# Aggregation of Linear Dynamic Models: An Application to Life-Cycle Consumption Models Under Habit Formation*

M. Hashem Pesaran
Trinity College, Cambridge

May 2002

**Abstract**

This paper provides a general framework for aggregating linear dynamic models by deriving the aggregate model as the optimal prediction (in the minimum mean-squared error sense) of the aggregate variable of interest with respect to an aggregate information set generated by current and past values of available aggregate observations. The approach is applied to a number of aggregation problems that have been considered in the literature. It is shown how the results in much of the literature can be readily obtained using the proposed forecasting approach, and a number of important extensions and generalizations are provided. Our approach does not require the assumption of independence of the micro distributed lag coefficients from the other micro coefficients, and establishes that in general the long-run coefficients obtained from the optimal aggregate relation are equal to the averages of the long-run coefficients from the micro relations. The approach is then applied to life-cycle consumption decision rules under habit formation and the implications of the heterogeneity in habit formation coefficients across individuals for the analysis of aggregate consumption is investigated. Using stochastic simulations it is shown that the estimates of the habit persistence coefficient are likely to be biased downward if they are based on analogue aggregate consumption functions, which could partly explain the excess smoothness and excess sensitivity puzzles in terms of neglected heterogeneity.

**Key Words:** Aggregation, Heterogeneous Dynamic Models, Long Memory, Life Cycle Models under Habit Formation.

**JEL Classifications:** C10, C43, D91, E21.

*Correspondences to: M. Hashem Pesaran, Faculty of Economics and Politics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD, England. The author wishes to thank Kees Jan van Garderen and particularly Michael Binder and Paolo Zaffaroni for their helpful comments on a preliminary version of this paper. The simulations reported in Section 6.2 were carried out by Andreas Pick.

# 1 Introduction

The aggregation problem is an inevitable aspect of applied research in economics. It arises primarily because behavioral relations in economics are generally derived as decision rules of individual economic agents, while many of the relations that applied economists are interested in studying are subject to aggregation across commodities, households, firms, regions, or time, and often over all these dimensions. Naturally, the problem is more pervasive in the case of macroeconomic research, but it also tends to be present in applied microeconomic analysis.[1]

There are a variety of reactions to the aggregation problem. At one extreme many investigators have chosen to ignore it altogether, arguing implicitly that empirically the aggregation problem is of second order importance. At the other extreme to avoid the problem some have opted for a highly disaggregated general equilibrium approach where the possibility of deriving testable restrictions on observable time series is extremely limited if not impossible. The focus of this paper is on a number of intermediate positions between these two extremes. Currently, there are two basic constructive approaches to the aggregation problem: the "deterministic" approach originally explored by Gorman (1953), Klein (1953), Theil (1954), Malinvaud (1970), and Muellbauer (1975); and the "statistical" or "stochastic" approach advanced by Kelejian (1980), Stoker (1984), Lippi (1988), Lewbel (1994), and Forni and Lippi (1997). [2]

The deterministic approach to aggregation is unduly restrictive and requires the aggregate function to match exactly the sum of the micro functions for *all* realizations of the disaggregate variables. The statistical (stochastic) approach is less restrictive and induces relationships between the population aggregates from the joint probability distribution of the micro variables and the parameters of the micro equations. While this is clearly an advance over the deterministic approach, it is nevertheless rather complicated to apply in practice and has not been found to be directly suitable for econometric analysis. There is also no guarantee that the induced aggregate relation should always exist or be unique.

With the aim of developing a unified theory of aggregation for econometric analysis, I shall propose an "optimal aggregate forecasting approach". This approach views aggregation as a forecasting problem where the focus of the analysis is on the optimal prediction (with respect to a particular loss function of interest, here minimum mean-squared error) of the *aggregate* variables

[1] For example, in the case of microeconometric studies of household consumption, the issue of commodity aggregation and the associated index number problem has for long been the subject of intense research. See Gorman (1953) or Muellbauer (1975). Similar considerations also arise in the microeconometric analysis of households' labour supply, firms' investment and employment decisions, and governments' expenditure decision.

[2] A precursor to the statistical approach to the aggregation problem can be found in Houthakker (1955/56) and Johansen (1972) who consider the problem of aggregation of technologies across production units. For an excellent review of the aggregation literature see Stoker (1993). The introduction in Barker and Pesaran (1990) could also be of interest.

conditional on available *aggregate* information. This approach starts with the probabilistic formulation of the statistical approach, but has the virtue that the aggregate function derived as the conditional optimal forecast exists under relatively weak assumptions concerning the existence of conditional expectations of the micro relations and furthermore is unique.[3] The paper shows that under the assumption that the disaggregated model is correctly specified, the mean-squared error of the optimal aggregate forecast is larger than the corresponding mean-squared error of forecasting the aggregate based on the disaggregated model, but smaller than the mean-squared error of forecasting the aggregate based on an *ad hoc* aggregate function, such as the macro analogue of the micro relations.

The optimal aggregate forecasting approach can be applied to a variety of problems. This paper focuses on aggregation of linear autoregressive distributed lag models in general, and in specific detail on aggregation of life-cycle decision rules under habit formation. The former problem has attracted considerable attention in the time series literature. (See, for example, Granger and Morris (1976), Rose (1977), Granger (1980), Trivedi (1985), Lippi (1988), Lewbel (1994), and Zaffaroni (2001), and the contributions of Granger and Forni and Lippi in Barker and Pesaran (1990)). It will be shown how the various results in this literature can be readily obtained using the approach advanced in this paper.[4] A number of extensions and generalizations will also be provided. For example, it is typically assumed that the micro disturbances are independently distributed across the micro units, and that the micro distributed lag coefficients are distributed independently from the other micro coefficients in the disaggregated model. See, for example, Lewbel (1994). The statistical approach followed by some of the above contributions abstracts from aggregation errors and thus overlooks their significance for empirical analysis. We relax these assumptions and derive optimal aggregate functions in a more general setting, paying particular attention to the aggregation errors that are inevitably involved, and show that the long-run coefficients obtained from the optimal aggregate equation are in fact equal to the averages of the long-run coefficients from the micro relations. A correspondence result is also established between the mean lag from the aggregate function and the average of the mean lags from the micro relations. This equivalence, however, requires the independence of the long-run effects and the mean lags at the micro level.

In view of the recent interest in habit formation models of consumption, Section 6 presents an analysis of aggregation of the life-cycle consumption decision rules under habit formation allowing for possible heterogeneity in the habit formation coefficients across individual consumers. It is shown that the optimal aggregate consumption function can be markedly different from the analogue function based on a representative consumer. In particular, unlike the analogue function the optimal

[3]It is clearly possible to derive optimal aggregate functions with respect to non-quadratic loss functions of the type discussed, for example, in Christoffersen and Diebold (1996) and Granger and Pesaran (2000).

[4]The optimal aggregate forecasting approach can also be applied to non-linear systems. For an application to static non-linear models with random coefficients see Garderen, Lee, and Pesaran (2000).

aggregate function in general cannot be represented as a finite-order autoregressive distributed lag model in consumption and labour income. The quantitative implications of the use of aggregate consumption data for the estimation of the structural parameters is also explored by means of stochastic simulations. It is shown that under heterogeneous habit formation the estimates of the structural parameters based on the analogue function can be very misleading, while the use of the optimal aggregate function generally leads to estimates that are quite close to their true values. Monte Carlo simulations suggest that the estimates of habit formation coefficients obtained from the analogue (representative agent) aggregate consumption function can be seriously biased downward, providing a possible explanation for the "excess smoothness" and "excess sensitivity" puzzles encountered in the empirical consumption function literature.

## 2 A General Framework for Micro (Disaggregate) Behavioral Relationships

Aggregation of behavioral or technical relations across individuals becomes a problem when there is some form of heterogeneity across individuals' relations. When individuals are identical in every respect and the associated micro relations are homogeneous, aggregation will not be a problem. This is, however, extremely unlikely to be the case in practice. Sources of heterogeneity include:

- input variables (heterogeneous initial endowments)
- micro parameters (heterogeneous coefficients)
- micro functionals (heterogeneous preferences and/or production functions)

Let the micro behavioral relationship be represented as

$$\mathbf{y}_{it} = \mathbf{f}_i\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right), \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{2.1}$$

where $\mathbf{y}_{it}$ denotes the vector of decision variables, $\mathbf{x}_{it}$ is a vector of observable variables, $\mathbf{u}_{it}$ is a vector of unobservable variables, and $\boldsymbol{\theta}_i$ denotes the vector of unknown parameters.

**Example 2.1** *When the source of heterogeneity is different inputs (or endowments) across individuals only, we have*

$$\mathbf{y}_{it} = \mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}\right), \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T. \tag{2.2}$$

*Such a scenario may arise in the analysis of nonlinear Engel curves,*

$$w_{it} = a_0 + a_1 \log x_{it} + a_2 \left(\log x_{it}\right)^2 + u_{it}, \tag{2.3}$$

*or in the analysis of (Cobb Douglas) production functions of the form*

$$y_{it} = AL_{it}^{\alpha} K_{it}^{1-\alpha} e^{u_{it}}. \tag{2.4}$$

*For this type of heterogeneity, aggregation clearly will not be a problem when the micro relations are linear.*

**Example 2.2** *When the input variables as well as the parameters differ across individuals, we have*

$$\mathbf{y}_{it} = \mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right), \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T. \tag{2.5}$$

*In the analysis of nonlinear Engel curves, such a scenario arises, for example, if the model is given by*

$$w_{it} = a_{0i} + a_{1i} \log x_{it} + a_{2i} \left(\log x_{it}\right)^2 + u_{it}. \tag{2.6}$$

**Example 2.3** *It is also possible that there is heterogeneity in the functional form of the micro relations, for example a production function of the form*

$$y_{it} = \left(\lambda_i L_{it}^{-\delta_i} + \left(1 - \lambda_i\right) K_{it}^{-\delta_i}\right)^{-1/\delta_i} e^{u_{it}}. \tag{2.7}$$

In this paper I consider the case where $\mathbf{f}(\cdot)$ is the same across individuals, but the input variables $\mathbf{x}_{it}$ and $\mathbf{u}_{it}$, and/or the parameters $\boldsymbol{\theta}_i$ differ across individuals. The analysis can also be easily extended to account for observed and unobserved macro (or aggregate) effects on individual behavior, namely

$$\mathbf{y}_{it} = \mathbf{f}\left(\mathbf{x}_{it}, \mathbf{z}_t, \mathbf{u}_{it}, \mathbf{v}_t, \boldsymbol{\theta}_i\right), \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{2.8}$$

where $\mathbf{z}_t$ represents a vector of observed macro effects, and $\mathbf{v}_t$ represents a vector of unobserved macro effects.

# 3 Alternative Notions of Aggregation

## 3.1 Deterministic Aggregation

This approach, employed for example by Gorman (1953) and Theil (1954), treats all the input variables and parameters as given and asks whether an aggregate function exists which is identical to the function that results from the aggregation of the micro relations. Let $\mathbf{Y}_t = N^{-1} \sum_{i=1}^{N} \mathbf{y}_{it}$. Then aggregating (2.1) under $\mathbf{f}_i(\cdot) = \mathbf{f}(\cdot)$ across all $i$, taking $\mathbf{x}_{it}$, $\mathbf{u}_{it}$, and $\boldsymbol{\theta}_i$ as given, we have

$$\mathbf{Y}_t = N^{-1} \sum_{i=1}^{N} \mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right). \tag{3.9}$$

An aggregation problem is said to be present if the aggregate function $\mathbf{F}(\mathbf{X}_t, \mathbf{U}_t, \boldsymbol{\theta}_a)$ (with $\mathbf{X}_t = N^{-1}\sum_{i=1}^{N}\mathbf{x}_{it}$, $\mathbf{U}_t = N^{-1}\sum_{i=1}^{N}\mathbf{u}_{it}$, and where $\boldsymbol{\theta}_a$ is the vector of parameters of the aggregate function) differs from $N^{-1}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i)$. Perfect aggregation holds if

$$\left\| \mathbf{F}(\mathbf{X}_t, \mathbf{U}_t, \boldsymbol{\theta}_a) - N^{-1}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i) \right\| = 0 \tag{3.10}$$

for all $\mathbf{x}_{it}$, $\mathbf{u}_{it}$, and $\boldsymbol{\theta}_i$, where $\|\mathbf{a} - \mathbf{b}\|$ denotes a suitable norm discrepancy measure between $\mathbf{a}$ and $\mathbf{b}$. This requirement turns out to be extremely restrictive and is rarely met in applied economic analysis, except for linear models with identical coefficients. Condition (3.10) is not satisfied when $\mathbf{f}(\cdot)$ is a nonlinear function of $\mathbf{x}_{it}$ and $\mathbf{u}_{it}$, even if $\boldsymbol{\theta}_i$ is identical across individuals.

## 3.2 A Statistical Approach to the Aggregation Problem

The restrictive nature of the deterministic aggregation condition (3.10) arises primarily because it requires the condition to be satisfied for all realizations of $\mathbf{x}_{it}$, $\mathbf{u}_{it}$, and $\boldsymbol{\theta}_i$, no matter how remote the possibility of their occurrence. An alternative and less restrictive approach would be to require that (3.10) holds "on average". More precisely, let $\boldsymbol{\mu}_{\mathbf{y}}(t)$ and $\boldsymbol{\mu}_{\mathbf{x}}(t)$ be the means of $\mathbf{y}_{it}$ and $\mathbf{x}_{it}$ across individuals at a point in time or over a given period of time (depending on whether the variables are stocks or flows) and define a macro (or aggregate) relation as one that links $\boldsymbol{\mu}_{\mathbf{y}}(t)$ to $\boldsymbol{\mu}_{\mathbf{x}}(t)$ at a point in time $t$. This approach was suggested by Kelejian (1980) and rigorously formalized by Stoker (1984). It treats $\mathbf{x}_{it}$, $\mathbf{u}_{it}$, and $\boldsymbol{\theta}_i$ across individuals as stochastic, having a joint probability distribution function $P(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i; \boldsymbol{\phi}_t)$ with parameter vector $\boldsymbol{\phi}_t$ that could vary over time, but not across individuals. Then

$$\boldsymbol{\mu}_{\mathbf{y}}(t) = \Psi_{\mathbf{y}}(\boldsymbol{\phi}_t) = \int \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\theta})\; P(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\theta}; \boldsymbol{\phi}_t)\; d\mathbf{x}_t\, d\mathbf{u}_t\, d\boldsymbol{\theta}, \tag{3.11}$$

and

$$\boldsymbol{\mu}_{\mathbf{x}}(t) = \Psi_{\mathbf{x}}(\boldsymbol{\phi}_t) = \int \mathbf{x}\, P(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\theta}; \boldsymbol{\phi}_t)\; d\mathbf{x}_t\, d\mathbf{u}_t\, d\boldsymbol{\theta}. \tag{3.12}$$

Let $\boldsymbol{\phi}_t = (\boldsymbol{\phi}_{1t}, \boldsymbol{\phi}_{2t})'$, where $\boldsymbol{\phi}_{2t}$ has the same dimension as $\mathbf{x}_{it}$, for all $i$, and suppose that for a given $\boldsymbol{\phi}_{1t}$ there is a one-to-one relationship between $\boldsymbol{\phi}_{2t}$ and $\boldsymbol{\mu}_{\mathbf{x}}(t)$. Then

$$\boldsymbol{\phi}_{2t} = \Psi_{\mathbf{x}}^{-1}(\boldsymbol{\phi}_{1t}, \boldsymbol{\mu}_{\mathbf{x}}(t)), \tag{3.13}$$

and

$$\boldsymbol{\mu}_{\mathbf{y}}(t) = \Psi_{\mathbf{y}}\left[\boldsymbol{\phi}_{1t}, \Psi_{\mathbf{x}}^{-1}(\boldsymbol{\phi}_{1t}, \boldsymbol{\mu}_{\mathbf{x}}(t))\right] = \mathbf{F}(\boldsymbol{\mu}_{\mathbf{x}}(t), \boldsymbol{\phi}_{1t}). \tag{3.14}$$

The relationship between $\boldsymbol{\mu}_{\mathbf{y}}(t)$ and $\boldsymbol{\mu}_{\mathbf{x}}(t)$ is then defined as the exact aggregate equation.

This is clearly an improvement over the deterministic approach, but it is still rather removed from direct empirical analysis,[5] and does not adequately focus on the inevitably approximate nature of econometric analysis. Also, perhaps more importantly, due to its reliance on unconditional means, this approach is not suitable for the analysis of dynamic systems.

## 3.3 A Forecasting Approach to the Aggregation Problem

Once again consider the exact aggregation condition (3.10) specified for all $\mathbf{x}_{it}$, $\mathbf{u}_{it}$, and $\boldsymbol{\theta}_i$, but now require that conditional on the aggregate information set $\Omega_t = \{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, ...; \mathbf{X}_t, \mathbf{X}_{t-1}, \ldots\}$ the mean of

$$\left\| \mathbf{F}\left(\Omega_t, \boldsymbol{\theta}_{at}\right) - N^{-1} \sum_{i=1}^{N} \mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right) \right\|$$

be as small as possible. For expositional simplicity denote the aggregate function $\mathbf{F}\left(\Omega_t, \boldsymbol{\theta}_{ta}\right)$ by $\mathbf{F}_t$, and $\mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right)$ by $\mathbf{f}_{it}$. Also note that the parameters of the aggregate function, $\boldsymbol{\theta}_{ta}$, will typically include first and higher moments of the joint distribution of $\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right)$ across $i$, and could be time-dependent.

Suppose that $\|a-b\|$ is quadratic, namely $\|a-b\| = (a-b)^2$, where $a$ and $b$ are scalars. Then,

$$\begin{aligned} E\left[\left(\mathbf{F}_t - \mathbf{Y}_t\right)^2 | \Omega_t\right] &= E\left\{\left[\left(\mathbf{F}_t - E\left(\mathbf{Y}_t|\Omega_t\right)\right) - \left(\mathbf{Y}_t - E\left(\mathbf{Y}_t|\Omega_t\right)\right)\right]^2 | \Omega_t\right\} \\ &= E\left\{\left[\mathbf{F}_t - E\left(\mathbf{Y}_t|\Omega_t\right)\right]^2 | \Omega_t\right\} + E\left\{\left[\mathbf{Y}_t - E\left(\mathbf{Y}_t|\Omega_t\right)\right]^2 | \Omega_t\right\} \\ &\quad -2E\left\{\left[\mathbf{F}_t - E\left(\mathbf{Y}_t|\Omega_t\right)\right]\left[\mathbf{Y}_t - E\left(\mathbf{Y}_t|\Omega_t\right)\right] | \Omega_t\right\}, \end{aligned} \tag{3.15}$$

and therefore the prediction that minimizes $E\left[\left(\mathbf{F}_t - \mathbf{Y}_t\right)^2 | \Omega_t\right]$ is given by

$$\mathbf{F}_t = E\left(\mathbf{Y}_t|\Omega_t\right) = N^{-1} \sum_{i=1}^{N} E\left[\mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right) | \Omega_t\right], \tag{3.16}$$

which I will refer to as the "optimal aggregator function" (in the mean-squared error sense). The orthogonal projection used (implicitly or explicitly) by Granger (1980), Lütkepohl (1984), and Lippi (1988) for aggregation of linear time series is a special case of this optimal aggregator which is more widely applicable. For an application to aggregation of static non-linear models see Garderen, Lee, and Pesaran (2000).

This choice of $\mathbf{F}_t$ globally minimizes $E\left[\left(\mathbf{F}_t - \mathbf{Y}_t\right)^2 | \Omega_t\right]$, but does not reduce it to zero, which is what (3.10) requires. We rather have

$$E\left[\left(\mathbf{F}_t - \mathbf{Y}_t\right)^2 | \Omega_t\right] = Var\left(\mathbf{Y}_t|\Omega_t\right) \neq 0, \tag{3.17}$$

unless, of course, $E\left(\mathbf{Y}_t|\Omega_t\right) = \mathbf{Y}_t$.

[5] For empirical analysis it is also required that $\phi_{1t}$ is time invariant.

It is also possible to define an aggregate prediction function, based on individual prediction of $\mathbf{y}_{it}$, conditional on information on *all* the observed disaggregate variables at time $t$. Let

$$\Phi_{it} = \{\mathbf{y}_{it-1}, \mathbf{y}_{it-2}, \ldots ; \mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots\} \tag{3.18}$$

denote the information set specific to individual $i$, and as before denote the information common to all individuals by

$$\Omega_t = \{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots ; \mathbf{X}_t, \mathbf{X}_{t-1}, \ldots\} \tag{3.19}$$

Then

$$\Psi_{it} = \Phi_{it} \cup \Omega_t \tag{3.20}$$

contains the information on the variables in the $i$-th equation, and

$$\Psi_t = \cup_{i=1}^{N} \Psi_{it} \tag{3.21}$$

all information available in the disaggregate model. Then the aggregate forecast, $\mathbf{Y}_{td}$, based on the universal information set, $\Psi_t$, is given by

$$\mathbf{Y}_{td} = N^{-1} \sum_{i=1}^{N} E\left[\mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right) | \Psi_t\right], \tag{3.22}$$

which in most cases simplifies to

$$\mathbf{Y}_{td} = N^{-1} \sum_{i=1}^{N} E\left[\mathbf{f}\left(\mathbf{x}_{it}, \mathbf{u}_{it}, \boldsymbol{\theta}_i\right) | \Psi_{it}\right]. \tag{3.23}$$

Then we have

$$E\left[\left(\mathbf{Y}_t - \mathbf{Y}_{td}\right)^2 | \Psi_t\right] \leq E\left\{\left(\mathbf{Y}_t - E\left(\mathbf{Y}_t | \Omega_t\right)\right)^2 | \Psi_t\right\}, \tag{3.24}$$

and hence

$$E\left(\mathbf{Y}_t - \mathbf{Y}_{td}\right)^2 \leq E\left(\mathbf{Y}_t - E\left(\mathbf{Y}_t | \Omega_t\right)\right)^2, \tag{3.25}$$

which is basically saying that the optimal predictors $\mathbf{Y}_{td}$ that utilize information on micro variables on average are expected to do better than the optimal predictors based on the aggregate information only.

## 4 Cross-Sectional Aggregation of Static Linear Models

The problem of aggregation of static linear models is straightforward and has been extensively discussed in the literature.[6] Here we provide a brief account using the optimal aggregator function developed in the previous section. Consider the linear micro relation

$$y_{it} = \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{4.26}$$

with random coefficients $\boldsymbol{\beta}_i$ (of dimension $k \times 1$) distributed independently of $\mathbf{x}_{it}$ (conditional on $\mathbf{x}_t$). Then denoting the cross-sectional averages of $y_{it}$ by $Y_t = N^{-1} \sum_{i=1}^{N} y_{it}$ we have

$$Y_t = N^{-1} \sum_{i=1}^{N} \boldsymbol{\beta}_i' \mathbf{x}_{it} + N^{-1} \sum_{i=1}^{N} u_{it}. \tag{4.27}$$

Our approach to deriving an optimal aggregate forecast involves two steps: First, we take conditional expectations of the aggregated relation, in this example given by (4.27), with respect to the universal information set, $\Psi_t = \cup_{i=1}^{N} \Psi_{it}$, defined by (3.21), and make use of the assumption that $E(\boldsymbol{\beta}_i \mid \Psi_t) = E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$. In the case of the present application this yields

$$\begin{aligned} E\left(Y_t | \Psi_t\right) &= N^{-1} \sum_{i=1}^{N} E(\boldsymbol{\beta}_i' \mid \Psi_t) \mathbf{x}_{it} + N^{-1} \sum_{i=1}^{N} E\left(u_{it} | \Psi_t\right) \\ &= \boldsymbol{\beta}' \mathbf{X}_t + N^{-1} \sum_{i=1}^{N} E\left(u_{it} | \Psi_t\right). \end{aligned}$$

In the second step, to obtain the forecast of the aggregates, $Y_t$, with respect to the aggregate information set, $\Omega_t$, we now take conditional expectations of the above relation with respect to $\Omega_t \subset \Psi_t$. This now yields

$$E\left(Y_t | \Omega_t\right) = \boldsymbol{\beta}' \mathbf{X}_t + N^{-1} \sum_{i=1}^{N} E\left(u_{it} | \Omega_t\right). \tag{4.28}$$

Assuming that $E\left(u_{it} | \Omega_t\right) = 0$ for all $i$ and $t$, the optimal forecast of the aggregate series $\{Y_t\}$ will be

$$E\left(Y_t | \Omega_t\right) = \boldsymbol{\beta}' \mathbf{X}_t, \tag{4.29}$$

which can be equivalently written in the form of the following aggregate regression function:

$$Y_t = \boldsymbol{\beta}' \mathbf{X}_t + \varepsilon_t, \tag{4.30}$$

where

$$\varepsilon_t = Y_t - E\left(Y_t | \Omega_t\right). \tag{4.31}$$

[6] See, for example, Theil (1954).

By construction $E(\varepsilon_t|\Omega_t) = 0$, and noting that $\mathbf{X}_t \subset \Omega_t$, we also have $E(\varepsilon_t|\mathbf{X}_t) = 0$. Therefore, the least squares regression of $Y_t$ on $\mathbf{X}_t$ will yield a consistent estimator of $\boldsymbol{\beta}$.

In the present case where he $\mathbf{x}_{it}$'s do not contain lagged values of $y_{it}$, the errors of the aggregate regression function, (4.30), are serially uncorrelated, but heteroskedastic. To see this note that for $j \geq 1$

$$E(\varepsilon_t\varepsilon_{t-j}|\Omega_t) = E\left(\varepsilon_t\left(Y_{t-j} - \boldsymbol{\beta}'\mathbf{X}_{t-j}\right)|\Omega_t\right) = \left(Y_{t-j} - \boldsymbol{\beta}'\mathbf{X}_{t-j}\right)E(\varepsilon_t|\Omega_t) = 0, \tag{4.32}$$

and therefore $E(\varepsilon_t\varepsilon_{t-j}) = 0$ for $j \geq 1$. Consider now the conditional variance of $\varepsilon_t$:

$$Var(\varepsilon_t|\Omega_t) = E\left(\varepsilon_t^2|\Omega_t\right) = E\left(Y_t^2|\Omega_t\right) - \boldsymbol{\beta}'\mathbf{X}_t\mathbf{X}_t'\boldsymbol{\beta}, \tag{4.33}$$

where, if the $\mathbf{x}_{it}$'s and $u_{jt}$'s are independently distributed for all $i$,$j$, it may be shown that

$$\begin{aligned} E\left(Y_t^2|\Omega_t\right) &= N^{-1}\left(\boldsymbol{\beta}'\boldsymbol{\mu}_t\boldsymbol{\mu}_t'\boldsymbol{\beta}\right) + N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\mu}_t'\Psi_{\beta,ji}\boldsymbol{\mu}_t \\ &\quad + N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N} tr\left(\Psi_{x,ij}\Psi_{\beta,ji}\right) + N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}, \end{aligned} \tag{4.34}$$

where $\boldsymbol{\mu}_t = E(\mathbf{x}_{it}|\Omega_t)$, $\Psi_{\beta,ij} = E\left(\boldsymbol{\beta}_i\boldsymbol{\beta}_j'|\Omega_t\right) - \boldsymbol{\beta}\boldsymbol{\beta}'$, $\Psi_{\mathbf{x},ij} = E\left(\mathbf{x}_{it}\mathbf{x}_{jt}'|\Omega_t\right) - \boldsymbol{\mu}_t\boldsymbol{\mu}_t'$, and $\sigma_{ij} = E(u_{it}u_{jt}|\Omega_t)$.

# 5 Aggregation of ARDL Models

Consider the simple autoregressive-distributed lag (ARDL) model

$$y_{it} = \lambda_i y_{i,t-1} + \beta_i x_{it} + u_{it}, \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{5.35}$$

and assume that $N$ is large.[7]

**Assumption A.1**: $(\lambda_i, \beta_i)$ are identically and independently distributed of $x_{jt}$ and $u_{jt}$, for all $i, j$ and $t$.

**Assumption A.2:** $|\lambda_i| < 1$ with probability 1 for all $i$, and the micro processes, (5.35), have been initialized at time $t \to -\infty$.

**Assumption A.3:** $x_{is}$'s have finite second-order moments and are distributed independently of $u_{jt}$ for all $i, j$, $t$, and $s \leq t$.

**Assumption A.4:** micro disturbances, $u_{it}$, are serially uncorrelated with mean zero and a finite variance, and admit the following decomposition

$$u_{it} = \varphi_i\eta_t + \xi_{it}, \tag{5.36}$$

[7] A group-specific intercept term can also be included in (5.35).

where $\eta_t$ is the component which is common across all micro units, and $\xi_{it}$ is the idiosyncratic component assumed to be distributed independently across $i$, with a mean zero and a finite variance.

Assumption A.1 is standard in the aggregation and panel literature with random coefficients. The stability conditions, $|\lambda_i| < 1$, for all $i$, can be relaxed at the expense of additional assumptions on the way the micro processes are initialized. Assumption A.3 is required for consistent estimation of the parameters of the aggregate equation and can be relaxed. Assumption A.4 is quite general and allows a considerable degree of dependence across the micro disturbances, $u_{it}$. Also it does not require $\xi_{it}$ and $\varphi_i \eta_t$ to be independently distributed.

To derive the optimal aggregator function, $E(Y_t|\Omega_t)$, one possibility would to work with the autoregressive distributed lag representations, (5.35). But this would involve deriving expectations such as $E(\lambda_i y_{i,t-j}|\Omega_t)$ which is complicated by the fact that $\lambda_i$ and $y_{i,t-j}$ are not independently distributed. To see this notice that under Assumption A.2, (5.35) may be solved for

$$y_{it} = \beta_i \sum_{j=0}^{\infty} \lambda_i^j x_{i,t-j} + \sum_{j=0}^{\infty} \lambda_i^j u_{i,t-j}, \ i = 1, 2, ..., N, \tag{5.37}$$

which makes the dependence of $y_{i,t-j}$ on $\lambda_i$ and $\beta_i$ explicit, and suggests that it might be more appropriate to work directly with the distributed lag representations, (5.37). This is the approach that we shall follow below:

Aggregating (5.37) across all $i$, we have

$$Y_t = N^{-1} \sum_{j=0}^{\infty} \sum_{i=1}^{N} \beta_i \lambda_i^j x_{i,t-j} + N^{-1} \sum_{j=0}^{\infty} \sum_{i=1}^{N} \lambda_i^j u_{i,t-j}, \tag{5.38}$$

where as before $Y_t = N^{-1} \sum_{i=1}^{N} y_{it}$. Introduce the new information set $\Upsilon_{it} = \{x_{it}, x_{i,t-1}, ...\} \cup \Omega_t$ which excludes the individual-specific information on lagged values of $y_{it}$, and let $\Upsilon_t = \cup_{i=1}^{N} \Upsilon_{it}$. Suppose also that $N$ is large enough so that $y_{i,t-j}$, $j = 1, 2, ..$ can not be revealed from the aggregates $Y_{t-1}, Y_{t-2}, ....$ Now, under Assumptions A.1 and A.4

$$E\left(\beta_i \lambda_i^j \mid \Upsilon_t\right) = E\left(\beta \lambda^j\right) = a_j, \tag{5.39}$$

$$E\left(\lambda_i^j \mid \Upsilon_t\right) = E\left(\lambda^j\right) = b_j, \tag{5.40}$$

and

$$E\left(\lambda_i^j u_{i,t-j} | \Upsilon_t\right) = E\left(\lambda_i^j \mid \Upsilon_t\right) E\left(u_{i,t-j}|\Upsilon_t\right).$$

Taking conditional expectations of both sides of (5.38) with respect to $\Upsilon_t$ we now have

$$\begin{aligned} E(Y_t|\Upsilon_t) &= N^{-1} \sum_{j=0}^{\infty} \sum_{i=1}^{N} E\left[\left(\beta_i \lambda_i^j\right) x_{i,t-j} | \Upsilon_t\right] + N^{-1} \sum_{j=0}^{\infty} \sum_{i=1}^{N} E\left(\lambda_i^j u_{i,t-j} | \Upsilon_t\right), \\ E(Y_t|\Upsilon_t) &= N^{-1} \sum_{j=0}^{\infty} \sum_{i=1}^{N} x_{i,t-j} E\left[\left(\beta_i \lambda_i^j\right) | \Upsilon_t\right] + N^{-1} \sum_{j=0}^{\infty} \sum_{i=1}^{N} E\left(\lambda_i^j \mid \Upsilon_t\right) E\left(u_{i,t-j}|\Upsilon_t\right). \end{aligned}$$

Hence, using (5.39) and (5.40) we have

$$E(Y_t|\Upsilon_t) = \sum_{j=0}^{\infty} a_j X_{t-j} + \sum_{j=0}^{\infty} b_j E(U_{t-j}|\Upsilon_t), \tag{5.41}$$

where $X_t = N^{-1}\sum_{i=1}^N x_{it}$ and $U_t = N^{-1}\sum_{i=1}^N u_{it}$. This result provides the forecast of the aggregate series $\{Y_t\}$ conditional on $\Upsilon_t$ that involves disaggregated observations on $x_{it}'s$. To obtain the aggregate forecast function we need to take expectations of both sides of (5.41) with respect to $\Omega_t$. Noting that $\Omega_t$ is contained in $\Upsilon_t$ we now have

$$E\left(Y_t|\Omega_t\right) = \sum_{j=0}^{\infty} a_j X_{t-j} + \sum_{j=0}^{\infty} b_j E\left(U_{t-j}|\Omega_t\right). \tag{5.42}$$

The aggregate predictor function, $E\left(Y_t|\Omega_t\right)$, is composed of a predetermined component, $\sum_{j=0}^{\infty} a_j X_{t-j}$, and a random component, $\sum_{j=0}^{\infty} b_j E\left(U_{t-j}|\Omega_t\right)$. To learn more about the random component, using (5.36) first note that

$$U_t = \varphi\ \eta_t + \mathcal{Z}_t,$$

where

$$\varphi = N^{-1}\sum_{i=1}^{N}\varphi_i, \text{ and } \mathcal{Z}_t = N^{-1}\sum_{i=1}^{N}\xi_{it}.$$

Namely, the aggregate error term, $U_t$, is itself composed of a common component, $\eta_t$, and an aggregate of the idiosyncratic shocks, $\mathcal{Z}_t$. Under Assumptions A.3 and A.4, $\eta_t$ and $\mathcal{Z}_t$ are serially uncorrelated and independently distributed of $x_{it}$'s, and hence (noting that $Y_t$ is not contained in $\Omega_t$) we have

$$E\left(U_t|\Omega_t\right) = \varphi\ E\left(\eta_t|\Omega_t\right) + E\left(\mathcal{Z}_t|\Omega_t\right) = 0. \tag{5.43}$$

Using this result in (5.42) now yields

$$E\left(Y_t|\Omega_t\right) = \sum_{j=0}^{\infty} a_j X_{t-j} + \sum_{j=1}^{\infty} b_j V_{t-j}, \tag{5.44}$$

where

$$V_{t-j} = E\left(U_{t-j}|\Omega_t\right) = \varphi\ E\left(\eta_{t-j}|\Omega_t\right) + E\left(\mathcal{Z}_{t-j}|\Omega_t\right),\ j = 1, 2, ... \tag{5.45}$$

The optimal aggregate dynamic model corresponding to the micro relations, (5.35), is now given by

$$Y_t = \sum_{j=0}^{\infty} a_j X_{t-j} + \sum_{j=1}^{\infty} b_j V_{t-j} + \varepsilon_t, \tag{5.46}$$

or

$$Y_t = \sum_{j=0}^{\infty} a_j X_{t-j} + \varphi \sum_{j=1}^{\infty} b_j E\left(\eta_{t-j}|\Omega_t\right) + \sum_{j=1}^{\infty} b_j E\left(\mathcal{Z}_{t-j}|\Omega_t\right) + \varepsilon_t, \tag{5.47}$$

where $\varepsilon_t = Y_t - E\left(Y_t|\Omega_t\right)$. By construction $\varepsilon_t$ is orthogonal to $\{X_t, X_{t-1,...}\}$ and $\{V_{t-1}, V_{t-2}, ...\}$. But, as in the static case, the contemporaneous errors of the aggregate equation, $\varepsilon_t$, are likely to be heteroskedastic. The above aggregate specification is optimal in the sense that $E\left(Y_t|\Omega_t\right)$ minimizes $E[Y_t - E\left(Y_t|\Omega_t\right)]^2$ with respect to the aggregate information set, $\Omega_t$.[8]

The terms $V_{t-1}, V_{t-2}, ...$ in addition to being orthogonal to the aggregate disturbances, $\varepsilon_t$, are in fact serially uncorrelated with zero means and a finite variance. First, it is easily seen that

$$E(V_{t-j}) = E\left[E\left(U_{t-j}|\Omega_t\right)\right] = E(U_{t-j}) = 0.$$

Also, for $j > 0$

$$\begin{aligned} E(V_{t-j}V_{t-j-1}|\Omega_{t-j-1}) &= V_{t-j-1}E(V_{t-j}|\Omega_{t-j-1}) \\ &= V_{t-j-1}E\left[E\left(U_{t-j}|\Omega_t\right)|\Omega_{t-j-1}\right] \\ &= V_{t-j-1}E(U_{t-j}|\Omega_{t-j-1}). \end{aligned}$$

But $U_{t-j}$ is a serially uncorrelated process with zero mean. Hence, $E(V_{t-j}V_{t-j-1}|\Omega_{t-j-1}) = 0$, which also implies that $E(V_{t-j}V_{t-j-1}) = 0$. Using a similar line of reasoning it is also easily established that $E(V_{t-j}V_{t-j-s}) = 0$, for all $s \geq 0$. Finally, since by Assumptions A.3 and A.4, $x_{is}$ and $u_{it}$ have finite variances, the random variables $V_{t-1}, V_{t-2},..$, being linear functions of $x_{is}$ and $u_{it}$, will also have finite variances. Clearly, the same arguments also apply to the components of $V_{t-j}$, namely $V_{t-j}^{\eta} = E\left(\eta_{t-j}|\Omega_t\right)$ and $V_{t-j}^{z} = E\left(\mathcal{Z}_{t-j}|\Omega_t\right)$, namely $V_{t-j}^{\eta}$ and $V_{t-j}^{z}$ have zero means, are serially uncorrelated with finite variances.

The aggregate function, (5.46), holds irrespective of whether the shocks to the underlying micro relations contain a common component. But the contribution of the idiosyncratic shocks, $\mathcal{Z}_t$, to the aggregate function will depend on the rate at which the distributed lag coefficients, $b_j$, decay as $j \to \infty$. Although, under assumption A.4 $\mathcal{Z}_t \xrightarrow{p} 0$, this does not necessarily mean that the contribution of the idiosyncratic shocks, given by $\sum_{j=1}^{\infty} b_j E\left(\mathcal{Z}_{t-j}|\Omega_t\right)$, will also tend to zero as $N \to \infty$. Heuristically, this is due to the fact that under Assumptions A.3 and A.4 the variance of $V_{t-j}^{z}$ is of order of $\sum_{j=1}^{\infty} b_j^2/N$ and need not tend to zero if the coefficients, $b_j$, do not decay sufficiently fast.[9] An example of such a possibility was first discussed by Granger (1980). We

[8] Notice that $\{X_t, X_{t-1}, ...\}$ and $\{V_{t-1}, V_{t-2}, ...\}$ are contained in $\Omega_t$.

[9] See also Zaffaroni (2001) who makes a similar point in relation to the time series properties of the aggregates, $Y_t$, using a spectral density approach assuming that the common and the idiosyncratic shocks are independently distributed. Note, however, that our analysis focuses on the time series properties of the aggregate forecasting function, $E\left(Y_t|\Omega_t\right)$, which need not necessarily have the same properties as the aggregates, $Y_t$, themselves.

now turn to this and other examples and show how a number of results in the literature can be obtained from the optimal aggregator function given by (5.46). In the general case where the micro relations are subject to both common and idiosyncratic shocks the effect of the common shocks on the aggregate forecast, $E(Y_t|\Omega_t)$, will dominate as $N \to \infty$. Hence for forecasting purposes the effects of idiosyncratic shocks could be ignored.

### 5.1 Cross-Sectional Aggregation of AR(1) Processes

This problem has been addressed by Granger (1980). Using a spectral density approach Granger shows that the aggregate series may have a long-memory component even if the underlying micro relations are covariance stationary. This result can be readily derived from (5.46). In the case of the aggregation of pure AR(1) models $\beta_i = 0$, and the stochastic process of the aggregate series will be given by

$$Y_t = \sum_{j=1}^{\infty} b_j V_{t-j} + \varepsilon_t, \tag{5.48}$$

where as before $b_j$ is the $j$-th order moment of $\lambda$, $V_{t-1}, V_{t-2}, ...$ are serially uncorrelated with zero means and finite variances, distributed independently of $\varepsilon_t$. In this simple case it is also reasonable to expect $V_t$ and $\varepsilon_t$ to have constant variances, although this assumption is not required for the analysis of the long-memory properties of the $Y_t$ process.

It is clear from (5.48) that the time series properties of $Y_t$ will critically depend on the coefficients $b_j$, and therefore the probability distribution of $\lambda$. Following Granger (1980) we assume that $\lambda$ has a Beta distribution of the second type on the range $(0, 1)$:

$$f(\lambda) = \frac{2}{B(p,q)} \lambda^{2p-1}(1-\lambda^2)^{q-1}, \qquad 0 \leq \lambda \leq 1, \tag{5.49}$$

where $p > 0$ and $q > 0$ are the parameters of the Beta distribution.[10] It is now easily seen that

$$b_j = E\left(\lambda^j\right) = \frac{B(p + j/2, q)}{B(p,q)}, \tag{5.50}$$

and for large $j$

$$b_j \approx (p + j/2)^{-q}. \tag{5.51}$$

Hence, for $q > 1$, the sequence $\{b_j\}$ will be absolute summable and the aggregate series, $Y_t$, will not be long memory. But for values of $0 < q \leq 1$, $\{b_j\}$ will not be absolute summable and the aggregate series will be long memory. As pointed out by Granger (1980) the long-memory property of the aggregate series only depend on whether $0 < q \leq 1$, and does not depend on $p$. Also, irrespective of whether $q$ exceeds unity, an exact finite order autoregressive-moving average representation for $Y_t$ does not seem to exist, unless of course $\lambda$ has a degenerate distribution.

[10] See, for example, Johnson, Kotz and Balakrishnan (1995, Ch. 25).

### 5.2 Cross-Sectional Aggregation of ARDL Models

Initially we consider the relatively simple case where the coefficients $\beta_i$ and $\lambda_i$ are independently and identically distributed across $i$. This case is discussed in Lewbel (1994) where he makes the additional assumptions that the distributions of $\beta_i$ and $x_{it}$ are uncorrelated and that $\lambda_i$ and $\beta_i x_{it}+u_{it}$ are independently distributed.[11] Under these assumptions and adopting the statistical approach described in Section 3.2, Lewbel derives the following aggregate infinite-order autoregressive specification

$$\mu_y(t) = \sum_{j=1}^{\infty} c_j \mu_y(t-j) + \beta\mu_x(t) + \mu_u(t), \tag{5.52}$$

where $\mu_y(t)$, $\mu_x(t)$, and $\mu_u(t)$ are the cross-sectional means of $y_{it}$, $x_{it}$, and $u_{it}$, respectively.[12]

Assuming the above infinite-order autoregressive representation exists, Lewbel shows that the coefficients $c_s$ satisfy the recursions

$$b_s = \sum_{r=0}^{s-1} b_r c_{s-r}, \tag{5.53}$$

with $b_j = E\left(\lambda^j\right)$, as before. It is then easily seen that $c_1 = b_1 = E(\lambda)$, $c_2 = E(\lambda - b_1)^2 = Var(\lambda)$, which establishes that the autoregressive component of the aggregate specification must at least be of second order; otherwise the distribution of $\lambda$ will be degenerate with all agents having the same lag coefficient.

Lewbel's result and a number of its generalizations can be derived from the optimal aggregate specification given by (5.46). Our approach also provides the conditions that ensure the existence of Lewbel's infinite order autoregressive representation. In the simple case considered by Lewbel, where $\beta_i$ and $\lambda_i$ are assumed to be independently distributed, we have $E\left(\beta_i \lambda_i^j\right) = E\left(\beta_i\right) E\left(\lambda_i^j\right) = \beta b_j$, and (5.46) simplifies to[13]

$$Y_t = \beta \sum_{j=0}^{\infty} b_j X_{t-j} + \sum_{j=1}^{\infty} b_j V_{t-j} + \varepsilon_t, \tag{5.54}$$

where as before $b_j = E\left(\lambda^j\right)$. To see the relationship between (5.54) and Lewbel's result, (5.52) first note that

$$Y_t = B(L)\left[\beta X_t + V_t\right] + \varepsilon_t - V_t, \tag{5.55}$$

where $B(L) = \sum_{j=0}^{\infty} b_j L^j$. Whether it is possible to write (5.55) as an infinite-order autoregressive specification in $Y_t$, depends on whether $B(L)$ is invertible and this in turn depends on the probability

[11] The consequences of relaxing some of these assumptions are briefly discussed by Lewbel (1994, Section IV).

[12] Cross-sectional mean of $y_{it}$, for example, is defined as the limit of $\frac{1}{N}\sum_{i=1}^{N} y_{it}$ as $N \to \infty$.

[13] Recall that here we are assuming that there are no common components in the micro shocks, $u_{it}$, and hence $V_t \xrightarrow{p} 0$.

distribution of $\lambda$. It is, for example, clear from our discussion in the previous section that if $\lambda$ has a Beta distribution of the second type with $0 < q \leq 1$, then $\{b_j\}$ will not be absolute summable and $B(L) = \sum_{j=0}^{\infty} b_j L^j$ may not be inverted. Therefore, under this distributional assumption Lewbel's autoregressive representation may not exist. But if $\{b_j\}$ is absolute summable, $B(L)$ can be inverted and (5.55) can be written as

$$Y_t = \sum_{j=1}^{\infty} c_j Y_{t-j} + \beta X_t + \sum_{j=1}^{\infty} c_j V_{t-j} + C(L)\varepsilon_t, \tag{5.56}$$

where $C(L) = 1 - \sum_{j=1}^{\infty} c_j L^j$. The coefficients $c_j$ are obtainable from the polynomial identity $B(L)C(L) \equiv 1$, and it is easily verified that they in fact satisfy the recursive relations (5.53) derived by Lewbel (1994).

In the more general (and realistic) case where $\beta_i$ and $\lambda_i$ are allowed to be statistically dependent, the optimal aggregate specification does not simplify to (5.56) and will be given by (5.46). In this more general setting there seems little gain in re-writing the resultant distributed lag model in the infinite-order autoregressive form favoured by Lewbel (1994).

## 5.3 Relationships Between Micro and Macro Parameters

In general, the optimal aggregate specification, (5.46), is still subject to the so-called aggregation problem, in the sense that not all the parameters of the cross-sectional distribution of the micro parameters can be recovered from the parameters of these aggregate relations. But some of the parameters of interest can still be obtained from the lag coefficients $a_j$ and $b_j$. A prominent example is the long-run impacts of $x_{it}$ on $y_{it}$, averaged across the micro units; namely

$$\theta_N = \frac{1}{N} \sum_{i=1}^{N} \frac{\beta_i}{1 - \lambda_i},$$

or the average of individual "mean lags" defined by

$$\tau_N = \frac{1}{N} \sum_{i=1}^{N} \frac{\lambda_i}{1 - \lambda_i}.$$

Using (5.46), it is easily seen that

$$\sum_{j=0}^{\infty} a_j = \sum_{j=0}^{\infty} E\left(\beta_i \lambda_i^j\right) = E\left(\sum_{j=0}^{\infty} \beta_i \lambda_i^j\right) = E\left(\frac{\beta_i}{1 - \lambda_i}\right).$$

Therefore, for sufficiently large $N$, $\theta_N$ is well approximated by $\sum_{j=0}^{\infty} a_j$, and the cross-sectional mean of the micro long-run coefficients can be estimated by the long-run coefficient of the associated optimal aggregate model. Notice that this result holds even if $\beta_i$ and $\lambda_i$ are not independently distributed, and irrespective of whether the micro shocks, $u_{it}$, contain a common component. The coefficient of $X_t$ in the aggregate specification can also be used to estimate $E(\beta_i)$, the average impact effects of $x_{it}$ on $y_{it}$. But to obtain separate estimates of $E(\beta_i)$ and $E(\lambda_i)$ from the aggregate model, further restrictions are needed. The assumption that $\beta_i$ and $\lambda_i$ are independently distributed, as shown by Lewbel (1994), is sufficient for separate identification of the parameters of the cross-sectional distributions of $\beta_i$ and $\lambda_i$ from the estimates of the aggregate equation.

The mean lag, $\tau_N$, can also be recovered from the coefficients of the aggregate specification if we assume that the individual-specific long-run coefficients and mean lags, namely $\beta_i/(1-\lambda_i)$ and $\lambda_i/(1-\lambda_i)$, are identically and independently distributed.[14] Under this assumption, the estimate of the mean lag based on the aggregate specification, (5.46), is given by

$$\tau_a = \frac{\sum_{j=0}^{\infty} j a_j}{\sum_{j=0}^{\infty} a_j}.$$

Substituting $E(\beta\lambda^j)$ for $a_j$ in the above expression it is easily seen that

$$\tau_a = \frac{E(\frac{\beta}{1-\lambda}\frac{\lambda}{1-\lambda})}{E(\frac{\beta}{1-\lambda})}.$$

Therefore, under the assumption that $\beta/(1-\lambda)$ and $\lambda/(1-\lambda)$ are independently distributed we have $E(\frac{\lambda}{1-\lambda}) = \tau_a$, and since $\lim_{N\to\infty} \tau_N = E(\frac{\lambda}{1-\lambda})$, then for sufficiently large $N$, the mean lag of the aggregate specification, $\tau_a$, can be used to estimate the cross-sectional average of the mean-lag of the micro relations, $\tau_N$.

Another approach to a resolution of the aggregation problem is to adopt a parametric specification of the cross-sectional distribution of the micro coefficients, $\beta_i$ and $\lambda_i$, and then directly estimate the unknown parameters of the cross-sectional distribution from the aggregate specification. As an example, suppose $\beta_i$ and $\lambda_i$ are independently distributed, and $\lambda_i$ follows a standard Beta distribution defined by[15]

$$f(\lambda) = \frac{1}{B(p,q)}\lambda^{p-1}(1-\lambda)^{q-1}, \qquad 0 \le \lambda \le 1,\ p>0,\ q>0. \tag{5.57}$$

Under this set-up, the optimal aggregate equation is given by

$$Y_t = \beta\sum_{j=0}^{\infty}\frac{B(p+j,q)}{B(p,q)}X_{t-j} + \sum_{j=1}^{\infty}\frac{B(p+j,q)}{B(p,q)}V_{t-j} + \varepsilon_t.$$

The unknown parameters of the distribution of the micro coefficients, $\beta, p$, and $q$, can be computed by estimating the above distributed lag model by, for example, the maximum likelihood method. For this purpose it seems reasonable to assume that $V_t$ and $\varepsilon_t$ are normally distributed with zero

[14]It is more likely that this assumption holds than the assumption of independently distributed $\beta_i$ and $\lambda_i$ used in Lewbel (1994).

[15]This specification of the Beta distribution is simpler to work with than the Beta distribution of the second type used earlier to ensure comparability with Granger's work.

means and constant variances, $\sigma_v^2$ and $\sigma_\varepsilon^2$, respectively. The problem of the truncation remainder that arises due to missing observation on $X$, can be resolved, for example, along the lines suggested for the estimation of the geometric distributed lag models by Pesaran (1973). The truncation remainder problem is likely to be more serious in the present application where the weights $b_j = B(p+j,q)/B(p,q)$ will be declining with $j$ at a much slower rate than in the case of the geometrically distributed lag models discussed in the literature.

Having estimated $p$ and $q$, the mean and variance of $\lambda$ can then be estimated by

$$E(\lambda) = \frac{p}{p+q}, \quad \text{and} \quad Var(\lambda) = \frac{qp}{(p+q)^2(p+q+1)}.$$

It can also be shown that the long-run impact of $X_t$ on $Y_t$ (or equivalently the cross-sectional mean of the long-run coefficients of the underlying micro equations) is given by (assuming $q > 1$)

$$\lim_{N\to\infty} \theta_N = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\beta_i}{1-\lambda_i} = \frac{\beta(p+q-1)}{q-1}.$$

# 6 Aggregation of Life-Cycle Consumption Decision Rules Under Habit Formation[16]

In the recent life-cycle literature habit formation has been emphasized as a potentially important factor that may help resolve a number of empirical puzzles. Deaton (1987), among others, argues that habit formation could help explain "excess smoothness" and "excess sensitivity" of aggregate consumption expenditures.[17] Carroll and Weil (1994) suggest that the reverse causality between growth and saving often observed in aggregate data could be due to the neglect of habit formation in consumption behaviour. Fuhrer (2000) maintains that the dynamics of aggregate consumption decisions as represented by autocovariance functions can be much better understood using a model with habit formation than using a model with standard time-separable preferences. A problem common to all these studies using representative agent frameworks is that the coefficient of habit formation needed to reconcile the model with the data is typically deemed implausibly high. In this section we consider the aggregate implications of allowing for heterogeneity in habit formation coefficients across individuals and investigate the extent to which empirical puzzles observed in

[16] In formulation of the aggregation problem in this section I have benefited from the joint work that I am currently conducting with Michael Binder on econometric implications of a wide-range of life-cycle models under habit formation. See Binder and Pesaran (2002).

[17] Excess smoothness refers to the situation where contrary to the prediction of the permanent income hypothesis changes in aggregate consumption do not vary closely with unanticipated changes in labour income. Excess sensitivity refers to the situation where changes in aggregate consumption respond to anticipated changes in labour income, whilst the theory predicts otherwise. For a review of the empirical literature on excess smoothness and excess sensitivity see, for example, Muellbauer and Lattimore (1995).

aggregate consumption data are due to the aggregation problem. Using stochastic simulations we show that the estimates of the habit persistence coefficient are likely to be seriously biased downward if they are based on analogue aggregate consumption functions, which could partly explain the excess smoothness and excess sensitivity puzzles in terms of neglected heterogeneity.[18]

We consider an economy composed of a large number of consumers, where each consumer indexed by $i$, $i = 1, 2, \ldots, N$, at the beginning of period $t$ is endowed with an initial level of financial wealth, $a_{i,t-1}$. His/her labour income over the period $t-1$ to $t$, $y_{it}$, is generated according to the following geometric random walk model

$$\log y_{it} = \alpha_i + \mu t + \sum_{s=1}^{t} v_s + \xi_{it}, \tag{6.58}$$

where $\alpha_i$ is the time-invariant individual-specific component, $\mu$ is an economy-wide drift term, $v_t$ is the economy-wide random component, and $\xi_{it}$ is the residual random component. The random components $\alpha_i$, $v_t$, and $\xi_{it}$ are assumed to be mutually independent, $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots$, and distributed identically as normal variates with zero means and constant variances:

$$\alpha_i \sim iid\, N\left(\alpha, \sigma_\alpha^2\right), \quad v_t \sim iid\, N\left(0, \sigma_v^2\right), \quad \text{and} \quad \xi_{it} \sim iid\, N\left(0, \sigma_\xi^2\right). \tag{6.59}$$

This formulation allows labour incomes at the individual and the economy-wide levels to exhibit geometric growth and at the same time yields a plausible steady state size distribution for labour incomes.[19] Each individual solves the following intertemporal optimization problem:

$$\max_{\{c_{i,t+s}\}_{s=0}^{\infty}} E\left[\sum_{s=0}^{\infty} \delta^s u(c_{i,t+s}, c_{i,t+s-1}) | \Phi_{it}\right] \tag{6.60}$$

subject to the period-by-period budget constraints,

$$a_{i,t+s} = (1+r)a_{i,t+s-1} + y_{i,t+s} - c_{i,t+s}, \qquad s = 0, 1, \ldots \tag{6.61}$$

the transversality condition,

$$\lim_{s\to\infty} (1+r)^{-s} E(a_{i,t+s} | \Phi_{it}) = 0, \tag{6.62}$$

and given initial consumption levels, $c_{i,t-1}$, as well as initial wealth levels, $a_{i,t-1}$, for all $i$. In equations (6.60)–(6.62) $u_{it} = u(c_{it}, c_{i,t-1})$ represents individuals $i$'s current-period utility function for period $t$, $\delta = 1/(1+\rho)$ represents a constant discount factor, $r$ is the constant real rate of

[18] In a different attempt at resolving the excess smoothness and excess sensitivity puzzles, Binder and Pesaran (2001) argue that social interactions when combined with habit formation can also help.

[19] For a more detailed discussion of the relative merits of the labour income specification (6.58) over the usual arithmetic formulations employed in the literature see Binder and Pesaran (2000).

interest, and $E(\cdot|\Phi_{it})$ denotes the mathematical conditional expectations operator with respect to the information set available to the individual at time $t$:

$$\Phi_{it} = \{c_{it}, c_{i,t-1}, \dots ; y_{it}, y_{i,t-1}, \dots ; a_{it}, a_{i,t-1}, \dots\}. \tag{6.63}$$

Given the focus of our analysis on aggregation of linear models we consider the case where the current period utility function is quadratic, namely

$$u_{it} = \frac{-1}{2}(c_{it} - \lambda_i c_{i,t-1} - \bar{c}_i)^2, \; 0 < \lambda_i < 1, \tag{6.64}$$

where $\lambda_i$ is the habit formation coefficient and $\bar{c}_i$ is the saturation coefficient. For simplicity we also assume that $\rho = r$, so that individuals are time-indifferent.[20] For each individual the consumption decision rule for time period $t$ that solves the above intertemporal optimization problem is given by:

$$\Delta c_{it} = \lambda_i \Delta c_{i,t-1} + \beta_i y_{it} + \gamma_i \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2) \left[\tilde{Y}_t - (1+r)\tilde{Y}_{t-1}\right], \tag{6.65}$$

where $\tilde{Y}_t$ is the economy-wide component of labour income,

$$\tilde{Y}_t = \exp(\mu t + \sum_{s=1}^{t} v_s), \tag{6.66}$$

$$\beta_i = \frac{r(1+r-\lambda_i)}{(1+r)^2}, \tag{6.67}$$

$$\gamma_i = \frac{r(1+r-\lambda_i)(1+g)}{(1+r)^2(r-g)}, \tag{6.68}$$

and $g$ is the rate of growth of labour income

$$g = \exp(\mu + \frac{1}{2}\sigma_v^2) - 1. \tag{6.69}$$

Notice that the labour income of individual $i$ can be decomposed as

$$y_{it} = \tilde{Y}_t \exp(\alpha_i + \xi_{it}). \tag{6.70}$$

Defining economy-wide average labor income as $Y_t = (1/N)\sum_{i=1}^{N} y_{it}$, then under (6.59) as $N \to \infty$ we have

$$Y_t \xrightarrow{p} \tilde{Y}_t \exp(\alpha + \frac{\sigma_\alpha^2}{2} + \frac{\sigma_\xi^2}{2}). \tag{6.71}$$

[20]For more details and relevant references to the literature see Binder and Pesaran (2000, 2002).

Also, aggregating the budget constraints, (6.61), yields

$$A_{t+s} = (1+r)A_{t+s-1} + Y_{t+s} - C_{t+s}, \qquad s = 0, 1, \ldots$$

where $A_t = (1/N)\sum_{i=1}^{N} a_{it}$ and $C_t = (1/N)\sum_{i=1}^{N} c_{it}$.

There will be an aggregation problem only when the habit formation coefficients, $\lambda_i$, differ across individuals. In the case where $\lambda_i = \lambda$ for all $i$ we have:

$$\Delta c_{it} = \lambda \Delta c_{i,t-1} + \beta y_{it} + \gamma \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2)\left[\tilde{Y}_t - (1+r)\tilde{Y}_{t-1}\right], \tag{6.72}$$

$$\beta = \frac{r(1+r-\lambda)}{(1+r)^2}, \tag{6.73}$$

$$\gamma = \frac{r(1+r-\lambda)(1+g)}{(1+r)^2(r-g)}, \tag{6.74}$$

and using (6.71) and noting that $\frac{1}{N}\sum_{i=1}^{N}\exp(\alpha_i) \xrightarrow{p} \exp(\alpha + \frac{1}{2}\sigma_\alpha^2)$ yields the perfect aggregate model

$$\Delta C_t = \lambda \Delta C_{t-1} + \beta Y_t + \frac{r(1+r-\lambda)(1+g)}{(1+r)^2(r-g)}\left[Y_t - (1+r)Y_{t-1}\right], \tag{6.75}$$

or equivalently

$$\Delta C_t = \lambda \Delta C_{t-1} + \frac{r(1+r-\lambda)}{(1+r)(r-g)}\left[Y_t - (1+g)Y_{t-1}\right]. \tag{6.76}$$

This specification is perfect in the sense that it yields aggregate forecasts of $\Delta C_t$ (or $C_t$) based only on aggregate time series observations $\Omega_t = \{C_{t-1}, C_{t-2}, ...; Y_t, Y_{t-1}, ...\}$ that have zero mean-squared errors and are indistinguishable from forecasts of aggregate consumption based on the individual-specific decision rules, (6.72) (using individual-specific consumption and labour income data).

Consider now the empirically more interesting case where yhe $\lambda_i$'s are allowed to vary across the individuals. Since $|\lambda_i| < 1$ for all $i$, then

$$\Delta c_{it} = \beta_i \sum_{j=0}^{\infty} \lambda_i^j y_{i,t-j} + \gamma_i \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2)\sum_{j=0}^{\infty}\lambda_i^j(\tilde{Y}_{t-j} - (1+r)\tilde{Y}_{t-j-1}). \tag{6.77}$$

Aggregating across $i$, we have

$$\begin{aligned}\Delta C_t \quad = \quad & \frac{1}{N}\sum_{j=0}^{\infty}\sum_{i=1}^{N}\beta_i \lambda_i^j y_{i,t-j} \\ & + \frac{1}{N}\sum_{j=0}^{\infty}\sum_{i=1}^{N}\gamma_i \lambda_i^j \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2)\left[\tilde{Y}_{t-j} - (1+r)\tilde{Y}_{t-j-1}\right].\end{aligned} \tag{6.78}$$

Assume that the $\lambda_i$'s are *iid* draws from a distribution with finite moments of all orders defined on the unit interval, and take conditional expectations of both sides of (6.78) with respect to $\Upsilon_t = \cup_{i=1}^{N} \Upsilon_{it}$, where $\Upsilon_{it} = \{y_{it}, y_{i,t-1}, \dots\} \cup \{Y_t, Y_{t-1}, ...\}$. Now,

$$\begin{aligned} E\left(\Delta C_t | \Upsilon_t\right) &= \frac{1}{N} \sum_{j=0}^{\infty} \sum_{i=1}^{N} E\left(\beta_i \lambda_i^j | \Upsilon_t\right) y_{i,t-j} + \\ &\quad + \frac{1}{N} \sum_{j=0}^{\infty} \sum_{i=1}^{N} E\left(\gamma_i \lambda_i^j | \Upsilon_t\right) \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2) \left[\tilde{Y}_{t-j} - (1+r)\tilde{Y}_{t-j-1}\right] \end{aligned}$$

Since $E\left(\beta_i \lambda_i^j | \Upsilon_t\right) = E\left(\beta_i \lambda_i^j\right) = a_j$ and $E\left(\gamma_i \lambda_i^j | \Upsilon_t\right) = E\left(\gamma_i \lambda_i^j\right) = b_j$, for all $i$, then we have[21]

$$E\left(\Delta C_t | \Upsilon_t\right) = \sum_{j=0}^{\infty} a_j Y_{t-j} + \left(\frac{1}{N} \sum_{i=1}^{N} \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2)\right) \sum_{j=0}^{\infty} b_j \left[\tilde{Y}_{t-j} - (1+r)\tilde{Y}_{t-j-1}\right]. \tag{6.79}$$

But as noted earlier, for $N$ sufficiently large

$$\frac{1}{N} \sum_{i=1}^{N} \exp(\alpha_i + \frac{1}{2}\sigma_\xi^2) \xrightarrow{p} \exp(\alpha + \frac{\sigma_\alpha^2}{2} + \frac{\sigma_\xi^2}{2}),$$

and in view of (6.71) we have

$$E\left(\Delta C_t | \Upsilon_t\right) = \sum_{j=0}^{\infty} a_j Y_{t-j} + \sum_{j=0}^{\infty} b_j \left[Y_{t-j} - (1+r)Y_{t-j-1}\right]$$

Also using (6.67) it is easily seen that

$$\begin{aligned} a_j &= E\left(\beta_i \lambda_i^j\right) = E\left[\frac{r(1+r-\lambda_i)\lambda_i^j}{(1+r)^2}\right] \\ &= \frac{r}{1+r} m_j - \frac{r}{(1+r)^2} m_{j+1}, \end{aligned} \tag{6.80}$$

and $m_j = E(\lambda^j)$ is the $j-$th order moment of $\lambda_i$. Similarly, using (6.68) and (6.80) we have

$$b_j = E\left(\gamma_i \lambda_i^j\right) = \frac{(1+g)}{(r-g)} a_j. \tag{6.81}$$

Now taking conditional expectations of (6.79) with respect to the aggregate information set $\Omega_t = \{Y_t, Y_{t-1}, ...; C_{t-1}, C_{t-2}, ...\}$

$$\begin{aligned} & E(\Delta C_t | \Omega_t) \\ &= \sum_{j=0}^{\infty} a_j Y_{t-j} + \frac{1+g}{r-g} \sum_{j=0}^{\infty} a_j \left[Y_{t-j} - (1+r)Y_{t-j-1}\right] \\ &= \left(\frac{1+r}{r-g}\right) \left\{ a_0 Y_t + \sum_{j=1}^{\infty} [a_j - (1+g)a_{j-1}] Y_{t-j} \right\}. \end{aligned}$$

[21] Recall that $\frac{1}{N}\sum_{i=1}^{N} \exp(\alpha_i) \xrightarrow{p} \exp(\alpha + \frac{1}{2}\sigma_\alpha^2)$.

The optimal aggregate consumption function can therefore be written as

$$\Delta C_t = \left(\frac{1+r}{r-g}\right) \sum_{j=0}^{\infty} a_j \left[Y_{t-j} - (1+g)Y_{t-j-1}\right] + \varepsilon_t, \tag{6.82}$$

where $\varepsilon_t$ is the aggregation error and by construction satisfies the orthogonality condition

$$E(\varepsilon_t \mid \Omega_t) = 0.$$

The aggregation errors are serially uncorrelated with zero means; but in general are not homoskedastic. The above optimal aggregate function is directly comparable to the aggregate model, (6.76), obtained under homogeneous habit formation coefficients. It is easily seen that (6.82) reduces to (6.76) if $\lambda_i = \lambda$ for all $i$. Also the aggregation errors, $\varepsilon_t$'s, vanish if and only if $\lambda_i = \lambda$. Finally, unless the habit formation coefficients are homogeneous the optimal aggregate model can not be written as a finite order ARDL model in $\Delta C_t$ and $Y_t - (1+g)Y_{t-1}$.

## 6.1 Some Monte Carlo Results

The above analysis suggests that in the present example the appropriate econometric specification for the analysis of aggregate consumption is given by (6.82), which is far more involved than the micro analogue specification, (6.76) employed in the literature. Using the optimal aggregate specification, (6.82), and its micro analogue we now employ stochastic simulation techniques to ascertain the quantitative importance of the aggregation error and to examine the extent to which neglected parameter heterogeneity could explain the excess smoothness and excess sensitivity puzzles observed in the empirical literature.

At the micro level we generated individual income and consumption time series using equations (6.65) and (6.70). The aggregate component of income, $\tilde{Y}_t$, was generated using (6.66). We consider two sets of experiments. The first represents calibration to "annual" observations and a second set provides calibration to "quarterly " observations. Under both sets we set $N = 5,000$ and experimented with different values of $T = (50, 100, 150)$, and $p = q = (0.5, 0.8, 1.0, 1.2, 1.50)$, where $p$ and $q$ are the parameters of the Beta distribution assumed for the cross-sectional distribution of $\lambda_i$, the habit formation coefficient. In the case of annual observations we used the parameter values $\sigma_\alpha^2 = 0.20$, $\sigma_v^2 = 0.01$, $\sigma_\xi^2 = 0.01$, $r = 0.04$, $g = 0.02$. For quarterly observations we selected the parameter values $r = (1.04\hat{}.25) - 1$, $g = (1.02\hat{}.25) - 1$, $\sigma_\alpha^2 = 0.20$, $\sigma_v^2 = 0.001$, $\sigma_\xi^2 = 0.001$. Recall that the rate of decay of the distributed lag coefficients of the optimal aggregate consumption function, (6.82), varies inversely with $q$. The aggregation bias is likely to be most serious when $q$ is relatively small.

To study the quantitative importance of aggregation for the estimation of the structural parameters ($r$, $g$, $p$, and $q$) we estimated the analogue aggregate model (6.76) and the optimal aggregate model (6.82) after deflating both sides of these relations by $Y_t$ to achieve stationarity. In particular, we estimated the following analogue aggregate consumption function by the Ordinary Least Squares method:

$$\frac{\Delta C_t}{Y_t} = \kappa_0 + \kappa_1 \frac{\Delta C_{t-1}}{Y_t} + \kappa_2 \frac{Y_{t-1}}{Y_t} + u_t,$$

where

$$\begin{aligned} \kappa_0 &= \frac{r(1+r-\lambda)}{(1+r)(r-g)}, \\ \kappa_1 &= \lambda, \quad \kappa_2 = \frac{-r(1+r-\lambda)(1+g)}{(1+r)(r-g)}. \end{aligned}$$

The structural parameters are exactly identified from the OLS estimates of $\kappa_0, \kappa_1$, and $\kappa_2$.

The optimal aggregate model was estimated, assuming a standard $Beta(p, q)$ distribution for the habit formation coefficients, $\lambda_i$ (see (5.57)). The computations were carried out iteratively by minimizing

$$Q(r, g, p, q) = \sum_{t=2}^{T} \left\{ \frac{\Delta C_t}{Y_t} - \left( \frac{1+r}{r-g} \right) \sum_{j=0}^{t-1} a_j G_{t-j} \right\}^2, \tag{6.83}$$

where

$$G_{t-j} = (Y_{t-j}/Y_t) - (1+g)(Y_{t-j-1}/Y_t), \tag{6.84}$$

with respect to the unknown parameters, $r, g$, $p$ and $q$. The decay rate of the distributed lag coefficients, $a_j$, is determined by $q$, and will exhibit long-memory properties for values of $q < 1$. However, since

$$\begin{aligned} a_j &= \frac{r}{1+r} \left( E(\lambda^j) - \frac{1}{1+r} E(\lambda^{j+1}) \right) \\ &= \frac{r}{1+r} \left\{ \frac{B(p+j, q)}{B(p, q)} - \frac{1}{1+r} \frac{B(p+j+1, q)}{B(p, q)} \right\} \end{aligned} \tag{6.85}$$

the long-memory properties of $E(\lambda^j)$ will be somewhat counteracted by those of $E(\lambda^{j+1})$, for small to moderate values of $j$. But as $j \to \infty$ since

$$E(\lambda^j) \approx (p+j)^{-q},$$

it is then easily seen that

$$\begin{aligned} a_j &\approx \frac{r}{1+r} \left\{ (p+j)^{-q} - \frac{1}{1+r} (p+j+1)^{-q} \right\} \\ &\approx \left( \frac{r}{1+r} \right)^2 (p+j)^{-q}, \text{ for } j \to \infty, \end{aligned}$$

and hence $a_j$ has the same asymptotic rate of decay as $m_j = E(\lambda^j)$.[22] Therefore, for most values of $j$, the overall rate of decay of $a_j$ could still be quite rapid even for values of $q < 1$. This is readily seen in Figures 1 and 2. The plot of $E(\lambda^j)$, given in Figure 1, exhibits a very slow rate of convergence for small values of $q$, but the same is not true of the plot of $a_j$ given in Figure 2. This suggests that the estimation of the optimal aggregate consumption model may be feasible even for values of $q < 1$.

The summary results for "annual" observations are given in Tables 1 and 2, and for the "quarterly" observations are given in Tables 3 and 4. Each table gives the mean, the median, standard error and mean square errors of the estimates across $1,000$ replications. Perhaps not surprisingly, the estimates based on the analogue aggregate model (in Tables 1 and 3) show a substantial degree of bias, with the bias being most serious in cases where $q$ is relatively small. Even for large values of $q$, the estimates based on the analogue aggregate model are still rather poor.[23] For example, for $q = p = 1.5$ and $T = 150$, the mean of the habit formation coefficient, $E(\lambda)$, is estimated to be 0.42 in the case of the annual observations and 0.40 in the case of quarterly observations as compared to its true value of 0.50.

In contrast, the estimates based on the optimal aggregate model are generally much closer to their true values. (See Tables 2 and 4.) The match between the estimates and the true values is particularly good for $T = 150$, irrespective of the value of $q$. Once again, as to be expected, the quality of the estimates improve as $q$ is increased.

## 7 Conclusions

This paper proposes an optimal forecasting approach to the analysis of aggregation and argues that aggregate functions ought to be derived as optimal forecasts with respect to a loss function of interest. In the case of quadratic loss functions routinely used in the econometric literature the optimal aggregate function is given by the conditional expectations of the aggregate variable of interest (formed as the average or the weighted average of the underlying micro decision rules) with respect to the available aggregate information set. This approach is particularly suited to the aggregation of dynamic models where parameter heterogeneity and dynamics interact in a complicated manner. This is illustrated by a re-examination of the aggregation of linear autoregressive models as discussed in the literature by for example Granger (1980), Lippi (1988) and Lewbel (1994). We have shown how the results in the literature can be derived under more general assumptions and note that the aggregate function derived by Lewbel following the statistical approach does not fully take

[22] I am grateful to Paolo Zaffaroni for pointing this out to me.

[23] Notice that the degree of heterogeneity of the habit formation coefficient under $p = q$, as measured by the variance of $\lambda$, is given by $Var(\lambda) = \frac{1}{4(2q+1)}$ and is inversely related to $q$. Therefore, one would expect that the estimates based on the aggregate analogue model would perform better for larger values of $q$.

account of the aggregation errors involved. We have also applied the forecasting approach to the problem of aggregating life-cycle consumption decision rules subject to habit formation and have shown that, in general, the optimal aggregate consumption function cannot be represented as a finite-order autoregressive distributed lag model in income and consumption. Under heterogeneity the distributed lag coefficients on labour income decay much more slowly than the geometric rate obtained for the analogue aggregate consumption function. The quantitative importance of this finding is examined by means of stochastic simulations. It is shown that the estimates of the habit formation coefficient based on the analogue aggregate model are biased downwards (in some cases very significantly so), while the same is not true of the estimates based on the optimal aggregate model. This result is relevant to the excess smoothness and excess sensitivity puzzles recently discussed in the consumption literature and suggests that the puzzle could partly be due to aggregation errors. However, a satisfactory empirical examination of the quantitative importance of the aggregation bias in this literature is beyond the scope of the present paper.

## References


Barker, T. and M.H. Pesaran (eds.) (1990): *Disaggregation in Econometric Modelling*, Routledge, London.

Binder, M., and M.H. Pesaran (2001): Life-Cycle Consumption Under Social Interactions, *Journal of Economic Dynamics and Control*, 25, 35-83.

Binder, M., and M.H. Pesaran (2002): Cross-Country Analysis of Saving Rates and Life-Cycle Models, Under Preparation.

Carroll, C.D., and D.N. Weil (1994): Saving and Growth: A Reinterpretation, *Carnegie-Rochester Conference Series on Public Policy*, 40, 133-192.

Christoffersen, P.F. and F.X. Diebold (1996): Further Results on Forecasting and Model Selection under Asymmetric Loss. *Journal of Applied Econometrics*, 11, 561-571.

Deaton, A. (1987): Life-Cycle Models of Consumption: Is the Evidence Consistent with the Theory? in: T.F. Bewley (ed.): *Advances in Econometrics: Fifth World Congress*, Vol. II, Cambridge: Cambridge University Press, 121-148. Forni, M. and M.Lippi (1997): *Aggregation and Microfoundations of Dynamic Macroeconomics*, Oxford, Oxford University Press.

Fuhrer, J.C. (2000): Habit Formation in Consumption and Its Implications for Monetary-PolicyModels *American Economic Review*, 90, 367-390.

Garderen, K.J. van, K. Lee, and M.H. Pesaran (2000): Cross-Sectional Aggregation of Non-Linear Models, *Journal of Econometrics, 95, 285-331.*

Gorman, W.M. (1953), Community Preference Fields, *Econometrica*, 21, 63-80.

Granger, C.W.J. (1980): Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics*, 14, 227-238.

Granger, C.W.J., and Z. Ding (1996): Varieties of Long Memory Models, *Journal of Econometrics*, 73, 61-77.

Granger, C.W.J., and M.J. Morris (1976): Time Series Modelling and Interpretation, *Journal of the Royal Statistical* Association, Series A, 139, 246-257.

Granger, C.W.J. and M.H. Pesaran (2000): A Decision Theoretic Approach to Forecast Evaluation, in W.S. Chan, W.K. Li and H. Tong (eds), *Statistics and Finance: An Interface*, Imperial College Press, London, 2000, chapter 15, pp.261-278.

Houthakker, H.S. (1955/56): The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis, *Review of Economic Studies* 23, 27-31.

Johanson, L. (1972): Simple and General Nonsubstitution Theorems for Input-Output Models, *Journal of Economic Theory*, 5, 383-394.

Johnson, N.L., S. Kotz, and N. Balakrishnan (1995): *Continuous Univariate Distributions,* Vol. 2, New York: John Wiley.

Kelejian, H.J. (1980): Aggregation and Disaggregation of Nonlinear Equations, in: J. Kmenta and J.B. Ramsey (eds.): *Evaluation of Econometric Models*, New York: Academic Press, 135-153.

Klein, L.R. (1953): *A Textbook of Econometrics,* Row Peterson and Company.

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

Lippi, M. (1988): On the Dynamic Shape of Aggregated Error Correction Models, *Journal of Economic Dynamics and Control*, 12, 561-585.

Lütkepohl, H. (1984): Linear Transformations of Vector ARMA Processes, J*ournal of Econometrics, 26, 283-93.*

Malinvaud, E. (1970): The Consistency of Nonlinear Regressions, *Annals of Mathematical Statistics,* 41, 956-969.

Muellbauer, J. (1975): Aggregation, Income Distribution and Consumer Demand, *Review of Economic Studies,* 42, 525-543.

Muellbauer, J., and R. Lattimore (1995): The Consumption Function: A Theoretical and Empirical Overview, in: M.H. Pesaran and M.R. Wickens (eds.): *Handbook of Applied Econometrics: Macroeconomics*, Oxford: Basil Blackwell, 221- 311.

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

Rose, D.E. (1977): Forecasting Aggregate ARIMA Processes, *Journal of Econometrics*, 5, 323-345.

Stoker, T.M. (1984): Completeness, Distribution Restrictions, and the Form of Aggregate Functions, *Econometrica,* 52, 887-907.

Stoker, T.M. (1986): Simple Tests of Distributional Effects on Macroeconomic Equations, *Journal of Political Economy,* 94, 763-795.

Stoker, T.M. (1993): Empirical Approaches to the Problem of Aggregation over Individuals, *Journal of Economic Literature,* 31, 1827-1874.

Theil, H. (1954), *Linear Aggregation of Economic Relations*, North Holland, Amsterdam.

Trivedi, P.K. (1985): Distributed Lags, Aggregation and Compounding: Some Econometric Implications, *Review of Economic Studies*, 52, 19-35.

Zaffaroni (2001): Contemporaneous Aggregation of Linear Dynamic Models in Large Economies, Mimeo, Banca d'Italia.

**Figure 1:** [1] **Trajectories of Distributed Lag Coefficients:** $m_i$

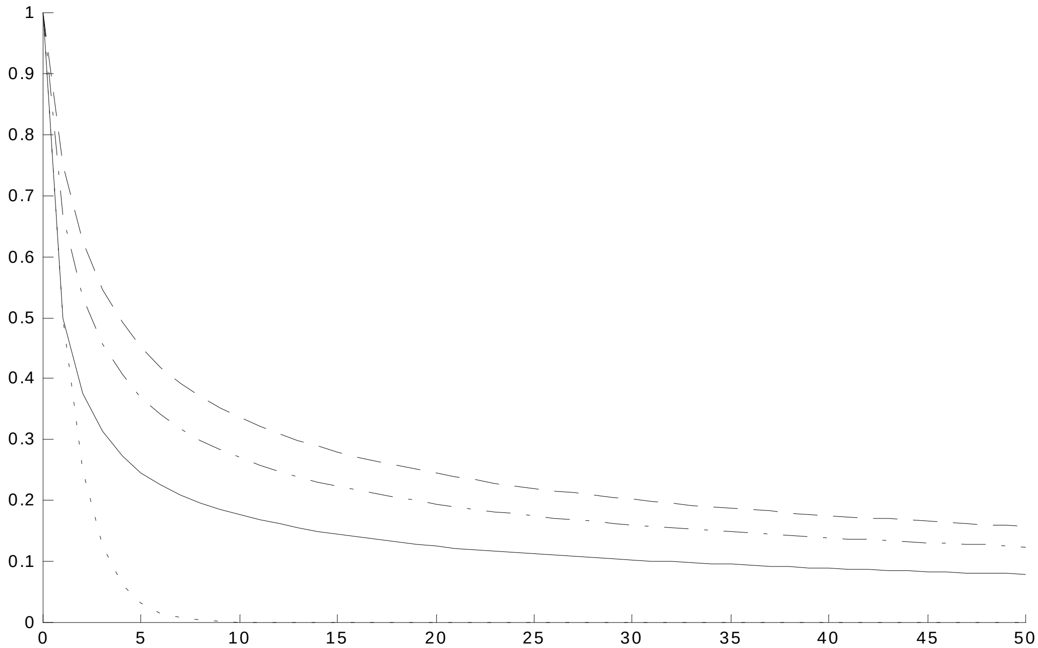
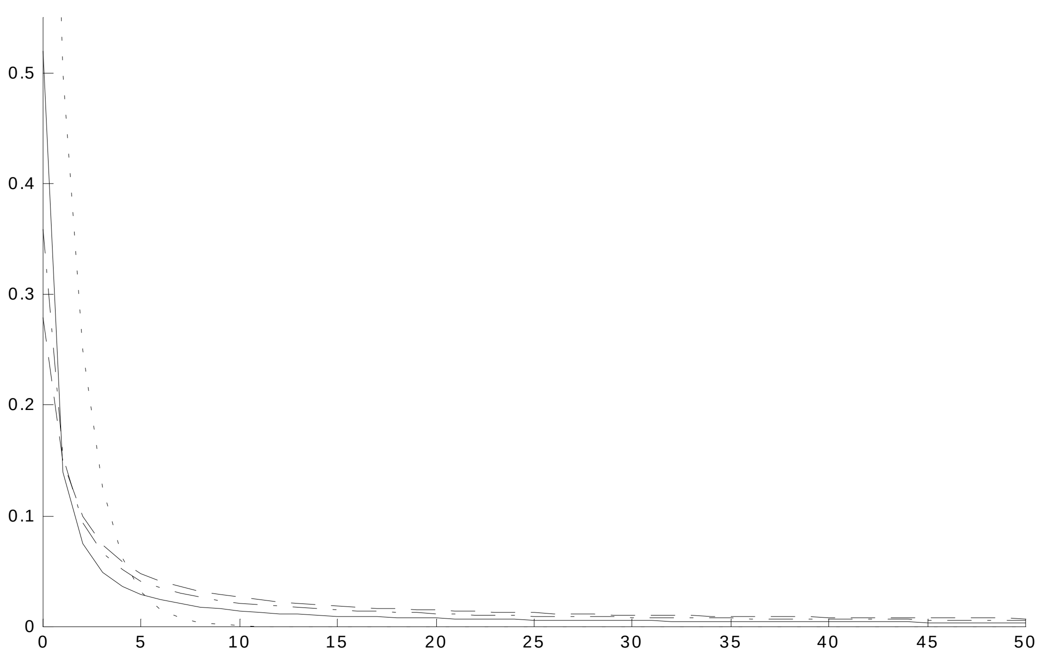



**Figure 2: Trajectories of Distributed Lag Coefficients:** $a_i(1+r)/r$




---

[1] For both Figure 1 and Figure 2 the solid line represents the case where p=q=.5, the dashdotted line the case where p=q=1, and the dashed line the case where p=q=1.5. The dotted line represents a geometric decay with coefficient .5.

Table 1: OLS Estimates Based on Analogous "Annual" Linear Aggreagte Model

| | | r | g | E(λ) |
|---|---|---|---|---|
| | true values | 0.04 | 0.02 | 0.5 |
| | mean | 0.1237 | 0.03053 | 0.2602 |
| $T = 50$ | median | 0.1235 | 0.03053 | 0.2602 |
| $p = q = 0.5$ | s.d. | 0.004788 | 0.0002734 | 0.005189 |
| | mse | 0.08384 | 0.01053 | 0.2399 |
| | mean | 0.0902 | 0.02793 | 0.3191 |
| $T = 50$ | median | 0.0901 | 0.02793 | 0.319 |
| $p = q = 0.8$ | s.d. | 0.002291 | 0.0002118 | 0.005039 |
| | mse | 0.05025 | 0.007937 | 0.181 |
| | mean | 0.07941 | 0.02679 | 0.3448 |
| $T = 50$ | median | 0.07936 | 0.02679 | 0.3447 |
| $p = q = 1$ | s.d. | 0.001712 | 0.000189 | 0.004875 |
| | mse | 0.03945 | 0.006794 | 0.1552 |
| | mean | 0.07235 | 0.02592 | 0.3645 |
| $T = 50$ | median | 0.07234 | 0.02592 | 0.3644 |
| $p = q = 1.2$ | s.d. | 0.001356 | 0.0001699 | 0.004555 |
| | mse | 0.03238 | 0.005925 | 0.1356 |
| | mean | 0.06551 | 0.02497 | 0.3862 |
| $T = 50$ | median | 0.06549 | 0.02497 | 0.3861 |
| $p = q = 1.5$ | s.d. | 0.001097 | 0.0001563 | 0.004367 |
| | mse | 0.02553 | 0.004973 | 0.1139 |
| | mean | 0.09941 | 0.02711 | 0.275 |
| $T = 100$ | median | 0.09932 | 0.02711 | 0.2748 |
| $p = q = 0.5$ | s.d. | 0.002848 | 0.000183 | 0.004993 |
| | mse | 0.05948 | 0.007117 | 0.2251 |
| | mean | 0.07498 | 0.02493 | 0.3326 |
| $T = 100$ | median | 0.07493 | 0.02492 | 0.3327 |
| $p = q = 0.8$ | s.d. | 0.00135 | 0.000127 | 0.00478 |
| | mse | 0.035 | 0.004928 | 0.1674 |
| | mean | 0.06726 | 0.02408 | 0.3574 |
| $T = 100$ | median | 0.06725 | 0.02408 | 0.3573 |
| $p = q = 1$ | s.d. | 0.000998 | 0.0001074 | 0.004597 |
| | mse | 0.02728 | 0.004077 | 0.1427 |
| | mean | 0.06226 | 0.02347 | 0.376 |
| $T = 100$ | median | 0.06226 | 0.02347 | 0.3761 |
| $p = q = 1.2$ | s.d. | 0.0007761 | 9.086e-05 | 0.004295 |
| | mse | 0.02227 | 0.003472 | 0.124 |
| | mean | 0.05747 | 0.02284 | 0.3965 |
| $T = 100$ | median | 0.05746 | 0.02285 | 0.3965 |
| $p = q = 1.5$ | s.d. | 0.0006089 | 7.827e-05 | 0.004053 |
| | mse | 0.01748 | 0.002845 | 0.1036 |
| | mean | 0.09576 | 0.03168 | 0.3305 |

continued ...

. . . continued

| | | | | |
|---|---|---|---|---|
| $T = 150$ | median | 0.09572 | 0.03168 | 0.3304 |
| $p = q = 0.5$ | s.d. | 0.001731 | 0.0002502 | 0.005755 |
| | mse | 0.05578 | 0.01169 | 0.1696 |
| | mean | 0.07549 | 0.02812 | 0.3713 |
| $T = 150$ | median | 0.07547 | 0.02812 | 0.3713 |
| $p = q = 0.8$ | s.d. | 0.001044 | 0.0001854 | 0.005111 |
| | mse | 0.0355 | 0.00812 | 0.1288 |
| | mean | 0.06826 | 0.0267 | 0.3896 |
| $T = 100$ | median | 0.06828 | 0.0267 | 0.3897 |
| $p = q = 1$ | s.d. | 0.0008333 | 0.0001605 | 0.004803 |
| | mse | 0.02828 | 0.006697 | 0.1105 |
| | mean | 0.06338 | 0.02568 | 0.4035 |
| $T = 150$ | median | 0.06338 | 0.02567 | 0.4036 |
| $p = q = 1.2$ | s.d. | 0.0006804 | 0.0001369 | 0.004389 |
| | mse | 0.02339 | 0.005678 | 0.09657 |
| | mean | 0.05855 | 0.02462 | 0.4191 |
| $T = 150$ | median | 0.05855 | 0.02462 | 0.4191 |
| $p = q = 1.5$ | s.d. | 0.000553 | 0.0001178 | 0.004049 |
| | mse | 0.01856 | 0.004621 | 0.08103 |

Note: The simulation results in Tables 1 and 2 are based on growth and discount rates calibrated to "annual" observations. The other parameter values not specified in the body of the tables are $\sigma_\alpha^2 = 0.20$, $\sigma_v^2 = 0.01$, $\sigma_\xi^2 = 0.01$. The simulations are based on 1000 replications.

Table 2: Estimates Based on the "Annual" Optimal Aggregate Model

| | | r | g | E(λ) | p | q |
|---|---|---|---|---|---|---|
| | true values | 0.04 | 0.02 | 0.5 | varies - see panels | |
| | mean | 0.05712 | 0.03324 | 0.4214 | 0.6359 | 0.8733 |
| $T = 50$ | median | 0.05706 | 0.03318 | 0.4215 | 0.6353 | 0.8721 |
| $p = q = 0.5$ | s.d. | 0.001236 | 0.001039 | 0.0066 | 0.02418 | 0.03877 |
| | mse | 0.01716 | 0.01328 | 0.07886 | 0.138 | 0.3754 |
| | mean | 0.05192 | 0.02888 | 0.4467 | 1.018 | 1.262 |
| $T = 50$ | median | 0.0519 | 0.02886 | 0.4467 | 1.018 | 1.261 |
| $p = q = 0.8$ | s.d. | 0.000906 | 0.0007394 | 0.005836 | 0.03884 | 0.05532 |
| | mse | 0.01195 | 0.008915 | 0.05363 | 0.2219 | 0.4652 |
| | mean | 0.05001 | 0.02733 | 0.4562 | 1.27 | 1.514 |
| $T = 50$ | median | 0.04998 | 0.02731 | 0.4561 | 1.27 | 1.513 |
| $p = q = 1$ | s.d. | 0.0007791 | 0.0006283 | 0.005335 | 0.05025 | 0.0676 |
| | mse | 0.01004 | 0.007354 | 0.04412 | 0.2748 | 0.5188 |
| | mean | 0.0487 | 0.02628 | 0.4628 | 1.523 | 1.768 |
| $T = 50$ | median | 0.0487 | 0.02627 | 0.4631 | 1.522 | 1.763 |
| $p = q = 1.2$ | s.d. | 0.0006918 | 0.0005527 | 0.004911 | 0.06181 | 0.08108 |
| | mse | 0.00873 | 0.006301 | 0.03749 | 0.3284 | 0.5733 |
| | mean | 0.04743 | 0.02527 | 0.4695 | 1.901 | 2.148 |
| $T = 50$ | median | 0.04743 | 0.02526 | 0.4696 | 1.899 | 2.146 |
| $p = q = 1.5$ | s.d. | 0.0006128 | 0.0004853 | 0.004328 | 0.08355 | 0.1036 |
| | mse | 0.007454 | 0.005291 | 0.03078 | 0.4098 | 0.6566 |
| | mean | 0.04455 | 0.02328 | 0.477 | 0.5452 | 0.5979 |
| $T = 100$ | median | 0.04454 | 0.02328 | 0.4768 | 0.5447 | 0.5976 |
| $p = q = 0.5$ | s.d. | 0.0005466 | 0.0004434 | 0.006592 | 0.01678 | 0.02167 |
| | mse | 0.004582 | 0.003315 | 0.02393 | 0.04823 | 0.1003 |
| | mean | 0.04294 | 0.02203 | 0.4866 | 0.8616 | 0.9093 |
| $T = 100$ | median | 0.04294 | 0.02203 | 0.4865 | 0.8615 | 0.91 |
| $p = q = 0.8$ | s.d. | 0.0003972 | 0.0003201 | 0.005575 | 0.02563 | 0.03074 |
| | mse | 0.002965 | 0.002055 | 0.01454 | 0.06675 | 0.1136 |
| | mean | 0.04242 | 0.02163 | 0.4896 | 1.07 | 1.116 |
| $T = 100$ | median | 0.04241 | 0.02162 | 0.4897 | 1.07 | 1.117 |
| $p = q = 1$ | s.d. | 0.0003401 | 0.0002744 | 0.005115 | 0.03238 | 0.038 |
| | mse | 0.00244 | 0.001652 | 0.01163 | 0.07752 | 0.1223 |
| | mean | 0.04208 | 0.02138 | 0.4915 | 1.28 | 1.324 |
| $T = 100$ | median | 0.04208 | 0.02137 | 0.4916 | 1.278 | 1.323 |
| $p = q = 1.2$ | s.d. | 0.0002978 | 0.0002411 | 0.004673 | 0.03889 | 0.04508 |
| | mse | 0.002106 | 0.001397 | 0.009722 | 0.08873 | 0.1322 |
| | mean | 0.04179 | 0.02116 | 0.4933 | 1.592 | 1.635 |
| $T = 100$ | median | 0.04178 | 0.02114 | 0.4935 | 1.592 | 1.633 |
| $p = q = 1.5$ | s.d. | 0.0002623 | 0.0002128 | 0.004157 | 0.05076 | 0.05693 |
| | mse | 0.001812 | 0.001175 | 0.007855 | 0.105 | 0.1466 |
| | mean | 0.04228 | 0.0215 | 0.4919 | 0.5119 | 0.5288 |

continued ...

| . . . continued | | | | | | |
|---|---|---|---|---|---|---|
| $T = 150$ | median | 0.04228 | 0.0215 | 0.4918 | 0.5115 | 0.5284 |
| $p = q = 0.5$ | s.d. | 0.0001222 | 9.496e-05 | 0.005387 | 0.01392 | 0.01217 |
| | mse | 0.002281 | 0.001501 | 0.009765 | 0.01828 | 0.03123 |
| | mean | 0.04174 | 0.02111 | 0.4945 | 0.8209 | 0.8392 |
| $T = 150$ | median | 0.04174 | 0.02111 | 0.4945 | 0.8206 | 0.8384 |
| $p = q = 0.8$ | s.d. | 0.0001054 | 7.936e-05 | 0.004824 | 0.021 | 0.01913 |
| | mse | 0.001741 | 0.001111 | 0.007332 | 0.02964 | 0.04363 |
| | mean | 0.04156 | 0.02099 | 0.4951 | 1.027 | 1.048 |
| $T = 150$ | median | 0.04156 | 0.02099 | 0.4952 | 1.027 | 1.047 |
| $p = q = 1$ | s.d. | 0.000101 | 7.515e-05 | 0.004535 | 0.02637 | 0.02488 |
| | mse | 0.001565 | 0.0009879 | 0.006655 | 0.03799 | 0.05364 |
| | mean | 0.04144 | 0.0209 | 0.4956 | 1.236 | 1.258 |
| $T = 150$ | median | 0.04144 | 0.0209 | 0.4956 | 1.235 | 1.256 |
| $p = q = 1.2$ | s.d. | 9.832e-05 | 7.273e-05 | 0.004123 | 0.03095 | 0.02935 |
| | mse | 0.001447 | 0.0009062 | 0.006048 | 0.04714 | 0.06462 |
| | mean | 0.04133 | 0.02082 | 0.4962 | 1.546 | 1.57 |
| $T = 150$ | median | 0.04133 | 0.02082 | 0.4962 | 1.545 | 1.568 |
| $p = q = 1.5$ | s.d. | 9.63e-05 | 7.072e-05 | 0.003786 | 0.03976 | 0.0386 |
| | mse | 0.00133 | 0.0008264 | 0.005372 | 0.06089 | 0.07979 |

Table 3: OLS Estimates Based on "Quarterly" Analogous Linear Aggregate Model

| | | r | g | $\mathsf{E}(\lambda)$ |
|---|---|---|---|---|
| | true values | $(1.04^{.25}) - 1$ | $(1.02^{.25}) - 1$ | 0.5 |
| | mean | 0.0281 | 0.006725 | 0.2436 |
| $T = 50$ | median | 0.02805 | 0.006723 | 0.2436 |
| $p = q = 0.5$ | s.d. | 0.001048 | 4.927e-05 | 0.005053 |
| | mse | 0.01828 | 0.001835 | 0.2564 |
| | mean | 0.0214 | 0.006484 | 0.3056 |
| $T = 50$ | median | 0.02136 | 0.006483 | 0.3056 |
| $p = q = 0.8$ | s.d. | 0.0005269 | 4.227e-05 | 0.004955 |
| | mse | 0.01156 | 0.001594 | 0.1945 |
| | mean | 0.01911 | 0.006334 | 0.333 |
| $T = 50$ | median | 0.01909 | 0.006333 | 0.3328 |
| $p = q = 1$ | s.d. | 0.0003992 | 3.91e-05 | 0.004816 |
| | mse | 0.009262 | 0.001444 | 0.1671 |
| | mean | 0.01756 | 0.006202 | 0.3539 |
| $T = 50$ | median | 0.01755 | 0.006202 | 0.3538 |
| $p = q = 1.2$ | s.d. | 0.0003208 | 3.693e-05 | 0.004522 |
| | mse | 0.007711 | 0.001312 | 0.1462 |
| | mean | 0.01601 | 0.006041 | 0.3772 |
| $T = 50$ | median | 0.01601 | 0.006041 | 0.377 |
| $p = q = 1.5$ | s.d. | 0.0002635 | 3.541e-05 | 0.004365 |
| | mse | 0.006167 | 0.001151 | 0.1229 |
| | mean | 0.02274 | 0.005913 | 0.2569 |
| $T = 100$ | median | 0.02271 | 0.005912 | 0.2568 |
| $p = q = 0.5$ | s.d. | 0.0006418 | 2.41e-05 | 0.004784 |
| | mse | 0.0129 | 0.001023 | 0.2431 |
| | mean | 0.01771 | 0.005711 | 0.3187 |
| $T = 100$ | median | 0.01769 | 0.005711 | 0.3188 |
| $p = q = 0.8$ | s.d. | 0.0003085 | 1.903e-05 | 0.004648 |
| | mse | 0.007859 | 0.0008212 | 0.1813 |
| | mean | 0.01605 | 0.005614 | 0.3455 |
| $T = 100$ | median | 0.01605 | 0.005614 | 0.3453 |
| $p = q = 1$ | s.d. | 0.0002283 | 1.692e-05 | 0.004497 |
| | mse | 0.006203 | 0.0007235 | 0.1546 |
| | mean | 0.01496 | 0.005536 | 0.3656 |
| $T = 100$ | median | 0.01496 | 0.005537 | 0.3656 |
| $p = q = 1.2$ | s.d. | 0.000178 | 1.523e-05 | 0.004231 |
| | mse | 0.005107 | 0.0006459 | 0.1344 |
| | mean | 0.01389 | 0.005448 | 0.3878 |
| $T = 100$ | median | 0.01389 | 0.005448 | 0.3879 |
| $p = q = 1.5$ | s.d. | 0.0001402 | 1.396e-05 | 0.004014 |
| | mse | 0.004038 | 0.0005575 | 0.1123 |

continued . . .

. . . continued

| | | | | |
|---|---|---|---|---|
| | mean | 0.02396 | 0.00687 | 0.28 |
| $T = 150$ | median | 0.02393 | 0.006869 | 0.2799 |
| $p = q = 0.5$ | s.d. | 0.0005813 | 4.245e-05 | 0.004945 |
| | mse | 0.01411 | 0.00198 | 0.22 |
| | mean | 0.01877 | 0.006474 | 0.3386 |
| $T = 150$ | median | 0.01876 | 0.006474 | 0.3386 |
| $p = q = 0.8$ | s.d. | 0.0002982 | 3.313e-05 | 0.004705 |
| | mse | 0.008924 | 0.001584 | 0.1615 |
| | mean | 0.017 | 0.006279 | 0.3632 |
| $T = 150$ | median | 0.017 | 0.006281 | 0.3633 |
| $p = q = 1$ | s.d. | 0.0002266 | 2.943e-05 | 0.004527 |
| | mse | 0.007147 | 0.001389 | 0.1369 |
| | mean | 0.0158 | 0.006124 | 0.3816 |
| $T = 150$ | median | 0.0158 | 0.006124 | 0.3816 |
| $p = q = 1.2$ | s.d. | 0.0001814 | 2.659e-05 | 0.004213 |
| | mse | 0.005946 | 0.001233 | 0.1185 |
| | mean | 0.0146 | 0.005946 | 0.4015 |
| $T = 150$ | median | 0.0146 | 0.005946 | 0.4016 |
| $p = q = 1.5$ | s.d. | 0.000145 | 2.382e-05 | 0.003934 |
| | mse | 0.004752 | 0.001056 | 0.09859 |

Note: The simulation results in Tables 3 and 4 are based on growth and discount rates calibrated to "quarterly" observations. The other parameter values not specified in the body of the tables are $\sigma_\alpha^2 = 0.20$, $\sigma_v^2 = 0.001$, $\sigma_\xi^2 = 0.001$. The simulations are based on 1000 replications.

Table 4: Estimates Based on the "Quarterly" Optimal Aggregate Model

| | | r | g | E($\lambda$) | p | q |
|---|---|---|---|---|---|---|
| | true values | $(1.04^{.25}) - 1$ | $(1.02^{.25}) - 1$ | 0.5 | varies - see panels | |
| | mean | 0.01247 | 0.007012 | 0.4346 | 0.593 | 0.7729 |
| $T = 50$ | median | 0.01246 | 0.007001 | 0.4345 | 0.5914 | 0.7698 |
| $p = q = 0.5$ | s.d. | 0.0003447 | 0.0003001 | 0.009949 | 0.0268 | 0.05482 |
| | mse | 0.002635 | 0.002142 | 0.06619 | 0.0968 | 0.2783 |
| | mean | 0.0118 | 0.006367 | 0.4573 | 0.9492 | 1.127 |
| $T = 50$ | median | 0.01179 | 0.006365 | 0.4574 | 0.9496 | 1.129 |
| $p = q = 0.8$ | s.d. | 0.0002481 | 0.0002079 | 0.007649 | 0.04245 | 0.07016 |
| | mse | 0.001958 | 0.001491 | 0.04337 | 0.1551 | 0.3349 |
| | mean | 0.01157 | 0.006156 | 0.4649 | 1.189 | 1.369 |
| $T = 50$ | median | 0.01157 | 0.006151 | 0.4648 | 1.188 | 1.368 |
| $p = q = 1$ | s.d. | 0.0002045 | 0.0001684 | 0.006523 | 0.05239 | 0.07837 |
| | mse | 0.001729 | 0.001277 | 0.0357 | 0.1957 | 0.3772 |
| | mean | 0.0114 | 0.006004 | 0.4703 | 1.428 | 1.609 |
| $T = 50$ | median | 0.0114 | 0.006002 | 0.4705 | 1.426 | 1.605 |
| $p = q = 1.2$ | s.d. | 0.0001812 | 0.0001472 | 0.005832 | 0.06402 | 0.0909 |
| | mse | 0.00156 | 0.001123 | 0.03024 | 0.2369 | 0.4192 |
| | mean | 0.01123 | 0.005854 | 0.4758 | 1.788 | 1.97 |
| $T = 50$ | median | 0.01123 | 0.00585 | 0.4758 | 1.783 | 1.966 |
| $p = q = 1.5$ | s.d. | 0.0001578 | 0.0001264 | 0.004953 | 0.08512 | 0.1114 |
| | mse | 0.001388 | 0.0009719 | 0.0247 | 0.2999 | 0.4833 |
| | mean | 0.01047 | 0.005376 | 0.4824 | 0.5272 | 0.5663 |
| $T = 100$ | median | 0.01046 | 0.005374 | 0.4821 | 0.5263 | 0.5655 |
| $p = q = 0.5$ | s.d. | 0.0001874 | 0.0001655 | 0.009207 | 0.01839 | 0.03168 |
| | mse | 0.0006407 | 0.0005126 | 0.0199 | 0.03281 | 0.07347 |
| | mean | 0.01027 | 0.00518 | 0.4913 | 0.836 | 0.8661 |
| $T = 100$ | median | 0.01026 | 0.005178 | 0.4914 | 0.8358 | 0.865 |
| $p = q = 0.8$ | s.d. | 0.0001169 | 0.0001015 | 0.006655 | 0.02733 | 0.03844 |
| | mse | 0.0004279 | 0.0003068 | 0.01097 | 0.04518 | 0.07642 |
| | mean | 0.01021 | 0.005128 | 0.4937 | 1.041 | 1.068 |
| $T = 100$ | median | 0.01021 | 0.005129 | 0.4937 | 1.04 | 1.069 |
| $p = q = 1$ | s.d. | 9.559e-05 | 8.241e-05 | 0.00581 | 0.03406 | 0.04465 |
| | mse | 0.0003727 | 0.0002515 | 0.008597 | 0.05296 | 0.08105 |
| | mean | 0.01018 | 0.005097 | 0.4951 | 1.246 | 1.271 |
| $T = 100$ | median | 0.01018 | 0.005096 | 0.4952 | 1.244 | 1.268 |
| $p = q = 1.2$ | s.d. | 8.125e-05 | 6.976e-05 | 0.005188 | 0.04047 | 0.05097 |
| | mse | 0.0003392 | 0.0002179 | 0.00712 | 0.0611 | 0.08715 |
| | mean | 0.01016 | 0.005072 | 0.4965 | 1.552 | 1.574 |
| $T = 100$ | median | 0.01015 | 0.005069 | 0.4966 | 1.552 | 1.574 |
| $p = q = 1.5$ | s.d. | 6.819e-05 | 5.835e-05 | 0.004479 | 0.05207 | 0.06176 |
| | mse | 0.0003117 | 0.0001904 | 0.005698 | 0.07342 | 0.09649 |
| | mean | 0.01017 | 0.00512 | 0.4928 | 0.5115 | 0.5266 |

continued ...

. . . continued

| | | | | | | |
|---|---|---|---|---|---|---|
| $T = 150$ | median | 0.01017 | 0.005119 | 0.4928 | 0.5114 | 0.5264 |
| $p = q = 0.5$ | s.d. | 6.696e-05 | 5.715e-05 | 0.006454 | 0.01484 | 0.01783 |
| | mse | 0.0003228 | 0.0002366 | 0.009702 | 0.01875 | 0.03201 |
| | mean | 0.01016 | 0.005098 | 0.4945 | 0.8219 | 0.8403 |
| $T = 150$ | median | 0.01016 | 0.005096 | 0.4944 | 0.8217 | 0.8402 |
| $p = q = 0.8$ | s.d. | 4.743e-05 | 3.885e-05 | 0.005309 | 0.02241 | 0.02483 |
| | mse | 0.0003087 | 0.0002106 | 0.007657 | 0.0313 | 0.04731 |
| | mean | 0.01015 | 0.005087 | 0.4951 | 1.029 | 1.049 |
| $T = 150$ | median | 0.01015 | 0.005087 | 0.4952 | 1.028 | 1.048 |
| $p = q = 1$ | s.d. | 4.141e-05 | 3.305e-05 | 0.004823 | 0.02828 | 0.03049 |
| | mse | 0.0003021 | 0.0001995 | 0.006884 | 0.04029 | 0.05785 |
| | mean | 0.01015 | 0.005078 | 0.4956 | 1.237 | 1.259 |
| $T = 150$ | median | 0.01015 | 0.005078 | 0.4956 | 1.236 | 1.257 |
| $p = q = 1.2$ | s.d. | 3.761e-05 | 2.949e-05 | 0.004361 | 0.03358 | 0.03585 |
| | mse | 0.0002946 | 0.00019 | 0.006206 | 0.04986 | 0.069 |
| | mean | 0.01014 | 0.005068 | 0.4963 | 1.547 | 1.571 |
| $T = 150$ | median | 0.01014 | 0.005068 | 0.4962 | 1.547 | 1.57 |
| $p = q = 1.5$ | s.d. | 3.39e-05 | 2.599e-05 | 0.003917 | 0.04264 | 0.04463 |
| | mse | 0.0002849 | 0.0001794 | 0.005416 | 0.06379 | 0.08371 |