ISSN 1518-3548

BANCO CENTRAL DO BRASIL

Working Paper Series

4

An Information Theory Approach to the Aggregation of Log-Linear Models

Pedro H. Albuquerque

July, 2000

ISSN 1518-2665

CGC 00.038.166/0001-05

Working Papers Series	Brasília	n. 4	Jul	2000	P. 1 – 26

Working Paper Series

Edited by:

Research Department (Depep)
(e-mail: conep.depep@bcb.gov.br)

Reproduction permitted only if source is stated as follows: Working Paper Series n. 4.

General Control of Subscription:

Banco Central do Brasil
Demap/Disud/Subip
SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo
70074-900 – Brasília (DF)
Telefone (61) 414-1392
Fax (61) 414-3165

Number printed: 450 copies

The views expressed in this work are those of the authors and do not reflect those of the Banco Central or its members.

Although these Working Papers often represent preliminary work, citation of source is required when used or reproduced.

Banco Central do Brasil Information Bureau

Address:	Secre/Surel/Dinfo
	Edifício-Sede, 2º subsolo
	SBS - Quadra 3, Zona Central
	70074-900 - Brasília (DF)
Phones:	(61) 414 (....) 2401, 2402, 2403, 2404, 2405, 2406
DDG:	0800 992345
FAX:	(61) 321 9453
Internet:	http://www.bcb.gov.br
E-mail:	cap.secre@bcb.gov.br
	dinfo.secre@bcb.gov.br

An Information Theory Approach to the Aggregation of Log-Linear Models

Pedro H. Albuquerque
Banco Central do Brasil

JEL Classification: C43
Keywords: Aggregation, Log-Linear, Information Theory

Abstract

In this paper, an unrestricted aggregation method for heterogeneous log-linear functions is presented. It employs inequality measures derived from information theory in the construction of an exact representation of the aggregate behavior of the economy. A condition for the identification of average micro parameters is proposed. It is shown that the method leads to previous results in the field when adequate restrictions are imposed. Two macroeconomic applications are discussed: the aggregation of the Lucas supply function and the time-inconsistent behavior of an egalitarian social planner facing agents with heterogeneous discount rates.

The author would like to thank W. Brock, J. Faria, Y. Kitamura, A. Lewbel, R. Manuelli, A. de Oliveira, H. Pesaran and K. West for helpful comments. Existing errors are nevertheless the sole responsibility of the author.

> The reaction of one man could be forecast by no known mathematics; the reaction of a billion is something else again.
>
> – *Foundation and Empire*, Isaac Asimov (1952)

1 Introduction

Although the search for microeconomic underpinnings of macroeconomic models has led to an increased interest in aggregation issues, the representative agent hypothesis is yet common in macroeconomic research. One possible reason is the technical difficulty that arises with heterogeneity.

On one hand, numerical simulation has been used to deal with the problem complexity. It is now common to find works that use simulation to analyze among other things how heterogeneity across economic units affects the aggregate behavior of the economy. On the other hand, symbolic approaches have been considered among others by Theil (1954, 1967 and 1971), Deaton and Muellbauer (1980), Stoker (1984, 1986 and 1993), Lewbel (1992) and Garderen, Lee and Pesaran (2000).

Following Garderen, Lee and Pesaran (2000), the literature is mainly divided between the *aggregation problem* and the *model selection problem*. The first deals with the conditions for the identification of micro parameters from aggregate macroeconomic models. The second deals with the optimal choice between aggregate and disaggregate models.

Although following the aggregation approach, this paper tries to go beyond identification conditions. It is shown that, at least in the heterogeneous log-linear case, an aggregate equation can represent the exact relationship among aggregate variables, as long as it includes additional terms containing inequality measures from information theory. Note that the log-linear case is of special interest, since it represents the usual modeling approach in theoretical and empirical macroeconomics.

The paper has the following structure: it starts presenting the information theory concepts that will be used throughout the paper, followed by the aggregation method. Then it states the condition that conduces to the identification of micro parameters (or its means) in the case of a log-linear model of the aggregate variables. Previous findings in the field are verified through the imposition of adequate restrictions.

Two macroeconomic applications are provided in the last section. First, it is shown how the Lucas supply function changes when its micro components are aggregated not in logs, as usual, but rather in levels, as in reality. Second, it is verified that the behavior of an egalitarian social planner is time-inconsistent when the planner faces agents with heterogeneous exponential discount rates, since in this case the resulting aggregate discount rate is not exponential.

2 Aggregation

To aggregate log-linear equations, consider the use of an information theory concept called *expected information of an indirect message*:

$$I(Y:X)=\int f_Y(\lambda)\ln\left[\frac{f_Y(\lambda)}{f_X(\lambda)}\right]d\lambda,\quad Y\sim f_Y(\lambda),\quad X\sim f_X(\lambda),\tag{1.1}$$

which measures the expected information carried by a message that transforms a prior probability density function f_X to a posterior f_Y.[1]

Theil (1967, 1971) used the expected information concept to propose inequality measures. One of them is known as *Theil's second measure*, which is defined as[2]

$$L(Y)=I(Y:X^*)=\int f_Y(\lambda)\ln\left(\frac{E[Y]}{\lambda}\right)d\lambda,\quad Y\sim f_Y(\lambda),\quad X^*\sim\frac{\lambda f_Y(\lambda)}{E[Y]}.^3$$

---

[1] See, for example, Theil (1967, p. 27, 1971, p. 641). It is also called *mean information*; see Kullback (1968, p. 5).

[2] The notation for this measure is L in Bourguignon (1979), I_0 in Shorrocks (1980), T_2 in Nygård & Sandström (1981, pp. 146 and 251), and I_{-1} in Maasoumi (1986). Bourguignon (1979) offers the following comment on Theil's second measure: "That the inequality measure L has seldom been used in applied works on income distribution is somewhat surprising because it has very much to commend it. Besides the fact that it is decomposable … and satisfies the basic properties of an inequality measure, L lends itself to a very simple interpretation in terms of social welfare. In the utilitarian framework, the social welfare function is the sum of identical concave individual utility function. If we choose the logarithm form for those utility functions, L is simply the difference between the maximum social welfare for a given total income, which corresponds to the equalitarian distribution, and the actual social welfare."

The sample analog of this measure is given by

$$\overline{L}(\mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{\overline{Y}}{Y_n}\right) = \ln\left[\frac{\overline{Y}}{\tilde{Y}}\right], \tag{1.2}$$

where

$$\overline{L} : R_+^N \to [0,\infty), \quad \mathbf{Y} = [Y_1 \cdots Y_N]', \quad \overline{Y} = \frac{1}{N}\sum_{n=1}^{N} Y_n\,, \quad \tilde{Y} = \left(\prod_{n=1}^{N} Y_n\right)^{\frac{1}{N}}.$$

Note that $\mathbf{Y}$ is a vector of N values taken by a variable Y, $\overline{Y}$ is the arithmetic mean of $\mathbf{Y}$, and $\tilde{Y}$ is the geometric mean of $\mathbf{Y}$. Theil's second measure represents the degree of relative dispersion among the units.

Consider now the following:

Proposition 1: *Given I+1 vectors representing the values taken by I+1 variables for N units at time t, and given I parameter vectors,*[4] *as described below:*

$$\mathbf{Y}_t = \begin{bmatrix} Y_{1t} \\ \vdots \\ Y_{nt} \\ \vdots \\ Y_{Nt} \end{bmatrix}, \qquad \mathbf{X}_{it} = \begin{bmatrix} X_{i1t} \\ \vdots \\ X_{int} \\ \vdots \\ X_{iNt} \end{bmatrix}, \quad \mathbf{a}_i = \begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \\ \vdots \\ a_{iN} \end{bmatrix},$$

$$Y_{nt} > 0,\ X_{int} > 0,\ i = 1,\ldots,I,\ n = 1,\ldots,N,\ \forall t\,.$$

If a log-linear functional form with heterogeneous parameters across units describes the microeconomic relationships in the economy:

$$Y_{nt} = X_{1nt}^{a_{1n}} X_{2nt}^{a_{2n}} \cdots X_{Int}^{a_{In}}\,, \tag{1.3}$$

---

[3] Note that *Theil's first measure* would be defined accordingly as $T(Y) = I(X^*{:}Y)$.
[4] Parameters could be time variant, with the same results.

then the relationship among the aggregate variables $\overline{Y}_t$ and $\overline{X}_{it}$ at each period t will be given by

$$\overline{Y}_t = \overline{X}_{1t}^{\overline{a}_1} \overline{X}_{2t}^{\overline{a}_2} \cdots \overline{X}_{It}^{\overline{a}_I} D(\Phi_t), \tag{1.4}$$

where

$$D(\Phi_t) = \exp\left\{\overline{L}(\mathbf{Y}_t) - \sum_{i=1}^{I}\left[\overline{a}_i \overline{L}(\mathbf{X}_{it}) - \overline{\operatorname{cov}}(\mathbf{x}_{it}, \mathbf{a}_i)\right]\right\}, \tag{1.5}$$

$$\overline{Y}_t = \frac{1}{N}\sum_{n=1}^{N} Y_{nt}\,,\ \overline{X}_{it} = \frac{1}{N}\sum_{n=1}^{N} X_{int}\,,\ \overline{a}_i = \frac{1}{N}\sum_{n=1}^{N} a_{in}\,,$$

$$\overline{\operatorname{cov}}(\mathbf{x}_{it}, \mathbf{a}_i) = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_{int} - \overline{\mathbf{x}}_{it})(\mathbf{a}_{in} - \overline{\mathbf{a}}_i),$$

$$\Phi_t = (\mathbf{X}_{1t}, \ldots, \mathbf{X}_{It}, \mathbf{a}_1, \ldots, \mathbf{a}_I), \qquad \mathbf{x}_{it} = \left[\ln(X_{i1t}) \ \ \cdots \ \ \ln(X_{iNt})\right]',$$

$$\mathbf{Y}_t = \mathbf{X}_{1t}^{\mathbf{a}_1} * \cdots * \mathbf{X}_{It}^{\mathbf{a}_I}, \qquad \mathbf{X}_{it}^{\mathbf{a}_i} = \left[X_{i1t}^{a_{i1}} \ \ \cdots \ \ X_{iNt}^{a_{iN}}\right]',$$

and the symbol * represents the Hadamard product (the component-by-component product). Note that $\overline{\operatorname{cov}}(\mathbf{x}_{it}, \mathbf{a}_i)$ is the cross-sectional sample covariance between the logarithm of the variables and their respective parameters.

The proof of Proposition 1 is given in Appendix 1.

Except for the term $D(\Phi_t)$, the aggregate equation is analogous to a representative unit equation. The dispersions of the variables and parameters are relevant here, as expected, with effects represented by $D(\Phi_t)$. The proposition shows, however, that it is possible to isolate the effects of these dispersions from the effects of the aggregate variable. The aggregate variable $\overline{Y}_t$ depends not only on the aggregate variables $\overline{X}_{1t}$ to $\overline{X}_{It}$, but also on inequality measures of each variable vector. Additionally, note that Proposition 1 does not rely on any hypothesis regarding statistical properties of variables and parameters.

Proposition 1 leads to an intuitive interpretation of the aggregation problem. The inequality measures can be seen as aggregate measures of the loss of information caused by the representative agent assumption. Given that, an exact aggregate model can be established by taking these measures in consideration when formulating the aggregate equation.

3 Identification of Micro Parameters

Proposition 1 provides a simple condition for the identification of micro parameters or, in the case where the units have heterogeneous parameters, for the identification of average micro parameters.

Start adding a disturbance term to equation (1.3):

$$Y_{nt} = X_{1nt}^{a_{1n}} X_{2nt}^{a_{2n}} \cdots X_{Int}^{a_{In}} e^{u_{nt}} .$$

From Proposition 1, it is known that the aggregate equation is

$$\overline{Y}_t = \overline{X}_{1t}^{\overline{a}_1} \overline{X}_{2t}^{\overline{a}_2} \cdots \overline{X}_{It}^{\overline{a}_I} D(\Phi_t) e^{\overline{u}_t} , \tag{2.1}$$

where

$$\overline{u}_t = \frac{1}{N} \sum_{n=1}^{N} u_{nt} .$$

Note that $\overline{L}(\mathbf{X}) = 0$ when all components of $\mathbf{X}$ are equal. Also, note that $\overline{\text{cov}}(\mathbf{x}, \mathbf{a}) = 0$ when all components of either $\mathbf{X}$ or $\mathbf{a}$ are equal. If these two properties are applied to equation (2.1) then it is easy to verify that $D(\Phi_t)$ continues defined as in equation (1.5) despite the introduction of a disturbance term in the micro equations. Now, consider the logarithmic representation of equation (2.1):

$$y_t = \overline{a}_1 x_{1t} + \overline{a}_2 x_{2t} + \cdots + \overline{a}_I x_{It} + d(\Phi_t) + \overline{u}_t , \tag{2.2}$$

where

$$y_t = \ln(\overline{Y}_t),\ x_{it} = \ln(\overline{X}_{it}),\ d(\Phi_t) = \overline{L}(\mathbf{Y}_t) - \sum_{i=1}^{I} \left[\overline{a}_i \overline{L}(\mathbf{X}_{it}) - \overline{\text{cov}}(\mathbf{x}_{it}, \mathbf{a}_i) \right].$$

First, note that if the values of $d(\Phi_t)$ were observed for each observation t in a time series sample of aggregates, then they could be used to estimate the parameters in equation (2.2). In general, however, the inequality measures are not known. In this case, the condition for obtaining consistent estimators of the average micro parameters using a log-linear equation of aggregate variables would be

$$\operatorname{plim} \overline{\operatorname{cov}}\left(\begin{bmatrix} x_{i1} & \cdots & x_{iT} \end{bmatrix}', \begin{bmatrix} \left(d(\Phi_1) + \overline{u}_1\right) & \cdots & \left(d(\Phi_T) + \overline{u}_T\right) \end{bmatrix}' \right) = 0, \ \forall i. \qquad (2.3)$$

This condition is obtained, for example, when the aggregate disturbance $\overline{u}_t$ and the population heterogeneity measured by $d(\Phi_t)$ (a combination of relative dispersion measures) are both independent from every aggregate variable x_{it} across time. Note that an instrumental-variable estimator can be employed when this condition is not valid. In terms of economic meaning, it is necessary to find instruments that explain the aggregate variables but have no relationship in the covariance sense with $d(\Phi_t) + \overline{u}_t$. For example, consider a consumption model where consumption is a function of income. An instrument could be a variable that, although capable of explaining the aggregate income, has zero covariance with the dispersions of income and consumption, as measured by the components of $d(\Phi_t)$.

4 Special Cases of Log-Linear Aggregation

Some well-known log-linear aggregation properties can be reproduced using equation (1.4) and imposing adequate restrictions. Note that proving tends to be simpler when the approach developed here is employed. Consider the following examples.

4.1 Scale-Invariant Distributions

Lewbel (1992) and Garderen, Lee and Pesaran (2000) discuss the case of *scale-invariant distributions*.[5] They are defined as

$$p_t\left(Z_{1nt},\ldots,Z_{Int}\middle|\overline{X}_{1t},\ldots,\overline{X}_{It},\Theta_t\right)=p_t\left(Z_{1nt},\ldots,Z_{Int}\middle|\Theta_t\right),\quad \forall n,t\,,$$

$$Z_{int}=\frac{X_{int}}{\overline{X}_{it}},\quad \forall i,n,t\,,$$

where $\overline{X}_{it}$ is assumed to be independent of Θ_t over t for every i.

Assume that the microeconomic functions are defined as

$$Y_{nt}=X_{1nt}^{a_1}X_{2nt}^{a_2}\cdots X_{Int}^{a_I}e^{u_{nt}}\,. \tag{3.1}$$

The equation above is equivalent to equation (1.3) with homogeneous parameters. The unobserved disturbances are distributed in accordance with a density

$$q_t\left(u_{nt}\middle|\overline{X}_{1t},\ldots,\overline{X}_{It},\Theta_t\right)=q_t\left(u_{nt}\middle|\Theta_t\right),\quad \forall n,t\,,$$

and their expected values obey to the condition

$$E_t\left[u_{nt}\middle|\overline{X}_{1t},\ldots,\overline{X}_{It},\Theta_t\right]=E_t\left[u_{nt}\middle|\Theta_t\right]=0\,.$$

Lewbel (1992) has shown that, given a scale-invariant distribution, the micro parameters in equation (3.1) can be recovered from the log-linear aggregate equation. This result can be verified using Proposition 1. Consider random realizations of Z_{int} for every n, i and t in a sample, drawn from $p_t\left(Z_{1nt},\ldots,Z_{Int}\middle|\Theta_t\right)$. Note that, since the micro parameters are constant across units, equation (1.5) simplifies to

$$D(\Phi_t)=\exp\left\{\overline{L}(\mathbf{Y}_t)-\sum_{i=1}^{I}a_i\overline{L}(\mathbf{X}_{it})\right\}.$$

[5] They are also called *mean-scaled distributions*.

Now, observe that

$$\overline{L}(\mathbf{X}_{it}) = \overline{L}(\overline{X}_{it}\mathbf{Z}_{it}) = \overline{L}(\mathbf{Z}_{it}),$$

$$\overline{L}(\mathbf{Y}_t) = \overline{L}\left(\mathbf{X}_{1t}^{a_1} * \cdots * \mathbf{X}_{It}^{a_I} * e^{\mathbf{u}_t}\right) = \overline{L}\left(\overline{X}_{1t}^{a_1}\mathbf{Z}_{1t}^{a_1} * \cdots * \overline{X}_{It}^{a_I}\mathbf{Z}_{It}^{a_I} * e^{\mathbf{u}_t}\right) \Rightarrow$$

$$\overline{L}(\mathbf{Y}_t) = \overline{L}\left(\mathbf{Z}_{1t}^{a_1} * \cdots * \mathbf{Z}_{It}^{a_I} * e^{\mathbf{u}_t}\right),$$

meaning that $\overline{L}(\mathbf{Y}_t)$ and $\overline{L}(\mathbf{X}_{it})$ are independent from $\overline{X}_{it}, \forall i$. The independence is due to two reasons. First, the vectors $(\mathbf{Z}_{1t}, \ldots, \mathbf{Z}_{It})$ and the disturbance vector $\mathbf{u}_t$ have been drawn from distributions that are independent from $\overline{X}_{it}$. Second, the function $\overline{L}(\mathbf{X}_{it})$ is homogeneous of degree zero. Hence, after applying Proposition 1 to equation (3.1) and taking the logarithm, it turns out that

$$y_t = a_1 x_{1t} + a_2 x_{2t} + \cdots + a_I x_{It} + d(\Phi_t) + \overline{u}_t, \tag{3.2}$$

$$\Phi_t = \left(\mathbf{Z}_{1t}, \ldots, \mathbf{Z}_{It}, e^{\mathbf{u}_t}, a_1, \ldots, a_I\right),\ d(\Phi_t) = \overline{L}\left(\mathbf{Z}_{1t}^{a_1} * \ldots * \mathbf{Z}_{It}^{a_I} * e^{\mathbf{u}_t}\right) - \sum_{i=1}^{I} a_i \overline{L}(\mathbf{Z}_{it}).$$

Note that condition (2.3) is valid since

$$\operatorname{cov}\left(x_{it}, \left(d(\Phi_t) + \overline{u}_t\right)\right) = \operatorname{cov}\left(\ln \overline{X}_{it}, \left(d(\Phi_t) + \overline{u}_t\right)\right) = 0, \quad \forall i.$$

The term $d(\Phi_t)$ was shown above to be independent from $x_{it} = \ln \overline{X}_{it}$, and the term $\overline{u}_t$ is independent by assumption. Therefore, Lewbel's result can be verified using Proposition 1: the micro parameters can be recovered from the log-linear aggregate equation when the distributions are scale invariant.

4.2 Lognormal Distributions

An important simplification arises when the distributions are lognormal. First, consider the relation between Theil's second measure, the characteristic function $\phi_{g(X)}(\lambda)$, the moment generating function $M_{g(X)}(\tau)$, and the cumulant generating function $K_{g(X)}(\tau)$:[6]

$$\phi_{g(X)}(\lambda = -1) = M_{g(X)}(\tau = 1) = E\left[e^{g(X)}\right] = e^{\ln E[X] - E[\ln X]} = \exp[L(X)],$$

---

[6] See Theil (1971, p. 367) and Amemiya (1985, p. 91).

$$K_{g(X)}(\tau=1)=\ln\left[M_{g(X)}(\tau=1)\right]=\sum_{j=1}^{\infty}\frac{\kappa_j}{j!}=L(X),\quad g(X)=x-E[x],\quad x=\ln(X),$$

where κ_j is the *j*-th cumulant of $g(X)$, for example:

$$\kappa_1=E[g(X)]=0,\ \kappa_2=E\left[(g(X)-\kappa_1)^2\right]=E\left[g(X)^2\right]=\mathrm{var}(x),$$

$$\kappa_3=E\left[(g(X)-\kappa_1)^3\right]=E\left[g(X)^3\right],\ \kappa_4=E\left[(g(X)-\kappa_1)^4\right]-3\kappa_2^2=E\left[g(X)^4\right]-3\,\mathrm{var}(x)^2\ .$$

From the relations above, the values of $L(X)$ for different distributions of X can be calculated using a characteristic function or cumulant table.

For example, if each element of **X** is randomly drawn from an independent lognormal distribution, it means, $\ln X_n \sim N\left(\mu,\sigma^2\right)$ for $\forall n$, then

$$\mathrm{plim}\,\overline{L}(\mathbf{X})=\frac{\sigma^2}{2}. \tag{3.3}$$

Consider now the following *fixed distribution* case:[7]

$$Y_{nt}=X_{1nt}^{a_{1n}}X_{2nt}^{a_{2n}}\cdots X_{Int}^{a_{In}}, \tag{3.4}$$

where the variables for each unit and time are drawn from a joint lognormal distribution where

$$p(\ln Z_{1nt},\ldots,\ln Z_{Int})\sim N(\mathrm{M},\Sigma),\ \forall n,t,$$

$$\mathrm{M}=\begin{bmatrix}\mu_1\\ \vdots\\ \mu_I\end{bmatrix},\ \Sigma=\begin{bmatrix}\sigma_1^2 & \cdots & \sigma_{1I}\\ \vdots & \ddots & \vdots\\ \sigma_{I1} & \cdots & \sigma_I^2\end{bmatrix},\ X_{int}=\theta_{it}Z_{int},$$

and where

$$\mathrm{plim}\,\overline{\mathrm{cov}}(\mathbf{a}_i,\mathbf{a}_j)=\mathrm{cov}(a_i,a_j)=\sigma_{aij},\quad \mathrm{plim}\,\overline{\mathrm{var}}(\mathbf{a}_i)=\mathrm{var}(a_i)=\sigma_{ai}^2,$$

[7] Fixed distributions are discussed e.g. in Stoker (1984).

meaning that

$$\operatorname{plim}\overline{\operatorname{cov}}(\mathbf{a}_i, \mathbf{x}_{it}) = 0\,, \tag{3.5}$$

$$\operatorname{plim}\overline{\operatorname{cov}}(\mathbf{a}_i * \mathbf{x}_{it}, \mathbf{a}_j * \mathbf{x}_{jt}) = \sigma_{aij}\sigma_{ij} + \sigma_{ij}\overline{a}_i\overline{a}_j + \sigma_{aij}\overline{x}_{it}\overline{x}_{jt} \Rightarrow$$

$$\operatorname{plim}\overline{\operatorname{cov}}(\mathbf{a}_i * \mathbf{x}_{it}, \mathbf{a}_j * \mathbf{x}_{jt}) = \sigma_{aij}\sigma_{ij} + \sigma_{ij}\overline{a}_i\overline{a}_j + \sigma_{aij}\left(x_{it} - \frac{\sigma_i^2}{2}\right)\left(x_{jt} - \frac{\sigma_j^2}{2}\right), \tag{3.6}$$

and $\operatorname{plim}\overline{\operatorname{var}}(\mathbf{a}_i * \mathbf{x}_{it}) = \sigma_{ai}^2\sigma_i^2 + \sigma_i^2\overline{a}_i^2 + \sigma_{ai}^2\overline{x}_{it}^2 \Rightarrow$

$$\operatorname{plim}\overline{\operatorname{var}}(\mathbf{a}_i * \mathbf{x}_{it}) = \sigma_{ai}^2\sigma_i^2 + \sigma_i^2\overline{a}_i^2 + \sigma_{ai}^2\left(x_{it} - \frac{\sigma_i^2}{2}\right)^2, \tag{3.7}$$

where (1.2) was used to obtain the relationship $\overline{x}_{it} = x_{it} - \overline{L}(\mathbf{X}_{it})$.

Using Proposition 1, and assuming that $N \to \infty$, it is shown in Appendix 2 that the aggregate equation in its logarithmic form is

$$y_t = c + \sum_{i=1}^{I}\left[\left(\overline{a}_i - \frac{\sigma_{ai}^2\sigma_i^2}{2}\right)x_{it} + \frac{\sigma_{ai}^2}{2}x_{it}^2\right]$$
$$+ \sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left[\sigma_{aij}x_{it}x_{jt} - \frac{\sigma_{aij}\sigma_j^2}{2}x_{it} - \frac{\sigma_{aij}\sigma_i^2}{2}x_{jt}\right] + r_t\,, \tag{3.8}$$

where c is a constant.

Note that the aggregate equation above is not log-linear: it includes squares and cross products of the aggregate variables, and it can have higher-order terms through r_t. It can be seen from equation (3.8) that, as noted in Garderen, Lee and Pesaran (2000), “When the parameters of the log-linear specifications differ across micro units, the slope coefficients (or their means) are no longer identifiable from the [log-linear] aggregate model.”

Finally, consider the particular case where the parameters do not vary across units. In this case, equation (3.8) simplifies to

$$y_t = a_1x_{1t} + \cdots + a_Ix_{It} + d(\Phi_t),$$

where

$$d(\Phi_t) = \sum_{i=1}^{I}\left[a_i(a_i - 1)\frac{\sigma_i^2}{2}\right] + \sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left[\sigma_{ij}a_i a_j\right]$$

is constant.

Therefore, the aggregate equation mimics the microeconomic equation in the case of homogeneous parameters. This is a simple example of the scale-invariant distribution case discussed before, where the micro parameters can be identified from a log-linear aggregate equation.

The results of this section show that, under the hypothesis of fixed distributions, the recoverability of the parameters from a log-linear aggregate equation is possible only when the parameters are homogeneous across the economy. This is in agreement with the aggregation properties of log-linear models that are discussed, for example, in Garderen, Lee and Pesaran (2000). The parameters would be identifiable, however, if Theil's second measures of the variables were known and were included in the equation.

5 Applications to Macroeconomic Modeling

5.1 The Lucas Supply Function

In this section, Proposition 1 is used to aggregate the Lucas supply function, which was originally developed in Lucas (1973), this time not in its logarithm version as usual but rather in its level version.[8]

Consider the microeconomic function in the Lucas model:

$$y_{nt} = \beta\left(p_{nt} - E\left[\overline{p}_t \middle| I_t\right]\right),\ \beta = \frac{b\sigma_z^2}{\sigma_p^2 + \sigma_z^2},\ p_{nt} = \overline{p}_t + z_{nt},$$

[8] Following Blanchard and Fischer (1989, p. 366), "This is another instance in which the use of logarithms creates problems. Output is implicitly defined as the product of individual outputs rather then their sum."

where y_{nt} is the log of output for market *n* and time *t*, p_{nt} is the log of price level for market *n*, $\overline{p}_t$ is the log of the geometric mean of price levels in all markets (the average of logs), I_t is the information before *t*, *b* is the slope of the supply function for each market, σ_z^2 is the variance of idiosyncratic price shocks z_{nt}, $z_{nt} \sim IN\left(0, \sigma_z^2\right)$, and σ_p^2 is the variance of the prior distribution of $\overline{p}_t$, $\overline{p}_t \sim IN\left(E\left[\overline{p}_t \middle| I_t\right], \sigma_p^2\right)$.

The usual aggregation procedure is to sum over the logs, yielding a per capita function of the type

$$\overline{y}_t = \beta\left(\overline{p}_t - E\left[\overline{p}_t \middle| I_t\right]\right).$$

This is the Lucas supply function. Note that all the aggregate variables here are constructed as geometric means.

If Proposition 1 is used to aggregate the microeconomic functions over the levels, a more realistic approach, a different result is found. Start with

$$y_{nt} = \beta\left(p_{nt} - E\left[\overline{p}_t \middle| I_t\right]\right),$$

and then using Proposition 1:

$$y_t = d(\Phi_t) + \beta\left(p_t - E\left[\overline{p}_t \middle| I_t\right]\right),$$

where $d(\Phi_t) = \overline{L}(\mathbf{Y}_t) - \beta\overline{L}(\mathbf{P}_t)$, and

$$\mathbf{Y}_t = \begin{bmatrix} e^{y_{1t}} & \cdots & e^{y_{Nt}} \end{bmatrix}', \quad \mathbf{P}_t = \begin{bmatrix} e^{p_{1t}} & \cdots & e^{p_{Nt}} \end{bmatrix}'.$$

Given the assumptions about the distributions, it is known that

$$\operatorname{plim} \overline{L}(\mathbf{P}_t) = \frac{\sigma_z^2}{2}, \ \operatorname{plim} \overline{L}(\mathbf{Y}_t) = \operatorname{plim} \overline{L}\left(\mathbf{P}_t^{\beta}\right) = \beta^2 \frac{\sigma_z^2}{2},$$

$$\operatorname{plim} E\left[\overline{p}_t \middle| I_t\right] = \operatorname{plim}\left\{E\left[p_t \middle| I_t\right] - \overline{L}(\mathbf{P}_t)\right\} = E\left[p_t \middle| I_t\right] - \frac{\sigma_z^2}{2},$$

where $\mathbf{P}_t^{\beta} = \begin{bmatrix} e^{\beta p_{1t}} & \cdots & e^{\beta p_{Nt}} \end{bmatrix}'$.

Therefore, for a large enough *N*, the aggregate supply function is given by

$$y_t = \beta^2 \frac{\sigma_z^2}{2} + \beta\left(p_t - E\left[p_t \middle| I_t\right]\right).$$

Consequently, the aggregation over levels introduces an intercept in the equation that will be affected by changes in the prior distribution of p_t. The result indicates that the long-run equilibrium level of aggregate supply in this economy will be affected by changes in the governmental policies. A change in policy that increases σ_p^2 will also decrease the equilibrium level of output. This is, however, a purely aggregational result, generated by the presence of nonlinearities in the microeconomic reaction functions.

5.2 The Time-Inconsistent Egalitarian Social Planner

Suppose that the household utility function in a heterogeneous economy commanded by an egalitarian social planner is given by

$$U_{nt} = \sum_{j=0}^{\infty} \beta_n^j C_{nt+j}^{\sigma}, \quad 0 < \beta_n < 1,\ 0 < \sigma < 1 \text{, or}$$

$$U_{nt} = \sum_{j=0}^{\infty} e^{-j\theta_n} C_{nt+j}^{\sigma}, \quad \theta_n > 0,\ \beta_n = e^{-\theta_n}.$$

The discount rates of the agents are assumed, as an example, to be randomly drawn from an exponential distribution with density

$$f_\theta(\lambda) = \begin{cases} \frac{1}{\xi} e^{-\frac{\lambda - \theta_{\min}}{\xi}} & \text{if } \lambda \geq \theta_{\min} \\ 0 & \text{if } \lambda < \theta_{\min} \end{cases},$$

$$E[\theta] = \theta_{\min} + \xi, \quad \operatorname{var}(\theta) = \xi^2, \quad \theta_{\min} > 0, \quad \xi > 0.$$

Note that the discount rate of each agent is the only random element in this setup. Assume that the number of families is large ($N \to \infty$). The social planner knows the distribution of the discount rates, but cannot identify the discount rate of any specific agent, being thereafter egalitarian, $C_{nt} = C_t, \forall n$.[9] Given these restrictions, the social planner will try to maximize the utility function of a representative agent given by

$$U_t = \operatorname{plim} \frac{1}{N} \sum_{n=1}^{N} \sum_{j=0}^{\infty} e^{-j\theta_n} C_{t+j}^{\sigma} = \sum_{j=0}^{\infty} \left(\operatorname{plim} \frac{1}{N} \sum_{n=1}^{N} e^{-j\theta_n} \right) C_{t+j}^{\sigma} .$$

Using a characteristic function table, as in Abramowitz and Stegun (1972, p. 930), it can be easily shown that

$$\operatorname{plim} \overline{L} \left(\begin{bmatrix} e^{-j\theta_1} & \cdots & e^{-j\theta_N} \end{bmatrix}' \right) = \ln \left(\frac{e^{j\xi}}{1 + j\xi} \right),$$

and hence, using Proposition 1:

$$\operatorname{plim} \frac{1}{N} \sum_{n=1}^{N} e^{-j\theta_n} = \operatorname{plim} \exp \left[\overline{L} \left(\begin{bmatrix} e^{-j\theta_1} & \cdots & e^{-j\theta_N} \end{bmatrix}' \right) - j\overline{\theta} \right] = \operatorname{plim} \frac{e^{j(\xi - \overline{\theta})}}{1 + j\xi} \Rightarrow$$

$$\operatorname{plim} \frac{1}{N} \sum_{n=1}^{N} e^{-j\theta_n} = \frac{e^{-j\theta_{\min}}}{1 + j\xi} = \frac{\beta_{\max}^{j}}{1 + j\xi}, \; \beta_{\max} = e^{-\theta_{\min}} .$$

Thereafter

$$U_t = \sum_{j=0}^{\infty} \beta(j) C_{t+j}^{\sigma}, \quad \beta(j) = \frac{\beta_{\max}^{j}}{1 + j\xi},$$

meaning that the social planner faces a representative agent utility function with a nonexponential discount rate. Nonexponential discount rates are associated with time-inconsistent behavior,[10] implying that, when dealing with heterogeneity, the egalitarian social planner would choose time-inconsistent consumption plans.

6 Conclusions

It is shown in this paper that inequality measures obtained from information theory can be used to construct an exact and unrestricted aggregate representation of an economy with log-linear microfoundations. This is true even when the units are heterogeneous both on variable and parameter values.

---

[9] The distribution is obtained, for example, through an anonymous poll.
[10] See, for example, Loewenstein and Thaler (1989), and Laibson (1997).

The method can be used to verify previous propositions regarding the aggregation of this class of functions, and it seems to simplify proving. The cases of scale-invariant distributions and lognormal distributions, for example, were considered in this paper.

The choice of log-linear functions as the subject of analysis comes from the broad use of log-linear aggregate macroeconomic models in theoretical and empirical macroeconomics. Having a precise description of the relationship between those models and their microeconomic foundations, as the paper tried to accomplish, seemed therefore to be a relevant task.

The results obtained here can be used both in empirical and theoretical analysis. Aggregate log-linear econometric models can be used to estimate micro parameters when specific dispersion measures are included in the equation. The method can also be used to generate analytical solutions to theoretical aggregation problems.

Two applications were presented. The Lucas supply function was aggregated not in its logarithmic form, as usual, but rather in its level form, leading to a different solution for the aggregate function. It was also shown that, given agents with heterogeneous exponential discount rates, an egalitarian social planner faces an aggregate utility function that has a nonexponential discount rate, which generates time-inconsistent plans.

The examples tried to demonstrate that the method could be helpful not only to econometricians but also to macroeconomists. Maybe, given its relative simplicity, it could also be useful to instructors willing to discuss aggregation issues at introductory levels.

Appendix 1

Starting with

$$Y_{nt} = X_{1nt}^{a_{1n}} X_{2nt}^{a_{2n}} \cdots X_{Int}^{a_{In}}\,,\; Y_{nt} > 0,\;\; X_{int} > 0,\;\; i = 1,\ldots, I,\;\; n = 1,\ldots, N, \forall t\,,$$

take the logarithm of each unit such that

$$y_{nt} = a_{1n} x_{1nt} + \cdots + a_{In} x_{Int}\,,$$

where lowercase represents the log of the variable.

Now, averaging this expression over units:

$$\overline{y}_t = \frac{1}{N}\sum_{n=1}^{N} y_{nt} = \frac{1}{N}\sum_{n=1}^{N}\left(a_{1n}x_{1nt} + \cdots + a_{In}x_{Int}\right) \Rightarrow$$

$$\overline{y}_t = \overline{a}_1 \overline{x}'_{1t} + \ldots + \overline{a}_I \overline{x}'_{It}\,,$$

where

$$\overline{a}_i = \frac{1}{N}\sum_{n=1}^{N} a_{in}\,,\;\; \overline{x}'_{it} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{a_{in}}{\overline{a}_i}\cdot x_{int}\right).$$

Taking the exponential of the equation above gives

$$\widetilde{Y}_t = \widetilde{X}'^{\,\overline{a}_1}_{1t} \ldots \widetilde{X}'^{\,\overline{a}_I}_{It}\,, \tag{1.i}$$

where

$$\widetilde{Y}_t = \left(\prod_{n=1}^{N} Y_{nt}\right)^{\frac{1}{N}},\;\; \widetilde{X}'_{it} = \left(\prod_{n=1}^{N} X_{int}^{a_{in}}\right)^{\frac{1}{N\overline{a}_i}}.$$

Using definition (1.2):

$$\exp\left[\overline{L}\left(\mathbf{Y}_t\right)\right] = \frac{\overline{Y}_t}{\widetilde{Y}_t},\;\; \exp\left[\overline{L}\left(\mathbf{X}_{it}\right)\right] = \frac{\overline{X}_{it}}{\widetilde{X}_{it}}\,, \tag{1.ii}$$

and noting that

$$\exp\left[\frac{\overline{\text{cov}}(\mathbf{x}_{it},\mathbf{a}_i)}{\overline{a}_i}\right]=\frac{\tilde{X}'_{it}}{\tilde{X}_{it}},\ \tilde{X}_{it}=\left(\prod_{n=1}^{N}X_{int}\right)^{\frac{1}{N}},\ \mathbf{x}_{it}=\begin{bmatrix}\ln(X_{i1t})\\ \vdots\\ \ln(X_{iNt})\end{bmatrix},$$

then

$$\frac{\exp\left[\overline{L}(\mathbf{X}_{it})\right]}{\exp\left[\overline{\text{cov}}(\mathbf{x}_{it},\mathbf{a}_i)/\overline{a}_i\right]}=\frac{\overline{X}_{it}}{\tilde{X}'_{it}}. \tag{1.iii}$$

Substituting (1.ii) and (1.iii) into (1.i), it follows that

$$\frac{\overline{Y}_t}{\exp\left[\overline{L}(\mathbf{Y}_t)\right]}=\left\{\frac{\overline{X}_{1t}\exp\left[\overline{\text{cov}}(\mathbf{x}_{1t},\mathbf{a}_1)/\overline{a}_1\right]}{\exp\left[\overline{L}(\mathbf{X}_{1t})\right]}\right\}^{\overline{a}_1}\cdots\left\{\frac{\overline{X}_{1t}\exp\left[\overline{\text{cov}}(\mathbf{x}_{It},\mathbf{a}_I)/\overline{a}_I\right]}{\exp\left[\overline{L}(\mathbf{X}_{It})\right]}\right\}^{\overline{a}_I}\Rightarrow$$

$$\overline{Y}_t=\overline{X}_{1t}^{\overline{a}_1}\overline{X}_{2t}^{\overline{a}_2}\cdots\overline{X}_{It}^{\overline{a}_I}D(\Phi_t),$$

where $\Phi_t=(\mathbf{X}_{1t},\ldots,\mathbf{X}_{It},\mathbf{a}_1,\ldots,\mathbf{a}_I)$, and

$$D(\Phi_t)=\exp\left\{\overline{L}(\mathbf{Y}_t)-\sum_{i=1}^{I}\left[\overline{a}_i\overline{L}(\mathbf{X}_{it})-\overline{\text{cov}}(\mathbf{x}_{it},\mathbf{a}_i)\right]\right\},\ \text{Q.E.D.}$$

Appendix 2

Applying Proposition 1 to equation (3.4), the aggregate equation in its logarithmic form turns out to be

$$y_t=\overline{a}_1x_{1t}+\overline{a}_2x_{2t}+\cdots+\overline{a}_Ix_{It}+d(\Phi_t), \tag{2.i}$$

where

$$d(\Phi_t)=\overline{L}(\mathbf{Y}_t)-\sum_{i=1}^{I}\left[\overline{a}_i\overline{L}(\mathbf{X}_{it})-\overline{\text{cov}}(\mathbf{x}_{it},\mathbf{a}_i)\right].$$

Taking the limit and using equations (3.3), (3.5), (3.6) and (3.7), this expression simplifies to

$$\text{plim}\, d(\Phi_t) = \text{plim}\, \overline{L}(\mathbf{Y}_t) - \frac{1}{2}\sum_{i=1}^{I} \overline{a}_i \sigma_i^2 \ . \tag{2.ii}$$

Note that generally Y_{nt} will not follow a lognormal distribution:

$$\text{plim}\, \overline{L}(\mathbf{Y}_t) = \sum_{j=1}^{\infty} \frac{\kappa_j}{j!} = \text{plim}\, \frac{1}{N}\sum_{n=1}^{N} \frac{(y_{nt} - \overline{y}_t)^2}{2} + r_t \Rightarrow$$

$$\text{plim}\, \overline{L}(\mathbf{Y}_t) = \text{plim}\, \frac{1}{N}\sum_{n=1}^{N} \frac{\left(a_{1n}x_{1nt} - \overline{a_1 x_{1t}} + \cdots + a_{In}x_{Int} - \overline{a_I x_{It}}\right)^2}{2} + r_t ,$$

$$r_t = \sum_{j=3}^{\infty} \frac{\kappa_j}{j!} ,\ \overline{a_i x_{it}} = \frac{1}{N}\sum_{n=1}^{N} (a_{in} x_{int}),$$

where r_t summarizes the effects of remaining cumulants, implying:

$$\text{plim}\, \overline{L}(\mathbf{Y}_t) = \text{plim}\left[\frac{1}{2}\sum_{i=1}^{I} \overline{\text{var}}(\mathbf{a}_i * \mathbf{x}_{it}) + \sum_{i=1}^{I-1}\sum_{j=i+1}^{I} \overline{\text{cov}}(\mathbf{a}_i * \mathbf{x}_{it}, \mathbf{a}_j * \mathbf{x}_{jt})\right] + r_t \Rightarrow$$

$$\text{plim}\, \overline{L}(\mathbf{Y}_t) = \frac{1}{2}\sum_{i=1}^{I}\left[\sigma_{ai}^2\sigma_i^2 + \overline{a}_i^2\sigma_i^2 + \sigma_{ai}^2\left(x_{it} - \frac{\sigma_i^2}{2}\right)^2\right]$$

$$+ \sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left[\sigma_{aij}\sigma_{ij} + \sigma_{ij}\overline{a}_i\overline{a}_j + \sigma_{aij}\left(x_{it} - \frac{\sigma_i^2}{2}\right)\left(\overline{x}_{jt} - \frac{\sigma_j^2}{2}\right)\right] + r_t \ .$$

From (2.ii)

$$\text{plim}\, d(\Phi_t) = \frac{1}{2}\sum_{i=1}^{I}\left[\sigma_{ai}^2\sigma_i^2 - \overline{a}_i(1 - \overline{a}_i)\sigma_i^2 + \sigma_{ai}^2\left(x_{it} - \frac{\sigma_i^2}{2}\right)^2\right]$$

$$+ \sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left[\sigma_{aij}\sigma_{ij} + \sigma_{ij}\overline{a}_i\overline{a}_j + \sigma_{aij}\left(x_{it} - \frac{\sigma_j^2}{2}\right)\left(x_{jt} - \frac{\sigma_j^2}{2}\right)\right] + r_t ,$$

and from (2.i), assuming that $N \to \infty$,

$$y_t = \sum_{i=1}^{I}\left\{\left[\sigma_{ai}^2 - \overline{a}_i\left(1-\overline{a}_i\right)\right]\frac{\sigma_i^2}{2} + \overline{a}_i x_{it} + \frac{\sigma_{ai}^2}{2}\left(x_{it} - \frac{\sigma_i^2}{2}\right)^2\right\}$$

$$+\sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left\{\sigma_{aij}\sigma_{ij} + \sigma_{ij}\overline{a}_i\overline{a}_j + \sigma_{aij}\left(x_{it} - \frac{\sigma_i^2}{2}\right)\left(x_{jt} - \frac{\sigma_j^2}{2}\right)\right\} + r_t \Rightarrow$$

$$y_t = c + \sum_{i=1}^{I}\left[\left(\overline{a}_i - \frac{\sigma_{ai}^2\sigma_i^2}{2}\right)x_{it} + \frac{\sigma_{ai}^2}{2}x_{it}^2\right]$$

$$+\sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left[\sigma_{aij}x_{it}x_{jt} - \frac{\sigma_{aij}\sigma_j^2}{2}x_{it} - \frac{\sigma_{aij}\sigma_i^2}{2}x_{jt}\right] + r_t, \text{ Q.E.D.}$$

References

Abramowitz, M. and I. A. Stegun (1972), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Washington: National Bureau of Standards.

Amemiya, T. (1985), *Advanced Econometrics*. Cambridge: Harvard.

Blanchard, O. J. and S. Fischer (1989), *Lectures on Macroeconomics*. Cambridge: The MIT Press.

Bourguignon, F. (1979), “Decomposable Income Inequality Measures,” *Econometrica*, 47, 901-920.

Deaton, A. S. and J. Muellbauer (1980), *Economics and the Consumer Behavior*. Cambridge: Cambridge Universtiy Press.

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

Kullback, S. (1968), *Information Theory and Statistics*. New York: Dover.

Laibson, D. (1997), “Golden Eggs and Hyperbolic Discounting,” *The Quarterly Journal of Economics*, 112, 443-477.

Lewbel, A. (1992), “Aggregation with Log-Linear Models,” *Review of Economic Studies*, 59, 635-642.

Loewenstein, G. and R. H. Thaler (1989), “Anomalies: Intertemporal Choice,” *Journal of Economic Perspectives*, 3-4, 181-193.

Lucas, R. E. Jr. (1973), “Some International Evidence on Output-Inflation Tradeoffs,” *American Economic Review*, 63, 326-334.

Maasoumi, E. (1986), “The Measurement and Decomposition of Multi-Dimensional Inequality,” *Econometrica*, 54, 991-997.

Nygård, F. and A. Sandström (1981), *Measuring Income Inequality*. Stockholm: Almqvist & Wiksell.

Shorrocks, A. F. (1980), "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48, 613-625.

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

__________ (1986), "Simple Tests of Distributional Effects on Macroeconomic Equations," *Journal of Political Economy*, 94, 763-795.

__________ (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*. Amsterdam: North-Holland.

__________ (1967), *Economics and Information Theory*. Amsterdam: North-Holland.

__________ (1971), *Principles of Econometrics*. New York : John Wiley & Sons.