t+1}) = IRF_{dis}(t, 1)$$

$$\begin{aligned} IRF_{AG}(t, 2) &= A_{1,t+2}E(a_{t+1}) + A_{2,t+2} \\ &= E(a_{t+2})E(a_{t+1}) + E(a_{t+2}a_{t+1}) - E(a_{t+2})E(a_{t+1}) \\ &= E(a_{t+2}a_{t+1}) = IRF_{dis}(t, 2), \end{aligned}$$

and, in general,

$$IRF_{AG}(t, h) = IRF_{dis}(t, h), \text{ for all } h \geq 0$$

as before.

It follows that the relationship between IRFs across aggregation levels not only holds for linear processes but also for any arbitrary group of heterogeneous processes  $y_{it}$ . This result might be very useful in applications where heterogeneity and non-linearities are suspected at the individual level but the shape of the aggregate process is difficult to get. In these situations, averages of the IRF estimates obtained from the individual non-linear specifications can be used to calibrate the aggregate model.

### 4.3. Higher order AR processes

A further extension would be to allow higher order lags at the disaggregate level.<sup>11</sup> To simplify the exposition, we next analyze the AR(2) case. The AR(p) one, though notationally cumbersome, can be analyzed in the same manner. Suppose the individual agents  $i$  behave according to the model,

$$y_{it} = a_{1i}y_{it-1} + a_{2i}y_{it-2} + \nu_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (33)$$

where  $\nu_{it}$  is defined as in (2). In addition to the assumptions stated in Section 3, we now require the existence of all the moments of  $a_2$  and all the cross-moments between  $a_1$  and

---

<sup>11</sup>The model analyzed here is just a particular case of that analyzed in 4.2., and thus, the results obtained there apply here as well. For the sake of completeness, however, we present the calculations pertaining to this case as well.

$a_2$ . Notice that we do not need to assume independence between  $a_1$  and  $a_2$ . The aggregate model is obtained by taking expectations in (33),

$$Y_t = E_I(a_1 y_{t-1}) + E_I(a_2 y_{t-2}) + u_t, \quad t = 1, \dots, T.$$

which can be written as an AR( $\infty$ ) process

$$Y_t = \sum_{j=0}^{\infty} C_j Y_{t-j} + u_t. \quad (34)$$

In order to obtain the coefficients  $C_j$ , define  $\alpha_{j1} = a_j$ ,  $A_{jk} = E(\alpha_{jk})$  and  $\alpha_{jk+1} = (\alpha_{jk} - E(\alpha_{jk})) a_j$ , for  $j = 1, 2$  and  $k = 1, 2, \dots$ , and notice that

$$\begin{aligned} E_I(a_1 y_{t-1}) &= E_I(a_1 + E_I(a_1) - E_I(a_1) y_{t-1}) = \\ &E_I(a_1) Y_{t-1} + E_I(\alpha_{12} y_{t-2}) + E_I((a_1 - E_I(a_1)) a_2 y_{t-3}), \end{aligned}$$

and

$$\begin{aligned} E_I(a_2 y_{t-2}) &= E_I(a_2 + E_I(a_2) - E_I(a_2) y_{t-2}) = \\ &E_I(a_2) Y_{t-2} + E_I((a_2 - E_I(a_2)) a_1 y_{t-3}) + E_I(\alpha_{22} y_{t-4}), \text{ etc.} \end{aligned}$$

Iterating the procedure above, after some algebra, one can obtain that<sup>12</sup>

$$\begin{aligned} C_1 &= E_I(\alpha_{11}) = E_I(a_1) \\ C_2 &= E_I(\alpha_{12}) + E_I(\alpha_{21}) = Var(a_1) + E_I(a_2) \\ C_3 &= E_I(\alpha_{13}) + 2cov(a_1, a_2) = Cov(a_1^2, a_1) + 2cov(a_1, a_2) \\ C_4 &= E_I(\alpha_{14}) + E_I(\alpha_{22}) + 3cov(a_1 a_2, a_1), \text{ etc.} \end{aligned} \quad (35)$$

Next, we check that the disaggregate and aggregate IRFs are the same. The IRF for each of the micro-units can be easily computed from definition (3) and, taking expectations, it is obtained that

$$\begin{aligned} IRF_{dis}(t, 1) &= E_I(a_1); \quad IRF_{dis}(t, 2) = E_I(a_1^2) + E_I(a_2); \\ IRF_{dis}(t, 3) &= E_I(a_1^3) + 2E_I(a_1 a_2); \quad IRF_{dis}(t, 4) = E_I(a_1^4) + 3E_I(a_1^2 a_2) + E_I(a_2^2), \end{aligned}$$

---

<sup>12</sup>Notice that these coefficients are different from the ones provided by Lewbel (1994) (Equations 6' and 16) since those expressions contain some typos.

and, in general,

$$IRF_{dis}(t, h) = E_I(IRF(t, h)) = E_I(a_1 IRF(t, h-1) + a_2 IRF(t, h-2)). \quad (36)$$

It is not difficult, although algebraically tedious, to check that the aggregate IRF associated with the aggregate process (33) coincides with the one obtained from averaging the individual IRFs. For instance, consider the first values of the aggregate IRF that are given by

$$\begin{aligned} IRF_{AG}(t, 1) &= C_1 = E_I(a_1) = IRF_{dis}(t, 1) \\ IRF_{AG}(t, 2) &= C_1^2 + C_2 = E_I(a_1^2) + E_I(a_2) = IRF_{dis}(t, 2) \\ IRF_{AG}(t, 3) &= C_1^3 + 2C_1C_2 + C_3 = E_I(a_1^3) + 2E_I(a_1a_2) = IRF_{dis}(t, 3) \\ IRF_{AG}(t, 4) &= C_1^4 + 3C_1^2C_2 + 2C_3C_1 + C_2^2 + C_4 \\ &= E_I(a_1^4) + 3E_I(a_1^2a_2) + E_I(a_2^2) = IRF_{dis}(t, 4), \text{ etc.} \end{aligned} \quad (37)$$

Hence, the aggregate IRF is just the expected value of the individual IRFs, as shown in Section 3 for the simple AR(1) case. The same result also holds when more general AR( $p$ ) dynamics are considered.

#### 4.4. Vector autoregressive processes

Finally, we consider vector autoregressive (VAR) systems at the disaggregate level. To do so, equation (1) is reinterpreted so that  $y_{it}$  and  $\nu_{it} = \rho_i u_t + \varepsilon_{it}$  are now  $J$  vectors of random variables,  $a_i$  is a  $J \times J$  matrix of random coefficients,  $\rho_i$  is a diagonal  $J \times J$  matrix verifying that  $E_I(\rho_i) = I_J$ , where  $I_J$  is the identity matrix of order  $J$ , and for simplicity,  $b_i$  is set to 0 for all  $i$ .

The individual IRF associated with this model can be obtained by

$$IRF_i(t, h) = E(y_{it+h}|u_t = \theta; z_{it-1}) - E(y_{it+h}|u_t = \bar{0}; z_{it-1}) \quad (38)$$

where  $z_{it-1} = \left( y_{it-1}, y_{it-2}, \dots \right)'$ ,  $\bar{0}$  is a  $J$  vector of zeroes and  $\theta$  are the relevant experimental shocks. Under the assumptions of Section 3, it is easy to check that, in the vector case, the individual IRF is given by

$$IRF_i(t, h) = a_i^h \rho_i \theta, \text{ for } h \geq 0. \quad (39)$$

and, so, the average response can be defined as

$$IRF_{dis}(t, h) = E_I(IRF(t, h)) = E_I\left(a^h\right)\theta, \text{ for } h \geq 0. \quad (40)$$

The aggregate VAR can be obtained using a similar strategy as in the scalar case. Taking expectations in (1) and adding and subtracting the matrix  $A_1 = E(a_1)$ , we get

$$Y_t = E_I(y_t) = A_1 Y_{t-1} + E_I((A_1 - E(A_1))(ay_{t-2} + \nu_{t-1})) + u_t.$$

Defining the  $J \times J$  matrix  $A_s = E(\alpha_s)$ , where  $\alpha_s = (\alpha_{s-1} - A_{s-1})a$  and  $\alpha_1 = a$ , one can iterate the procedure above to obtain the aggregate model

$$Y_t = \sum_{s=1}^{\infty} A_s Y_{t-s} + u_t.$$

The aggregate IRF can be obtained as,

$$IRF_{AG}(t, h) = E(Y_{t+h}|u_t = \theta; Z_{t-1}) - E(Y_{t+h}|u_t = \bar{0}; Z_{t-1}),$$

where  $Z_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$ , which yields the expression  $IRF_{AG}(t, 0) = \theta$  and

$$IRF_{AG}(t, h) = \sum_{j=1}^h A_j IRF_{AG}(h-j), \text{ for } h \geq 1. \quad (41)$$

As in the scalar case, it is easy to show that the relations  $A_s = m_s - \sum_{r=1}^{s-1} m_{s-r} A_r$  and  $m_s = \sum_{r=0}^{s-1} m_r A_{s-r}$  also hold in the matrix case, where  $m_s$ , defined as  $m_s = E_I(a^s)$ , is of order  $J \times J$ . Thus, it is straightforward to check that

$$IRF_{AG}(t, h) = E_I\left(a^h\right)\theta. \quad (42)$$

So, comparing (42) and (40), one can conclude that in the vector case, the average response to an economic shock is the same, regardless of the level of aggregation at which it is considered.

## 5. FINITE SAMPLE PROPERTIES OF IRF ESTIMATORS UNDER INDIVIDUAL HETEROGENEITY

This section explores whether it is possible to obtain good approximations to the population relationships established in the previous sections when the corresponding sample

counterparts are employed. In order to do that, we have carried out two Monte Carlo experiments that estimate the impulse response to an aggregate shock using both aggregate and disaggregate data and standard estimation techniques. We focus on linear models exclusively.

The first simulation explores the finite-sample properties of estimated IRFs as the degree of heterogeneity of the underlying processes increases for a given level of persistence, while the second focuses on assessing efficiency when estimation is carried out at different levels of aggregation.

In order to isolate the impact of an increase in heterogeneity from an increase in average persistence, different degrees of heterogeneity have been considered while keeping the level of persistence constant. To achieve this, the following approach was adopted. The data has been generated according to the model:

$$y_{it} = a_i y_{it-1} + \rho_i u_t + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (43)$$

where  $T = \{100, 400\}$ ,  $N = \{100, 200\}$ ,  $u_t$  and  $\varepsilon_{it}$  are  $i.N(0, 1)$  and  $i.N(0, \sigma_i^\varepsilon)$  random variables, respectively and  $\sigma_i^\varepsilon$  and  $\rho_i$  are draws from two independent uniform distributions in the interval  $(0.5, 1.5)$ . The autoregressive parameter  $a$  is distributed as a  $U(\omega_1, \omega_2)$ , for different values of  $(\omega_1, \omega_2)$ . These values determine both the degree of heterogeneity of the individual units and the level of persistence of the aggregate. In this exercise, the degree of heterogeneity will be determined by the standard deviation of  $a$ ,  $\sigma_a$ , while the level of persistence will be measured by the cumulated impulse response up to lag 100,  $CIR(100)$ .

To select  $(\omega_1, \omega_2)$ , the following approach has been pursued. Firstly, we have set three levels of persistence, denoted as LP I, II and III, corresponding to the case where there is no heterogeneity and  $a$  takes the values  $a = \{0.85, 0.90, 0.95\}$ , respectively. The persistence implied by a model like (43) in these cases, as measured by the  $CIR(100)$ , yields

$$\begin{aligned} LP \text{ I:} \quad CIR(100) &= \sum_{j=0}^{100} 0.85^j = 6.67. \\ LP \text{ II:} \quad CIR(100) &= \sum_{j=0}^{100} 0.9^j = 10. \\ LP \text{ III:} \quad CIR(100) &= \sum_{j=0}^{100} 0.95^j = 19.89. \end{aligned}$$

Secondly, four values of  $\sigma_a$  have been considered, namely  $\sigma_a = \{0, 0.01, 0.05, 0.1\}$ . Next, the values  $(\omega_1, \omega_2)$  are chosen in such a way that the desired  $\sigma_a$  is obtained and the resulting set of heterogeneous processes imply a similar value of the CIR(100) as LP I to LP III. Under heterogeneity,  $\text{CIR}(100) = \sum_{j=0}^{100} E(a^j) = \left(\frac{1}{\omega_2 - \omega_1}\right) \sum_{j=0}^{100} (\omega_2^{j+1} - \omega_1^{j+1}) / j + 1$ . Then, the restriction on the standard deviation implies that  $(\omega_2 - \omega_1) = \sqrt{12}\sigma_a$  while the condition on the CIR allows us to identify unique values for  $(\omega_1, \omega_2)$ . Table I reports these values for each of the levels of persistence and heterogeneity considered in this experiment.

TABLE I. DISTRIBUTION OF  $a$

		$a \sim U(\omega_1, \omega_2)$							
		$\sigma_a = 0.0$		$\sigma_a = 0.01$		$\sigma_a = 0.05$		$\sigma_a = 0.1$	
	CIR(100)	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$
LP I	6.67	0.85	0.85	0.83	0.866	0.7470	0.9202	0.6155	0.9619
LP II	10	0.90	0.90	0.882	0.9166	0.7900	0.9632	0.6455	0.9919
LP III	19.89	0.95	0.95	0.9300	0.9646	0.826	0.992	0.6700	1.00

The aggregate process,  $Y_t$ , has been computed as the simple average of the  $y'_{it}$ s

$$Y_t = \frac{\sum_{i=1}^N y_{it}}{N}. \quad (44)$$

As regards the estimation of  $Y_t$ , difficulties arise because it contains an infinite number of parameters. Berk (1974) and Lewis and Reinsel (1985) have shown that  $\sqrt{T}$ -consistent and asymptotically normally distributed estimates can be obtained by approximating the  $\text{AR}(\infty)$  process by an  $\text{AR}(k)$  model, where  $k$  does not increase too quickly or too slowly.<sup>13</sup>

The lag length,  $k$ , is the key parameter in implementing procedures that approximate  $\text{AR}(\infty)$  models in applications. Ng and Perron (1995) have argued that standard selection criteria (the AIC and the BIC) choose values of  $k$ ,  $\hat{k}$ , that are proportional to  $\log T$  and, so, do not verify the lower bound condition stated above.<sup>14</sup> Kuersteiner (2005) has shown

<sup>13</sup>More specifically,  $k$  should verify an upper bound condition,  $k^3/T \rightarrow 0$ , and a lower bound one,  $T^{1/2} \sum_{j=k+1}^{\infty} |A_j| \rightarrow 0$  as  $k, T \rightarrow \infty$ . These results are valid when  $Y_t$  is a short memory process. Related results for the case where  $Y_t$  is a long memory process have been recently presented by Godet (2008).

<sup>14</sup>In fact, bias terms arising as a consequence of the asymptotic misspecification of the model when these

that the general-to-specific (GTS) approach (Ng and Perron, 1995) can be used to produce a data-dependent selection rule such that the parameters obtained in the AR( $k$ ) model are consistent and asymptotically normal for the parameters of the underlying AR( $\infty$ ) model. Then, the consistency and the asymptotic normality of aggregate IRF estimates in AR( $\infty$ ) models follow from an application of the delta method.

AR( $k$ ) processes have been fitted to  $Y_t$ , where  $k$  was chosen according to the AIC and the GTS. Following Ng and Perron (1995), the maximum value of  $k$ ,  $k_{max}$ , was set according to the rule  $k_{max}(Y_T) \approx 10 * (T/100)^{0.25}$ , which yields  $k_{max} = \{10, 14\}$  for  $T = \{100, 400\}$ , respectively. A similar strategy has been followed to fit models for the individual  $y'_{it}$ s. This time,  $k_{max}$  was set to  $k_{max}(Y_T)/3$ .

For each replication  $r$ , the estimated disaggregate and aggregate models have been employed to compute the corresponding IRFs, denoted as  $\widehat{IRF}_{dis,r}$  and  $\widehat{IRF}_{AG,r}$ , respectively.  $\widehat{IRF}_{dis,r}$  has been obtained as the mean of the individual IRFs while  $\widehat{IRF}_{AG,r}$  is the sample analog of (16).

Table II reports the average mean squared error (MSE) of  $\widehat{IRF}_{dis}$  and  $\widehat{IRF}_{AG}$ , which is defined as

$$\overline{MSE}_{dis} = R^{-1} \sum_{r=1}^R \left( h^{-1} \sum_{j=0}^h \left( \widehat{IRF}_{dis,r}(j) - IRF_{true,r}(j) \right)^2 \right), \quad (45)$$

and

$$\overline{MSE}_{AG} = R^{-1} \sum_{r=1}^R \left( h^{-1} \sum_{j=0}^h \left( \widehat{IRF}_{AG,r}(j) - IRF_{true,r}(j) \right)^2 \right), \quad (46)$$

where  $R = 1000$  is the number of replications,  $h = 20$  is the horizon of the IRF and  $IRF_{true,r}$  is the true IRF in replication  $r$ , given by

$$IRF_{true,r}(j) = N^{-1} \sum_{i=1}^N a_{i,r}^j, \quad j = 1, \dots, h.$$

Only values obtained by selecting  $k$  using the GTS approach and  $N = 100$  are reported. Those obtained by using the AIC and  $N = 200$  were similar so they are omitted for the sake of brevity.

---

criteria are employed are of order  $T^{-1/2}$ , considerably more severe than the usual finite sample biases that are typically of order  $T^{-1}$ .

TABLE II. MEAN SQUARED ERROR, GTS SELECTION METHOD

	$\sigma_a = 0.0$		$\sigma_a = 0.01$		$\sigma_a = 0.05$		$\sigma_a = 0.1$	
	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$	$\overline{MSE}_{dis}$	$\overline{MSE}_{AG}$
T=100								
LP I	0.063	0.234	0.064	0.259	0.061	0.239	0.056	0.210
LP II	0.168	0.461	0.166	0.470	0.153	0.456	0.101	0.328
LP III	0.569	1.091	0.570	1.138	0.303	0.748	0.138	0.432
T=400								
LP I	0.013	0.061	0.014	0.061	0.0121	0.057	0.010	0.055
LP II	0.028	0.102	0.028	0.106	0.023	0.095	0.016	0.077
LP III	0.065	0.183	0.067	0.192	0.040	0.141	0.018	0.090

Data has been generated according to (43) and (44). The number of processes,  $N$ , is equal to 100.

$\overline{MSE}_{dis}$  and  $\overline{MSE}_{AG}$  are defined in (45) and (46), respectively.

Figures 2 to 4 present the average over the number of replications of  $IRF_{true}$ ,  $\widehat{IRF}_{dis}$  and  $\widehat{IRF}_{AG}$  for LP I-III, respectively, for  $T = 100$ . Confidence bands at the 5% significance level are also reported. Four graphs are presented in each figure, corresponding to the different degrees of heterogeneity considered. Graphs in the upper left corner correspond to the case where there is no heterogeneity ( $\sigma_a = 0$ ), while the remaining graphs display estimated IRFs under positive  $\sigma'_a$ 's.

Several interesting conclusions can be drawn from inspecting Figures 2 to 4 and Table II. Firstly, the figures show that the average bias is in general small and fairly similar for the  $\sigma_a = 0$  and the  $\sigma_a > 0$  cases and for  $\widehat{IRF}_{dis}$  and  $\widehat{IRF}_{AG}$ . Furthermore, confidence bands do not seem to vary much as  $\sigma_a$  increases. This suggests that the existence of individual heterogeneity does not hurt much the properties of the estimators. Confidence bands are considerably narrower for  $\widehat{IRF}_{dis}$  than for  $\widehat{IRF}_{AG}$ , which implies that the use of disaggregate information can bring about important efficiency gains in the estimation of aggregate quantities. Similar conclusions are drawn by analyzing the average MSE:  $\overline{MSE}_{dis}$  is always smaller than  $\overline{MSE}_{AG}$ , even in the absence of heterogeneity.

Secondly, increasing heterogeneity for a given level of persistence does not affect much

$\overline{MSE}_{dis}$  unless persistence is high (LP III), in which case  $\overline{MSE}_{dis}$  tends to decrease as heterogeneity increases. A similar pattern is observed for  $\overline{MSE}_{AG}$ : for moderate levels of persistence (LP I), increasing  $\sigma_a$  deteriorates  $\overline{MSE}_{AG}$ . However the opposite effect is found in more persistent situations. For instance, for LP III with  $T = 100$ , the  $\overline{MSE}_{AG}$  for  $\sigma_a = 0.1$  is less than half the  $\overline{MSE}_{AG}$  obtained for the case of no heterogeneity ( $\sigma_a = 0$ ). This evidence is surprising and at odds with the common belief that individual heterogeneity hurts the properties of aggregate estimators.

Thirdly, increasing the level of persistence always increases the MSE. This increase may be due to the well-known downward bias affecting the OLS estimator of AR coefficients when persistence is high. Interestingly, the increase tends to be smaller in the presence of heterogeneity. For instance, for  $\sigma_a = 0$  and  $T = 100$ , the MSE of  $\widehat{IRF}_{dis}$  ( $\widehat{IRF}_{AG}$ ) increases by a factor of 9 (4.5) when moving from LP I to LP III. However, for  $\sigma_a = 0.1$ , the MSE of  $\widehat{IRF}_{dis}$  ( $\widehat{IRF}_{AG}$ ) rises only by a factor of 2.5 (2) from LP I to LP III. Finally, both the bias and the MSE drop considerably when larger sample sizes are considered.

**(Figure 2 about here)**

**(Figure 3 about here)**

**(Figure 4 about here)**

An interesting conclusion of the previous exercise is that important efficiency gains are obtained when disaggregate data is employed to estimate the aggregate IRF. However in applications, highly disaggregate data is sometimes not available. We have carried out a second experiment to assess how efficiency improves as more disaggregate data is used. To do that, we have considered a similar DGP as before where  $N$ , the number of individual processes, is set equal to 200. Next, this data has been aggregated in several steps: Level of aggregation 5 (LAgg 5) contains 100 processes, each computed as the simple average of two of the original processes. LAgg 4 to 1 have been computed in an analogous way, such that they contain 50, 10, 5 and 1 process, respectively. Thus, LAgg 1 is simply the aggregate process defined in (44). Table III presents the MSE obtained by estimating these data following a similar strategy as in the previous exercise. We focus on LP III and consider the same levels of heterogeneity as before, namely,  $\sigma_a = \{0, 0.01, 0.05, 0.1\}$ .

Not surprisingly, the MSE decreases considerably as more disaggregate data is employed. Comparing the MSE of LAgg 1 with that of LAgg 5 for  $T = 100$ , it is seen that the latter is around 40 to 60% smaller. Interestingly, even when highly aggregate data is considered, as in LAgg 2, where each of the series is the result of aggregating 40 AR(1) heterogeneous processes, efficiency gains of around 20% are obtained with respect to LAgg 1. This fact stresses the convenience of using disaggregate data whenever is available.

TABLE III. MSE AT DIFFERENT AGGREGATION LEVELS

	$\sigma_a = 0.0$	$\sigma_a = 0.01$	$\sigma_a = 0.05$	$\sigma_a = 0.1$
T=100				
LAgg 1 (1 process)	1.144	1.071	0.774	0.436
LAgg 2 (5 processes)	0.946	0.948	0.631	0.353
LAgg 3 (10 processes)	0.919	0.883	0.605	0.333
LAgg 4 (50 processes)	0.769	0.738	0.462	0.226
LAgg 5 (100 processes)	0.666	0.639	0.374	0.168
T=400				
LAgg 1 (1 process)	0.213	0.173	0.146	0.094
LAgg 2 (5 processes)	0.188	0.159	0.133	0.091
LAgg 3 (10 processes)	0.180	0.152	0.122	0.079
LAgg 4 (50 processes)	0.133	0.111	0.072	0.037
LAgg 5 (100 processes)	0.101	0.084	0.047	0.022

Data has been generated according to (43) and LP III. N is equal to 200.

LAgg 1 to 5 aggregate the original data in 1, 5, 10, 50 and 100 series, respectively.

## 6. EMPIRICAL ILLUSTRATION

Monetary authorities and central banks are very interested in knowing how sluggishly inflation returns to its long-run equilibrium level after the arrival of a shock. Recently, it has been argued that the persistence of shocks to inflation tends to increase with the level of aggregation. For instance, one of the conclusions of the Inflation Persistence Network, created by the European Central Bank with the aim of analyzing the patterns of European

inflation persistence, was that there is clear evidence of large differences across sectors and *“that measures of the degree of inflation persistence increase with the level of aggregation. Individual or highly disaggregate price series are, on average, much less persistent than aggregate ones”* (see Angeloni et al., 2007, and the references therein).<sup>15</sup> Similar findings have been reported for U.S. inflation (Clark, 2006).

Most of the empirical studies that have compared the persistence of inflation shocks across different aggregation levels have relied on averages of the SAC (or other scalar measures such as the LAR) as measures of disaggregate persistence (see Clark, 2006, Altissimo et al., 2006a, and the references therein). However, this approach can be problematic for the reasons discussed in Section 3.

This section compares estimates of U.S. inflation persistence computed at different levels of aggregation. In addition to the SAC, cross-sectional averages of the IRFs are considered as measures of shock response.

We use a similar data set to Clark (2006). Price indexes and nominal expenditures for all components of consumption, as measured in the NIPA accounts, have been obtained from the webpage of the Bureau of Economic Analysis (BEA). This dataset permits breakdowns at various levels of aggregation. We focus on core inflation, which excludes food and energy prices. Then, the aggregate variable, denoted as Level 1, is core inflation. We also report results for data broken into several levels of disaggregation, each spanning all the core inflation. The most disaggregate level (that we will refer to as Level 4) contains 109 disaggregate prices. Level 3 and Level 2 aggregate these 109 series into 46 and 11 categories, respectively. See Clark (2006) for details on the construction of these variables. The data is quarterly and covers the period 1976 to 2002.<sup>16</sup>

To illustrate the existence of heterogeneity in inflation data, the mean, the standard deviation and the sum of the autoregressive coefficients (with and without small-sample bias correction) have been computed for each of the 109 series in Level 4. Table IV summarizes

---

<sup>15</sup>These conclusions are based on a number of studies that can be found in the network’s webpage at <http://www.ecb.int/events/conferences/html/inflationpersistence.en.html>.

<sup>16</sup>Clark (2006) analyzed the period 1959-2002. Nevertheless, in order to avoid problems derived from the existence of structural breaks around the 1973 crisis, which would have a great impact on persistence estimates (see Perron, 1989), we have preferred to avoid this period by considering only post-crisis data.

the results by presenting some descriptive statistics of the quantities obtained. More specifically, it reports the average, the weighted average, the three quantiles and the minimum and maximum values of the individual statistics detailed above. The weights employed to construct the weighted averages represent the listed components shares of core PCE nominal spending in 2001, in percentage terms. To compute the SAC, AR( $k$ ) processes have been fitted, where  $k$  has been chosen according to the GTS. Bias-corrected estimates (denoted as SAC<sub>B-C</sub> in the table) have been calculated following Kilian's (1998) method.

TABLE IV. DESCRIPTIVE STATISTICS (LEVEL 4)

	Mean inflation	Std inflation	SAC	SAC <sub>B-C</sub>
Average	3.70	5.53	0.660	0.771
W. average	3.95	5.69	0.741	0.832
S. deviation	1.50	4.09	0.218	0.185
First quantile	2.88	3.45	0.538	0.678
Second quantile	4.04	4.36	0.729	0.804
Third quantile	4.75	5.89	0.810	0.911
Min	-1.91	2.25	-0.240	-0.141
Max	7.64	29.4	0.955	1.00

Inspection of Table IV reveals that disaggregate inflation data is highly heterogeneous. The mean inflation values over the period considered range from -1.91 to 7.46 with a mean (median) of 3.70 (4.04) and a standard deviation of 1.5. Individual dispersion also shows important disparities across series, as shown by the minimum and maximum values (2.25 and 29, respectively). More importantly, the dynamics of inflation series, as measured by the SAC, also seem to be very heterogeneous. Values of the SAC (bias-corrected SAC) range from -0.24 (-0.14) to 0.95 (1.00), with a mean of 0.66 (0.77) and a standard deviation of 0.21 (0.18).

Table V reports several measures of persistence computed at different levels of aggregation (Levels 1 to 4). To compute impulse response functions, AR( $k$ ) processes have been estimated where the order  $k$  has been chosen according to the GTS and to the AIC.<sup>17</sup>

<sup>17</sup>The maximum number of lags was set to 20, 16, 12 and 8, for aggregation levels 1 to 4, respectively.

The first four rows of Table V present the sum of the first  $h$  values of the IRFs relative to aggregation levels 1 to 4, for  $h = \{4, 8, 12, 16 \text{ and } 20\}$ , that is, the cumulated response of inflation from 1 to 5 years after the shock occurs. The bottom row of Table V reports the average value of the SAC for the four aggregation levels. Confidence intervals have been computed using bootstrap methods. Figure V, in turn, depicts the estimated IRFs associated to aggregation levels 1 to 4.<sup>18</sup>

TABLE V. PERSISTENCE MEASURES

IRFS AT DIFFERENT AGGREGATION LEVELS								
Level 1		Level 2		Level 3		Level 4		
$\sum_{i=1}^h IRF_{Level\ 1}(h)$		$\sum_{i=1}^h IRF_{Level\ 2}(h)$		$\sum_{i=1}^h IRF_{Level\ 3}(h)$		$\sum_{i=1}^h IRF_{Level\ 4}(h)$		
	GTS	AIC	GTS	AIC	GTS	AIC	GTS	AIC
$h = 4$	1.77	2.08	2.05	2.02	1.89	1.68	1.85	1.69
$h = 8$	3.06	3.70	3.28	3.18	3.07	2.51	2.97	2.50
$h = 12$	4.01	4.77	4.08	3.98	3.97	3.11	3.81	3.09
$h = 16$	4.58	5.56	4.67	4.57	4.70	4.48	4.46	3.50
$h = 20$	4.86	6.13	5.20	4.60	5.19	4.99	5.01	3.85
SAC AT DIFFERENT AGGREGATION LEVELS								
	0.89	0.87	0.80	0.82	0.69	0.69	0.67	0.66
	(0.85,0.95)	(0.84,0.94)	(0.75,0.83)	(0.76,0.84)	(0.65,0.71)	(0.65,0.71)	(0.65,0.69)	(0.64,0.68)

In agreement with the theoretical results, Table V shows that impulse responses computed from aggregate and sectoral data are very close at all the considered horizons when the GTS selection method is employed. As can be seen from Figure V, the four IRFs present approximately the same values and the same pattern of decay, so they imply a very similar degree of shock persistence, in line with the results presented in Table V. It is also remarkable that the GTS performs better than the AIC in obtaining more homogeneous values across aggregation levels, stressing the importance of fitting long autoregressions when individual

The significance level for applying the general-to-specific criterion was 10%. Small sample-bias corrected estimates have also been computed but they are not reported since results are qualitatively identical.

<sup>18</sup>The GTS method has been employed in the calculations. Confidence bands at the 5% significance level have been obtained using bootstrap and correspond to  $IRF_{Level\ 4}$ .

heterogeneity is present.

The values of the SAC, however, vary considerably across aggregation levels, reproducing what has been found in previous studies. Values of the SAC range from the 0.66 corresponding to aggregation level 4 (with a confidence interval of (0.64, 0.68)) to the 0.86 for level 1 (with a C.I. of (0.82, 0.92)). Therefore, from only looking at these figures, one would conclude that the response to a shock is higher, the higher the level of aggregation at which it is measured. However, a more detailed analysis of the evolution of the shock as described by the IRF suggest the opposite conclusions.

**(Figure 5 about here)**

## 7. CONCLUSION

This paper examines the relations among shock persistence measures computed at different levels of aggregation in a context where individual dynamic heterogeneity may exist. It is shown that the average of the individual IRFs equals the aggregate IRF at all horizons in a variety of situations, that include both linear and non-linear models. A similar relationship also holds for some scalar measures of persistence such as the CIR. However, other tools as the SAR or the LAR have poor properties in this framework. In particular, they tend to yield higher values the higher the level of aggregation considered. Since these measures are the most employed in applications, it is not surprising that different average shock behavior has been reported in many empirical papers. An empirical application using U.S. inflation data that illustrates the theoretical results has also been provided.

Our results have some important implications on macroeconomic modelling. Some authors have warned against the practice of using averages of microparameters to calibrate macro models (Altissimo et al., 2009). We qualify this statement by showing that the adequacy of this procedure depends on the type of microparameter employed. Thus, averages of individual IRFs can be used in macro calibration independently of whether there is heterogeneity or not at the individual level or whether models are linear or not.

This paper opens several avenues for future research. Our results imply that both macro (aggregate) or micro (disaggregate) data can be used to estimate the impulse response func-

tion associated to an aggregate shock. Furthermore, they suggest that important efficiency gains can be obtained by using micro information. Thus, a thorough examination of the large and finite sample properties of micro and macro estimators is needed.

The relation between the micro and the aggregate IRFs can also be exploited in other ways. If disaggregate data is available, it is possible to completely determine the dynamics of the aggregate process. If the latter is stationary, the disaggregate IRF defined in (5) equals the coefficients of the Wold representation of the aggregate process. Then, the macro dynamics can be identified very simply from the (average) of the micro impulse responses (and vice versa). If the aggregate process is not stationary, its Wold representation does not exist. However, the coefficients of the polynomial  $A(z) = M(z)^{-1}$ , where  $M(z) = \sum_{j=0}^{\infty} m_j z^j$  are  $m_j$  are the coefficients of the disaggregate IRF, are those corresponding to the AR( $\infty$ ) representation of the aggregate process. These relationships could potentially be used to improve the accuracy of the estimation combining micro and macro data, in the spirit of Imbens and Lancaster (1994).

It would also be possible to recover micro information when only aggregate data is available. Since the aggregate IRF identifies all the moments of the distribution of the autoregressive parameter,  $a$ , it is possible to infer some interesting information about the distribution of this parameter only from the aggregate data. For instance, probability bounds for  $a$  can be easily computed from the aggregate data by applying Markov and Chebichev inequalities. Furthermore, under certain conditions, the whole distribution of the autoregressive parameter  $a$  can be recovered using only aggregate data. This is the well known “moment problem”, that consists of inverting the mapping that takes a probability measure to the sequences of moments. This problem already has a long tradition, begun by Stieljes in 1894 and developed by subsequent authors (see, for instance, Lin, 1997 for a description of the conditions needed to recover the density from the moments).

## REFERENCES

- Altissimo, F., B. Mojon and P. Zaffaroni (2009), “Can aggregation explain the persistence of inflation?,” *Journal of Monetary Economics* 56, 231-241.
- Altissimo, F., L. Bilke, A. Levin, T. and B. Mojon (2006a), “Sectoral and Aggregate Inflation Dynamics in the Euro,” *Journal of the European Economic Association*, Vol. 4, pp. 585-593.
- Altissimo, F., M. Ehrmann and F. Smets (2006b), “Inflation persistence and price setting behaviour in the euro area”, Occasional paper series, N° 46, ECB.
- Andrews, D. and H. Chen (1994), “Approximately median-unbiased estimation of autoregressive models,” *Journal of Business and Economic Statistics*, 12, pp. 187-204.
- Angeloni, I., L. Aucremanne, M. Ehrmann, J. Galí, A. Levin and Frank Smets, (2005), “New Evidence on Inflation Persistence and Price Stickiness in the Euro Area: Implications for Macro Modelling,” Economics Working Papers 910, Department of Economics and Business, Universitat Pompeu Fabra.
- Berk, K.N. (1974), “Consistent autoregressive spectral estimates,” *The Annals of Statistics* 2, pp. 489-502.
- Bilke, L. (2005), “Break in the mean and persistence of inflation: a sectoral analysis of French CPI,” ECB Working Paper 463.
- Caballero, R.J. and E. M.R.A. Engel (2007), “Price Stickiness in Ss models: New interpretations and new results,” *Journal of Monetary Economics* 54, pp. 100–121.
- Clark, T. (2006), “Disaggregate evidence on the persistence of consumer price inflation”, *Journal of Applied Econometrics*, Vol. 21, 5, pp. 563-587.
- Crucini, M. J. and M. Shintani (2008), “Persistence in Law-of-One-Price Deviations: Evidence from Micro-data,” *Journal of Monetary Economics*, 55, pp. 629-644.
- Forni M. and M. Lippi (1997), *Aggregation and the microfoundations of dynamic macroeconomics*, Oxford University Press Inc., New York.
- Gadea, L. and L. Mayoral (2009), “Aggregation is not the solution: the PPP puzzle strikes back,” *Journal of Applied Econometrics*, 24, 875-894.
- Godet, F. (2007), “Linear prediction of long-memory processes: asymptotic results on

mean-squared errors,” Mimeo.

Granger, C.W.J. (2008), “Non-Linear Models: Where Do We Go Next. Time Varying Parameter Models?,” *Studies in Nonlinear Dynamics & Econometrics*, 2008 Article 1.

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

Haque, N., M.H. Pesaran and S. Sharma (2000), “Neglected heterogeneity and dynamics in cross-country savings regressions,” in (eds.) J. Krishnakumar and E. Ronchetti, *Panel Data Econometrics- Future Directions: papers in honor of Pietro Balestra*, Elsevier, Amsterdam.

Hsiao, C. and M. H. Pesaran (2007), “Random coefficient panel data models,” in L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data*, Kluwer Academic Publishers.

Hu. X. and S. Ng (2004), “Estimation covariance structures of dynamic heterogeneous panels,” Mimeo.

Imbens, G. W. and T. Lancaster (1994), “Combining micro and macro data in microeconomic models,” *Review of Economic Studies* 61, pp. 631-653.

Imbs, J, A. Muntaz, M.O. Ravn and H. Rey (2005), “PPP Strikes Back: Aggregation and the Real Exchange Rate,” *Quarterly Journal of Economics* 120, pp. 1-43.

Kuersteiner, G. M. (2005), “Automatic Inference for Infinite Order Vector Autoregressions”, *Econometric Theory*, 21, pp. 85-115.

Lewbel, A. (1994), “Aggregation and Simple Dynamics,” *American Economic Review* 84, pp. 905-918.

Lewis, R. and G. Reinsel (1985), “Prediction of multivariate time series by autoregressive model fitting,” *Journal of Multivariate Analysis* 64, pp. 393-411.

Lin, G. D. (1997) “On the moment problems”, *Statist. Probab. Lett.* 35 85-90.

Lünnemann, P. and T. Matha (2004), “How persistent is disaggregate inflation? An analysis across EU Countries and HIPC Sub-indices,” ECB Working paper 415.

Kelejian, H. J. (1980), “Aggregation and disaggregation of nonlinear equations,” in: J. Kmenta and J. B. Ramsey (eds.): *Evaluation of Econometric Models*, New York: Academic Press, 135-153.

Koops, G., M. Pesaran and S.M. Potter (1996), “Impulse response analysis in nonlinear

multivariate models,” *Journal of Econometrics* 74 (1), pp. 199-47.

Judd, K. L. (1985), “The law of large number with a continuum of IID random variables,” *Journal of Economic Theory* 35, 19-25.

Ng, S. and P. Perron, (1995), “Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag,” *Journal of the American Statistical Association* 90, pp. 268-281.

Pesaran, M. H. (2006), “Estimation and inference in large heterogeneous panels with a multifactor error structure.” *Econometrica* 74.

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

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

Stoker, T. M. (1984), “Completeness, distribution restrictions and the form of the aggregate function,” *Econometrica*, Vol. 52, N<sup>o</sup> 4, pp.887-907.

Teräsvirta,T. (2006), “Forecasting economic variables with non-linear models,” in G. Elliot, C.W.J. Granger and A. Timmermann, eds, *Handbook of Economic Forecasting* , Vol 1, Amsterdam: Elsevier, 413-457.

Uhlig, H. (1996), “A law of large numbers for large economies,” *Economic Theory* 8, 41-50.

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

Zhang, P. and D. Small (2006), “Bayesian Inference for random coefficient dynamic panel data models,” Mimeo.

## APPENDIX

### AGGREGATION AND THE LAW OF LARGE NUMBERS

Throughout the paper it is assumed that the population aggregate model is defined as the expected value of the individual processes. However, aggregate data, denoted as  $\bar{Y}_{Nt}$  henceforth, is usually constructed as (a possibly weighted) average of the individual data. Hence,  $\bar{Y}_{Nt}$  would approximately follow the aggregate model in (13) if a LLN relating  $Y_t$  and  $\bar{Y}_{Nt}$  holds. However, it has been argued that such a LLN might not hold (see, for instance, Forni and Lippi, 1997, p. 17). Since the applicability of the results obtained in Section 3 relies on this convergence, it is worth considering this issue in more detail.

For simplicity, we assume that the aggregate data  $\bar{Y}_{Nt}$  is constructed as a simple average of a large number of individual processes

$$\bar{Y}_{Nt} = \frac{\sum_{i=1}^N y_{it}}{N},$$

where  $y_{it}$  is defined as in (1) with  $b_i = 0$  for all  $i$ .  $\bar{Y}_{Nt}$  can be written as the sum of two terms

$$\bar{Y}_{Nt} = \frac{1}{N} \left( \sum_{i=1}^N \frac{\varepsilon_{it}}{(1 - a_i L)} + \sum_{i=1}^N \frac{\rho_i u_t}{(1 - a_i L)} \right), \quad (47)$$

that will be referred to as the idiosyncratic and common components, respectively. Then,  $\bar{Y}_{Nt}$  would be a good approximation of the aggregate model derived in Section 3 provided that a LLN applies, such that  $\bar{Y}_{Nt}$  and  $Y_t = E_I(y_t)$  are close for large  $N$ .

The holding of a LLN relating  $\bar{Y}_{Nt}$  and  $Y_t$  hinges on whether the limit of  $\bar{Y}_{Nt}$  when  $N \rightarrow \infty$  is stationary or not. So, before considering the convergence of  $\bar{Y}_{Nt}$  and  $Y_t$ , the asymptotic properties of  $\bar{Y}_{Nt}$  as  $N$  increases should be reviewed.

This issue has been analyzed by Zaffaroni (2004) and we briefly summarize the results that are relevant to the problem considered here. The asymptotic behavior of  $\bar{Y}_{Nt}$  critically depends on the properties of the distribution of  $a$  around 1. As shown by Granger (1980), if the support of  $a$  is given by  $[\omega_1, \omega_2]$  with  $\omega_2 < 1$ , the corresponding aggregate process is I(0), for any shape of the distribution of  $a$ . On the other hand, if  $\omega_2 = 1$  and the distribution of  $a$  is such that  $P(a = 1) > 0$ , then  $\bar{Y}_{Nt}$  converges to an I(1) random variable. An interesting intermediate case arises whenever  $\omega_2 = 1$  and  $a$  belongs to a family of absolutely

continuous distributions such that  $P(a = 1) = 0$ . To characterize the convergence in this case, Zaffaroni (2004) considers the following semiparametric specification of the density of  $a \in (0, 1)$  around unity,<sup>19</sup>

$$f(a) \sim c_b (1 - a)^{-b}, \text{ as } a \rightarrow 1, \quad 0 < c_b < \infty, \quad b \in [0, 1)$$

In this case,  $\bar{Y}_{Nt}$  converges to a stationary random variable provided  $b < 0.5$  and to a nonstationary one otherwise. Interestingly, if  $0 \leq b < 0.5$ , the limit of  $\bar{Y}_{Nt}$  is a long-memory process and if  $b > 0$ , the limit process can be characterized as a fractionally integrated process with order of integration  $d = b$ .

Under similar assumptions as the ones adopted in this paper, Zaffaroni (2004) shows that provided the limit of  $\bar{Y}_{Nt}$  is stationary, a strong LLN holds and  $\bar{Y}_{Nt} \xrightarrow{L_2} E_t(y_t) = Y_t$ . In this case, the idiosyncratic component converges almost surely to zero while the common component converges in  $L_2$  to the corresponding expectation<sup>20</sup>

$$N^{-1} \sum_{i=1}^N \frac{\rho_i u_t}{(1 - a_i L)} \xrightarrow{L_2} E_I \left( \frac{\rho u_t}{1 - aL} \right) = \sum_{j=0}^{\infty} E_I(a^j) u_{t-j}.$$

However, whenever the limit of  $\bar{Y}_{Nt}$  is a nonstationary random variable, the convergence above fails: the idiosyncratic component no longer vanishes because the variance of  $N^{-1} \sum_{i=1}^N \frac{\varepsilon_{it}}{(1 - a_i L)}$  tends to infinity. On the other hand, the common component does not converge to its expected value because neither the Bochner nor the Pettis integral of this component exist.

In principle, this could be a major drawback for the results established in Section 3. If nonstationary variables are observed,  $\bar{Y}_{Nt}$  might not be a good proxy for  $Y_t$ . Then, one should not expect the persistence estimates obtained with aggregate data,  $\bar{Y}_{Nt}$ , to be close to those obtained with the corresponding disaggregate variables. Notice, however, that this

---

<sup>19</sup>This condition is semiparametric because the behavior of the density for any given interval  $[0, \gamma_2]$  with  $\gamma_2 < 1$  is unspecified. Standard distributions, such as the Uniform or the Beta, are contained in this specification by setting  $b = 0$  and  $b \geq 0$ , respectively.

<sup>20</sup>The expectation of the idiosyncratic component is taken with respect to the Pettis integral (see Uhlig, 1986). This is because the Bochner integral (which extends the definition of the Lebesgue integral to functions taking values in a Banach space) of that component may not exist. This is the well known *measurability problem* (Judd, 1985).

problem has an easy solution. Taking first differences from the original aggregate data,  $\bar{Y}_{Nt}$ , we obtain

$$(1-L)\bar{Y}_{Nt} = \frac{1}{N} \left( \sum_{i=1}^N \frac{(1-L)\varepsilon_{it}}{(1-a_iL)} + \sum_{i=1}^N \frac{(1-L)\rho_i u_t}{(1-a_iL)} \right), \quad (48)$$

and, in this case, the same results as in the case where the limit of  $\bar{Y}_{Nt}$  is stationary are recovered, that is, the idiosyncratic component in (48) converges to zero while the common one converges to the corresponding expectation. Thus, it holds that

$$(1-L)\bar{Y}_{Nt} \xrightarrow{L_2} (1-L)Y_t, \quad (49)$$

where  $(1-L)Y_t$  is the first difference of  $Y_t = E_I(y_t)$  and is a stationary process. Thus, whenever nonstationarity is detected, the usual procedure of first differentiating the data would be sufficient in order to guarantee the convergence of  $(1-L)\bar{Y}_{Nt}$  to  $(1-L)Y_t$ . The IRF of  $Y_t$  can be estimated by first estimating the IRF associated with  $(1-L)\bar{Y}_{Nt}$ , and, then, cumulating the corresponding values. That is,

$$\widehat{IRF}_{AG}(t, h) = \sum_{j=1}^h \widehat{IRF}_{(1-L)\bar{Y}_N}(j, t),$$

where  $\widehat{IRF}_{AG}(t, h)$  and  $\widehat{IRF}_{(1-L)\bar{Y}_N}(j, t)$  denote the estimates of the IRFs associated with  $Y_t$  and with  $(1-L)Y_t$ , respectively.

# FIGURES

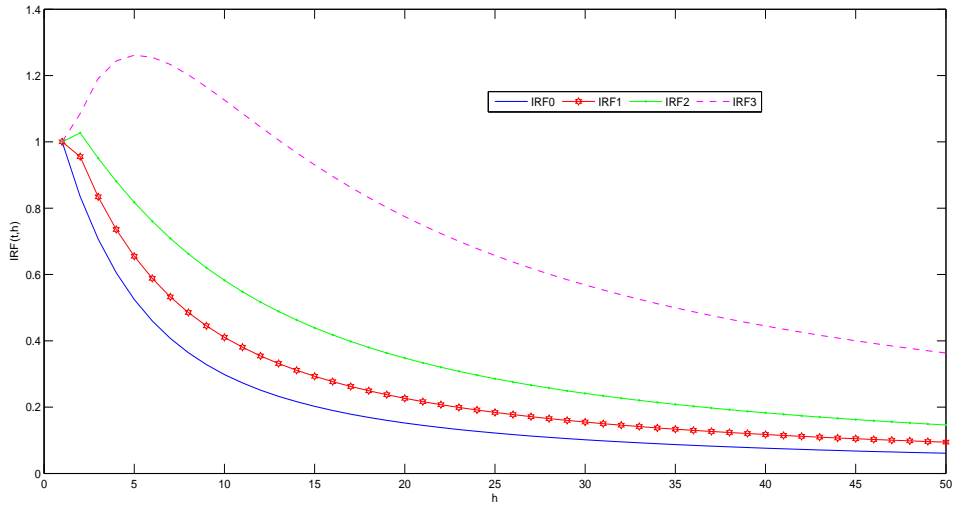


FIG 1. IRFs under correlation between  $a$  and  $\rho$ .

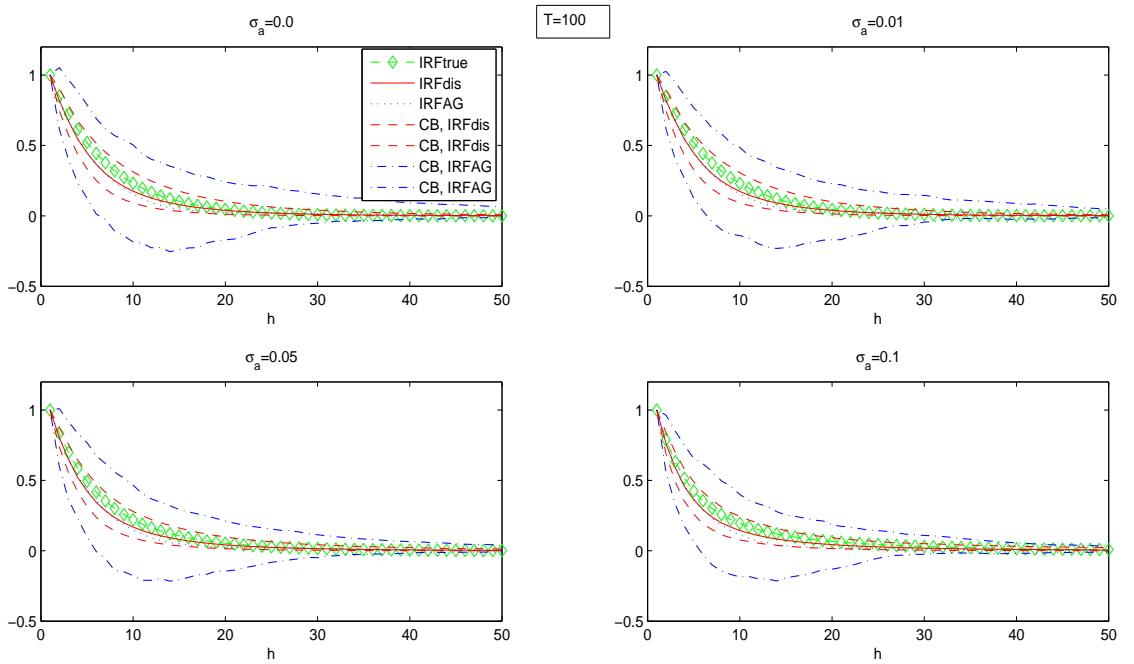


FIG 2. Estimated IRFs, LP I.  $T=100$ .

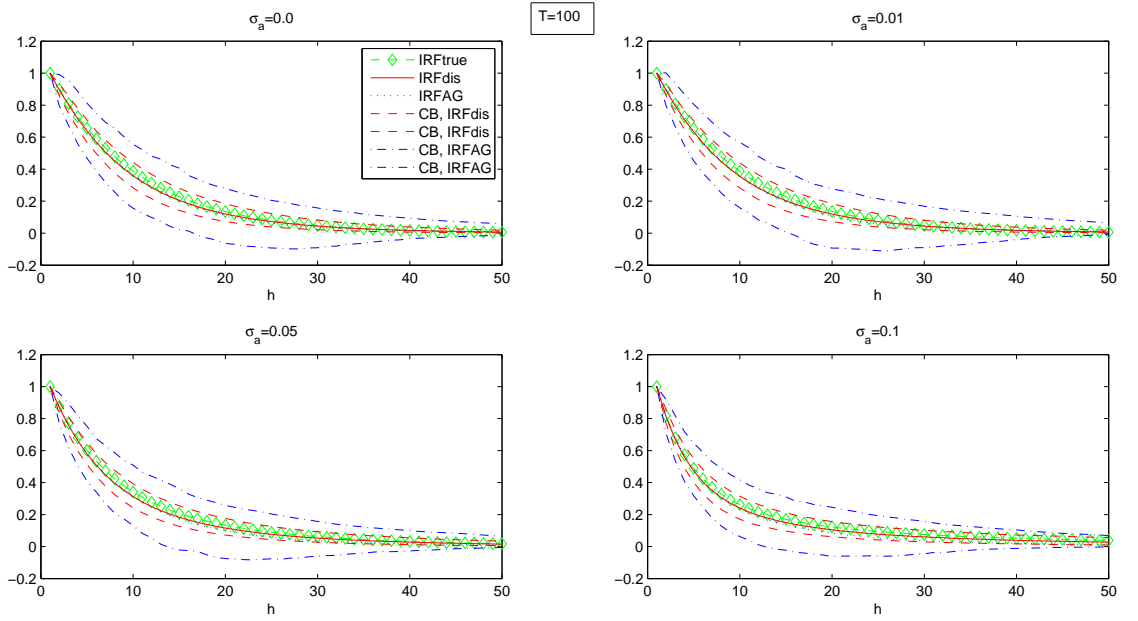


FIG 3. Estimated IRFs, LP II.  $T=100$ .

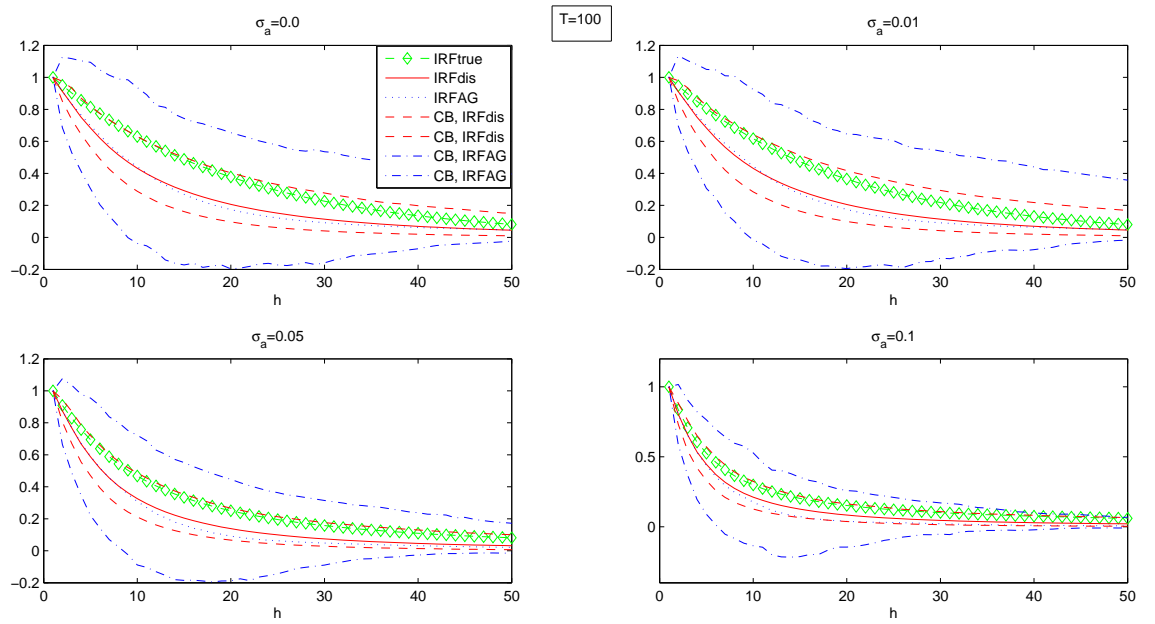


FIG 4. Estimated IRFs, LP III  $T=100$ .

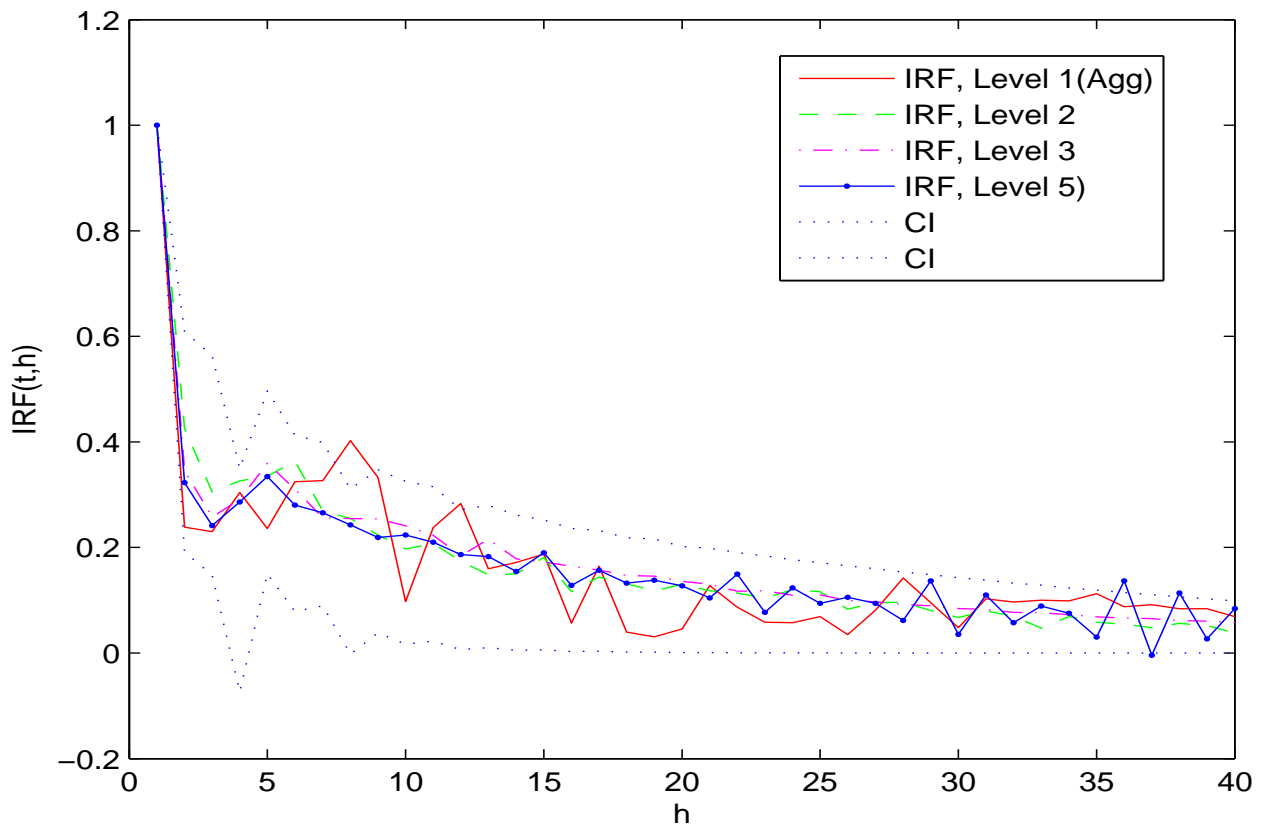


FIG 5. Impulse responses of U.S. inflation computed at different aggregation levels