Towards understanding the relationship between aggregate fluctuations and individual heterogeneity

Filippo Altissimo[*] **and Paolo Zaffaroni**[**]

[*] **European Central Bank** [**]**Banca d'Italia**

Preliminary, July 2003

1 Introduction

Understanding the cause of aggregate fluctuations and, therefore, being able to assess whether relevant macroeconomic variables possess a stochastic (ST) or a deterministic trend (DT), namely models for which the effect of random shocks persist for ever or models for which this effect is eventually absorbed as time passes converging towards a steady state, represents one of the main issues in empirical macroeconomics. The number of papers dedicated to this issue is considerably large. The *pure* ST model, or more generally models where the size of the random walk component is *large*, dominated for more than a decade since seminal work of Nelson and Plosser (1982). To date, however, the dominant paradigm oscillates between the so-called *we don't know* viewpoint (see, in particular, Christiano and Eichenbaum (1990) and Rudebusch (1993)) and the DT model (see, among others, Diebold and Senhadji (1996)). In the class of DT models we, agnosticly, include models which are practically in-distinguishable from it, such as long memory models (see Diebold and Rudebusch (1989)), where the effect of the shock diminishes slowly according to a power law, or even ST models with a *small* random walk component (see Cochrane (1988)).

The mainstream approaches all used aggregate macro economic data, mainly real GNP data. This has limited the ability of the various statistical methods to disentangle between the DT and the ST hypothesis, both because long spans of data are not available and because these methods typically exhibit low power against the contiguous DT and ST alternatives. These limitations represent the ultimate causes of this long-lasting debate.

This paper reexamines the nature of the fluctuations of real income per capita combining the information stemming from aggregate and individual (panel) data. There is no reason to assume that agents do perceive only shocks common to all individuals and employ the same model of economic behaviour. Instead, it is reasonable to allow for idiosyncratic random shocks and, even if retaining the same model, assume different parameter values across agents. Although, theoretically, one can allow for a great deal of heterogeneity in shocks and parameters, empirical estimation of such *rich* models entails simplification of the degree of heterogeneity. This typically takes the form of simplifying, if not cancelling, heterogeneity across parameters and maintaining heterogeneity across shocks (see Abowd and Card (1989) and Pischke (1995)).

In this paper we consider *simple*, in the sense of linear, dynamic models of individual behaviour but un-restricting heterogeneity across shocks and parameters. Linearity is justified by the fact that the reduced form of dynamic general equilibrium models is typically well approximated by auto-regressive moving average (ARMA) models with exogenous regressors. (The approximation error is zero only for linear-quadratic models of intertemporal optimization.) Accounting for full heterogeneity matches, on one hand, empirical observation, as noted above. This permits not to impose, for instance, that all individual incomes have an exact unit root but rather leave the data say whether this holds or not, case by case. On the other, a number of statistical results emphasizes the relevance of heterogeneity in imparting the statistical properties of the aggregate and that even mild difference in the form of heterogeneity at individual level can imply relevant differences at the aggregate level (see Robinson (1978), Granger (1980) and Forni and Lippi (1997)). In particular, it has been pointed out that stationary processes for individuals can nevertheless give rise to nonstationarity at the aggregate level. Moreover, although the individual processes are modelled as (stationary) ARMA, aggregate process will not be an ARMA anymore in general. Finally, the impact of purely idiosyncratic shocks might have a non-degenerate impact on aggregate fluctuations (see Lippi and Zaffaroni (1998) and Zaffaroni (2003) for details). In words, this means that aggregation could drastically modify the impulse response function at the aggregate level compared with the ones of individual processes. The crucial ingredient turns out to be the shape of the cross-sectional distribution of the auto-regressive roots. These results explicitly provide the taxonomy for linking the parameters dictating aggregate dynamics the aggregate with parameters representing cross-sectional heterogeneity.

The main empirical findings, based on a a panel of real income extracted from the PSID data set for roughly a thousand households over twenty-six years, are the following. Real GNP per capita significantly belongs to the class of trend stationary processes. In particular, the impact of random shocks does dissipate as time passes albeit slowly, according to a power law. Technically, the estimated power spectrum of aggregate data exhibit mass at zero frequency, well approximated by a long memory process with a small memory parameter. The propagation mechanism appears, therefore, markedly different from the one characterizing individual income processes. The latter turn out to be well approximated, in great majority, by stationary ARMA models (around a deterministic trend), for which the impact of random shocks dissipates exponentially fast. A small fraction of units, instead, does exhibit a unit auto-regressive coefficient. The estimates indicate a significant degree of heterogeneity across individual parameters, besides significant heterogeneity for the shocks perceived by individuals. The estimated degree of memory of the aggregate series mimics strikingly well the memory induced by the shape of the cross-sectional distribution of the auto-regressive roots estimated from individual data.

Uncovering the nature of business cycle fluctuations has relevant implications. Among many, it has a say on whether aggregate consumption should or should not exhibit *excess smoothness*. When the income process has a ST, consumption will be more volatile than income if the permanent income hypothesis holds (see Campbell and Deaton (1989)). Empirically, the opposite has bee observed, justifying the terminology of *excess smoothness*. Several explanations have been formulated, nicely surveyed in Attanasio (1999), all retaining the ST assumption for income but imposing some frictions which attenuate model-driven *excess smoothness*. It turns out that allowing for full heterogeneity in parameters and random shocks implies, under general circumstances, that consumptions is predicted to be less volatile than income. Therefore there would be no *excess smoothness* paradox at all. This result does not require to impose any frictions, which in any case would not alter our result.

Although the main goal of this paper is empirical, estimating the relationship between individual heterogeneity and the dynamic properties of the aggregate macro economic time series, prompted few but relevant methodological issues. First, we developed a formal statistical procedure to test whether heterogeneity across individuals is also due to non-negligible variation of parameters across individuals, and not simply due to the presence of an idiosyncratic component of random shocks hitting individuals. Second, the model for individual income is formally an ARMA$(1,1)$ factor model. Estimation methods recently proposed (see Stock and Watson (1999) and Forni, Hallin, Lippi, and Reichlin (2000)) do not not require a parametric specification of the model but impose other restrictions which rule out auto-regressive structures, such as ours. Therefore, we devised a simple estimation method, fully exploiting the ARMA$(1,1)$ structure. Third, heterogeneity across coefficients is described assuming that parameters are independent and identically distributed (*i.i.d.*) realizations of some underlying (multivariate) random variable, henceforth denominated the cross-sectional distribution. Its distribution is estimated nonparametrically. However, due to the small temporal dimension of the sample, a large bias is likely to occurr, especially for the distribution of the auto-regressive coefficients. Therefore, we envisage an estimation method, tailored to our model, that takes into full account the bias problem. This aspect is crucial since the shape of the cross-sectional distribution represents the key ingredient driving the effect of the aggregation mechanism.

This paper develops as follows. Section 2 recalls a number of results on aggregation of heterogeneous time series, used throughout the paper. A preliminary analysis of the data on individual income is described in section 3. Section 4 analyzes the features of a possible factor structure in the data.

The econometric model, and the related estimation method, are introduced in section 5. The empirical result, concerning the degree of persistence of aggregate income, is reported in section 6. Section 7 explores the theoretical and empirical implications of heterogeneity for the consumption smoothing phenomenon. Concluding remarks are in section 8. A more detailed analysis of the statistical properties of the test for heterogeneity are described in Appendix A. Versions of the PIH model with frictions, when allowing for heterogeneity, are exemplified in Appendix B.

2 Some facts on aggregation of heterogeneous AR(1) models

In this section we examine, by means of a series of numerical examples, the most important results on contemporaneous aggregation of heterogeneous ARMA models, when the number of units gets arbitrarily large. Further details are reported in Zaffaroni (2003). We focus, for sake of simplicity, on the case where the behaviour for some characteristic of the ith agent is described by an AR(1) model:

$$x_{it} = \alpha_i x_{it-1} + \eta_{it}, \quad (1)$$

where both the coefficients and the random shocks vary across individuals. When considering an arbitrary large number of units, a convenient way to allow for heterogeneity is to assume that the coefficients α_i are *i.i.d.* random drawn from some underlying, finite dimensional, distribution $F(\alpha)$. Stationarity of x_{it} then requires

$$\mid \alpha_i \mid < 1 \;\; a.s., \quad (2)$$

or, alternatively, that $F(\alpha)$ has support $(-1, 1)$. Finally, we assume that the random shock represent the sum of a common and of an idiosyncratic component

$$\eta_{it} = u_t + \epsilon_{it}, \quad (3)$$

with the u_t being an *i.i.d.* sequence $(0, \sigma_u^2)$ and ϵ_{it} being an *i.i.d.* sequence $(0, \sigma_\epsilon^2)$. The $\epsilon_{i,t}$ are also assumed independent across individuals.

In view of (3) and linearity of the model one gets the following decomposition for the aggregate (L denotes the lag operator)

$$X_{n,t} = \frac{1}{n}\sum_{i=1}^{n} \frac{u_t}{1-\alpha_i L} + \frac{1}{n}\sum_{i=1}^{n} \frac{\epsilon_{it}}{1-\alpha_i L} = U_{n,t} + E_{n,t},$$

meaning that the aggregate could be separated in a common and idiosyncratic component.

The statistical properties of each unit are well-defined, given her realization for α_i. On the other hand, knowledge of the entire history of the x_{it}, or even of a finite number n of them, is completely uninformative on $F(\alpha)$. In that case, in fact, the randomness assumption of the α_i is unnecessary. However, when looking at an arbitrarily large number of units, $F(\alpha)$ will entirely determine the properties of the limit aggregate, meaning the limit (with respect to some metric) of the $X_{n,t}$ for $n \to \infty$. We aim at establishing the statistical properties of the so-called limit aggregate.

It is well know that summing a finite number of ARMA processes yields again an ARMA process. For example, the sum of n distinct AR(1) models, with different auto-regressive parameters, yields an ARMA$(n, n-1)$ (see Granger and Morris (1976)). However, when n diverges to infinity, it turns out that for absolutely continuous $F(\alpha)$, the limit of $X_{n,t}$ will not belong to the class of ARMA processes, in contrast to the individual x_{it}.

To illustrate the main results, let us consider two possible ways of parameterizing $F'(\alpha) = f(\alpha)$. For sake of comparison, we consider only distributions satisfying

$$E\alpha_i = \mu, \tag{4}$$

for some $0 \leq \mu \leq 1$ (together with (2)). Generalization to the case of negative μ is straightforward. However, we do not explore this case since the estimates of μ are significantly positive for the data set used in this paper.

First, consider case of the uniform distribution over the interval $[1-\epsilon, 1]$, some $0 \leq \epsilon \leq 2$. Imposing (4) yields

$$f_1(\alpha) = \begin{cases} \frac{1}{2(1-\mu)}, & 2\mu - 1 \leq \alpha \leq 1, \\ 0, & \text{otherwise} . \end{cases}$$

Second, consider the case of Beta distribution, with parameters $p, q > 0$,

$$f_2(\alpha) = \begin{cases} B^{-1}(p,q)\alpha^{p-1}(1-\alpha)^{q-1}, & 0 \le \alpha < 1, \\ 0, & \text{otherwise.} \end{cases}$$

where $B(p, q)$ denotes the Beta function of order p, q. Imposing (4) requires

$$p = \left(\frac{\mu}{1-\mu}\right) q. \tag{5}$$

Let us now focus on the common component. This can be written as

$$U_{n,t} = u_t + \hat{\mu}_1 u_{t-1} + \hat{\mu}_2 u_{t-2} + ... +, \tag{6}$$

setting

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} \alpha_i^k, \quad \text{real } k.$$

When n gets large, by the strong law of large numbers, each $\hat{\mu}_k$ will converge $a.s.$ to the population moments of the α_i:

$$\hat{\mu}_k \to \mu_k = E(\alpha_i^k) \;\; a.s. \quad \text{for } n \to \infty.$$

It turns out that (under suitable regularity conditions) the $U_{n,t}$ will converge in mean square to the limit aggregate

$$U_t = u_t + \mu_1 u_{t-1} + \mu_2 u_{t-2} + ... + \tag{7}$$

as n goes to infinity. This holds for the uniform distribution case and for the Beta distribution case when $q > 1/2$.

From (7), the dynamic pattern of the μ_k represents the impulse response of the common shocks u_t on the aggregate. We now make a numerical evaluation of the impulse response function of the limit aggregate. For the uniform density $f_1(\alpha)$ simple calculation yields

$$\mu_{1,k} = \frac{1}{(k+1)\, 2(1-\mu)} \left(1 - (2\mu - 1)^{k+1}\right), \tag{8}$$

whereas for the Beta density $f_2(\alpha)$

$$\mu_{2,k} = \frac{\Gamma(\frac{\mu}{1-\mu}q + k)}{\Gamma(\frac{\mu}{1-\mu}q)} \frac{\Gamma(\frac{q}{1-\mu})}{\Gamma(\frac{q}{1-\mu} + k)}, \tag{9}$$

where $\Gamma(x)$ indicates the Gamma function.

Table 0 reports the dynamic pattern of μ_k, both for the uniform and the Beta distribution case (the latter for various values of q). We compare the results with the case of homogeneous AR(1), setting the auto-regressive coefficient equal to μ. In this case, the impulse response will be given by μ^k. The first half of the table reports the results for $\mu = 0.8$. This is very close to the empirical value found for $E\alpha_i$ from the data. The second half of the table considers $\mu = 0.95$. In this way, we can compare the effect of aggregation with an homogeneous, yet very persistent, case. The effect of aggregation is dramatic: the impulse response function of the aggregate process (common component) decays towards zero very slowly, compared with the constant coefficient case. This is true even for unrealisticly large values of μ, such as $\mu = 0.95$. For the Beta distribution case, note that the smaller is q, the slower will the effect of random shocks fade away. Finally, choosing a different μ has a greater effect on the constant parameter cases rather than on the heterogeneous case.

Further insights can be obtained looking at the corresponding analytic results. The long-run dynamic pattern of the $\mu_{1,k}$ is evident from (8). Case $2\mu - 1 = 1$ is ruled out since this implies a degenerate (uniform) distribution on the interval $[1, 1]$. Case $2\mu - 1 = -1$ asks for $\mu = 0$. Ignoring these cases, $0 < (2\mu - 1) < 1$ and therefore,

$$\mu_{1,k} \sim \frac{c}{k+1} \quad \text{as } k \to \infty.$$

Therefore, for aggregate the impulse response displays an hyperbolically decaying pattern. A similar result is obtained for the Beta distribution case. For the $\mu_{2,k}$ one needs to use Stirling's formula, $\Gamma(x) \sim \sqrt{2\pi} e^{-x+1}(x-1)^{x-1/2}$ as $x \to \infty$, to (9) yielding

$$\mu_{2,k} \sim c k^{-q} \quad \text{as } k \to \infty.$$

Again, an hyperbolic behaviour arises whose intensity is now directly determined by the Beta parameter q, in agreement with the findings of Table 0. In particular, the smaller is q, the slower will the $\mu_{2,k}$ converge towards zero (recall than $q > 0$ always). Note, however, that the effect of the mean μ

Table 0: impulse response functions of limit aggregate U_t.

k	μ^k	$f_1(\alpha)$	$f_2(\alpha)$ $q=$	0.2	0.3	0.7	1	3
			$\mu = 0.8$					
1	0.8	0.8		0.8	0.8	0.8	0.8	0.8
2	0.64	0.65		0.72	0.70	0.67	0.66	0.65
5	0.33	0.39		0.61	0.57	0.48	0.44	0.37
10	0.11	0.23		0.54	0.47	0.33	0.28	0.17
50	1.4×10^{-5}	0.05		0.39	0.29	0.12	0.07	0.01
200	4.1×10^{-20}	0.01		0.36	0.26	0.09	0.05	0.01
			$\mu = 0.95$					
1	0.95	0.95		0.95	0.95	0.95	0.95	0.95
2	0.90	0.90		0.91	0.91	0.91	0.90	0.90
5	0.77	0.78		0.83	0.82	0.79	0.79	0.78
10	0.60	0.62		0.76	0.73	0.67	0.65	0.62
50	0.08	0.19		0.58	0.49	0.33	0.27	0.15
200	3.5×10^{-5}	0.05		0.54	0.45	0.27	0.22	0.11

$f_1(\alpha)$ denotes the uniform density and $f_2(\alpha)$ the Beta density with parameter q.

Table 00: asymptotic behaviour of $var_n(E_{n,t})$.

n	$f_1(\alpha)$	$f_2(\alpha)$ $q=$ 0.2	0.3	0.7	1	3
10	0.1	1.6	0.4	0.3	0.2	0.2
100	2.2×10^{-2}	1.4×10^{6}	5.4	3.1×10^{-2}	5.2×10^{-2}	1.6×10^{-3}
$1,000$	4.1×10^{-3}	3.6×10^{4}	8.3×10^{3}	1.4×10^{-2}	3.7×10^{-3}	1.5×10^{-3}
$5,000$	8.9×10^{-4}	2.8×10^{8}	1.8×10^{3}	4.3×10^{-3}	8.4×10^{-4}	3.1×10^{-4}
$10,000$	4.9×10^{-4}	∞	1.2×10^{4}	3.7×10^{-3}	4.2×10^{-4}	1.6×10^{-4}
$20,000$	3.1×10^{-4}	∞	4.6×10^{3}	1.6×10^{-3}	2.2×10^{-4}	7.9×10^{-5}

$f_1(\alpha)$ denotes the uniform density and $f_2(\alpha)$ the Beta density with parameter q. Symbol ∞ indicates a large value (computer outflow) for $\mathrm{var}_n(E_{n,t})$.

is completely annihilated asymptotically, as this impacts on the other Beta parameter p. In practice, as Table 0 shows, a non-negligible effect arises since we look at the impulse response for a finite horizon. In view of the linearity of the set-up, this characterization of the impulse response has a neat representation in terms of the auto-covariance function (acf) and spectral density of the limit aggregate U_t.

Let us focus on the idiosyncratic component $E_{n,t}$. Table 00 reports the asymptotic behaviour of

$$\mathrm{var}_n(E_{n,t}) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{1-\alpha_i^2},$$

for various values of n. Hereafter, $\mathrm{var}_n(.)$ indicates the variance operator, for given parameter values $(\alpha_1, ..., \alpha_n)$.

The most, somewhat striking, result that appears from Table 00 is that this component, made by averaging perfectly independent and stationary $(a.s.)$ units, does not necessarily vanish. In fact, $\mathrm{var}_n(E_{n,t})$ gets smaller as n increases for the uniform distribution case and for the Beta distribution case for $q = 0.7, 1, 3$. In contrast, $\mathrm{var}_n(E_{n,t})$ gets arbitrarily large for $q = 0.2, 0.3$. It turns out that this is precisely what the theory predicts.

One can then ask which component is the dominant one with respect to

Table 000: asymptotic behaviour of R_n.

n	$f_1(\alpha)$	$f_2(\alpha)$ $q=$	0.2	0.3	0.7	1	3
10	0.1		0.9	0.9	0.3	0.1	0.1
100	2.7×10^{-2}		0.51	–	4.9×10^{-2}	1.7×10^{-2}	1.1×10^{-2}
$1,000$	3.1×10^{-3}		0.49	–	1.2×10^{-2}	3.7×10^{-3}	1.1×10^{-3}
$5,000$	8.6×10^{-4}		0.48	–	3.6×10^{-3}	9.4×10^{-4}	2.3×10^{-4}
$10,000$	3.4×10^{-4}		0.17	–	1.5×10^{-3}	3.6×10^{-4}	1.1×10^{-4}
$20,000$	1.8×10^{-4}		–	–	9.9×10^{-3}	1.6×10^{-4}	5.7×10^{-5}

$f_1(\alpha)$ denotes the uniform density and $f_2(\alpha)$ the Beta density with parameter q. Symbol – indicates an undetermined value (computer outflow) for R_n.

aggregate dynamics. In Table 000, we look at the behaviour of the ratio

$$R_n = \frac{\text{var}_n(E_{n,t})}{\text{var}_n(U_{n,t})}$$

as n gets large.

It turns out that R_n gets smaller as n increases for the uniform distribution case and for the Beta distribution case for $q = 0.7, 1, 3$. For the cases where the idiosyncratic variance tends to explode ($q = 0.2$, 0.3), instead, R_n is stable, in the sense that it does not diverge nor converge towards zero. This suggests that also the variance of $U_{n,t}$ tends to explode and, moreover, at the same rate. This is to say that in the non-stationary case the common and the idiosyncratic component have precisely the same importance in determining aggregate fluctuations.

Model (1) represents an extremely simplified set-up. In fact, we consider below, a more complicated and realistic model than (1). However, it turns out that, as far as the analysis of aggregation is concerned, the results that apply to an AR(1) set-up equally apply to higher-dimensional models, which must include an auto-regressive component, with no qualitative differences.

Summarizing, we have found that allowing for heterogeneity across the auto-regressive parameter, has a dramatic effect on the behaviour of the impulse response function. In fact, the effect of random shocks to the aggregate decays much slowly compared with any individual AR(1), in particular according to a power law. Second, the effect of the idiosyncratic shocks might not be negligible on aggregate fluctuations. The key ingredient appears to be the shape of the cross-sectional distribution $f(\alpha)$ around 1, dictated by the Beta parameter q. For α approaching 1, $f_2(\alpha) \uparrow \infty$ when $q < 1$ and $f_2(\alpha) \downarrow 0$ when $q > 1$. When $q = 1$ then $f_1(\alpha) \to c$.

Such results havd been established by Zaffaroni (2003) for the general case where

$$f(\alpha) \sim c(1-\alpha)^{\beta}, \quad \text{as } \alpha \to 1^{-}, \tag{10}$$

for $\beta > -1$ and $0 < c < \infty$. Note that (10) only defines the local behaviour of the density around unity.

3 The data

To perform an empirical investigation of the effect of aggregation we utilize a panel of individual income processes drawn from the Panel Study of Income Dynamics (PSID). From the PSID, we select a sample of male individuals in the working age (16-65) in the period 1967-1992. We retain those individuals with positive annual income (earning plus assets income) in at least 24 consecutive years starting between 1967 and 1969. This allows us to construct a sample of 950 time series ($n = 950$) of individual income processes, $I_{i,t}$.

We start by deflating income by the consumes price index for urban consumers (base 1983) and the data sample mean and median are reported in Figure 1. The average age of the individuals in the sample is 27 years at the beginning of the sample in 1967 and ends at 52 in 1992. In order to capture the deterministic component of the income level as well as the effect of the change of the age structure in the sample over time the individual data have been regressed over a quadratic trend as:

$$\log(I_{i,t}) = a_i + b_i t + c_i t^2 + y_{i,t} \text{ with } i = 1, ..., n \tag{11}$$

where $I_{i,t}$ is the real individual income, $y_{i,t}$ is the regression residual and t is a time trend equal to one in 1967. Only the significant coefficients at the 95% confidence are retained. In Figure 2 the average deterministic component, defined as

$$\frac{\sum_i \exp(a_i + b_i t + c_i t^2)}{n},$$

appears as the smoothed bold line. In the figure it is plotted versus the average age of the cross-section, the estimated income age profile presents the characteristic hump shape with a peak in the mid 40s and decline afterwards. The residual $y_{i,t}$ of the above regression is our variable of interest.

There are two facts which are common knowledge about the behavior of individual income data, the first one is that individual income is much more variable than aggregate one, second that individual process are much less persistent than aggregate ones. In light of those facts we start by investigate the variance-covariance structure of the $y_{i,t}$. Along the line of Abowd and Card (1989) and Pischke (1995), we estimate the covariance and correlation matrix of the income residual $y_{i,t}$ along the cross-section at each point in time t between 1971-1990, as:

$$\frac{\sum_i (y_{i,t} - \overline{y}_t)(y_{i,t+h} - \overline{y}_{t+h})}{n}$$

with $h = 1, ..., 20$ and $\overline{y}_t$ the cross-section average at time t, and the results are in the 20×20 matrix in Table 1. Below the diagonal there is the covariance, while above the correlation. As a matter of comparison, Table 2 presents some descriptive statistics for relevant aggregate variables (GNP, personal income and consumption, both absolute and per capita)[1]; in particular mean, standard deviation and autocorrelation up to lag five for the log-change and standard deviation and autocorrelation up to lag five for the residual of the regression of the log value on a deterministic quadratic trend.

An immediate fact clearly emerge from the inspection of Tables 1-2, namely that the individual income variance is an order of magnitude larger that the one of the aggregate income. This finding is in line with similar one of Abowd and Card (1989) and Pischke (1995) and while some of this variation can be attributed to measurement errors, large part of it surely reflects the presence of large idiosyncratic income shocks. Cross section dispersion of $y_{i,t}$ is relatively high in early 70s and the 80s and relatively small in the second part of the 70s.

[1] All the aggregate data are NIPA in chained 1996 dollars.

Turning to the off-diagonal terms, the average first order correlation is positive and on average over time equal to 0.37 and the second one is still positive and equal to 0.15. However the off diagonal term in Table 1 are not pure autocorrelation coefficients of the data given that they do not account of the dependence in the cross-section average, $\overline{y}_t$.

To this end in Figure 3, the distribution over the individuals of the first four autocorrelations is displayed as well as the cumulative distribution form each of the autocorrelations over the individuals versus the cumulative density of the test of the null of zero correlations. These second plots are utilized to asses if the sample dispersion in the correlation coefficients would be coherent with the null of zero correlation assuming that each income process on which the correlation is calculated is a independent draw. The average correlation at lag one is equal to 0.25 (standard error of the mean 0.008) and 0.02, -0.05, -0.09 and -0.11 at lag $2-4$ respectively (standard error of the mean of 0.007, 0.006, 0.005 and 0.005).

Again the evidence is in line with the common knowledge that the individual process are less persistent than the aggregate one, as the comparison with the value of the autocorrelations in Table 2 point out. Moreover this rapid decay of the covariance points to the idea that the stochastic term y can be well described as a stationary process. This points is at odds with previous literature (as Abowd and Card (1989) and Pischke (1995)) which in estimating individual income process retained the idea of modelling the growth rates of individual income as a simple short MA process, which in the estimation turned out to have a negative first coefficients. In the estimation step, we will not impose the unit root but we will estimate a short ARMA process for individual income and in this way allowing the possibility of a unit root in the level of individual income.

4 Features of the common component in the cross section

The previous section pointed to the fact that the individual income process do present a very high variance respect to the aggregate data and also they show much lower persistence. However, these two facts can be the results of large idiosyncratic income shocks as well as of the presence of measurement errors at individual level. Here we investigate the key issue of the existence of a common component (or common shock) in the cross-section of 950 time series.

Retaining the assumption of linearity, we start modelling the individual income process as the sum of two components:

$$y_{it} = \Psi_i(L)u_t + \xi_{it} = f_t + \xi_{it}, \;\; i = 1, ..., n,$$

where u_t is an aggregate (common) shock and ξ_{it} a stationary idiosyncratic component, orthogonal to the common one, where Ψ_i is a lag polynomial term which represent the way in which the common shock affect the y_{it} process.[2]

Following Stock and Watson (1999) and Forni, Hallin, Lippi, and Reichlin (2000), the presence of common components u_t in the covariance structure of a large cross-section can be inferred by the relative size of the eigenvalues of the variance-covariance matrix of the data. In Table 3 the first twenty eigenvalues associated with the contemporaneous variance-covariance matrix of the data are reported. Clearly the first two dominating eigenvalues point to the presence of some common feature in the data.[3]

To this end, we consider the cross-sectional average of the income process y_i as our aggregate measure:

$$Y_{n,t} = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}. \tag{12}$$

In the terminology of Forni, Hallin, Lippi, and Reichlin (2000), the averaging operation is an aggregating sequence and, as the size of the cross-section increases, $Y_{n,t}$ will be only a function of the common shocks u_t. On the contrary, if the shocks underlying the individual process were independent

[2]The assumption of a unique common shocks is only for convenience of exposition but the same logic applies in the presence of multiple common shocks.

[3]The number of large eigenvalues in the contemporaneous variance-covariance matrix cannot be directly associated to the presence of more common shocks, see (Forni, Hallin, Lippi, and Reichlin (2000)).

across agent or if they are only correlated among a limited amount across agents, than the cross-sectional average of the stochastic components, $Y_{n,t}$, would converge to zero by a standard law of large number as the size of the cross-section increases, conditional to the fact that the individual income processes do not present large persistence, but this has been verified to be not the case as in the previous section.[4] This fact does not seem to be the case as Figure 4 shows where $Y_{n,t}$ is the bold line. The variable $Y_{n,t}$ is the aggregate measure of which we aim to disentangle the dynamic properties thought the characteristic of the individual process $y_{i,t}$.

There are two immediate concerns in this approach. First that the aggregation in (12) is performed on the logarithm of the variable, while it would be more appropriate to consider the logarithm of the cross-sectional average of the level in order to resemble as close as possible the aggregation process implicit in the national account macro data. Second element of uncertainty is the relation between the variable $Y_{n,t}$ and its aggregate counterpart, in particular the personal per-capita income coming from national accounts.

Relative to the first point, the relation between the

$$\log(\frac{1}{n}\sum_{i=1}^{n} I_{i,t})$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\log(I_{i,t}) = \frac{1}{n}\sum_{i=1}^{n}(a_i + b_i t + c_i t^2) + Y_{n,t}$$

is not linear given that the expectation of the log is different from the log of the expectation and an exact mapping between the quantities is function of the distribution of the idiosyncratic shocks and parameters. While it would be quite hard to take fully account of that, empirically in Figure 5, we plotted the two series in the level and they are remarkably similar and the correlations of their changes is above 0.90. Moreover it should be stressed that we are not interested in the exact estimation of structural relationships, where this type of non linear effect would result in a possible bias in the estimation of the structural parameters as noted in Attanasio

[4]See Stock and Watson (1999) and Forni, Hallin, Lippi, and Reichlin (2000).

(1999), but we are interested in understanding the dynamic property of the data.

Concerning the relation between $Y_{n,t}$ and the aggregate income per capita process the relation is more problematic issue that clearly relates to the representativety of the PSID sample and consequently of our sub-sample. It is very doubtful that a small sample of 950 income process can be representative of the population and that it remains representative over a twenty five year horizon, for this reason we are not stressing to much the relation with the national account data, but we are more interested in understanding the dynamic property of $Y_{n,t}$ per se given the properties of the underlying individual process. However in Figure 4, the residual of the regression of the log of the aggregate per capita income over a quadratic trend is plotted together with the $Y_{n,t}$ process. The graph is quite striking given that in the common sample the two time series are quite similar except in the very beginning of our sample, which can be mainly related to the feature of our PSID sub-sample. Since then the two data series provide the same signal, having the same cyclical behavior and a correlation in the period 1970-1992 of 0.90.

Under a set of conditions verified in the following section on the distribution of the parameters characterizing the individual income process, as result of the aggregation process the variable $Y_{n,t}$ will be function only of the common (or aggregate) shocks and not of the idiosyncratic components (including also idiosyncratic measurement errors). The dynamic property of the aggregate are synthesized in its auto-correlation function in Figure 6, where it is plotted together with the autocorrelation of the aggregate per capita income and also with the average of the individual autocorrelation. The first correlation of the aggregate $Y_{n,t}$ is much higher than the average of the correlations and quite similar to the one of national account data; after it declines to 0.3 at lag two and than goes to zero faster than the national account data one but remaining always higher than the average of the individual correlations.

In the aggregation process two elements play a role. On one side the aggregation decrease the effect of idiosyncratic innovations relative to the common ones and on the other side it average the different propagation mechanism of the common shocks. Here we are in the position of assessing one the question of the paper, namely if the propagation mechanism of the common shocks in the cross-section of individual is equal across agent, i.e. $\Psi_i(L) = \Psi(L)\ \forall i$, or if it is different. So we aim to understand if the dispersion of the correlation observed in the previous section is due to the presence of large idiosyncratic noise and measurement errors.

As a first step in Figure 7, the correlations of the individual income process $y_{i,t}$ with the aggregate one $Y_{n,t}$ at lags 0 to 3 are reported. Being the aggregate variable $Y_{n,t}$ just a function of the aggregate shocks (at least asymptotically) it follows that if the true data generating process is of the form:

$$M_0: \ \ y_{it} = \Psi(L)u_t + \xi_{it} = f_t + \xi_{it} \ with \ i = 1, ..., n$$

then the dispersion of the empirical correlation is only a result of the small sample variability. In Appendix A, we develop a formal testing procedures based on the above intuition which consider the distribution of the correlation at lag h between the aggregate variable Y and the individual. Then one can test the null of $\Psi_i(L) = \Psi(L)\ \forall i$ through a test statistic denominated T_h. Moreover a modification of the same test for the more general hypothesis of $\Psi_i(L) = \alpha_i \Psi(L)$ with α_i a positive scalar $\forall i$, called R_h is also proposed.

The results in Figure 8.1, 8.2 and 8.3 are the cumulative distribution function (cdf) associated with the statistics T_0, T_1 and T_2 while Figure 8.4 and 8.5 reports one relative to the statistics R_1 and R_2; in both cases they are compare with the cdf under the null. The results point to the rejection of the hypothesis of homogeneity in the propagation mechanism of the common shocks given that all the statistic proposed present cdf which are far away from the standard normal one. While the small sample size of the data can be element of concern in the estimation, in light of the simulation exercise in the Appendix A it seems that the small sample effect cannot by himself generate the sharp differences between the statistic cdf have the one presented above.

5 Model and estimation

In the actual estimation we have to consider the trade off between accommodating the possible heterogeneity in the data generating process and the need of a parsimonious specification due to the short data set in hand. We considered a model with an autoregressive part of order one and two moving average components, one for the common shock u_t and a second one for the idiosyncratic term ϵ_{it}. The specification for each i is as follows:

$$y_{it} = \frac{\gamma_{0i} + \gamma_{1i} L}{1 - \alpha_i L} u_t + \frac{1 + \delta_i L}{1 - \alpha_i L} \epsilon_{it} = y_{it}^{\epsilon} + y_{it}^{u}, \tag{13}$$

with u_t is an *i.i.d.* innovation having unit variance and ϵ_{it} another *i.i.d.* sequence with mean and variance $(0, \sigma_{\epsilon i}^2)$. Finally u_t and ϵ_{si} are independent at all t, s, i. For notational convenience we denote with θ_i the parameters vector $\{\alpha_i, \gamma_{0i}, \gamma_{1i}, \delta_i, \sigma_{\epsilon i}^2\}$. At the individual level the model is indistinguishable from a simple ARMA(1,1) process as:

$$(1 - \alpha_i L)\, y_{it} = (1 + \eta_i L) w_{it}, \tag{14}$$

with $w_{it} \sim (0, \sigma_{wi}^2)$ being a linear combination of the common and idiosyncratic innovations. However the cross sectional aggregation of the innovations over the agents can allow to recover the common shocks given that:

$$\frac{\sum w_{it}}{n} \xrightarrow{i} E(\gamma_0) u_t.$$

The estimation procedure for the model in (13) for each i hinges upon this and it is composed by the following steps:

1. The ARMA(1,1) model in (14) is estimated for every i and the innovation $\widehat{w}_{it}$ recovered;

2. The cross-sectional average of the $\widehat{w}_{it}$ delivers an estimates of the $E(\gamma_0)u_t$; after standardization $\widehat{u}_t$ are recovered.

3. The estimated common shock $\widehat{u}_t$ is utilized as regressor in the estimation of (13) for each i. This delivers an estimation of the parameters vector $\widehat{\theta}_i$ and of the idiosyncratic innovations $\widehat{\epsilon}_{it}$.

Given the new estimates of w_{it} as $\widehat{\gamma}_{0i}\widehat{u}_t+\widehat{\epsilon}_{it}$ the procedure can be iterated going to step 2 until convergence; the need for iteration derives from the improve in efficiency in the estimates of the common shock u_t. In the actual estimation, we iterated the procedure six times. This was enough to assure a correlation of 0.97 between the parameters estimates at the last two rounds.

Note that in the cross section there is a problem of identification of the sign of the common shocks and the sign of the loading parameters of the common shocks, namely γ_{0i} and γ_{1i}. To couple with this identification problem, we assumed that the average of γ_0 loading is positive.

Moving to the result of the estimation, Figure 9 shows the common shock u_t as at the last iteration of the estimation procedure; while a summary of the statistics of the estimated parameters is in Table 4. The autoregressive parameter α assumes values in the range [-1.36, 1.24] with mean 0.356 but median 0.529 pointing to a skewed distribution towards unit root. Concerning the factor loadings of the common shocks, they have mean respectively 0.0716 and 0.0253, but again the distribution is skewed to the left and the median is equal to 0.0349 and 0.011 respectively. Figure 10 shows the distribution of the autoregressive parameters α; the density is uni-modal and skewed to the right.

An additional problem arises when estimating autoregressive models with a small sample. It is well-known that a significative downward bias affects the least squares estimator of the autoregressive parameters, especially when the latter is close to unity. Bias correction methods do exist but only for a class of relatively simple models (see Andrews (1993)). Unfortunately, model (13) is ruled out from such class, due to its moving average component. Moreover, in our random coefficient framework, this bias problem asks for a correction of the entire parameters' distribution. The problem can be seen as follows. We can estimate $f_T(\theta)$, that is the empirical distribution of the θ_i for a sample of size T (henceforth the small-sample density). On the other hand, we are really interested in the population density $f(\widehat{\theta})$ which satisfies:

$$f_T(\widehat{\theta}) = \int K_T(\widehat{\theta}, \theta) f(\theta) d\theta \tag{15}$$

where $K_T(\widehat{\theta}, \theta)$ indicates the small-sample kernel which, for a given $f(.)$ yields $f_T(.)$ for a given sample of size T. We proceed using the following argument in order to recover $f(.)$.

Assume that the true density is $f_T(.)$. We can then seek its associated small-sample density, that is another function h_T such that:

$$h_T(\widehat{\theta}) = \int K_T(\widehat{\theta}, \theta) f_T(\theta) d\theta. \tag{16}$$

This is not equivalent to evaluate the bias for each point estimate but it means that we are in effect re-sampling from $f_T(\theta)$, generate the data, re-estimate and obtaining $h_T(\theta)$. Hence, let us define the change-of-measure $\frac{f_T(\theta)}{h_T(\theta)}$. When we applied this change of measure to $h_T(.)$ yields the population density under which $h_T(.)$ was generated, namely $f_T(.)$. Moreover if the realizations under $h_T(.)$ are reweighted by the change of measure $\frac{f_T(\theta)}{h_T(\theta)}$, they can be treated as realizations from $f_T(.)$.

Applying the same line of reasoning in order to recover the population density $f(.)$ associated with the small sample density $f_T(.)$, one would need the change of measure $\frac{f(\theta)}{f_T(\theta)}$. In that case the change of measure $\frac{f(\theta)}{f_T(\theta)}$ could be appiled to the realizations under $f_T(.)$ in order to recover the moments of the parameters under $f(.)$.

Of course, we do not have $\frac{f(\theta)}{f_T(\theta)}$ but rather $\frac{f_T(\theta)}{h_T(\theta)}$and we can use it as a first approximation. In fact properly re-arranging terms in (15) it yields:

$$f(\widehat{\theta}) - f_T(\widehat{\theta}) \frac{f_T(\widehat{\theta})}{h_T(\widehat{\theta})} = \int K_T(\widehat{\theta}, \theta) \left(\frac{f(\widehat{\theta})}{f_T(\widehat{\theta})} - \frac{f_T(\widehat{\theta})}{h_T(\widehat{\theta})} \right) f(\theta) d\theta, \tag{17}$$

and this saws that the corrections works, namely the left hand side converges to zero, as long as the term $\frac{f(\theta)}{f_T(\theta)} - \frac{f_T(\theta)}{h_T(\theta)}$ is small and asymptotically zero.

We computed $h_T(\theta)$ and the change $\frac{f_T(\theta)}{h_T(\theta)}$ as follow. We draws 1000 sample of θ from the empirical distribution $f_T(\theta)$ and given those parameters values we generated the associated data y_{it} with sample size equal to the one of the true data; on the generated data we applied our estimation procedure to recover $\widehat{\theta}$ and to have an estimate of $h_T(\theta)$. The procedure is repeated 100 times and the average of the empirical densities $h_T(\theta)$ is considered.

The results of the change of measure are visible in the Figure 10, where the distribution of the autoregressive parameter α is plotted together with the bias corrected one. In Table 4 the moment of the parameters of interested under the bias corrected densities are reported in the last three lines of the table. As expected the bias correction shift the distribution of the persistence parameters to the right versus the unit root; the mean increases to 0.430 while no major correction are present for the other parameters. Figure 11 reports the again the bias-corrected distribution together with its 90 per cent confidence band, based on the 100 bootstrap estimates of $f(\alpha)$.

6 Persistence of aggregate income

We now discuss the empirical results, focusing on the degree of memory of aggregate income. The key variable is the cross-sectional distribution of the autoregressive coefficients α_i. As indicated above, Figure 10 shows the nonparametric estimate of such distribution, both un-adjusted and adjusted for the small-sample bias. The following features emerge. First, the support of the distribution is $[-1, 1]$. Only 19 units exhibit an estimated value greater or equal than 1 and 6 are below -1. Second, the distribution appears unimodal, with the great majority of individuals displaying a positive autoregressive coefficient and a mode of 0.53 in the original density and 0.62 in the adjusted one. Finally, the behaviour near unity of such distribution satisfies

$$\hat{f}(\alpha) \sim c(1-\alpha)^{0.13} \text{ as } \alpha \to 1^-,$$

implying that (10) holds with

$$\hat{\beta} = 0.13. \tag{18}$$

The bootstrap standard deviation of $\hat{\beta}$ is 0.07.

Let the (log) aggregate be

$$g_{n,t} = \frac{1}{n}\sum_{i=1}^{n} \log(Y_{i,t}) = \log\left(\prod_{i=1}^{n} Y_{i,t}\right)^{\frac{1}{n}} = \log G_{n,t},$$

thus equal to the logarithm of the geometric mean $G_{n,t}$ of individual incomes. It is well known that the geometric mean is different from the arithmetic mean, in particular satisfying

$$Z_{n,t} = \frac{1}{n}\sum_{i=1}^{n} Y_{i,t} > G_{n,t}.$$

Note, however, that in terms of growth rates, use of these measures should be approximately equivalent. In fact, writing $Z_{n,t} = r_{n,t}G_{n,t}$ for some $1 < r_{n,t} < \infty$ yields, setting $z_{n,t} = \log Z_{n,t}$,

$$z_{n,t} - z_{n,t-1} = g_{n,t} - g_{n,t-1} + \log\frac{r_{n,t}}{r_{n,t-1}} \approx g_{n,t} - g_{n,t-1}$$

whenever $r_{n,t}$ is approximately constant across time as also the result of the previous section shows as well as Figure 5.

In our application n is roughly $1,000$ and $g_{n,t}$ will be well approximated by the limit aggregate g_t. Indeed, under (18)

$$g_{n,t} \to g_t \quad \text{as } n \to \infty$$

in mean square. Under (18) the acf of the g_t satisfies

$$\text{cov}(g_t, g_{t+u}) \sim c\,u^{-1.1} \quad \text{as } u \to \infty.$$

Therefore, the aggregate appear to be a stationary process around a deterministic trend. Technically, the aggregate displays short memory. However, the acf decays toward zero as a power law, a case of *quasi long memory* and thus markedly different from the behaviour of the individual income processes. This type of behaviour mimics extremely well a *pure long memory* process and, as a consequence, a unit root process.

Moreover, we estimate the memory parameter using now the NIPA data, based on $z_{n,t} = \log Z_{n,t}$. Although, as indicated above, there is no perfect correspondence between the $z_{n,t}$ and the $g_{n,t}$, the estimate of the memory parameter appear unchanged, with

$$\text{cov}(z_{n,t}, z_{n,t+u}) \sim c\,u^{-1.2} \quad \text{as } u \to \infty.$$

with standard deviation of the estimate of 0.24. Finally, note that under (18), the idiosyncratic component vanishes (in mean square) at the aggregate level.

7 Consumption smoothing

Aggregate consumption is smoother than GNP. This is evident looking at the figures on the second and eighth row of Table 5. This stylized fact was the principal motivation for the formulation of Friedman's permanent income hypothesis (PIH). However, the relationship between consumption and income fluctuations is far from obvious. In fact, based on the modern version of the PIH (see Sargent (1978)), it turns out that the model can predict a greater volatility of aggregate consumption than the one of aggregate income! Therefore, the excess smoothness (ES) phenomenon would then be in conflict with the theory, justifying the denomination of excess smoothness of aggregate consumption. The prediction of the PIH model depends entirely on the dynamic properties of the aggregate income process: smoothness of consumption will be excessive if the income process has a unit-root process and not so if the income process is stationary (see Campbell and Deaton (1989)). The views, empirically motivated, that consumption is smoother than income and that the latter is well described by a unit root process represent a serious challenge to the empirical importance of the PIH model. Therefore, the perfect information model has been suitably modified by means of different types of frictions which attenuate the model-implied volatility of aggregate consumption (see Attanasio (1999)). Pischke (1995) allows for heterogeneity of the income innovations but some frictions are needed in order to attenuate the excess smoothness phenomenon. In this paper we re-examine the PIH model but allowing for complete heterogeneity of both innovations and parameters. It turns out that the latter source of heterogeneity permits to reconcile the ES, in the sense that, thanks to the aggregation effect, the model-driven aggregate consumption is smoother than aggregate income. More importantly, this result holds within the basic, full information, framework. Introducing some frictions might exacerbate the result though. The implications of parameters' heterogeneity for Goodfriend (1992) and Pischke (1995) versions of the PIH model are reported in Appendix B.

We maintain the assumption that individual income is described by the dynamic factor model introduced in section 5:

$$y_{it} = \frac{\gamma_{0i} + \gamma_{1i}L}{1-\alpha_i L}u_t + \frac{1+\delta_i L}{1-\alpha_i L}\epsilon_{it} = y^u_{it} + y^\epsilon_{it}, \qquad (19)$$

with

$$\mid \alpha_i \mid < 1 \;\; a.s.$$

The idiosyncratic component vanishes (in mean square) whenever $\beta > -1/2$ (see Zaffaroni (2003)). When first-differencing, however, this holds for any shape of $f(\alpha)$:

$$\frac{1}{n}\sum_{i=1}^{n}(y^\epsilon_{it} - y^\epsilon_{it-1}) \rightarrow_2 0 \;\; as \;\; n \rightarrow \infty.$$

Therefore, for any shape of the distribution of the α_i, the model implied for the first-differenced aggregate income is

$$\frac{1}{n}\sum_{i=1}^{n}\Delta y_{it} \rightarrow_2 \Delta Y_t = \nu(L)\epsilon_t$$

setting $\nu(L) = \sum_{k=0}^{\infty} \nu_k L^k$ where

$$\nu_0 = E\gamma_{0i}, \;\; \nu_1 = E(\gamma_{0i}(\alpha_i - 1) + \gamma_{1i}), \;\; \nu_k = E\,(\gamma_{0i}\alpha_i + \gamma_{1i})(\alpha_i - 1)\alpha_i^{k-2}, \;\; k \geq 2.$$

Agents can distinguish between common and idiosyncratic component. Therefore, applying the modern PIH (see Campbell and Deaton (1989, eq (1))) to each of the two components yields

$$\Delta c_{it} = \frac{r}{1+r}\left(\frac{(1+r)\gamma_{0i} + \gamma_{1i}}{1+r-\alpha_i}\right)u_t + \frac{r}{1+r}\left(\frac{(1+r+\delta_i}{1+r-\alpha_i}\right)\epsilon_{it}.$$

Whenever $r > 0$

$$\frac{1}{n}\sum_{i=1}^{n}\Delta c_{it} \rightarrow_2 \Delta C_t = \nu(\frac{1}{1+r})u_t.$$

In general (see Zaffaroni (2003))

$$E(\frac{\alpha_i}{1+r})^k = (\frac{1}{1+r})^k E\alpha_i^k \sim c(\frac{1}{1+r})^k\, k^{-(\beta+1)}, \quad \text{as } k \rightarrow \infty.$$

Therefore, condition $r > 0$ ensures that

$$\left|\nu(\frac{1}{1+r})\right| < \infty, \qquad (20)$$

although the left-hand side of (20) will be larger than for the case of homogeneous parameters. However, when $r = 0$ then sum of the $E\alpha_i^k$ might not be summable anymore. Of course, a compensation occurs due to the terms $r/(1+r)$ which will tend to make the π_k arbitrarily small. Indeed, one can easily show that this interest rate effect is dominant, compared with the aggregation effect. In fact, considering the simple case of independent parameters,

$$\nu(\frac{1}{1+r}) = \frac{r}{1+r}\left(E\gamma_{0i} + \frac{E\gamma_{1i}}{1+r}\right)\sum_{k=0}^{\infty}\frac{E\alpha_i^k}{(1+r)^k}.$$

For

$$\left(\frac{1}{1+r}\right)^k \sim \cos(k\sqrt{r}) \quad \text{as } r \to 0^+$$

then

$$\sum_{k=0}^{\infty}\frac{E\alpha_i^k}{(1+r)^k} \sim c\, r^{\beta} \text{ as } r \to 0^+.$$

Thus

$$\text{var}(\Delta C_t) \sim c\, r^{2(1+\beta)} \text{ as } r \to 0^+, \tag{21}$$

equal to zero in the limit, recalling that it must be $\beta + 1 > 0$.

Aggregate consumption volatility is

$$\text{var}(\Delta C_t) = \sigma_\epsilon^2\left(\nu(\frac{1}{1+r})\right)^2$$

and aggregate income volatility

$$\text{var}(\Delta Y_t) = \sigma_\epsilon^2\left(\sum_{k=0}^{\infty}\nu_k^2\right)$$

Whenever

$$\frac{\left(\nu(\frac{1}{1+r})\right)^2}{(\sum_{k=0}^{\infty}\nu_k^2)} < 1 \tag{22}$$

then there would be no issue of ES, since the model would predict that in fact consumption is smoother than income, in line with the empirical evidence.

Table 5 reports the results of some numerical examples, to assess the extent to which (22) is verified, allowing for coefficients' heterogeneity. We assume that the γ_{0i}, γ_{1i} are independent from the α_i which are distributed according to the Beta distribution $f_2(\alpha)$ (eq. (9)) with mean $E(\alpha_i) = \mu$. Using (5), it easily follows that

$$\text{var}(\alpha_i) = \frac{\mu(1-\mu)}{q(\frac{1}{(1-\mu)} + \frac{1}{q})},$$

converging towards zero as $q \to \infty$, hence yielding $\alpha_i = \mu$ *a.s.*. Therefore, taking q large we are able to obtain the homogeneous parameters case as a special case of our set-up.

Table 5 shows the result obtained for various shape of the cross-sectional distribution of the α_i and

Table 5 permits to evaluate the effects of parameters' heterogeneity, compared with the case where homogeneous parameters case. This last case is represented by the last column, corresponding to $q = 100$. In this case the distribution of the α_i is (nearly) degenerate around $E(\alpha_i) = \mu$. The classical PIH model with (homogeneous) stationary ARMA income processes and unit-root income processes correspond, respectively, to the first and second portions of the table ($\mu = 0.8$ and 0.95) and to the last portion of the table ($\mu = 0.999$). The standard deviations ratios here obtained mimic closely the values reported in the literature under these two hypothesis (see, for example, Deaton (1992, p.111)). All the other columns correspond to the heterogeneous case. Recall that the smaller is q, the more dense is the distribution of the α_i around unity. For q below 0.5 the (log) income processes is nonstationary and for q between 0.5 and 1 is stationary but with long memory. Equation (22) is nevertheless always well defined since the ν_k are the MA coefficients of the first-differences (log) income, always stationary for any values of q. In all cases, aggregation changes the impulse response of the income processes as q diminishes and, as a consequence, the standard deviations ratio increases.

It turns out that for μ equal to 0.8 and 0.95 the standard deviations ratio is always below one, although increasing with r and with $1/q$. Only for μ very close to unity ($\mu = 0.999$), the model predicts that consumption is more volatility than income in most cases. The intuition underlying the result works as follows. Aggregation induces a substantial change in the

Table 5:

Ratio of consumption volatility versus income volatility

			$f_2(\alpha)$					
r	$E(\gamma_{1,i})$	$q =$	0.2	0.3	0.7	1	3	100
			$\mu = 0.800$					
.001	0.5		0.35	0.22	0.05	0.03	0.01	0.01
.01	0.5		0.55	0.42	0.21	0.15	0.08	0.06
.05	0.5		0.73	0.64	0.46	0.39	0.29	0.24
.001	1		0.38	0.23	0.06	0.03	0.01	0.01
.01	1		0.59	0.45	0.22	0.17	0.08	0.06
.05	1		0.77	0.67	0.48	0.41	0.30	0.25
			$\mu = 0.950$					
.001	0.5		0.50	0.36	0.14	0.09	0.04	0.03
.01	0.5		0.77	0.66	0.46	0.39	0.27	0.22
.05	0.5		0.97	0.91	0.80	0.76	0.69	0.64
.001	1		0.53	0.38	0.14	0.09	0.04	0.03
.01	1		0.81	0.69	0.48	0.41	0.29	0.23
.05	1		1.01	0.95	0.83	0.79	0.72	0.66
			$\mu = 0.999$					
.001	0.5		1.01	0.95	0.84	0.79	0.72	0.67
.01	0.5		1.24	1.23	1.22	1.22	1.21	1.21
.05	0.5		1.28	1.27	1.27	1.27	1.26	1.26
.001	1		1.06	0.99	0.88	0.84	0.76	0.71
.01	1		1.31	1.29	1.28	1.28	1.27	1.27
.05	1		1.32	1.32	1.31	1.31	1.31	1.30
			$\mu = 0.600$					
.001	0.04		0.23	0.13	0.03	0.11	0.01	0.01
.01	0.04		0.37	0.27	0.11	0.75	0.04	0.03
.05	0.5		0.51	0.43	0.27	0.22	0.15	0.12

The table reports the standard deviations ratio

$$\frac{\left|\nu(\frac{1}{1+r})\right|}{\left(\sum_{k=0}^{\infty}\nu_k^2\right)^{\frac{1}{2}}}$$

for several combination of the parameters. We set $E\gamma_{0,i} = 1$. for all cases but case $\mu = 0.6$ where we set $E\gamma_{0,i} = 0.1$.
$f_2(\alpha)$ denotes the Beta density with parameter q (and eq.(5)).

impulse response of the aggregate income process, as indicated in section 3, with respect to the exponential decaying impulse response of the individual income processes. Namely, the ν_k decay hyperbolically, as $k^{-(q+1)}$, for k going to infinity. On the other hand, the PIH implies that the π_k decay as $1/(1+r)^k\, k^{-q}$ where the first, exponential factor, dominates for $r > 0$. Therefore, the effect of aggregation on consumption dynamics is attenuated by the very nature of the model. Hence, even though parameter's heterogeneity does influence the predicted volatility ratio, the PIH model pins down the behaviour of aggregate consumption tightly.

The results are very sensitive to the assumptions made on r. The most reasonable values for the real interest rate are $r = 0.001$, virtually zero, and $r = 0.01$, resembling values used in the literature. In contrast, changes of $E\gamma_{1,i}$ have a negligible effect, and so do changes $E\gamma_{0,i}$, not reported for sake of simplicity.

Allowing for heterogeneity represents an intermediate case between the homogeneous stationary case ($\alpha_i = \alpha$, smaller than one in absolute value) and the homogeneous unit root case ($\alpha_i = 1$). In the latter case, maintaining heterogeneity of the MA coefficients,

$$\mathrm{var}(\Delta C_t) = \sigma_\epsilon^2 \left(E\gamma_{0i} + E\frac{\gamma_{1i}}{1+r} \right)^2,$$

and

$$\mathrm{var}(\Delta Y_t) = \sigma_\epsilon^2 \left(E^2\gamma_{0i} + E^2\gamma_{1i} \right).$$

Therefore, the convergence of r to zero does have only a negligible effect on aggregate consumption dynamics. Indeed, when $\gamma_{0i}\gamma_{1i} > 0$ $a.s.$ then aggregate consumption will certainly be more volatile than aggregate income. This effect is only mitigated in our case (cf. (21)), since $q = \beta + 1 > 0$, but not eliminated. In the homogeneous and stationary parameter case, instead, the interest rate effect dominates and $\mathrm{var}(\Delta C_t)$ converges to zero fast, with r^2.

The last panel of Table 5, corresponding to $\mu = 0.6$, shows the standard deviation ratio for parameter values that match estimated values of the median of the corresponding estimated cross-sectional distribution. Note the dramatic effect of parameter heterogeneity, with an implausibly small ratio for the homogeneous case. Note, though that the ratio is always smaller than the aggregate estimate for plausible values of the real interest rate.

8 Conclusion

This paper provides a framework able to estimate the relationship between aggregate dynamics and individual heterogeneity. In particular, we are interested in the mapping between the shape and degree of heterogeneity and the degree of memory of the aggregate. When applying such framework to data on individual income (PSID) and aggregate income (NIPA), we conclude that given the observed structure of the individual data, aggregate per capita income is well described by a trend stationary process (with a deterministic trend). The unit root proposition is highly rejected by the data.

We also explore the implications of individual heterogeneity for the consumption smoothness phenomenon.

Tables

	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
1971	**0.24**	0.36	0.12	-0.02	-0.08	-0.06	-0.14	-0.15	-0.20	-0.16	-0.16	-0.12	-0.06	-0.06	-0.13	-0.11	-0.11	0.01	0.00	-0.05
1972	0.08	**0.22**	0.38	0.19	0.09	0.10	-0.04	-0.08	-0.17	-0.13	-0.22	-0.17	-0.18	-0.21	-0.25	-0.15	-0.09	-0.09	-0.03	-0.05
1973	0.02	0.07	**0.17**	0.55	0.20	0.20	0.07	0.04	-0.11	-0.17	-0.18	-0.25	-0.23	-0.23	-0.26	-0.16	-0.19	-0.13	-0.04	-0.10
1974	0.00	0.03	0.08	**0.13**	0.32	0.27	0.16	0.09	0.01	-0.03	-0.11	-0.08	-0.17	-0.20	-0.19	-0.21	-0.23	-0.13	-0.09	-0.12
1975	-0.02	0.02	0.03	0.05	**0.16**	0.32	0.19	0.12	0.10	0.02	-0.08	-0.05	-0.12	-0.21	-0.17	-0.16	-0.26	-0.15	-0.10	-0.06
1976	-0.01	0.02	0.03	0.03	0.04	**0.12**	0.40	0.13	0.09	0.00	-0.03	-0.05	-0.14	-0.14	-0.15	-0.22	-0.22	-0.22	-0.13	-0.20
1977	-0.02	-0.01	0.01	0.02	0.03	0.05	**0.12**	0.37	0.25	0.07	0.04	-0.08	-0.11	-0.16	-0.15	-0.25	-0.17	-0.19	-0.17	-0.17
1978	-0.03	-0.01	0.01	0.01	0.02	0.02	0.05	**0.13**	0.43	0.13	0.13	-0.04	-0.17	-0.20	-0.16	-0.23	-0.21	-0.13	-0.16	-0.11
1979	-0.03	-0.03	-0.01	0.00	0.01	0.01	0.03	0.05	**0.11**	0.42	0.27	0.06	-0.10	-0.18	-0.19	-0.21	-0.22	-0.13	-0.15	-0.13
1980	-0.03	-0.02	-0.02	0.00	0.00	0.00	0.01	0.02	0.05	**0.11**	0.43	0.29	-0.03	-0.11	-0.14	-0.16	-0.23	-0.21	-0.19	-0.09
1981	-0.02	-0.03	-0.02	-0.01	-0.01	0.00	0.00	0.01	0.03	0.04	**0.09**	0.47	0.03	-0.03	-0.08	-0.21	-0.16	-0.27	-0.23	-0.12
1982	-0.02	-0.03	-0.04	-0.01	-0.01	-0.01	-0.01	0.00	0.01	0.03	0.05	**0.11**	0.19	0.02	-0.02	-0.17	-0.16	-0.17	-0.09	-0.05
1983	-0.01	-0.04	-0.04	-0.03	-0.02	-0.02	-0.02	-0.03	-0.01	0.00	0.00	0.03	**0.17**	0.33	0.16	0.02	-0.01	-0.11	-0.13	-0.11
1984	-0.01	-0.04	-0.04	-0.03	-0.03	-0.02	-0.02	-0.03	-0.02	-0.01	0.00	0.00	0.05	**0.15**	0.36	0.17	0.13	-0.05	-0.11	-0.06
1985	-0.02	-0.05	-0.04	-0.03	-0.03	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	0.00	0.03	0.06	**0.16**	0.44	0.17	-0.04	0.01	-0.04
1986	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.03	-0.03	-0.02	-0.03	-0.02	0.00	0.03	0.07	**0.16**	0.35	0.20	0.12	0.02
1987	-0.02	-0.02	-0.03	-0.03	-0.04	-0.03	-0.02	-0.03	-0.03	-0.03	-0.02	-0.02	0.00	0.02	0.03	0.06	**0.17**	0.35	0.10	0.06
1988	0.00	-0.01	-0.02	-0.02	-0.02	-0.03	-0.02	-0.02	-0.01	-0.03	-0.03	-0.02	-0.02	-0.01	-0.01	0.03	0.05	**0.13**	0.38	0.16
1989	0.00	-0.01	-0.01	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.02	-0.03	-0.01	-0.02	-0.02	0.00	0.02	0.02	0.05	**0.14**	0.31
1990	-0.01	-0.01	-0.01	-0.02	-0.01	-0.02	-0.02	-0.01	-0.02	-0.01	-0.01	-0.01	-0.02	-0.01	-0.01	0.00	0.01	0.02	0.04	**0.13**

Table 1 - Covariance and Correlation between Income: PSID 1971-1990.

		Income	Consumption	GNP	Income per cap	Consumption per cap	GNP per cap
delta-log	mean	0.0305	0.0347	0.0332	0.0177	0.0219	0.0204
	std	0.0257	0.0174	0.0242	0.0260	0.0180	0.0245
	ρ_1	0.1682	0.1696	-0.0028	0.1588	0.1996	-0.0015
	ρ_2	0.0358	-0.0713	-0.0378	0.0257	-0.0442	-0.0366
	ρ_3	-0.1438	-0.1158	-0.1893	-0.1582	-0.0923	-0.1938
	ρ_4	-0.0377	-0.0691	-0.0860	-0.0609	-0.0560	-0.0963
	ρ_5	0.0828	-0.0185	0.0454	0.0482	-0.0457	0.0249
log-detr.	std	0.0373	0.0283	0.0299	0.0387	0.0323	0.0309
	ρ_1	0.6306	0.7688	0.6576	0.6346	0.8100	0.6638
	ρ_2	0.3786	0.4916	0.3487	0.3951	0.5815	0.3764
	ρ_3	0.1737	0.2723	0.0721	0.1957	0.3880	0.1082
	ρ_4	0.1337	0.1264	-0.0307	0.1547	0.2399	0.0078
	ρ_5	0.0643	-0.0067	-0.1000	0.0708	0.0846	-0.0780

Table 2 - Aggregate data descriptive statistics
(NIPA source, chained 1996 dollars annual ranging from 1946 to 2001)

	Γ_0
λ_1	*0.14*
λ_2	*0.10*
λ_3	*0.08*
λ_4	*0.07*
λ_5	*0.06*
λ_6	*0.05*
λ_7	*0.05*
λ_8	*0.04*
λ_9	*0.03*
λ_{10}	*0.03*

Table 3 - Standardized eigenvalues associated to variance-covariance matrix

	α	γ_0	γ_1	δ	σ_ϵ
mean	*0.356*	*0.072*	*0.025*	*0.078*	*0.433*
median	*0.529*	*0.035*	*0.011*	*0.055*	*0.277*
std	*0.536*	*0.194*	*0.148*	*0.583*	*0.432*
mean	*0.430*	*0.075*	*0.026*	*0.047*	*0.489*
median	*0.616*	*0.040*	*0.013*	*0.072*	*0.275*
std	*0.559*	*0.185*	*0.130*	*0.489*	*0.593*

Table 4 - Moments of the empirical distribution of micro parameters

Figures

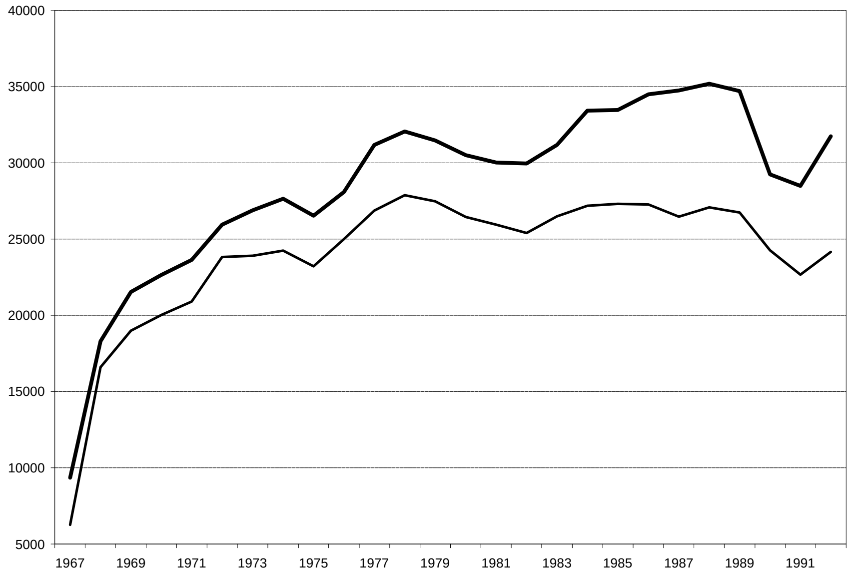

Figure 1 - Mean (bold) and median of real income.

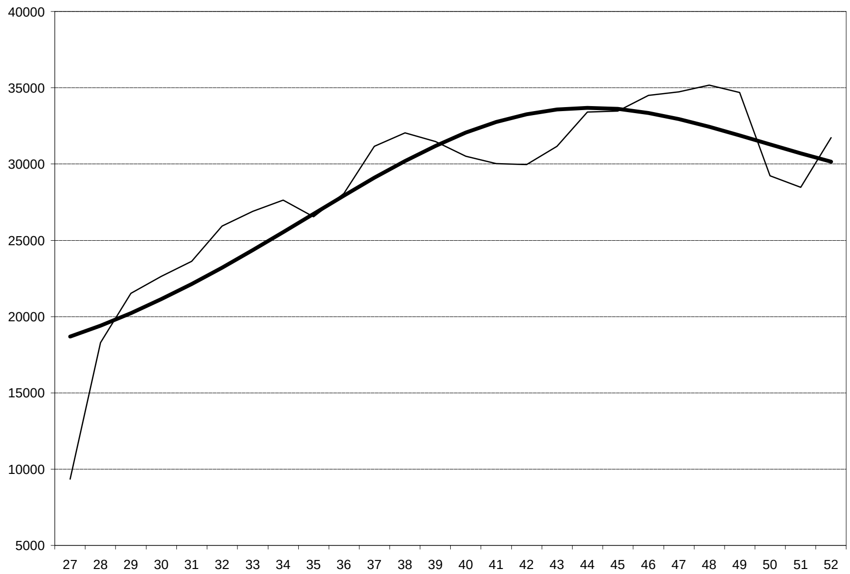

Figure 2 - Estimated deterministic component as function of average age.

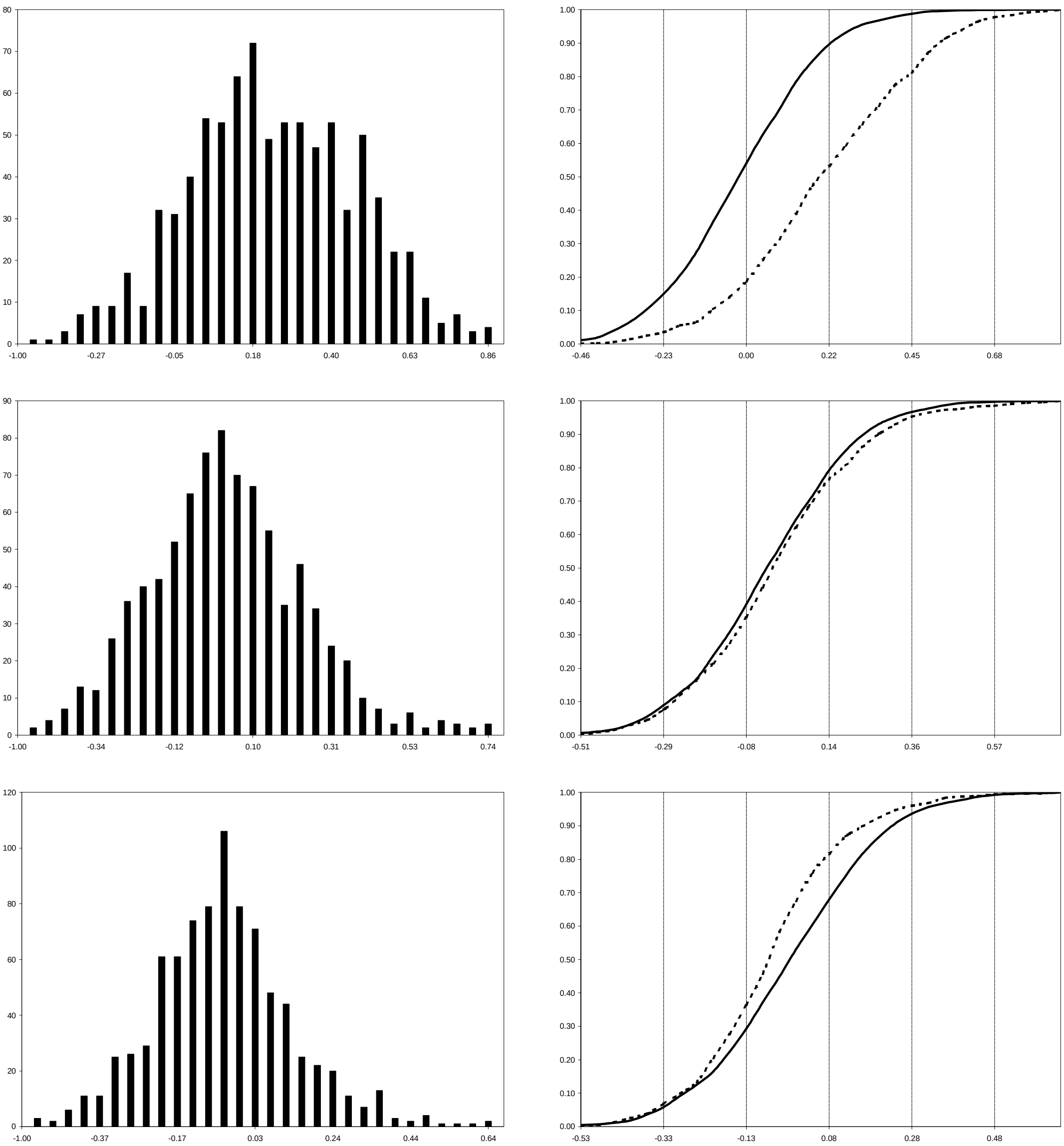

Figure 3 - Histogram of autocorrelations of agent i at lag 1-3 and their cdf versus the cdf of the test of the zero autocorrelation.

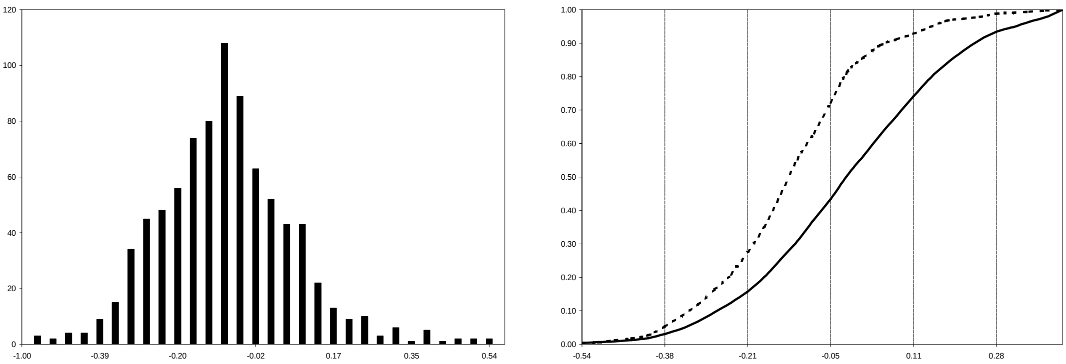

Figure 3 - Histogram of autocorrelations of agent i at lag and their cdf versus the cdf of the test of the zero autocorrelation.

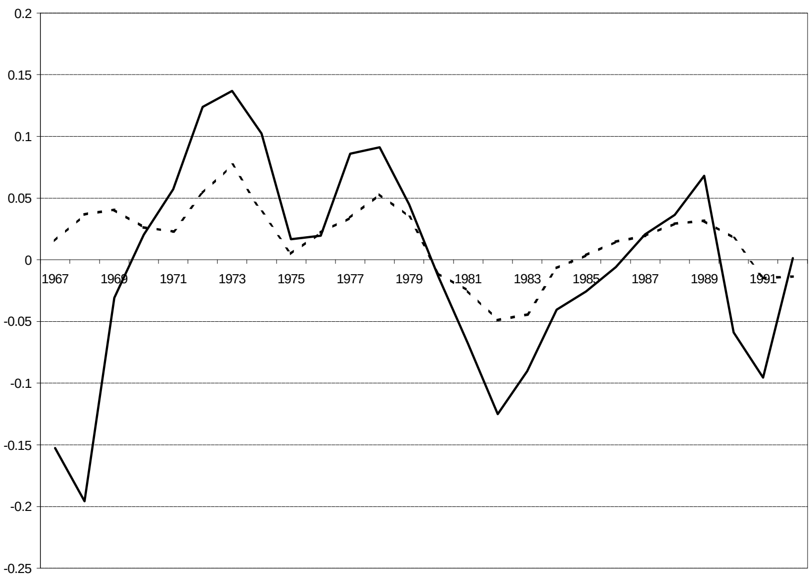

Figure 4 - Average of the PSID income process and NIPA aggregate (detrended).

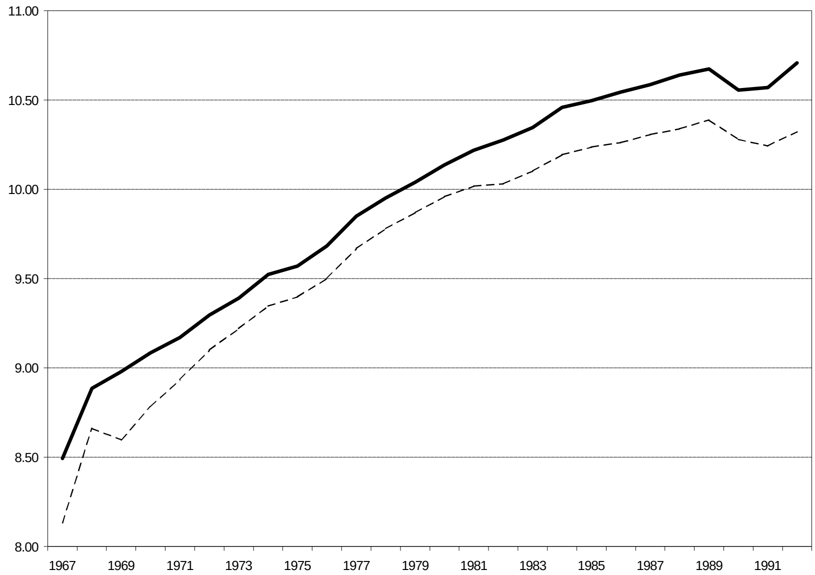

Figure 5 - Log of the average individual PSID income and average of the

logs.

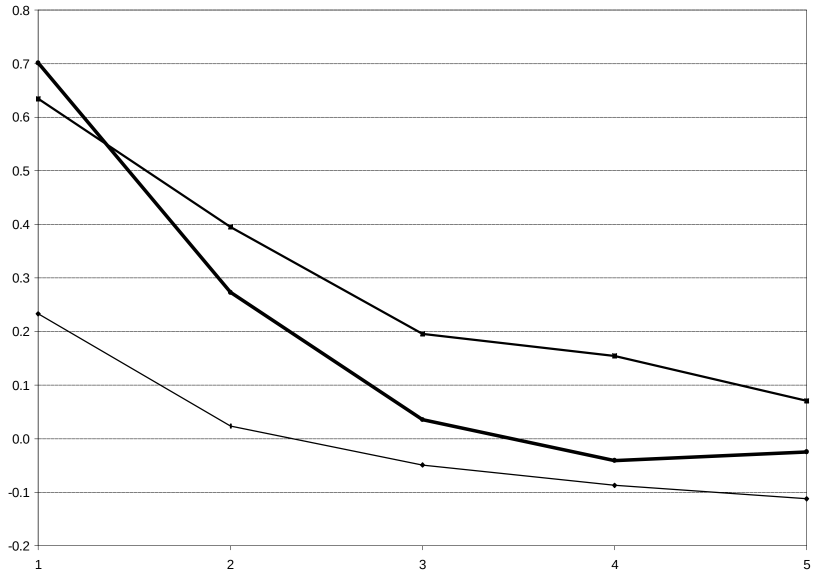

Figure 6 - Autocorrelation of the aggregate Y, avegage of the agent's autocorrelations and autocorrelation of the NIPA per capita personal income (quadratic detrended).

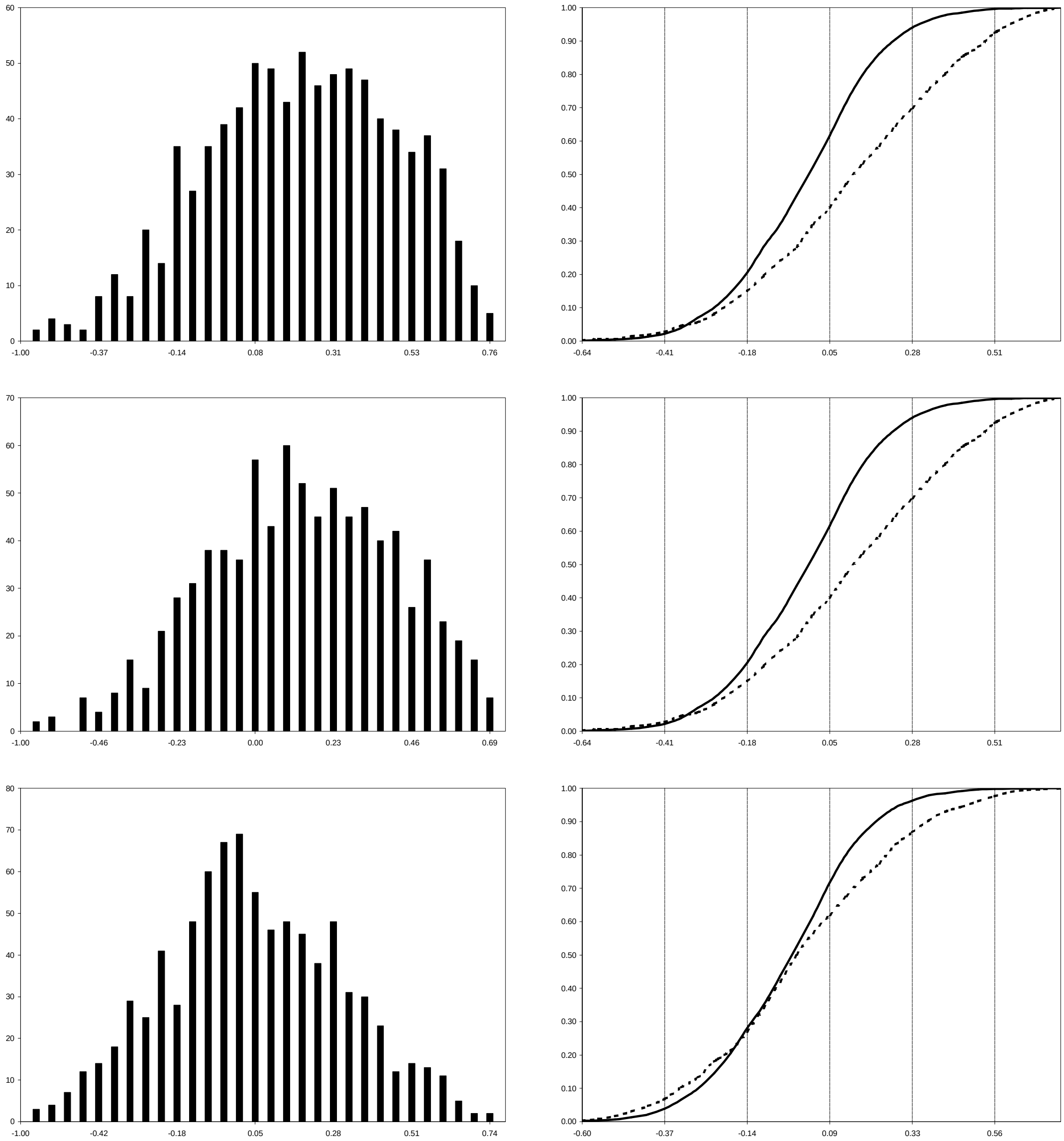

Figure 7 - Histogram of correlations of agent i with aggregate at lag 0-2 and their cdf versus the test of zero autocorrelation.

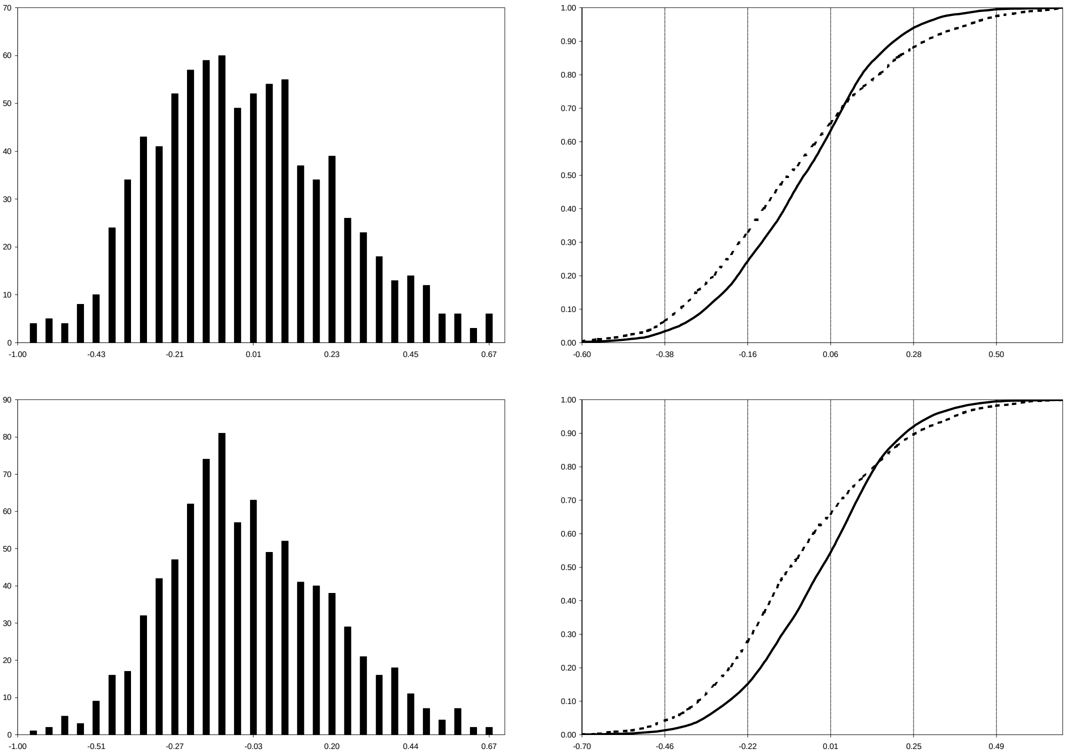

Figure 7 - Histogram of correlations of agent i with aggregate at lag 3-4 and their cdf versus the cdf of the test of the zero autocorrelation.

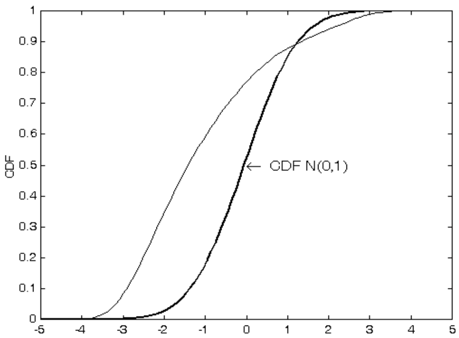

Fig. 8.1: T_0 on the PSID data

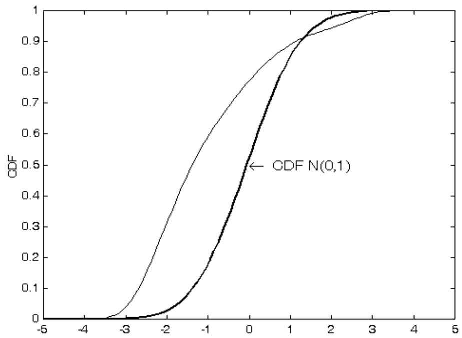

Fig. 8.2: T_1 on the PSID data

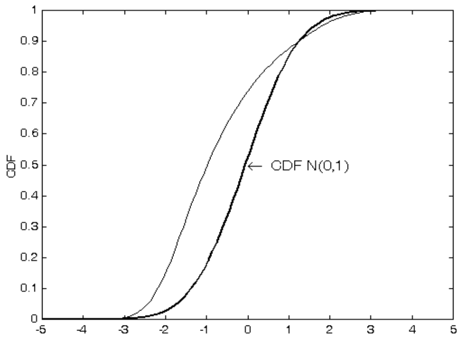

Fig. 8.3: T_2 on the PSID data.

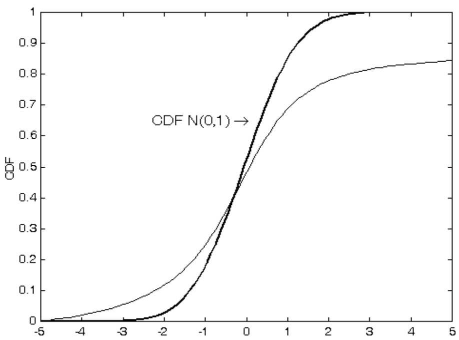

Fig. 8.4: R_1 on the PSID data

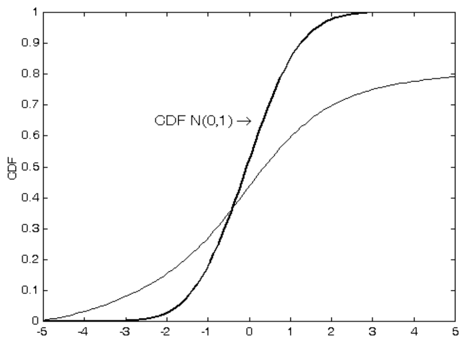

Fig. 8.5: R_2 on the PSID data

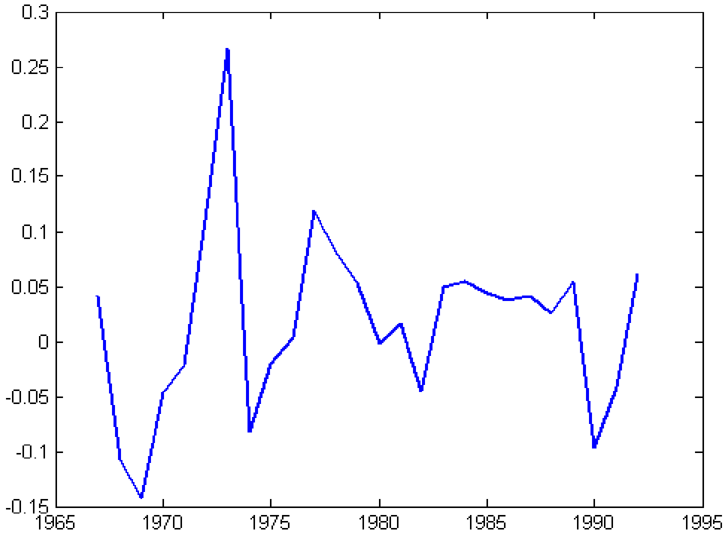

Figure 9 - Common shock u_t:

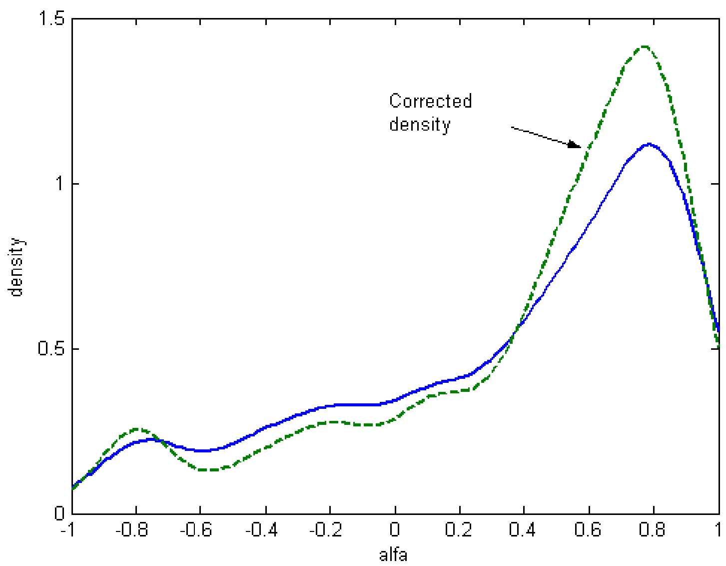

Figure 10 - Distribution of ®:

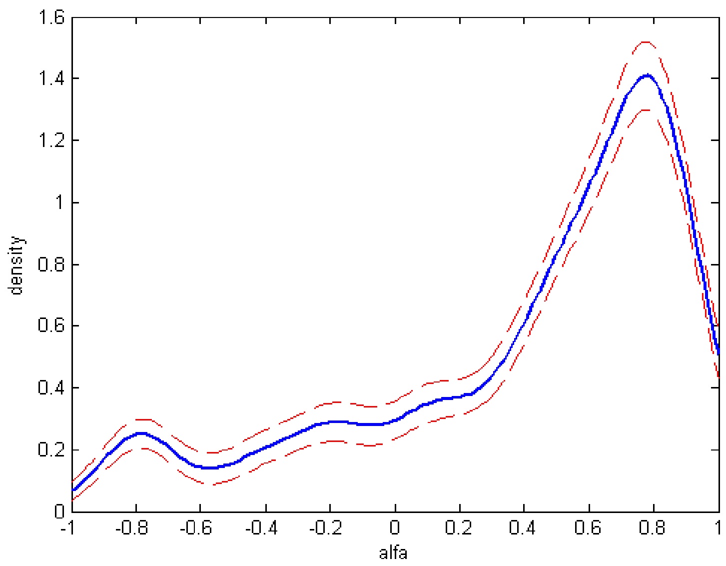

Figure 11 - Corrected density of ® and 90% con¯dence bands.

9 Appendix A - Tests of heterogeneity

Let us consider the model under the null as:

$$M_0: \;\; x_{it} = \Psi(L)u_t + \xi_{it} = f_t + \xi_{it} \; with \; i = 1, ..., n$$

with ξ_{it} a stationary idiosyncratic component such that $u_t \perp \xi_{is} \; \forall t, s$ and u_t is a $iid(0, \sigma_u^2)$ and $\Psi(L)$ is squared summable. The model under the null of the test is the one which imposes the homogeneity in the propagation of the commons shocks. Under the altervative the propagation of the common shock u_t is different across agents as:

$$M_{11}: \;\; x_{it} = \Psi_i(L)u_t + \xi_{it} = f_{i,t} + \xi_{it}.$$

Also we consider the following modification of the main alternative M_{11} as:

$$M_{12}: \;\; x_{it} = \alpha_i \Psi(L)u_t + \xi_{it} = \alpha_i f_t + \xi_{it}$$

which allows for different impact effect of the aggregate shocks but equal dynamic over time; clearly $M_{12} \subset M_{11}$.. Assume that conditions hold such that:

$$\frac{1}{n}\sum_{i=1}^{n} x_{i,t} = X_{n,t} \rightarrow_2 X_t \;\; as \;\; n \rightarrow \infty.$$

which their are automatically fulfill under M_0. Consider the contemporaneous cross-covariance of the income process of each agent i with the aggregate mean $X_{n,t}$:

$$\hat{h}_{i,0} = \frac{1}{T}\sum_{t=1}^{T} X_{n,t}(x_{i,t} - \overline{x}_i),$$

and the lag τ cross-covariance correspondent:

$$\hat{h}_{i,\tau} = \frac{1}{T}\sum_{t=1}^{T} X_{n,t}(x_{i,t+\tau} - \overline{x}_i).$$

Under specification M_0 :

$$\hat{h}_{i,\tau} = \frac{1}{T}\sum_{t=1}^{T} f_t(f_{t+\tau} + \xi_{i,t+\tau} - \bar{f} - \overline{\xi}_{i,.}) + \frac{1}{T}\sum_{t=1}^{T} \overline{\xi}_{.,t}(f_{t+\tau} + \xi_{i,t+\tau} - \bar{f} - \overline{\xi}_{i,.}),$$

and

$$\begin{aligned}
\bar{\hat{h}}_{.,\tau} &= \frac{1}{n}\sum_{i=1}^{n}\hat{h}_{i,\tau} \\
&= \frac{1}{T}\sum_{t=1}^{T} f_t(f_{t+\tau}+\bar{\xi}_{.,t+\tau}-\bar{f}-\bar{\bar{\xi}}_{.,.})+\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{.,t}(f_{t+\tau}+\bar{\xi}_{.,t+\tau}-\bar{f}-\bar{\bar{\xi}}_{.,.}),
\end{aligned}$$

setting

$$\bar{f}=\frac{1}{T}\sum_{t=1}^{T} f_t,\ \ \bar{\xi}_{.,t}=\frac{1}{n}\sum_{i=1}^{n}\xi_{i,t},\ \ \bar{\bar{\xi}}_{.,.}=\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{.,t},\ \text{and}\ \bar{\xi}_{i,.}=\frac{1}{T}\sum_{t=1}^{T}\xi_{i,t}.$$

As $n\to\infty$

$$\begin{aligned}
\bar{\hat{h}}_{.,\tau} &\overset{n}{\to} \frac{1}{T}\sum_{t=1}^{T}(f_t-\bar{f})(f_{t+\tau}-\bar{f}), \\
\hat{h}_{i,\tau} &\overset{n}{\to} \frac{1}{T}\sum_{t=1}^{T}(f_t-\bar{f})(f_{t+\tau}-\bar{f}+\xi_{i,t+\tau}).
\end{aligned}$$

Note first that the mean of $\hat{h}_{i,\tau}$ over i is the same both under the model M_0 and M_{11}, so the mean over i of the covariance does not allow to discriminate between the two specification. The variance of the estimate of the cross covariance of the agents i with the aggregate under the model M_0 is given by given that $f_t \perp \xi_{is}\ \forall t,s$:

$$\begin{aligned}
E(\hat{h}_{i,\tau}-\bar{\hat{h}}_{.,\tau})^2 &\overset{n}{\to} E\left(\frac{1}{T^2}\sum_{t,s}^{T}(f_t-\bar{f})(f_s-\bar{f})\xi_{i,t+\tau}\xi_{i,s+\tau}\right) \\
\simeq\ \frac{1}{T}\sum_{\tau=-\infty}^{\infty}\gamma_f(\tau)\gamma_{\xi_i}(\tau) \overset{T}{\to} 0.
\end{aligned}$$

We are interested in the case n large and T small. Thus, as $n\to\infty$, we can consider the statistic

$$\begin{aligned}
T_{i,\tau} &= \frac{\hat{h}_{i,\tau}-\bar{\hat{h}}_{.,\tau}}{\left(\frac{1}{T^2}\sum_{t,s}^{T}(f_t-\bar{f})(f_s-\bar{f})\xi_{i,t+\tau}\xi_{i,s+\tau}\right)^{\frac{1}{2}}} \\
&\overset{n}{\to} \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t-\bar{f})\xi_{i,t+\tau}}{\left(\frac{1}{T}\sum_{t,s}^{T}(f_t-\bar{f})(f_s-\bar{f})\xi_{i,t+\tau}\xi_{i,s+\tau}\right)^{\frac{1}{2}}}
\end{aligned}$$

which under model M_0 is asymptotically equivalent to:

$$\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t - \bar{f})\xi_{i,t+\tau}}{\sqrt{\sum_{\tau=-\infty}^{\infty}\gamma_f(\tau)\gamma_{\xi_i}(\tau)}},$$

clearly $N(0,1)$ for any i (under suitable conditions on f_t) as T goes to infinity. If T is large, we could compute the $T_{i,\tau}$ for each i in the cross-section and, under M_0, they should be considered as draw from the distribution under the null and so they should behave as a standard normal. On the other hand if T is small, we should boostrap but given the autocovariance structure of the ξ we cannot do it in general and so we resort on the use on asymptotic.

The idea is to estimate the test $T_{i,\tau}$ for the different agents i and then construct the empirical distribution of the $T_{i,\tau}$ over i for different τ. Under the null, asymptotically the statistic for each agent i should be a draw from a standard normal and so the distribution over i of the statistics should resemble the one of a standard normal.

Under the alternatives M_{11} and M_{12}, the numerator of the statistic will converge to:

$$\begin{aligned}\hat{h}_{i,\tau} - \bar{\hat{h}}_{.,\tau} &\xrightarrow{n} \frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t^i - \bar{f}^i - f_t + \bar{f} + \xi_{i,t+\tau}) \qquad (23)\\ &\xrightarrow{T} \sum_j (\psi_j^i - \psi_j)\psi_j\sigma_u^2;\end{aligned}$$

while the estimate of $\gamma_{\xi_i}(\tau)$ under the alternative will converge to:

$$\widehat{\gamma}_{\xi_i}(\tau) = \gamma_{\xi_i}(\tau) + \sum_j (\psi_j^i - \psi_j)(\psi_{j+\tau}^i - \psi_{j+\tau})\sigma_u^2.$$

So for each i under the alternative, the test will be distributed normally with mean as in (23) and variance not unitary. The distribution over the i agents of the asymptotic distribution will depend on the cross section distribution of the ψ^i parameters and so would not be standard normal in general.

Finally, the proposed test strategies will have power both versus M_{11} and also with respect to M_{12}. However model M_{12} is not an interesting alternative given that the difference across agents is only in term of impact of the aggregate shock on each agent but not in term of different dynamic. In order to design a test strategy which has power only versus the interesting alternative of different cross section dynamic, we propose the following:

$$S_{i,\tau} = \sqrt{T}\left(\frac{\hat{h}_{i,\tau}}{\hat{h}_{i,0}} - \frac{\bar{\hat{h}}_{\cdot,\tau}}{\bar{\hat{h}}_{\cdot,0}}\right)$$

for $\tau > 1$, under the null $M_0 \cup M_{11}$ the test will be distributed as $N(0, V)$ where

$$V = \frac{1}{\sigma_f^4}\sum_{k=-\infty}^{\infty}\gamma_f(k)\gamma_{\xi_i}(k+\tau).$$

9.1 Small sample performance

A small Montecarlo is performed in order to investigate the small sample properties of the testing strategy. The data generating process has the following specification:

$$M_{12}: \;\; x_{it} = a_i(u_t - \rho_i u_{t-1}) + \epsilon_{it} \text{ with } \epsilon_{it} \sim N(0,1)$$

where $a_i \sim U[0.5, 1.5]$ and $\rho_i \sim U[0, 1.0]$. The restricted model with equal dynamic is:

$$M_{11}: \;\; x_{it} = a_i(u_t - \rho u_{t-1}) + \epsilon_{it} \text{ with } \epsilon_{it} \sim N(0,1),$$

and, under the null, the model reduces to:

$$M_0: \;\; x_{it} = (u_t - \rho u_{t-1}) + \epsilon_{it} \text{ with } \epsilon_{it} \sim N(0,1)$$

with $\rho = 0.5$. We considered a population of 500 agents, $i = 1, ..., 500$ and the sample size is $T = 30$.

The simulation exercise is structured that at each Montecarlo replication we generated x_{it} $\forall i$ and computed the statistic $T_{i,0}$, $T_{i,1}$ and $R_{i,1}$. The statistics are then used to estimate their empirical distribution over i; the figures below reports the average of the cumulative empirical distribution of the three statistics over the Montecarlo replications (500) and its 90% confidence interval versus the standard normal cdf.

In Figure A, the data generating process is M_0; as expected they resemble very closely the asymptotic distribution. In Figure B, the data generating process is M_{11} and the average distribution over i of all three tests is far from the null one; this evidence is particularly true for the autocovariance based test T. Finally in Figure C, the DGP follows M_{12} and as expected the R test does not have power versus this form of heterogeneity which is instead detected in the generality of cases by the T test.

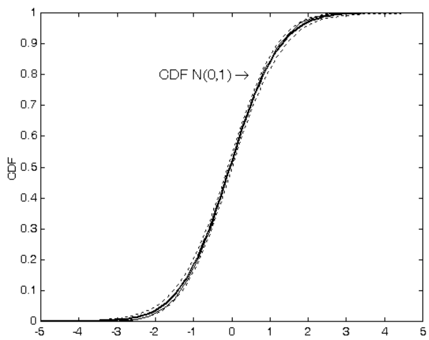

Fig A.1: T_0 statistic under model M_0

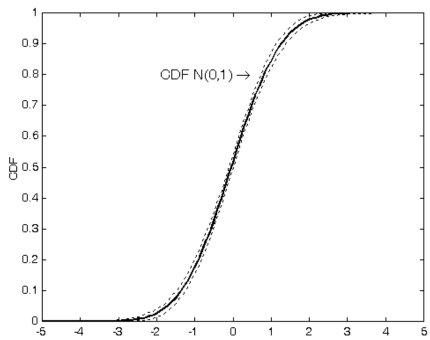

Fig A.2: T_1 statistic under model M_0

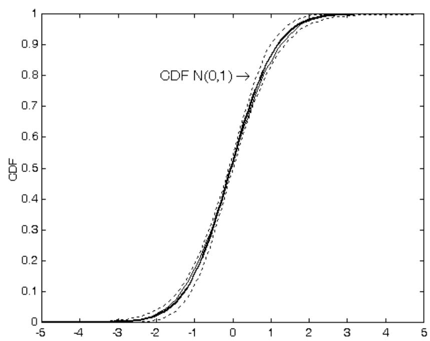

Fig A.3: R_1 statistic under model M_0

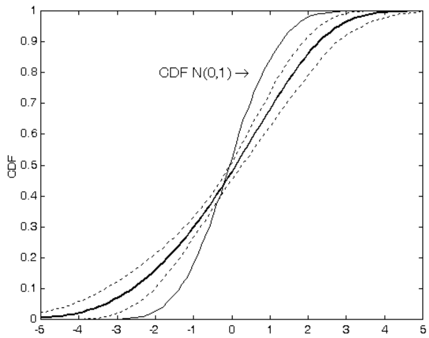

Fig B.1: T_0 statistic under model M_{11}

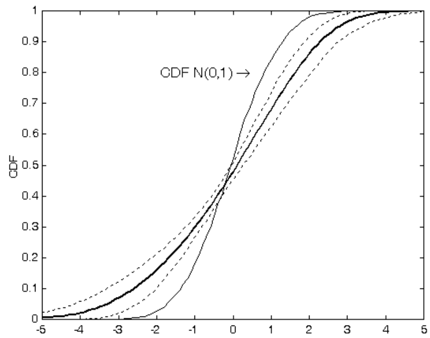

Fig B.3: T_1 statistic under model M_{11}

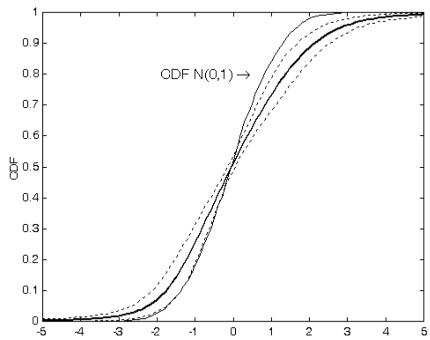

Fig B.3: R statistic under model M_{11}

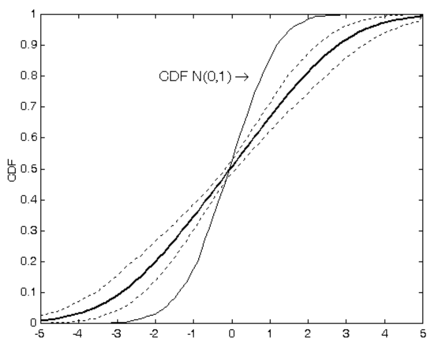

Fig C.1: T_0 statistic under model M_{12}

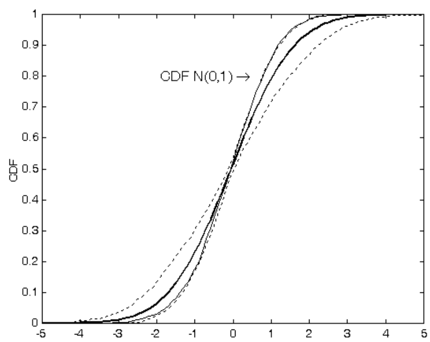

Fig C.2: T_1 statistic under model M_{12}

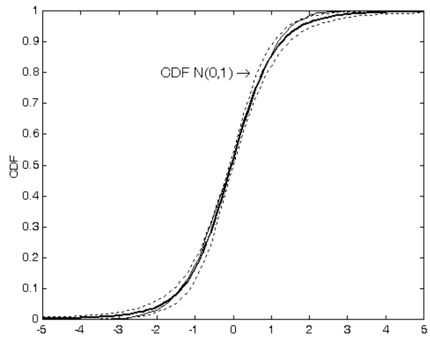

Fig C.3: R_1 statistic under model M_{12}

10 Appendix B - Imperfect information

We now discuss the effect of aggregation for two cases of imperfect information. In particular, we generalize Pischke (1995) and Goodfriend (1992) versions of the PIH.

Unobservable aggregate shock. Pischke (1995) considers the case when that agents are not able to disentangle common from idiosyncratic shocks but observe only y_{it}. This implies that, although the true dynamics of individual income is still given by (19), agents perceive it as

$$\Delta y_{it} = \Phi_i(L)\eta_{it};$$

where Δy_{it} is ARMA(1,2) with respect to the white noise sequence η_{it} and

$$\Phi_i(L) = (1 - L)\frac{(1 + \theta_i L)}{(1 - \alpha_i L)};$$

and the MA(1) coefficient is an exact, nonlinear, function of the $\alpha_i; \gamma_i; \delta_i$, obtainable comparing the acf, at lag 1, of $(1 - L)(1 + \theta_i L)\eta_{it}$ with the one of

$$(1 - L)(\gamma_{0i} + \gamma_{1i} L)\varepsilon_t + (1 - L)(1 + \delta_i L)u_{it}:$$

This can be re-written as

$$(1 - L)(1 + \frac{\gamma_{1i}}{\gamma_{0i}} L)\gamma_{0i}\varepsilon_t + (1 - L)(1 + \delta_i L)u_{it};$$

where the common innovation will now have heterogeneous variance $\sigma_\epsilon^2 \gamma_{0i}^2$. Setting

$$q_i = \frac{\gamma_{0i}^2 \sigma_\epsilon^2}{\sigma_{ui}^2}$$

and

$$\gamma_i = \frac{\gamma_{1i}}{\gamma_{0i}},$$

yields

$$\frac{\theta_i}{1+\theta_i^2} = \frac{q_i \gamma_{0i} + \delta_i}{q_i(1+\gamma_i^2) + (1+\delta_i^2)}.$$

Choosing the stable solution yields

$$\theta_i = \frac{1}{2(q_i \gamma_i + \delta_i)} \left(q_i(1+\gamma_i^2) + (1+\delta_i^2) - \sqrt{(q_i(1+\gamma_i^2) + (1+\delta_i^2))^2 - 4(q_i\gamma_i + \delta_i)^2} \right)$$

It turns out that

$$min[\delta_i, \gamma_i] \leq \theta_i \leq max[\delta_i, \gamma_i].$$

where the extremes are achieved only for

$$\lim_{q_i \to 0^+} \theta_i = min[\delta_i, \gamma_i]$$

and

$$\lim_{q_i \to \infty} \theta_i = max[\delta_i, \gamma_i].$$

Individual consumption follows from

$$\begin{aligned} \Delta c_{it} &= \Pi_i(\frac{1}{1+r})\eta_{it} \\ &= \Pi_i(\frac{1}{1+r}) \left(\Pi_i^{-1}(L)\frac{1+\gamma_i L}{1-\alpha_i L}(\epsilon_t - \epsilon_{t-1}) + \Pi_i^{-1}(L)\frac{1+\delta_i L}{1-\alpha_i L}(u_{it} - u_{it-1}) \right). \end{aligned}$$

Note that what matters is the distribution of the θ_i. But when $0 < q \leq q_i \leq Q < \infty$ the θ_i are always bounded away from one in modulus. For the aggregate, only the part involving the common shock matters, yielding

$$\frac{1}{n}\sum_{i=1}^{n} \Delta c_{it} \to_2 \Delta C_t = \sum_{k=0}^{\infty} \pi_k \epsilon_{t-k}, \ \ as \ \ n \to \infty,$$

setting

$$E\left(\Pi_i(\frac{1}{1+r})\frac{1+\gamma_i L}{1+\theta_i L}\right) = \sum_{k=0}^{\infty} \pi_k L^k.$$

This yields

$$\pi_0 = E\Pi_i(\frac{1}{1+r}), \;\; \pi_k = E\Pi_i(\frac{1}{1+r})(-\theta_i)^{k-1}(\gamma_i - \theta_i), \;\; k \geq 1.$$

Aggregate consumption will be smoother than aggregate income if

$$\sum_{k=0}^{\infty} \pi_k^2 < \sum_{k=0}^{\infty} \nu_k^2 \tag{24}$$

As pointed out by Pischke (1995), changes of aggregate consumption will not be martingale difference anymore. The effect of aggregation is limited, however. First, the α_i do not enter into the expression for the π_k. The autoregressive root is represented by θ_i. However, the π_k will decay exponentially fast towards zero since $\theta_i \leq \bar{\theta} < 1$ $a.s.$ whenever $0 < q \leq q_i \leq Q < \infty$, as indicated above. Therefore $E\theta_i^k = O(\bar{q}^k)$ ruling out the possibility of an hyperbolic rate of decay.

Lagged information about aggregate shocks. Agents observe at each period

$$\nu_{it} = \epsilon_t + u_{it}$$

and can distinguish between the two only for previous period. Then

$$\Delta y_{it} = \nu_{it} - u_{it-1}.$$

Following Pischke (1995, section 2.3) version of Goodfriend (1992) model

$$\begin{aligned} \Delta c_{it} \;\; = \;\; & \left(\omega_i \frac{1+r+\gamma_i}{1+r+\alpha_i} + (1-\omega_i)\frac{1+r+\delta_i}{1+r+\beta_i}\right)\nu_{it} \\ & +(1+r)\left(\frac{1+r+\gamma_i}{1+r+\alpha_i} - \frac{1+r+\delta_i}{1+r+\beta_i}\right)((1-\omega_i)\epsilon_{t-1} - \omega_i u_{it-1}) \end{aligned}$$

Aggregating yields

$$\Delta C_t = \theta_1(r)\epsilon_t + \theta_2(r)\epsilon_{t-1}$$

setting

$$\begin{aligned}\theta_1(r) &= E\left(\omega_i\frac{1+r+\gamma_i}{1+r+\alpha_i}+(1-\omega_i)\frac{1+r+\delta_i}{1+r+\beta_i}\right)\\ \theta_2(r) &= (1+r)E\left(\left(\frac{1+r+\gamma_i}{1+r+\alpha_i}-\frac{1+r+\delta_i}{1+r+\beta_i}\right)(1-\omega_i)\right).\end{aligned}$$

Consumption is smoother than income whenever

$$\theta_1^2(r)+\theta_2^2(r)<\sum_{s=0}^{\infty}\nu_s^2. \tag{25}$$

The effect of aggregation is limited since changes in aggregate consumption follow an MA(1) process and it is well-known that aggregation is harmless when aggregating finite order MA (see Zaffaroni (2003, section 3)).

References

Abowd, J., and D. Card (1989): "On the covariance structure of earnings and hours changes," *Econometrica*, 57, 411–445.

Andrews, D. (1993): "Exactly median-unbiased estimation of first order autoregressive/unit root models," *Econometrica*, 61, 139–165.

Attanasio, O. (1999): "Consumption," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1B, pp. 781–814. Amsterdam, NE: Elsevier.

Campbell, J., and A. Deaton (1989): "Why is consumption so smooth?," *Review of Economic Studies*, 56, 357–374.

Christiano, L., and R. Eichenbaum, M. (1990): "Unit roots in real GDP: Do we know, and Do We Care?," *Carnegie Rochester Series on Public Policy*, 32, 7–62.

Cochrane, J. (1988): "How big is the random walk in GDP?," *Journal of Political Economy*, 96, 893–920.

Deaton, A. (1992): *Understanding consumption.* Oxford: Oxford University Press.

Diebold, F., and G. Rudebusch (1989): "Long memory and persistence in aggregate output," *Journal of Monetary Economics*, 24, 189–209.

Diebold, F., and A. Senhadji (1996): "Deterministic vs. stochastic trend in U.S. GNP, yet again," *American Economic Review*, 86, 1291–1298.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The generalized factor model: identification and estimation," *The Review of Economics and Statistics*, 82, 540–554.

Forni, M., and M. Lippi (1997): *Aggregation and the microfoundations of dinamic macroeconomics*. Oxford: Oxford University Press.

Goodfriend, M. (1992): "Information-aggregation bias," *American Economic Review*, 82, 508–19.

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

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

Lippi, M., and P. Zaffaroni (1998): "Aggregation of simple linear dynamics: exact asymptotic results," Econometrics Discussion Paper 350, STICERD-LSE.

Nelson, C., and C. Plosser (1982): "Trends and random walks in a macroeconomic time series: some evidence and implications," *Journal of Monetary Economics*, 10, 139–162.

Pischke, J.-S. (1995): "Individual income, incomplete information, and aggregate consumption," *Econometrica*, 63, 805–40.

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

Rudebusch, G. (1993): "The uncertain unit root in real GDP," *American Economic Review*, 83, 264–272.

Sargent, T. J. (1978): "Estimation of dynamic labor demand under rational expectations," *Journal of Political Economy*, 86, 1009–44.

Stock, J., and M. Watson (1999): "Diffusion indexes," *Preprint*.

Zaffaroni, P. (2003): "Contemporaneous aggregation of linear dynamic models in large economies," *Journal of Econometrics* , forthcoming.