The Aggregation of Dynamic Relationships caused by Incomplete Information
Submission to the Sir Clive Granger Memorial Conference

Michael Thornton
School of Economics
University of Reading

April 2010

Abstract

We consider the aggregation of heterogeneous dynamic equations across a large population, as introduced by Granger (1980), where the dynamics arise because agents face a signal extraction problem caused by incomplete information. Agents are unable to tell the duration of the shocks they face and make forecasts that use their private history efficiently but do not utilise any other agents' history. Homogeneous versions of this theoretical model have been proposed in the literature on household consumption. We show that, under plausible assumptions, the observed changes in the cross-section aggregate shows long term persistence even though every individual micro-series follows a random walk. In doing so we are obliged to weaken the independence assumptions used previously in this aggregation literature, widening considerably the type of results that such models can produce to include aggregates which are not mean reverting.

Keywords. Incomplete Information; Aggregation of Dynamic Relationships; Long Memory; Household behaviour.

Acknowledgements: This paper is a modest revamp of chapter 1 of my University of Essex PhD thesis. I am extremely grateful to Marcus Chambers and Gordon Kemp for their careful supervision and to my examiners, Emma Iglesias and Karim Abadir, for helpful comments. This research was supported by Economic and Social Research Council studentship Award no PTA-030-2006-00395.

Address for Correspondence: Michael A Thornton, School of Economics, University of Reading, Whiteknights, Reading, RG6 6AW, England.
e-mail: m.a.thornton@reading.ac.uk.

1 Introduction

1.1 Aggregation and Long Memory

Among the many seams opened in economics by Sir Clive Granger, one of the richest has been the area of fractional integration. In a ground breaking 1980 paper, Granger and Roselyn Joyeaux studied the properties of a series, Y_t, $t = 0, 1, 2, 3, ...$, given by

$$(1 - L)^d Y_t = \epsilon_t \tag{1}$$

where ϵ_t is a white noise process and d, the order of integration of Y_t, is not an integer. The fractional differencing filter in (1) was also developed and studied independently by Hosking (1981).

This natural extension of the unit root leads to the property of long memory[1]: hyperbolic decay of the autocorrelation function. For the same series this takes the form

$$\gamma_j \equiv Cov(Y_t, Y_{t-j}) \sim cj^{2d-1}, \text{ as } j \to \infty, \tag{2}$$

for some constant c. The frequency domain analogue is that the spectral density of the series around the origin is characterised by

$$f(\lambda) \sim c|\lambda|^{-2d}, \text{ as } \lambda \to 0, \tag{3}$$

where λ denotes frequency. The parameter d reflects the degree of memory of the series. The autocovariances are not summable for $d > 0$. Despite this, a process is second order stationary, having first and second moments that are stable through time, if $-1/2 < d < 1/2$. Whereas, if $1/2 \leq d < 1$, the process is mean reverting but has unbounded variance: the impact of shocks decays at a slow rate. The singularity resulting from (1) dominates the spectral density around the zero frequency, even when ϵ_t is a finite order auto-regressive moving average (ARMA) process. This has enabled a range of semi-parametric estimators of d, such as those proposed by Geweke and Porter-Hudak (1983) and Robinson (1994), which are consistent, asymptotically normal and robust to short-run dynamics. These estimators have been applied in a number of empirical studies to test for the presence, and to estimate the degree, of long memory in macroeconomic variables. Amongst them, Gil Alana and Robinson (2001) test for fractional integration at seasonal frequencies in quarterly data for consumption and income in the U.K. and Japan: the same data sets used by Hylleberg, Engle, Granger and Yoo (1990) for the initial application of their test for seasonal unit roots. They find evidence to support a non-integer order of integration at some seasonal frequencies, despite being unable to reject the null hypothesis of a unit root at the zero

[1] See Guegan (2005) for a thorough overview of definitions of the term.

frequency in the U.K. data. Also looking at U.K. aggregate consumption data, Chambers (1998) finds a value of d of around 1.4.

The fact that shocks persist in long memory models for far longer than in conventional (autoregressive moving average) ARMA models has made them useful in modelling a range of economic variables. A good overview of applications in economics and finance is provided in Baillie (1996). Although fractionally integrated models are not the only models to produce long memory, see Granger and Ding (1996), they have proved very popular in both empirical and theoretical work. Part of this popularity arises from the possibility to generate them from the aggregation of heterogeneous dynamic (AR(1)) relationships. Suppose that

$$x_{i,t} = \theta_i x_{i,t-1} + \psi_i \epsilon_t + \upsilon_{i,t}, \tag{4}$$

where ϵ_t is a white noise shock that is common to all agents and $\upsilon_{i,t}$ is a white noise shock that is specific to household i, with $E\{\upsilon_{i,t}\upsilon_{j,t}\} = 0$, $\forall i \neq j$. These shocks are held to be independent of one another and the autoregressive parameter, θ_i. The properties of the aggregate $X_{N,t} = \sum_{i=1}^{N} x_{i,t}$ for large N were considered by Robinson (1978). While independently Granger (1980) showed that as the number of relationships to be aggregated becomes large, the aggregate can exhibit long memory. In that paper, the autoregressive parameter varies on $(0,1)$ with a type 2 beta density

$$f(\theta) = \frac{2}{B(p,q)}\theta^{2p-1}(1-\theta^2)^{q-1},\ p,q>0,\ 0 \leq \theta \leq 1,\ f(\theta) = 0,\ \text{elsewhere},$$

with $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ where $\Gamma(q) = \int_0^\infty y^{q-1}e^{-y}dy$ is the Gamma function. The memory parameter d is always less than one and depends negatively on the parameter q, which determines the shape of the density approaching one.

The aggregation of heterogeneous AR(1) relationships has also been studied by Gonçalves and Gourieroux (1998), who develop Granger's analysis and consider the behaviour of common and idiosyncratic components when the aggregate is an average across the population. They establish what has become known as the 'usual result' that idiosyncratic shocks tend to disappear from the aggregate asymptotically as the population gets larger. Pesaran (2003) also considers this problem in the context of establishing a framework for aggregating linear dynamic models based on providing an optimal predictor. Lippi and Zaffaroni (1998) and Zaffaroni (2004) provide a rigorous analysis of cross-sectional aggregation under weaker assumptions and, significantly, establish the conditions under which idiosyncratic shocks do not tend asymptotically to zero, although they always do after differencing. Oppenheim and Viano (2004) consider the aggregation of processes that experience no common shocks, and hence use the normalisation $N^{-1/2}$ to establish convergence of the aggregate to a stochastic process. Their model is formulated first in discrete time and then in continuous time as the sum of heterogeneous Ornstein-Uhlenbeck processes.

1.2 Incomplete information

Despite possessing many desirable empirical features, fractionally integrated models sometimes suffer from a serious shortcoming. As Granger (2000) observes

> There is also little or no basic economic theory leading to fractional $I(d)$ variables, unlike the efficient market theory for $I(1)$

This paper attempts to address this point, generating results where the micro-series are random walks, in line with 'the efficient market theory', but the macro-series will be I(d). We examine the case when agents optimise under imperfect information, imagining they face shocks of differing durations that they are unable to tell apart. Agents use their own information efficiently: they base their actions on optimal forecasts, derived using techniques in Granger and Newbold (1977). Their own information is, however, incomplete in the sense that they do not know any other agent's shocks. Models of incomplete information have a long history in economics for explaining behaviour both by firms and by households. As an example, we show that these conditions are created by a very plausible modification to the models of household consumption under uncertainty proposed by Goodfriend (1992) and Pischke (1995): heteroskedasticity of income shocks across the population. In the context of consumption, the capacity to generate long memory in an aggregate from a large number of I(1) series means that aggregate consumption can be I(d) even though every household obeys Hall's (1978) well-known random walk hypothesis.

In considering the aggregation, we adopt the general, semi-parametric, framework developed by Robinson (1978) and deployed in Lippi and Zaffaroni (1998) and Zaffaroni (2004), but we are forced to confront a number of issues not previously covered in this literature. The most important of these is that, because agents' behaviour is based on signal extraction between two types of shocks, their AR(1) parameter is a known function of the variances of those shock processes. These elements had generally, hitherto, been assumed to be independent[2]. Relaxing that assumption gives rise to some significant, and in some ways counter-intuitive, new results with clear implications for the modelling and forecasting of the aggregate. It becomes possible to generate orders of integration that are not bounded at one and idiosyncratic shocks which survive in the aggregate even after differencing.

This paper also sits within a wider literature linking the parameters of micro-series with those of the macro-series to which they aggregate, raising the possibility that information from one could enhance understanding of the other. Lewbel (1994) considers how the moments of the micro-distribution of a heterogeneous autoregressive parameter shape the autocorrela-

[2] There is a very brief discussion of a more general case in Zaffaroni (2004).

tion function of the aggregate, without studying long memory directly. Abadir and Talmain (2002) explore the properties of aggregate output from a real business cycle model of heterogeneous firms operating under monopolistic competition. They examine the autocorrelation properties of aggregate GDP within the model when the firms receive temporary productivity shocks and show that the series exhibits a form of long memory, displaying more persistence than a standard ARIMA model while also being mean reverting. To further this in our consumption example, the micro-model is estimated on a U.S. panel study and estimates of the behaviour of the distribution of the autoregressive parameters made.

Section 2 develops the basic model. Section 3 discusses how these same parameters would translate through to macroeconomic data, particularly in the memory of consumption. The model is then applied to U.S. panel data in section 4. Some extensions to the basic model are discussed in section 5 and section 6 concludes.

2 Economic Underpinnings

We adopt the straightforward conventions on behaviour and processes used widely in the literature on incomplete information. Assumption 1 ensures a closed form expressions for each agent's behaviour under uncertainty; the remainder outline the information set.

Assumption 1: Agent's preferences. *Agent's are faced with a long-run optimisation problem and exhibit preferences that imply smoothing and certainty equivalence.*

Assumption 2: Shock processes and level of information. *All agents are subject to shocks that are either permanent, I(1), or transitory, I(0), in nature. No agent knows which type of shock they are experiencing at any point in time, although they do know the variances of the permanent and transitory shock processes.*

Assumption 3: Cross-section impact of shock processes. *The permanent shocks are macroeconomic in nature, impacting to some degree on all agents in the economy. The transitory shocks are microeconomic in nature, unconstrained by any macroeconomic process, so that each agent's transitory shock is orthogonal to every other.*

Assumption 4: Variances of shock processes. *The variances of these shocks varies across agents.*

In order to fix ideas, we discuss these in the context of our example: household consumption. Assumption 1 translates to the set-up discussed in the well-known paper by Hall (1978), with household utility that consist of the (infinite) discounted sum of temporally separable quadratic felicity functions, non-stochastic tastes and a discount rate that is constant and equal to the rate of return on capital. The assumption that preferences are temporally separable is simplifies the analysis. It would certainly be possible to consider other types of discounting/ habit formation/ short-run persistence in the agent's consumption problem. The effect would be to introduce an additional autoregressive parameter into the model, complicating the analysis of aggregation unnecessarily. As discussed in Zaffaroni (2004), the properties of the aggregate of heterogeneous higher order autoregressive processes is determined by the root of each autoregressive polynomial closest to the unit circle.

It seems reasonably plausible that households should experience uncertainty about the persistence of the given shocks that they face, even though experience has taught them something about the relative scale of the two processes. In proposing this dilemma, Pischke (1995) imagines a worker being made redundant but being unable to forecast precisely the duration of their unemployment. Other work-related examples include a worker receiving the offer of overtime. At first sight this is a classic transitory shock, but the worker will adapt their consumption patterns in the light of revised inferences about the likelihood of future overtime, capacity within their industry, the tightness of their labour market and the possibility of future salary increases that the offer of overtime might indicate. Or a worker receiving a one-off performance-related bonus will come to a judgment that balances the stated 'one-off' with the greater likelihood of higher pay or promotion indicated by the 'performance-related'. The presence of additional full-information shocks, the persistence of which are recognisable by the household, is discussed as an extension.

In order to explore the effects of aggregation across the economy, we are forced to make assumptions about which shocks are common and which are idiosyncratic. In the main, we follow Pischke (1995) and Goodfriend (1992) in assuming that the permanent shock, ϵ_i, affects everyone in the economy, while all transitory shocks, υ_i, are idiosyncratic. This assumption is, however, stronger than is needed to drive many of the results in section 3 and variations are discussed in section 5.

We diverge from that literature only in assumption 4. In both Goodfriend (1992) and Pischke (1995) both shocks are assumed to be homoskedastic across the economy. This is restrictive. Within a large economy, different agents, maybe by choice, will experience different levels of income volatility. This could be due to differing exposure to macroeconomic fluctuations, it could be due to more household specific shocks, or it could be a combination of the two. Our model supposes that households are heterogeneous in their exposure to both permanent/ common and transitory/ idiosyncratic shocks, but the homoskedasticity of either one is contained as a special case. Many of our results still hold when only one type of shock is heteroskedastic.

2.1 Agent's Behaviour

Becoming more specific, denote the permanent shock, $\epsilon_{i,t}/(1-L)$, where L is the lag operator and $\epsilon_{i,t}$ is a white noise processes with zero mean and variance $\sigma^2_{\epsilon,i}$. While we maintain assumption 3, the permanent shock can be written as $\psi_i\epsilon_t/(1-L)$, $\psi_i \in (0,\infty)$, and its variance, without loss of generality as ψ_i^2.[3] Denote the transitory shock, $\upsilon_{i,t}$, which is a white noise process with zero mean and variance $\sigma^2_{\upsilon,i}$. Unable to observe these in isolation, agents are forced to forecast their sum. They notice that the first difference of this series[4] follows a stationary first order moving average process, that is

$$\{\epsilon_{i,t} + \upsilon_{i,t} - \upsilon_{i,t-1}\} = \eta_{i,t} - \theta_i\eta_{i,t-1}. \tag{5}$$

Since the spectral densities of the two sides are identical it follows that,

$$\frac{\sigma^2_{\eta,i}}{2\pi}\left|1-\theta_i e^{-i\lambda}\right|^2 = \frac{\sigma^2_{\epsilon,i}}{2\pi} + \frac{\sigma^2_{\upsilon,i}}{2\pi}\left|1-e^{-i\lambda}\right|^2,$$

where $i=\sqrt{-1}$. Equating these spectral densities over any two of: the points $\lambda = 0$; or, $\pi/2$; or the region, $0<\lambda<\pi/2$ and solving the resulting quadratic equation gives,

$$\begin{aligned}
\sigma^2_{\eta,i} &= \sigma^2_{\upsilon,i}/\theta_i = \sigma^2_{\epsilon,i}/(1-\theta_i)^2 = \psi_i^2\sigma^2_{\epsilon}/(1-\theta_i)^2, \\
\theta_i &= 1 + \frac{1}{2}\left[a_i - \sqrt{(a_i+2)^2-4}\right],
\end{aligned} \tag{6}$$

where $a_i = \sigma^2_{\epsilon,i}/\sigma^2_{\upsilon,i}$ and we have chosen this particular root as it delivers an invertible moving average.

Note that $\theta_i \in (0,1)$ provided neither υ_i or ϵ_i are degenerate and $\theta_i \to 1$ as $\sigma^2_{\epsilon,i}/\sigma^2_{\upsilon,i} \to 0$, that is for those agents whose shocks are more likely to be transitory in nature. Also note that θ_i is heterogeneous as long as at least one of the permanent and transitory shocks is heteroskedastic across the population. The parameters of the structural model $(\sigma^2_{\epsilon,i}, \sigma^2_{\upsilon,i})$ are fully recoverable from the parameters of the reduced model $(\theta_i, \sigma^2_{\eta,i})$. The major implication from inverting (6), which will feature later, is that

$$a_i = \sigma^2_{\epsilon,i}/\sigma^2_{\upsilon,i} = \frac{(1-\theta_i)^2}{\theta_i}. \tag{7}$$

In the absence of complete information, agents respond to the stochastic process modelled by the right hand side of (5). So in period t, agents forecast the change in their the stochastic process next period as $-\theta_i\eta_{i,t}$. The essential feature of this paper is in that each period agents base their decision only on their value of η for that period. They do so optimally, so the variable of interest does not depend on past or future values of η_i. Neither does it

[3] Where do not follow this convention it is to allow equations to be recycled in section 5 when assumption 3 is weakened.

[4] Either through experience of the observable income process or widespread familiarity with the rules of adding ARMA processes

depend on $\eta_{j,t}$, but since

$$E\{\eta_{i,t}\eta_{j,t}\} = E\left\{\frac{\epsilon_{i,t}\epsilon_{j,t}}{(1-\theta_i L)(1-\theta_j L)}\right\} = \frac{\psi_i\psi_j}{1-\theta_i\theta_j} \neq 0,$$

the two are not independent. This means that agent j's shock contains information relevent to agent i. It also means that the correct aggregation procedure treats all common and idiosyncratic shocks as dynamic components.

To fix ideas we return to our consumption example, where ϵ and υ are now income shocks. The household recognises that the change in its income is given by (5). As Pischke (1995) points out, its optimal behaviour is to increase consumption by $\eta_{i,t}$ plus the discounted value of the forecast change in income for next period. That is

$$\Delta c_{i,t} = \left(1 - \frac{\theta_i}{1+r}\right)\eta_{i,t} = \left(1 - \frac{\theta_i}{1+r}\right)\frac{\epsilon_{i,t} + (1-L)\,\upsilon_{i,t}}{1-\theta_i L}, \tag{8}$$

where $\Delta \equiv (1-L)$.

It is clear from the first part of (8) that household consumption, given their information set, is still a random walk. Since $\eta_{i,t}$ is not orthogonal to $\eta_{j,t}$ for $i \neq j$, however, equation (8) can only be aggregated in cross-section correctly using the final expression, and aggregate consumption will not be martingale. In addition, provided common and idiosyncratic shocks do not make up the same proportion of permanent and transitory shocks for each household then $\eta_{i,t}$ provides additional information on $\epsilon_{j,t}$ and $\upsilon_{j,t}$. If households had access to information on each other's shocks they would improve their own estimates of their permanent and transitory shocks. In Goodfriend's model this information is available to households with a one-period lag, which induces a moving average into the process for aggregate consumption. We follow Pischke (1995) in assuming that this information is never fully assimilated by households, either because of the difficulty of evaluating this information or because it is relatively costly to acquire and process in comparison with the benefits to future utility that it conveys.

3 Cross-sectional Aggregation and Long Memory

We now consider the properties of the aggregate of the variable of interest across N agents in the economy. The notation in this section is directly related to the consumption example from the previous section, but the analysis could easily be applied to other situations with incomplete information. We use $C_{N,t}$ to denote aggregate consumption adjusted for deterministic changes to preferences[5], so that

$$\begin{aligned} \Delta C_{N,t} &\equiv \sum_{i=1}^{N} \frac{1}{N} [\Delta c_{i,t}] \\ &= \sum_{i=1}^{N} \frac{1}{N} \left[\left(1 - \frac{\theta_i}{1+r} \right) \frac{\epsilon_{i,t} + (1-L)\upsilon_{i,t}}{1-\theta_i L} \right]. \end{aligned} \tag{9}$$

3.1 Heteroskedastic Income Shocks

The properties of $C_{N,t}$ depend on the assumptions made about the heterogeneity of income shocks υ_i or ϵ_i. If both types of shocks are completely idiosyncratic then $\Delta C_{N,t}$ is simply the sum of N independent white noise processes and is itself white noise.

It is unlikely, however, that all shocks occurring within an economy are unrelated or that all agents face equal exposure to them. We model this heteroskedasticity of risk across i in the levels of $\sigma^2_{\epsilon,i}$ and/or $\sigma^2_{\upsilon,i}$ which implies, by equation (6), heterogeneity in θ_i. If $\sigma^2_{\epsilon,i}$ and $\sigma^2_{\upsilon,i}$ are constrained to take a finite number of values, say p and q respectively, then θ_i can take only $x \leq pq$ values. In this case it is well known that (9) would have an ARMA$(x, x-1)$ structure.

A finite order ARMA model, however, will not result if θ_i has a distribution that is continuous over a non-zero interval. Following Granger (1980), provided $\theta_i \in (0,1)$ is not bounded strictly below some level $\kappa < 1$, then the aggregate $\Delta C_{N,t}$ can exhibit long memory. Granger's original assumed Beta (II) distribution for θ is stronger than we need to characterise the limit process of $C_{N,t}$. We follow Zaffaroni (2004) in making the following, semi-parametric, assumption about the distribution of θ.

Assumption 5: Distribution of θ close to 1. *θ is absolutely continuous on $(0,1)$ and is distributed as $\theta \sim c_b(1-\theta)^b$, as $\theta \to 1^-$, $b > -1$, with c_b an arbitrary constant.*

Assumption 5 accommodates a number of parametric forms of the distribution of θ, for example, a uniform distribution would translate to a value of $b = 0$.

As we shall see, the properties of $C_{N,t}$ are determined by the behaviour of the density of the θ_i's as they approach unity. Intuitively, the behaviour of the autocovariance function at long lags depends increasingly on those whose AR parameter is very close to one.

The properties of aggregates of heterogeneous AR(1) relationships, of the form shown in (4), has been studied extensively by Granger (1980), Gonçalves and Gourieroux (1998), Lippi and Zaffaroni (1998) and Zaffaroni (2004). Equation (9), however, differs from the type of problem studied previously in threee ways.

[5]Perhaps after the removal of a deterministic trend component reflecting the demography of the population.

1. there is a term which is a function of the AR parameter, $\left(1 - \frac{\theta_i}{1+r}\right)$ on the right hand side;
2. because it is derived as an optimal linear forecast, θ_i is itself a function of the variances ψ_i^2 and $\sigma_{\upsilon,i}^2$ and,
3. the 2 types of shock have different orders of integration.

The first and second points break with an assumption that has been common throughout the literature until now: that of independence between the heterogeneous AR parameters and the variances of the common and idiosyncratic shocks. Provided there is a positive rate of return on capital, $r > 0$, then $\left(1 - \frac{\theta_i}{1+r}\right)$ is bounded and non-zero and so will only affect the scale of the variance and spectral density of $C_{N,t}$, not whether or not they exist or the rate at which they head off toward a singularity.

The second point is more challenging. Following the relationship between $\sigma_{\epsilon,i}^2$ and $\sigma_{\upsilon,i}^2$ laid out in (7), we can imagine, without loss of generality

$$u_{i,t} = \frac{\psi_i\sqrt{\theta_i}}{1-\theta_i}\overline{u_{i,t}}, \tag{10}$$

where $\overline{u_{i,t}}$ is a zero mean white noise process with $E\left\{\overline{u_{i,t}u_{j,t}}\right\} = 1, \forall i = j$ and 0 otherwise. This translates the heteroskedastic processes for both idiosyncratic and permanent shocks into homoskedastic processes multiplied by functions of θ_i and ψ_i. Since θ_i is a function of ψ_i and $\sigma_{\upsilon,i}^2$, we make the following assumption.

Assumption 6: Boundedness of ψ_i and its relationship with θ_i. *ψ_i is strictly bounded, and strictly positive, for all corresponding $\theta_i \in (0, \gamma)$ for any $\gamma < 1$. $\psi_i \sim c_a(1-\theta_i)^a$ as $\theta \to 1^-$, with c_a an arbitrary constant.*

This mirrors assumption 5 in focusing on the processes affecting households with high values for θ_i. It does not attempt to describe the relationship between ψ_i and θ_i fully, only the part that is dominant as θ_i is in the locality of one. The parameter a determines whether ψ_i: remains positive and bounded ($a = 0$); heads off to infinity ($a < 0$); or, down to zero ($a > 0$), and the rate at which it does so. Assumption 6 precludes singularities corresponding to values of $\theta < 1$[6]. A value of $a = 0$ is necessary but not sufficient for ψ_i to be determined independently of θ_i. A value for $a > 0$ would indicate that a high θ_i, *i.e.* those with a relatively high ratio of transitory to permanent shock, resulted from being relatively insulated from the impact of macroeconomic shocks, and hence a narrow distribution of $\sigma_{\upsilon,i}^2$. Whereas a value for $a < 0$ would indicate that a high θ_i resulted from being relatively risk loving,

[6]Anyone who has a value of $\psi_i = 0$ is removed from the problem of incomplete information. Provided they are not so numerous as to weaken substantively arguments based on having a large population, they can be accommodated by remembering that their consumption is a random walk, see section 5.

heavily exposed to macroeconomic shocks but even more so to microeconomic shocks, and hence a more widespread distribution of $\sigma^2_{\upsilon,i}$.

We consider the aggregation of the permanent and transitory shocks separately. Taking account of the different levels of integration, we write

$$\begin{aligned}
C_{N,t} &= U_{N,t} + \frac{E_{N,t}}{1-L}, \text{ where,} \\
U_{N,t} &\equiv \sum_{i=1}^{N} \frac{1}{N}\left[\left(1-\frac{\theta_i}{1+r}\right)\frac{\upsilon_{i,t}}{1-\theta_i L}\right], \\
E_{N,t} &\equiv \sum_{i=1}^{N} \frac{1}{N}\left[\left(1-\frac{\theta_i}{1+r}\right)\frac{\epsilon_{i,t}}{1-\theta_i L}\right].
\end{aligned}$$

In the first instance we will investigate the existence of the variances of these processes, which will determine whether they remain in any aggregates, and the behaviour of their spectral densities, which will provide further insight into memory. Following (10), the idiosyncratic component, $U_{N,t}$ will have variance and spectral density respectively denoted

$$\begin{aligned}
V^U_{N,t} &= \sum_{i=1}^{N} \frac{1}{N^2}\left[\left(1-\frac{\theta_i}{1+r}\right)^2 \frac{\psi_i^2\theta_i}{(1-\theta_i)^2}\frac{1}{(1-\theta_i^2)}\right], \\
S^U_{N,t}(\lambda) &= \sum_{i=1}^{N} \frac{1}{2\pi N^2}\left[\left(1-\frac{\theta_i}{1+r}\right)^2 \frac{\psi_i^2\theta_i}{(1-\theta_i)^2}\frac{1}{\left|1-\theta_i e^{-i\lambda}\right|^2}\right], \qquad (11) \\
-\pi < \lambda \leq \pi.
\end{aligned}$$

The variances and spectral density of the differenced, cross-sectionally aggregated common shock, $E_{N,t}$, on the other hand, will also include the cross products between households reacting to the same shock. This will have variance and spectral density respectively denoted

$$\begin{aligned}
V^E_{N,t} &= \sum_{i,j=1}^{N} \frac{1}{N^2}\left[\left(1-\frac{\theta_i}{1+r}\right)\left(1-\frac{\theta_j}{1+r}\right)\frac{\psi_i\psi_j}{(1-\theta_i\theta_j)}\right], \\
S^E_{N,t}(\lambda) &= \frac{1}{2\pi N^2}\left|\sum_{i=1}^{N}\left(1-\frac{\theta_i}{1+r}\right)\frac{\psi_i}{1-\theta_i e^{-i\lambda}}\right|^2, \qquad (12) \\
-\pi < \lambda \leq \pi.
\end{aligned}$$

An unusual feature of processes comprising the sum of heterogeneous dynamic equations is that these variances and spectral densities are sums which depend on a parameter, θ_i, that is a random variable and as such become random variables themselves. Throughout this literature, these quantities are considered as functions of the parameters of the distributions of these random variables and their convergence, or otherwise, evaluated as the population size, N, gets large.

In considering these objects, we adopt the approach employed in Zaffaroni (2004), adapted to take account of the particular issues outlined above. The finiteness of these quantities will, in general, depend on the values of expressions of the form

$$\sum_{i=1}^{N} \frac{1}{N(1-\theta_i)^k}.$$

These will be finite in the limit and equal to

$$E\left\{\frac{1}{(1-\theta_i)^k}\right\},$$

as long as

$$\int_u^1 \frac{dF(\theta)}{(1-\theta)^k} < \infty, \tag{13}$$

for each arbitrary constant $0 < u < 1$, where $F(\theta)$ denotes the distribution of θ. Even where this is not finite, it will determine the rate at which the sum goes off to infinity. One crucial difference from the existing literature is that k will turn out to be determined jointly by the parameter a as well as the parameter b and the moment in question.

Finally we will evaluate the memory of the two processes. Following Granger (1980) and Zaffaroni (2004), we will show the potential for hyperbolic decay in the autocorrelation function of the idiosyncratic component. Following Lippi and Zaffaroni (1998) we will analyse the behaviour of the spectrum around the origin for the common component.

3.2 Idiosyncratic component

Idiosyncratic shocks are assumed to be I(0), but aggregate consumption also contains common shocks, which are assumed to be I(1), and so is likely to be differenced. As a result we consider the properties of the idiosyncratic shock both in levels and in differences, denoted $\Delta U_{N,t}$.

We remind that an AR(1) process $x_t = u_t/(1-\theta L) = \sum_{k=0}^{\infty} \theta^k u_{t-k}$, where u_t is white noise has a variance $Var\{x_t\} = Var\{u_t\}/(1-\theta^2)$ and an autocovariance at lag s given by $Cov\{x_t x_{t-s}\} = \theta^s Var\{u_t\}/(1-\theta^2)$. If that process is differenced, the resulting process is

$$\Delta x_t = (1-L)u_t/(1-\theta L) = u_t + \sum_{k=0}^{\infty} \theta^k(\theta-1)u_{t-k-1}.$$

Since the two terms do not overlap and u_t is white noise, this differenced AR(1) process has a variance, $Var(\Delta x)$

$$\begin{aligned} &= Var(u) + \sum_{k=0}^{\infty} \theta^{2k}(\theta-1)^2 Var(u) \\ &= \left\{(1-\theta^2) + (1-\theta)^2\right\} Var(x) \\ &= 2(1-\theta)Var(x), \end{aligned} \tag{14}$$

and an autocovariance, $E\{\Delta x_t \Delta x_{t-s}\}$

$$\begin{aligned}
&= E\left\{\left[u_t + \sum_{k=0}^{\infty}\theta^k(\theta-1)u_{t-k-1}\right]\times\right.\\
&\quad \left.\left[u_{t-s} + \sum_{k=0}^{\infty}\theta^k(\theta-1)u_{t-k-s-1}\right]\right\}\\
&= \theta^{s-1}(\theta-1)Var(u) + \theta^s\sum_{k=0}^{\infty}\theta^{2k}(\theta-1)^2 Var(u)\\
&= \theta^{s-1}\left\{(\theta-1) + \frac{\theta(\theta-1)^2}{1-\theta^2}\right\}Var(u)\\
&= -\theta^{s-1}(1-\theta)^2 Var(x). \qquad (15)
\end{aligned}$$

We evaluate the aggregation of the idiosyncratic component in the following propositions, considering the variance, spectral density and memory of $U_{N,t}$ and $\Delta U_{N,t}$. First, we consider the nature of the variance, the circumstances under which it is finite and under which the 'usual result' holds.

Proposition 1: As $N \to \infty$

(i) *Variance of* $U_{N,t}$

if $a + b > 1/2$ then $V^U_{N,t} \overrightarrow{a.s.}\ 0$,

whereas if $a + b \leq 1/2$ then $N^{\frac{2(a+b)-1}{b+1}} V^U_{N,t} \overrightarrow{d}\ S_\delta$,

(ii) *Variance of* $\Delta U_{N,t}$

if $a + b > 0$ then $V^{\Delta U}_{N,t} \overrightarrow{a.s.}\ 0$,

whereas if $a + b \leq 0$ then $N^{\frac{2(a+b)}{b+1}} V^{\Delta U}_{N,t} \overrightarrow{d}\ S_\delta$

where S_δ denotes a δ-stable random variable defined[7] as,

$S_\delta(1) + S_\delta(2) + ... + S_\delta(N) =^d N^{1/\delta} S_\delta + D_N$,

where the left hand side denotes the sum of $N \geq 2$ independent realisations of the process S_δ and D_N is a real constant depending on N. The parameter $0 < \delta \leq 2$ is the highest finite moment of the distribution.

Proof See Appendix

We note an extra normalisation required is for the convergence of $V^U_{N,t}$ and $V^{\Delta U}_{N,t}$ to a δ-stable random variable and that it is always a negative power of N. Under the normalisation N^{-2}, which is implicit in $V^U_{N,t}$, these quantities do not converge, even to a random variable. These conditions in proposition 2 should then be read as the circumstances in which the

[7] Variously, see Samorodnitsky and Taqqu section 1.1.

idiosyncratic component does not succumb to the law of large numbers and vanish in mean square - the failure of the 'usual result' as first noted in Lippi and Zaffaroni (1998). The fact that agents are signal processing between the two types of shock does, potentially, increase the possibility of this type of result. In the previous literature, the usual result failed only if $b < -1/2$, which required that the distribution of θ became more dense as it approached one. In this setting, however, it is perfectly possible for the usual result to fail even for positive b, when θ is becoming less dense as it approaches 1, provided $a + b \leq 1/2$. There are two reasons for this. The first is the reinforcement of the effect on the size of $V_{N,t}^U$ from higher values of θ if these come with higher $\sigma_{v,i}^2$ provided $a < 0$. The second is that the signal extraction in (5) in effect integrates the transitory shocks.

In Zaffaroni (2004), the differenced idiosyncratic component always disappears in mean square. Under our assumptions this will not happen if $a + b \leq 0$. This is noteworthy as it had previously been assumed to hold within the literature on consumption under incomplete information[8].

The proof of proposition 2 underlines the important role of discounting in this model. If $r = 0$, then the order of k in each part of the proof would be one lower, which would redetermine the critical parameter values. This point applies equally to all other propositions in this section.

Given the possibility of unbounded variance, even with differenced data, it is natural now to consider the spectral density of the idiosyncratic component $S_{N,t}^U(\lambda)$ at the origin and at other frequencies. This is the subject of proposition 2.

Proposition 2: As $N \to \infty$

(i) *Spectral density of $U_{N,t}$ away from the origin*

if $a + b > 0$ then $S_{N,t}^U(\lambda) \overrightarrow{a.s.}\ 0$, for $\lambda \neq 0$,

whereas if $a + b \leq 0$ then $N^{\frac{2(a+b)}{b+1}} S_{N,t}^U(\lambda) \overrightarrow{d}\ S_\delta$, for $\lambda \neq 0$,

(ii) *Spectral density of $U_{N,t}$ at the origin*

if $a + b \leq 1$ then $S_{N,t}^U(0) \overrightarrow{a.s.}\ \infty$,

whereas if $a + b > 1$ then $S_{N,t}^U(0) \overrightarrow{a.s.}\ 0$.

Proof See Appendix

Proposition 2 is very closely related to proposition 1. From part (i) we see that the spectral density will contain a continuum of singularities across some frequency band, i.e. that could

[8]In Goodfriend (1992) it is a matter of assumption, whereas Pischke (1995) generalises his example to cover a law of large numbers argument, for completely idiosyncratic shocks.

not be removed by differencing, if $a+b \leq 0$. It is not surprising that it is power in these non-zero bands that causes the failure of the 'usual result' even in differenced data and that they are of the same order of distribution as in proposition 1(ii).

For $0 < a+b \leq 1/2$, $U_{N,t}$ has an infinite variance but a spectral density that tends toward zero for $\lambda \neq 0$.

(ii) makes it clear that in the limit $S^U_{N,t}(\lambda)$ is either unbounded or zero at the origin. Comparing the values for the parameters over which these occur with those from proposition 1 we see that for $1/2 < a+b \leq 1$ $U_{N,t}$ has both a singularity at the origin and a finite variance (actually 0 in the limit). This suggests that $U_{N,t}$ could be a long memory process and it is that which we now consider.

The most straightforward way to derive the memory for the idiosyncratic component is from its autocorrelation function, using (2).

Since the variance of $U_{N,t}$ goes either to 0 or ∞, we normalise by the variance to get convergence to a stochastic process.

Proposition 3 As $N \to \infty$,

(i) *Memory of* $U_{N,t}$

$U_{N,t}/\sqrt{V^U_{N,t}}$ has memory $d_U = \frac{1}{2}(3-b-2a)$. If $2a+b > 2$, $U_{N,t}/\sqrt{V^U_{N,t}} \overrightarrow{d} U_{M,t}(d_U)$ where $U_{M,t}(d_U)$ is a Gaussian stochastic process.

(ii) *Memory of* $\Delta U_{N,t}$

$\Delta U_{N,t}/\sqrt{V^{\Delta U}_{N,t}}$ has memory $d_U - 1 = \frac{1}{2}(1-b-2a)$. If $a+b > 0$, $\Delta U_{N,t}/\sqrt{V^{\Delta U}_{N,t}} \overrightarrow{d} U_{M,t}(d_U - 1)$ where $dU_{M,t}$ is a Gaussian stochastic process.

Proof See Appendix

The condition that $a+b > 0$ ensuring that $NV^{\Delta U}_{N,t}$ exists is clearly linked to the finiteness of the spectrum away from the origin discussed in proposition 2.

Proposition 3 highlights the different effects of the a and b parameters on $U_{N,t}$. In determining the mean square of $U_{N,t}$, a and b carried equal weight, whereas in determining memory, a counts for twice b. This is because the effects of the ψ_i are normalised by N^{-1} in propositions 1 and 2, but by only $N^{-1/2}$ in proposition 3. One implication is that the link between survival in mean square and memory is more sophisticated than is the case when all the parameters are assumed to be independent of one another. There is no degree of memory for $U_{N,t}$ that guarantees that it escapes the usual result. Although, given $b > -1$, we can say that any idiosyncratic process that escapes the 'usual result' must have $d_U > 1/2$.

This analysis incorporates the model studied previously in the literature, where the parameters are assumed to be independent, as a special case. When $a = 1$, the effects of signal processing and weights in aggregate second moments cancel out and propositions 1, 2 and 3 translate back to the results in Zaffaroni (2004). This includes the case when households experience idiosyncratic shocks driven by the same homoskedastic process, meaning that all the heterogeneity comes from different exposures to common macroeconomic shocks.

Allowing heterogeneous ψ_i has, however, given rise to the possibility of an unusual and counterintuitive result. The interaction between a and b means that, unlike in the previous literature, the memory of the idiosyncratic component need not be bounded below 1. The model is capable of producing an idiosyncratic series, $U_{N,t}$, that is unit root or even explosive, despite being the aggregate of stationary AR(1) relationships. This is all the more remarkable given our starting point that consumption is a random walk for every household, given their information set. A theme to which we shall return after we have considered the common component.

3.3 Common component

We now consider the common shock process. For simplicity and for easy comparison with the idiosyncratic component, we do this by considering the properties of $E_{N,t}$. Since aggregate consumption contains $E_{N,t}/(1-L)$, it is clear that properties such as the finiteness of the variance are only relevant to data that has been differenced. We also note that the order of integration of the common component in consumption will then be one higher than that of $E_{N,t}$.

It is clear straightaway that, unlike in the idiosyncratic case $V^E_{N,t}$ will never go to zero. We adapt an expression used in Gonçalves and Gourieroux (1998) to write

$$E_{N,t} = \sum_{s=0}^{\infty} \left\{ \sum_{i=1}^{N} \frac{1}{N} \left[\left(1 - \frac{\theta_i}{1+r} \right) \psi_i \left(\theta_i^s \right) \right] \right\} \epsilon_{t-s}. \tag{16}$$

As $N \to \infty$, the expression in $\{.\}$ tends, almost surely, to its expectation, which we denote ν_s. We note that, by repeated integration by parts,

$$\begin{aligned} \int_0^1 \left(1 - \frac{\theta_i}{1+r} \right) \theta^{(s)} (1-\theta)^{a+b} d\theta = {} & \frac{\Gamma(s+1)\Gamma(a+b+1)}{\Gamma(s+a+b+2)} \\ & - \frac{1}{1+r} \frac{\Gamma(s)\Gamma(a+b+1)}{\Gamma(s+a+b+1)}, \end{aligned} \tag{17}$$

provided $r > 0$. Then for large values of s, by Stirling's formula,

$$\nu_s \sim c s^{-(a+b+1)}, \tag{18}$$

with c an arbitrary constant. We can then rewrite

$$
\begin{aligned}
V^{E}_{N,t} &= 1\left[1+\nu_1^2+\nu_2^2+\nu_3^2+\nu_4^2+....\right], \\
S^{E}_{N,t}(0) &= \frac{1}{2\pi}\left[1+\nu_1+\nu_2+\nu_3+\nu_4+....\right]^2.
\end{aligned}
$$

Now, from (18), $\sum_{s=0}^{\infty}\nu_s$ converges, indicating that $S^{E}_{N,t}(0)$ is bounded, if $a+b>0$, while $\sum_{s=0}^{\infty}\nu_s^2$ converges, indicating that $V^{E}_{N,t}$ is bounded, if $a+b>-1/2$. This is stated more formally in the next two propositions, which mirror proposition 1 and 2.

Proposition 4: As $N\to\infty$

Variance of $E_{N,t}$

if $a+b>-1/2$ then $V^{E}_{N,t} \overrightarrow{a.s.} V^{E}<\infty$,

whereas if $a+b\leq -1/2$ then $n^{\frac{2(a+b)+1}{b+1}}V^{E}_{N,t} \overrightarrow{d} S_{\delta}$,

where, as in proposition 1, S_{δ} denotes a δ stable random variable.

Proof See Appendix

As long as $a+b>-1/2$ as a condition then proposition 4 points out that the variance of the common component, $E_{N,t}$ tends to a non-zero constant. We now consider how this variance is split across the spectrum.

Proposition 5: As $N\to\infty$

(i)*Spectral Density of $E_{N,t}$ away from the origin*

if $a+b>-1$ then $S^{E}_{N,t}(\lambda) \overrightarrow{a.s.} s^{e}(\lambda)<\infty$, for $\lambda\neq 0$,

(ii)*Spectral Density of $E_{N,t}$ at the origin*

while if $a+b\leq 0$ then $S^{E}_{N,t}(0) \overrightarrow{a.s.} \infty$.

Proof See Appendix

Finally, we consider the impact of this on the series $C_{N,t}$, we investigate the memory properties of $E_{N,t}$.

Proposition 6

Behaviour of the spectral density of $E_{N,t}$ in the region of the origin

As $\lambda\to 0$

$$
S^{E}_{N,t}(\lambda)\sim \begin{array}{l} c_1\lambda^{2(a+b)},\ (a+b)<0, \\ c_2\log(\lambda^{-1}),\ (a+b)=0, \end{array}
$$

$$c_3, \ (a+b) > 0,$$

with c_1, c_2, c_3 arbitrary positive constants.

Proof See Appendix

This implies that $E_{N,t} \overrightarrow{d} E_t(d_E)$, where $E_t(d_E)$ is a stochastic process with memory parameter $d_E = -(a+b)$, following (3). As the shock is common, we cannot apply central limit theorem arguments to it and so E_t will only be Gaussian if ϵ_t is.

As with the idiosyncratic component, the aggregate common component, $E_{N,t}$, could be unit root or even explosive, even though it is the sum of strictly stationary series. When $a = 0$, the results of propositions 4, 5 and 6 are the same as those for the model with independent parameters. This includes the case when all households are equally vulnerable to macroeconomic shocks, meaning that the heterogeneity comes from household specific volatilities.

3.4 Modelling the Aggregate

These results raise some interesting questions concerning the modelling of the aggregate. Remembering that the idiosyncratic component is inherently one degree of integration below the common, and that it is likely that we will want to difference the data, we summarise the results of this section in the following 3 cases:

Case 1: $a + b \leq 0 \Rightarrow V^U_{N,t}, V^{\Delta U}_{N,mt} \rightarrow \delta stable$

Here the idiosyncratic component survives in mean square in both the aggregate and its first difference. The order of integration of $C_{N,t}$ is therefore $d = d_U = 3/2 - a - b/2$, which is greater than 1. The order of integration of $\Delta C_{N,t}$ is therefore $d_\Delta = d_U - 1 = 1/2 - a - b/2$, which is greater than 0.

Case 2: $0 < a + b \leq 1/2 \Rightarrow V^U_{N,t} \rightarrow \delta stable, V^{\Delta U}_{N,mt} \rightarrow 0$

Here the idiosyncratic component survives in mean square in the aggregate but not in its first difference. The order of integration of $C_{N,t}$ is therefore $d = d_U = 3/2 - a - b/2$, which is greater than 1/2. The order of integration of $\Delta C_{N,t}$ is $d_\Delta = d_E = -a - b$, which must be below 1/2.

Case 3: $a + b > 1/2 \Rightarrow V^U_{N,t}, V^{\Delta U}_{N,mt} \rightarrow 0;$

The 'usual result' holds for both levels and first differences. The order of integration of $C_{N,t}$ is therefore $d = d_E + 1 = 1 - a - b$, which is less than 1/2. The order of integration of $\Delta C_{N,t}$ is $d_\Delta = d_E = -a - b$, which must be below $-1/2$.

In the region covering case 1, the variable of interest has an infinite spectrum. Whereas in case 3, it is stationary, even in levels. So case 2 is perhaps the most plausible for many macroeconomic series. It is certainly the most interesting. Firstly, in case 2 we have the counter-intuitive result that differencing the aggregate will reduce its order of integration by more than one. This is a feature of all aggregations in which the 'usual result' fails to hold and is only exacerbated, not created, by signal processing shocks of different orders of integration. Secondly, this feature makes it possible to estimate a and b separately from estimates of d and d_Δ. If $d - d_\Delta > 1$, then it is easy to show that $b = 2(d - d_\Delta) - 3$ and $a = 3 - 2d + d_\Delta$.

Our results also have implications for the forecasting of the aggregate, which will be driven overwhelmingly by the common shocks. Forecasts of the impact of a sudden common shock will depend heavily on how persistent the shock is within the system, in other words the memory. As the memory of the idiosyncratic component is always larger than that of the common component, it will dominate estimates of the memory of aggregate consumption whenever the idiosyncratic component survives in mean square. Some caution is needed, therefore, in using estimates of the memory, for example from a log-periodogram type estimator, in projecting forward the effect of a shock based on the economy if there is evidence that either cases 1 and 2 apply. In case 1, this problem is inescapable: d_E is masked by d_U. We must do our best with an upper bound. In case 2, a forecast of consumption would be better based on an estimate of memory from the differenced series, that is $d_\Delta + 1$, than one estimated from the series in levels, d.

4 Estimation

4.1 Data

The data are taken from the University of Michigan's Panel Study of Income Dynamics (PSID). This data series has been used to estimate models of household consumption by numerous authors, including Hall and Mishkin (1982), Zeldes (1989), Lawrence (1991) and Dynan (2000). A description of the data set is in the appendix. We assume that the data are not subject to time aggregation. Estimates under time aggregation in both discrete and continuous time are in Thornton (2009).

4.2 Methodology

The models to be estimated are of the form

$$\Delta c_{i,t} = X_{i,t}\beta + \theta_i(\Delta c_{i,t-1} - X_{i,t-1}\beta) + \left(1 - \frac{\theta_i}{1+r}\right)\zeta_{i,t}. \tag{19}$$

The term $X_{i,t}\beta$ and its lag serve to adjust for non-stochastic changes in household preferences. Following those previous authors, the external regressors in $X_{i,t}$ are largely demographic: the change in food needs; age of the head; a dummy indicator for whether the head of the household is white plus a constant. In common with previous authors, we do not find evidence of persistent household effects and do not include them in our estimates. The coefficient β is estimated from a single dynamic equation covering the entire sample and retained in the estimates of the heterogeneous models. The choice of interest rate r is set at 5 per cent over the period. Changing this value made little difference to the estimates.

The error term, $\zeta_{i,t}$ contains both idiosyncratic and common shocks. In the model discussed above, it is not white noise. Note that the its correlation structure, $\Gamma_{i,j}(s) \equiv E\left\{\zeta_{i,t}\zeta_{j,t-s}\right\}$ is

$$\begin{aligned}
\Gamma_{i,j}(0) &= \sigma_{\epsilon,i}^2 + 2\sigma_{\upsilon,i}^2 = \left\{1 + \frac{2\theta_i}{(1-\theta_i)^2}\right\}\psi_i^2,\ i = j \\
&= \sigma_{\epsilon,i}\sigma_{\epsilon,j}\psi_i\psi_j,\ i \neq j, \\
\Gamma_{i,j}(1) &= -\sigma_{\upsilon,i}^2 = -\frac{\theta_i}{(1-\theta_i)^2}\psi_i^2,\ i = j, \\
&= 0, i \neq j, \\
\Gamma_{i,j}(s) &= 0, \forall |s| > 1.
\end{aligned} \tag{20}$$

The final equalities result from (7). Note that our parameters of interest determine the covariance structure of the error term, as often happens in models that are aggregated temporally. We therefore use a recursive procedure outlined in Bergstrom (1990) to compute the likelihood conditional on the first two observations in the sample. Even this procedure requires the Choleski decomposition of a (sparse) matrix of equal to the number of households multiplied by the number of time periods. For computational convenience the estimates come from 16 cohorts of 46 households.

4.3 Results

The distribution of the (preliminary) estimates of θ, based on the Epanechnikov kernel, are drawn in figure 1. These are summarised, along with estimates of the parameters a and b, in table 1.

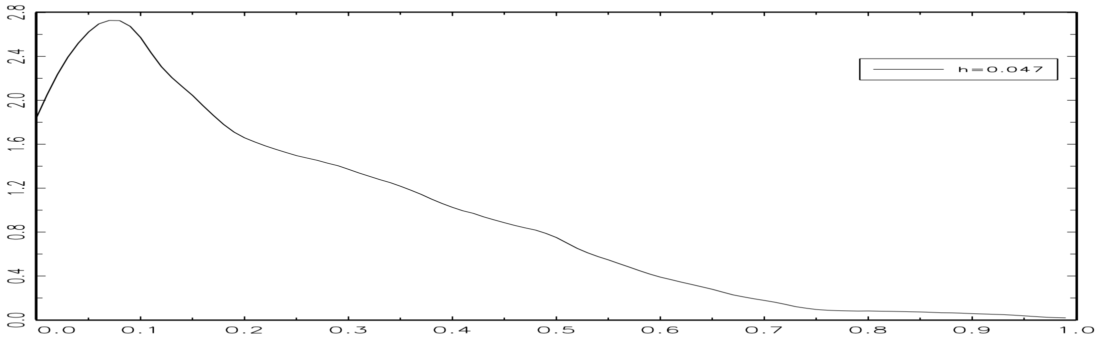

Figure 1: Estimated Density of θ under Incomplete Information

Table 1. Estimates of Heterogenous autoregressive parameter for household consumption under incomplete information preferences	
Mean of θ's	0.2264
Standard Deviation of θ's	0.1935
b as $\theta \to 1$	0.0369
a as $\theta \to 1$	0.0007
Memory of consumption p.c. at 0	1.4809
Memory of Δ consumption p.c. at 0	-0.0376
no. observations	8832

The estimates lie in the region $0 < a+b \leq 1/2$. This is the case where the idiosyncratic component survives in levels but not in differences and hence the estimated order of integration of consumption per capita is more than 1 higher than the estimated order of integration in the difference of consumption. Aggregate consumption would be integrated of order 1.4 to 1.3, which is quite close to the estimates in Chambers (1998) on U.K. on aggregate consumption data.

5 Extensions

We now introduce two generalisations to the model, focusing on their impact on the memory of the variable in question. Such generalisations will also affect the autocovariance structure of the micro data laid out in (20), mostly in straight-forward ways[9]. Once again the context of aggregate consumption is used to enhance intuition.

[9]For further discussion see Thornton (2009).

5.1 Partial Information

While it seems plausible that agents are unsure of the duration of many of the shocks that they face, it is also likely that there are some shocks whose persistence is well known. For example, gifts or lottery wins are clearly very different in persistence to promotions or attaining qualifications. Suppose that household i, in addition to the shocks analysed previously in the paper, received income shocks with a known persistence. We consider the impact of these when they are permanent or transitory, common or idiosyncratic. Assumption 1 implies that any permanent or transitory shock of known persistence will add a component integrated of order one to the aggregate, $C_{N,t}$. These shocks would not influence the distribution of θ_i, which depends on the unknown shocks. One effect therefore is that the aggregate must be integrated of at least order one, but may actually be higher under cases 2 and 3 above. If the additional known shock is common, then the first difference is at least integrated of order 0, but may be higher. If the additional known shocks are all idiosyncratic, then their aggregation and division by N produces a quantity which survives in mean square in levels (where it has unbounded variance) but not in differences (where it does not). In this case the order of integration of the first difference of the aggregate could be negative.

5.2 The Presence of Common, Transitory or Idiosyncratic, Permanent shocks

The precise mapping of common to permanent and idiosyncratic to transitory serves to simplify the analysis of the aggregate and provide direct comparison with earlier models. The mapping is not essential for the model to generate long memory, provided agents are faced with a signal extraction problem between permanent and transitory shocks and, as we shall see, there is relevant information in the aggregate that they do not use in solving that problem.

Attempting to minimise the additional notation required, suppose each agent's permanent shock process also contains idiosyncratic shocks, $\mu_{i,t}$, so that $\epsilon_{i,t} = \psi_i \epsilon_t + \mu_{i,t}$. While each agent's transitory shock process also contains a common shock, $\tau_i \nu_t$, so that $\upsilon_{i,t} = \xi_{i,t} + \tau_i \nu_t$. As before, all of the processes are zero mean and independent of one another and, without loss of generality, the variances of ϵ_t and ν_t are equal to one.

Equation (7) would still apply as before, leaving θ_i the same function of $\sigma^2_{\upsilon,i}/\sigma^2_{\epsilon,i}$, but these variances are each now the sum of the variances of the corresponding idiosyncratic and common shocks.

When considering aggregation we must now evaluate both the common and the idiosyncratic shock processes in first differences. We can model these as moving average processes using identical arguments to those used to establish (5). The first difference in the common shock is $\Delta\epsilon_i t = \psi_i\epsilon_t + (1-L)\tau_i\nu_t = \frac{\psi_i}{(1-\phi_i)^2}(1-\phi_i L)e_t$, with e_t a common white noise shock with unit variance. The scale parameter in the final expression is established using identical arguments to those used in equation (6). While we can also model the first difference of the idiosyncratic shock as $\Delta\upsilon_i t = \mu_{i,t} + (1-L)\xi_{i,t} = (1-\chi_i L)u_{i,t}$ with $u_{i,t}$ an idiosyncratic and heteroskedastic white noise shock with $Var(u_{i,t}) = Var(\xi_{i,t})/\chi_i$. Note that as long as $\mu_{i,t}$ is not degenerate[10], in which case χ_i would equal one, the order of integration of the idiosyncratic component has risen by one.

It is well known that moving average terms, in general, have no impact on the long term properties of a time series captured in its degree of memory or the finiteness of its variance, see Zaffaroni (2004). This is subject to the condition that neither of the moving average polynomials $(1-\phi_i L)$ or $(1-\chi_i L)$ cancel with the autoregressive polynomial $(1-\theta_i L)$. Cancellation is only brought about when the ratio of the variances of transitory to permanent shocks is the same for common and idiosyncratic shocks, this is if τ_i^2/ψ_i^2 equals $\sigma^2_{\xi,i}/\sigma^2_{\mu,i}$, and therefore both equal $\sigma^2_{\upsilon,i}/\sigma^2_{\epsilon,i}$. If this were the case, then the agent would gain no extra information about the duration of their own shocks by looking at shocks across the economy: being able to split their own current shock into common and idiosyncratic components would not change their forecast of the next period's realisation of the stochastic process for them. This reinforces the point that persistence depends very much on incomplete relevant information.

As before, the aggregation of both components should consider the heteroskedasticity of the common and idiosyncratic shocks, e_t and $u_{i,t}$. For the common shock, additional assumptions are needed about the behaviour of the new parameter ϕ_i, particularly as $\theta_i \to 1^-$. The assumption that ϕ_i is independent of θ_i is one candidate, but assumptions in which the two are related are possible. For the idiosyncratic shock and thus assumptions on $Var(\xi_{i,t})$ and χ_i are needed.

6 Conclusions

In analysing an autoregressive model derived from a signal extraction problem we have bridged the literatures on the empirical implications of the random walk hypothesis and the theoretical underpinnings of long memory in macroeconomic time series. We have shown, on

[10] There are meaningful permanent idiosyncratic shocks.

the one hand, that incomplete information in a very simple model can produce micro-series that are random walks but aggregate to something that can exhibit long memory. Moreover, earlier results in this literature that idiosyncratic shocks tend to zero in mean square are broken for some (not unreasonable) parameter values. At the same time, our autoregressive signal extraction model is unique within the literature on the aggregation of heterogeneous autoregressive models in being capable of producing: integration of order greater than 1; and, an idiosyncratic component that can survive in mean square even after differencing. Estimates of the heterogeneous parameters were made using maximum likelihood, which took account of the presence of the parameters of the model in both the autoregressive and moving average parts. Kernel techniques were used to study features of the underlying distributions. More sophisticated estimation techniques are worthy of investigation.

References

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

[2] Baillie, R. T. (1996)., Long memory processes and fractional integration in econometrics. Journal of Econometrics, 73: 5–59.

[3] Bergstrom, A. R (1990). Continuous Time Econometric Modelling, Oxford Universtiy Press.

[4] Chambers, M.J. (1998). Long Memory and Aggregation in Macroeconomic Time Series. International Economic Review 39: 1053–1072.

[5] Dynan, K. E. (2000). Habit Formation in Consumer Preferences: Evidence from Panel Data. American Economic Review 90:391–406

[6] Geweke, J. , Porter-Hudak, S. (1983). The estimation and application of long memory time series models. Journal of Time Series Analysis 4:221–238.

[7] Gil Alana, L. A., Robinson P.M. (2001). Testing Seasonal fractional integration in UK and Japanese consumption and income. Journal of Applied Econometrics, 16: 95–114.

[8] Gonçalves, E., Gourieroux, C., (1998) Aggregation de processes autoregressifs d'ordre 1. Annales d'Economie et Statistique 12,127–149.

[9] Goodfriend M. (1992) Information-aggregation bias. American Economic Review 82: 508–519.

[10] Granger, C. W. J., Newbold P. (1977). Forecasting economic time series, Academic Press.

[11] Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14:222–238.

[12] Granger, C. W. J., Joyeaux, R. (1980). An introduction to long memory time series models. Journal of Time Series Analysis 1:15–29.

[13] Granger, C. W. J., Ding, Z. (1996). Varieties of Long Memory Models. Journal of Econometrics 73:61–778.

[14] Granger, C. W. J.(2000). Current Perspectives on Long Memory Processes. Academia Economc Papers 28:1–16.

[15] Guegan, D., (2005). How can we define a concept of long memory? An econometric survey. Econometric Reviews 24: 113–149.

[16] Hall, R. E. (1978). Stochastic Implications of the Life Cycle - Permanent Income Hypothesis: Theory and Evidence. Journal of Political Economy 86:971–987.

[17] Hall, R. E., Mishkin, F. S. (1982). The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households. Econometrica 50:461–482

[18] Hosking, J.(1981) Fractional differencing. Biometrika 68: 165–178.

[19] Hylleberg, S. Engle, R. F., Granger, C. W. J., Yoo, B. S., (1990) Seasonal integration and cointegration. Journal of Econometrics 44: 215–238.

[20] Lawrence, E. (1991). Poverty and the Rate of Time Preference: Evidence from Panel Data. Journal of Political Economy 99:54–77.

[21] Lewbel, A. (1994) Aggregation and simple dynamics. American Economic Review 84: 95–918.

[22] Lippi, M. , Zaffaroni, P., (1998) Aggregation of simple linear dynamics: exact asymptotic results. Econometrics Discussion paper 350, STICERD-LSE.

[23] Oppenheim, G., Viano, M. C., (2004) Aggregation of random parameters Ornstein-Uhlenbeck or AR processes: some convergence results. Journal of Time Series Analysis 25: 335–350

[24] Pesaran M. H. (2003) Aggregation of linear dynamic models: an application to the life cycle hypothesis under habit formation. Economic Modeling 20: 385–435

[25] Pischke, J.-S., (1995) Individual income, incomplete information, and aggregate consumption. Econometrica 63: 805–840.

[26] Robinson, P. M. (1978). Statistical Inference for a random coefficient autoregressive model. Scandinavian Journal of Statistics, 5, 163–168.

[27] Robinson, P. M. (1994). Semiparametric analysis of long memory time series. Annals of Statistics, 22: 515–539.

[28] Samorodnitsky, G., Taqqu, M., (1994) Stable non-Gaussian Processes: Stochastic models with Infinite Variance. Chapman and Hall, New York, London.

[29] Thornton, M. A., (2009) Information and Aggregation: The Econometrics of Dynamic Models of Consumption under Cross-Sectional and Temporal Aggregation. University of Essex PhD thesis. Unpublished.

[30] Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. Journal of Econometrics. 120: 75–102.

[31] Zeldes, S. P. (1989). Consumption and Liquidity Constraints: An Empirical Investigation. Journal of Political Economy. 97:305–46.

Appendix

Proof of Proposition 1

(i) First note that the quantity

$$X_{N,t}^U = \sum_{i=1}^{N} \left[\frac{\psi_i^2 \theta_i}{(1-\theta_i)^2} \frac{1}{(1-\theta_i^2)} \right] \tag{A1}$$

will be finite as long as the contribution for households i whose θ_i is in the locality of one is finite. This will depend on whether,

$$\sum_{i=1}^{N} \left[\frac{c_a^2}{(1-\theta_i)^{2-2a}} \frac{1}{(1-\theta_i^2)} \right], \tag{A2}$$

is finite.

Then we note that

$$\sum_{i=1}^{N} \frac{1}{2} \left[\frac{1}{(1-\theta_i)^{3-2a}} \right] \leq \sum_{i=1}^{N} \left[\frac{1}{(1-\theta_i)^{2-2a}} \frac{1}{(1-\theta_i^2)} \right] \leq \sum_{i=1}^{N} \left[\frac{1}{(1-\theta_i)^{3-2a}} \right]. \tag{A3}$$

Given that c_a is a constant and θ_i is bounded at one, we can apply Zaffaroni (2004) lemma 1 (pp96), with $k = 3 - 2a$ and for $\delta = \frac{b+1}{k}$ to get

for $\delta > 1$ that is $2a + b > 2$

$N^{-1} X_{N,t}^U \overrightarrow{a.s.} \; c_a E \left\{ \frac{1}{(1-\theta_i)^{3-2a}} \right\} < \infty,$

and for $\delta \leq 1$,

$N^{-1/\delta} X_{N,t}^U \overrightarrow{d} \; S_\delta.$

In the first case, $\delta > 1$, we have a finite quantity when dividing by N, and so the first part of the proposition holds by Kolmogorov's law of large numbers. Indeed in the second case, $\delta \leq 1$, $N^{-2} X_{N,t}^U \to 0$ so long as $N^{1/\delta}$ is growing more slowly than N^2, in other words as long as $\delta > 1/2$.

Then we note that if,

$N^{-2}X^U_{N,t} \to 0$ then $V^U_{N,t} \overrightarrow{a.s.}\ 0$,

and if,

$N^{-1/\delta}X^U_{N,t} \ \overrightarrow{d}\ S_\delta$ then $N^{(2-1/\delta)}V^U_{N,t} \ \overrightarrow{d}\ S_\delta$,

since the additional elements in $V^U_{N,t}$ are bounded, and bounded above 0, when θ is in the locality of 1[11]: $(1-\frac{\theta_i}{1+r})^2$ is bounded by $(\frac{r}{1+r})^2$ and 1, θ_i is bounded above by one, and c_a, c_b are constants. The proof is completed by noting that $\delta > 1/2 \Leftrightarrow (b+1)/3 - 2a > 1/2 \Leftrightarrow a+b > 1/2$ and that $2 - 1/\delta = (2b+2a-1)/(b+1)$.

(ii) We now consider the variance of $\Delta U_{N,t}$. First we note that, following the expansion in (14), the equivalent quantity to (A1) is

$$X^{\Delta U}_{N,t} = \sum_{i=1}^{N} 2\frac{\psi_i^2\theta_i}{(1-\theta_i)^2}\frac{(1-\theta_i)}{(1-\theta_i^2)}.$$

Following an identical argument to part (i), this will be finite in the limit if

$$\sum_{i=1}^{N}\left[\frac{c_a^2}{(1-\theta_i)^{2-2a}}\frac{(1-\theta_i)}{(1-\theta_i^2)}\right], \tag{A4}$$

is finite.

Then we note that

$$\sum_{i=1}^{N}\left[\frac{1}{2(1-\theta_i)^{2-2a}}\right] \le \sum_{i=1}^{N}\left[\frac{1}{(1-\theta_i)^{2-2a}}\frac{(1-\theta_i)}{(1-\theta_i^2)}\right] \le \sum_{i=1}^{N}\left[\frac{1}{(1-\theta_i)^{2-2a}}\right]. \tag{A5}$$

We then apply Zaffaroni lemma 1 with $k = 2-2a$. This part goes to a delta stable random variable when $\delta = \frac{b+1}{2-2a} \le 1/2$, in other words when $a+b \le 0$. And in this case $2 - 1/\delta = 2 - (2-2a)/(b+1) = 2(a+b)/(b+1)$.

◇

Proof of Proposition 2

(i) Firstly we notice that

$$\sum_{i=1}^{N}\frac{1}{\left|1-\theta_i e^{-i\lambda}\right|^2}, \tag{A6}$$

is finite for $\lambda \neq 0$ and decreasing in λ. The finiteness of the spectral density away from the origin will then depend on the finiteness of

$$\sum_{i=1}^{N}\frac{1}{N^2}\left[\frac{c_a^2}{(1-\theta_i)^{2-2a}}\right]. \tag{A7}$$

The proof of (i) then follows similar arguments to those used in proposition 1 to establish

[11] Another, equivalent explanation is to note that, in keeping with Zaffaroni's lemma, following the transformation $z = (1-\theta)^{-1}$, any finite polynomial of θ is inevitably a slowly varying function of z

the properties of the variance $V^U_{N,t}$, now with $k = 2(1-a)$. Under this condition $\delta > 1/2 \Leftrightarrow (b+1)/2(1-a) > 1/2 \Leftrightarrow a+b > 0$ and that $2 - 2(1-a)/(b+1)/ = 2(a+b)/(b+1)$.

(ii) We note that the spectral density of the idiosyncratic component at the origin is

$$S^U_{N,t}(0) = \sum_{i=1}^{N} \frac{1}{2\pi N^2} \left[\left(1 - \frac{\theta_i}{1+r}\right)^2 \frac{\psi_i^2 \theta_i}{(1-\theta_i)^2} \frac{1}{(1-\theta_i)^2} \right]. \tag{A8}$$

The proof of (ii) then follows that for proposition 1 with $k = 2(2-a)$. Like the variance, the spectral density at the origin is either infinite or 0. The delineating factor is whether $\delta \leq 1/2 \Leftrightarrow (b+1)/2(2-a) \leq 1/2 \Leftrightarrow a+b \leq 1$.

◇

Proof of Proposition 3

First we define $x^u_{i,t} = \sum_{j=0}^{\infty} \theta_i^j \upsilon_{i,t-j}$.

(i) We consider first the case when $NV^U_{N,t}$ is bounded. Following the proof of proposition 1(i) (when $\delta > 1$), that is when $2a+b > 2$. And hence, as $N \to \infty$

$$\frac{U_{N,t}}{\sqrt{V^U_{N,t}}} = \left[NV^U_{N,t}\right]^{-1/2} \left[1/N^{1/2} \sum_{i=1}^{N} \left[\left(1 - \frac{\theta_i}{1+r}\right) \frac{(1-L^m)}{(1-L)}\right] x^u_{i,t}\right] \vec{d}\ U_t, \tag{A9}$$

where U_t are distributed N(0,1), by the Lindberg-Levy central limit theorem. The memory of this process will be determined by the autocovariance of the $x^u_{i,t}$, $Cov\left\{\frac{1}{N^{1/2}} x^u_{i,t}, \frac{1}{N^{1/2}} x^u_{i,t-s}\right\}$, which for large s will be determined by those individuals with values for θ_i in the region of one, since it is their shocks that persist longest. In this case we can write, as $N \to \infty$,

$$\begin{aligned} Cov&\left\{\frac{1}{N^{1/2}} \sum_{i=1}^{N} x^u_{i,t}, \frac{1}{N^{1/2}} \sum_{i=1}^{N} x^u_{i,t-s}\right\} = \sum_{i=1}^{N} \frac{\sigma^2_{\upsilon,i}}{N} \frac{\theta_i^{s+1}}{(1-\theta_i^2)} \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{c_a^2 \theta_i^{s+2}}{(1-\theta_i^2)(1-\theta_i)^{2-2a}} \to cE\left\{\frac{\theta_i^{s+2}}{(1-\theta_i)^{3-2a}}\right\}, \end{aligned} \tag{A10}$$

deploying the argument used to bound (A3) and with c an arbitrary constant. Since, for integer s, by repeated integration by parts,

$$\int_0^1 \theta^{s+2}(1-\theta)^{2a+b-3} d\theta = \frac{\Gamma(s+3)\Gamma(2a+b-2)}{\Gamma(s+2a+b+1)}, \tag{A11}$$

where $\Gamma(x)$ denotes the Gamma function, then by Stirling's formula for large s,

$$E\left\{\frac{\theta_i^{s+2}}{(1-\theta_i)^{3-2a}}\right\} = c\frac{\Gamma(s+3)\Gamma(2a+b-2)}{\Gamma(s+2a+b+1)} \sim cs^{(2-b-2a)}, \tag{A12}$$

with c an arbitrary constant that is not the same throughout. This equates to a memory parameter of $d_U = 1/2(3-2a-b)$.

(ii) Following the proof of proposition 1(ii), when $a+b > 0$, then $NV^{\Delta U}_{N,t}$ is bounded. We can then deploy the arguments used for part (i). Noting that, following (15),

$$Cov\left\{\frac{1}{N^{1/2}}\sum_{i=1}^{N}\Delta x_{i,t}^{u}, \frac{1}{N^{1/2}}\sum_{i=1}^{N}\Delta x_{i,t-s}^{u}\right\} = \sum_{i=1}^{N} -\frac{\sigma_{v,i}^{2}}{N}\frac{\theta_i^s(1-\theta_i)^2}{(1-\theta_i^2)}$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\frac{\theta_i^{s+1}(1-\theta_i)}{(1-\theta_i^2)(1-\theta_i)^{1-2a}} \to -cE\left\{\frac{\theta_i^{s+1}}{(1-\theta_i)^{2-2a}}\right\}, \tag{A13}$$

with c an arbitrary constant. It follows that for large s, applying Stirling's formula,

$$E\left\{\frac{\theta_i^{s+1}}{(1-\theta_i)^{1-2a}}\right\} = c\frac{\Gamma(s+2)\Gamma(2a+b)}{\Gamma(s+2a+b+2)} \sim cs^{(-b-2a)}, \tag{A14}$$

with c an arbitrary constant. This equates to a memory parameter of $1/2(1-2a-b) = d_U - 1$.

Note that the result from (ii) can be used to show that the memory of $U_{N,t}$ is $d_U = 1/2(3-2a-b)$ for cases where $a+b > 0$. The memory of $U_{N,t}$ when $a+b \leq 0$, when not even $NV_{N,t}^{\Delta U}$ is bounded, can be established using theorem 3 in Zaffaroni (2004) with $2a+b-2$ replacing b.

◇

Proof of Proposition 4

The proof is very close to that for proposition 1, with the obvious difference that, as ϵ_t is a common shock then its average does not follow the law of large numbers argument used with the idiosyncratic shock. The task is to determine when $V_{N,t}^{E}$ is finite and when it is not. First note that the quantity

$$X_{N,t}^{E} = \sum_{i,j=1}^{N}\left[\frac{\psi_i\psi_j}{(1-\theta_i\theta_j)}\right], \tag{A15}$$

will be finite as long as the contribution for households i, j with values θ_i, θ_j are in the locality of 1 is finite. This will depend on whether,

$$\sum_{i,j=1}^{N}\left[\frac{c_a^2(1-\theta_i)^a(1-\theta_j)^a}{(1-\theta_i\theta_j)}\right], \tag{A16}$$

is finite. To determine the order of k in the expression of the form (13) that translated to this expression, we notice that (A16) can be bounded by

$$\sum_{i=1}^{N}\left[\frac{c_a^2(1-\theta_i)^{2a}}{(1-\theta_i^2)}\right] \leq \sum_{i,j=1}^{N}\left[\frac{c_a^2(1-\theta_i)^a(1-\theta_j)^a}{(1-\theta_i\theta_j)}\right] \leq \left[\sum_{i=1}^{N}\frac{c_a(1-\theta_i)^a}{(1-\theta_i^2)^{1/2}}\right]^2. \tag{A17}$$

To determine finiteness, we first consider the lower bound. Following similar arguments to those used on equation (A3), we determine this expression to be of order $k = 1-2a$ and hence to be bounded if $\delta > 1/2 \Leftrightarrow \frac{b+1}{1-2a} > 1/2 \Leftrightarrow a+b > -1/2$. Otherwise it will go to a delta stable random variable of order $2 - 1/\delta = (2b+2a+1)/(b+1)$. Then we consider the upper bound. Since the quantity inside the [.] is to be squared, it will only be bounded if it goes to a delta stable variable of order > 1. That is it will be bounded provided $\delta > 1 \Leftrightarrow \frac{b+1}{1/2-a} > 1 \Leftrightarrow a+b > -1/2$. Otherwise it will go to a delta stable variable of order $2 - 2/\delta = (2b+2a+1)/(b+1)$. Since both the upper and lower bounds go to the same distribution for the same values of a and b, the result follows.

◇

Proof of Proposition 5

The finiteness of $S^E_{N,t}(\lambda)$ away from the origin and the $M-1$ zeros induced by temporal aggregation depends on the finiteness of

$$S^E_{N,t}(\lambda) = \frac{1}{2\pi N^2}\left|\sum_{i=1}^{N}\left(1-\frac{\theta_i}{1+r}\right)\frac{\psi_i}{1-\theta_i e^{-i\lambda}}\right|^2. \tag{A18}$$

$\left|1-\theta_i e^{-i\lambda}\right|^{-1}\left|1-\theta_j e^{i\lambda}\right|^{-1}$ is finite for all i,j for $\lambda \neq 0$ and decreasing in λ. The finiteness of the spectral density away from the origin will then depend on the finiteness of

$$\sum_{i,j=1}^{N}\frac{1}{N^2}\left[c_a^2(1-\theta_i)^a(1-\theta_j)^a\right]. \tag{A19}$$

The proof of (i) then follows similar arguments to those used in proposition 4, noting that

$$\sum_{i=1}^{N}\left[c_a^2(1-\theta_i)^{2a}\right] \leq \sum_{i,j=1}^{N}\left[c_a^2(1-\theta_i)^a(1-\theta_j)^a\right] \leq \left[\sum_{i=1}^{N}c_a(1-\theta_i)^a\right]^2. \tag{A20}$$

The lower bound is amenable to Zaffaroni (2004) lemma 1, and will be finite as long as $\frac{b+1}{-2a} > 1/2 \Leftrightarrow b+a > -1$. The upper bound, which is a squared term, will be finite as long as $\frac{b+1}{-a} > 1 \Leftrightarrow b+a > -1$. In both instances the bound for finiteness is the same.

The proof for (ii) is also similar. We note that the spectral density of the idiosyncratic component at the origin is

$$S^E_{N,t}(0) = \frac{1}{2\pi N^2}\left[\sum_{i=1}^{N}\left(1-\frac{\theta_i}{1+r}\right)\frac{\psi_i}{1-\theta_i}\right]^2. \tag{A21}$$

We then note that

$$\sum_{i=1}^{N}\left[c_a^2\frac{(1-\theta_i)^{2a}}{(1-\theta_i)^2}\right] \leq \sum_{i,j=1}^{N}\left[c_a^2\frac{(1-\theta_i)^a(1-\theta_j)^a}{(1-\theta_i)(1-\theta_j)}\right] \leq \left[\sum_{i=1}^{N}c_a\frac{(1-\theta_i)^a}{(1-\theta_i)}\right]^2. \tag{A22}$$

Similar arguments to part (i) lead to the conclusion that if $\frac{b+1}{2-2a} > 1/2$, $\frac{b+1}{1-a} > 1 \Leftrightarrow b+a > 0$ then $S^E_{N,t}(0)$ is finite.

◇

Proof of Proposition 6

The proof follows from Lippi and Zaffaroni (1998) Theorem 11, using (18) allowing $a+b$ to replace b.

◇

Data Source

The data are taken from the University of Michigan's Panel Study of Income Dynamics (PSID). The PSID is an annual survey of the households in its sample and any split-offs: new households including members who had been in households included in the sample. Household consumption data are restricted to questions on food: consumed inside the home; outside the home; and, purchased by food stamps. The longest unbroken spell for these data is the 14 years from 1974 to 1987. The sample considered here consists of those households who existed from 1972 to 1987, in order to provide instruments at sufficient lags, and live in the 48 contiguous states, in order to make common shocks easier to identify. In order to avoid additional variation due to a change in the main decision maker, the sample is restricted to those households who maintained the same head throughout this period. Households with a retired head are removed. As part of data collection, the degree of imputation in the responses given to some of the questions is recorded by the interviewer. Our sample is restricted to households whose answers needed minimal imputation. Since the latter part of the chapter examines heterogeneous dynamic models, upon which the extremes of the underlying distribution is most important, the only additional households who are removed as outliers are those who claim to spend nothing in one period. This leaves a balanced panel of 736 households. The data are weighted using the Bureau of Labor statistics series for food inside and outside the home, with 1984 as the base year.

The PSID also records a number of other relevant variables for households, including an estimate of the food needs of the household based on its composition. Other demographic variables of potential interest include the age[12], gender and race of the head. Figures for household income, employment, hours worked and whether the household received any lump-sum payments are also collected for the previous calendar year.

[12]Since data collection is only roughly annual it is possible for the head of the household to have zero, one or two birthdays between surveys. Where lagged demographic variables are used in estimates, however, the lag of age is set to the current age minus one. This ensures the perfect collinearity between age, constant and lag age that means that the latter can be omitted from any partialling procedures with no effect, rather than leaving three series that are very nearly collinear