Quaderni di Statistica
Vol. 11, 2009

Perfect aggregation in linear models: a geometrical insight

Silvia Terzi
Dipartimento di Economia, Università RomaTre
E-mail: terzi@uniroma3.it

Summary: Econometric literature has devoted a good deal of attention to the aggregation problem, especially within the framework of parameter estimation (Theil, 1954). In this context the question is: what is the relationship between the parameters of the aggregate relation and the parameters of the micro-relations? A second issue is concerned with prediction. In this context starting from Grunfeld and Griliches's pioneer work (1960) the focus is on whether to predict the aggregate dependent variable by means of macro or micro equations. If we define *perfect aggregation* as non contradiction between the two models, we would expect perfect aggregation conditions within the two contexts to be the same. However this is not the case. In fact perfect aggregation within prediction is implicitly defined as equivalence between aggregate and disaggregate models with respect to some goodness of fit criterion; thus giving rise to a less restrictive definition, inconsistent with Theil's *rule of perfection*. The aim of the present paper is to unify the estimation and the prediction approaches by defining a goodness of fit criterion that does not contradict Theil's findings.

Keywords: Aggregation, Linear prediction models, Perfect Aggregation.

1. Introduction

The aggregation problem has been thoroughly studied in econometrics mainly within the framework of parameter estimation. In this context the question is: what is the relationship between the parameters (**b**) of the aggregate regression and the parameters ($\boldsymbol{\beta}_i$, i= 1,…, m) of the micro-relations? It is well known that least squares (LS) estimation of the aggregate model leads, in general, to biased estimators. As Theil (1954) shows, in order for the LS estimator to be free of aggregation bias it has to be either:

$$\boldsymbol{\beta}_i = \mathbf{b} \qquad \forall i = 1,2,\ldots,m$$

or:

exact linear relations between independent variables of different microrelations[1].

A second issue is concerned with prediction. In this context starting from pioneer work of Grunfeld and Griliches (1960), herein GG, the focus is on whether to predict the aggregate dependent variable by means of macro or micro equations. These authors – followed successively by others (see for example Sasaki, 1978) – introduce a within sample goodness of fit criterion based on the sum of the squared residuals (R^2) and argue that whenever the goodness of fit of the aggregate model is greater than the goodness of fit of the model derived from the micro-equations, there is an aggregation gain.

More recently Pieraccini (2005) shows that direct estimation of the aggregate model leads in general to aggregation bias and that even in presence of perfect aggregation the disaggregate model is to be preferred. Thus, in reply to GG's question: "Is aggregation necessarily bad?" he asks: "Is aggregation ever necessary?"

In this paper – consistently with Pieraccini's findings - we argue that the selection criterion suggested by GG is biased: the expected goodness of fit of the aggregate model cannot be greater than that of the model derived from the correctly specified micro-relations.

In fact when predicting an aggregate variable ($\mathbf{y}_a = \sum_i \mathbf{y}_i$) by means of a macro relation we are projecting $\mathbf{y}_a$ on the subspace (S_a) spanned by the *k* aggregate independent variables (in other words, on a subspace whose dimension is at most *k*). Vice versa when we resort to micro-relations each micro dependent variable $\mathbf{y}_i$ is projected on a *k*-dimensional subspace (S_i); thus the sum of these projections can belong to a subspace (T) of greater dimension (Hefferson, 2008). Thus the goodness of fit of the model derived from well specified micro-relations will, in general, be greater than that of the aggregate model. Moreover, since $S_a \subseteq T$, prediction by means of the aggregate model will be as

[1] It i s also assumed – although not explicitly - that the aggregate independent variables are linearly independent.

good as prediction via the disaggregate model only if the two subspaces, S_a and T, have the same dimension. If we define *perfect aggregation* the equivalence between the two models, a necessary condition would thus be: $dim(T) = dim(S_a)$.

As we will see, this condition can be easily reconduced to Theil's *rule of perfection*, derived within the estimation context. Although it may seem an obvious requirement that conditions for perfect aggregation within estimation and prediction contexts be the same, this consistency requirement has rarely been pursued in literature. In fact perfect aggregation within prediction is often implicitly defined as non-contradiction between the two models with respect to some goodness of fit criterion. This gives rise to a less restrictive definition of perfection, but also to inconsistent consequences. For example Pesaran *et al.* (1989) while seeking a test for perfect aggregation within a prediction approach, explicitly leave aside the case of exact linear relations among variables.

The aim of the present paper is to define a goodness of fit criterion that does not contradict Theil's findings. First of all we will shed a light on the definition of perfect aggregation in order to unify the estimation and prediction approaches. Then we will define an appropriate goodness of fit criterion which – in contrast with GG's findings – leads to an unbiased selection criterion, and prove the non implementability of a test for perfect aggregation suggested by Pesaran *et al.* (1989).

2. Perfect aggregation

Let us consider the following micro-behavioural equations referring to n observations $(h = 1,2,...,n)$ of m micro-units $(i = 1, 2,\ldots,m)$ in which a dependent variable Y is expressed as a linear combination of k explanatory variables X_j $(j = 1,\ldots,k)$:

$$Y_{ih} = \sum_j \beta_{ij} X_{ijh} + u_{ih}$$

In matrix notation the same model can be written as:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{u}_i \qquad i = 1,2,\ldots,m \tag{1}$$

We assume model (1) to be correctly specified, so that:

$$E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}_i\,;\;\; E(\mathbf{u}_i\,\mathbf{u}_i') = \sigma_i^2\,\mathbf{I}\,;\; E(\mathbf{u}_i\,\mathbf{u}_j') = \sigma_{ij}^2\,\mathbf{I}$$

and $rank(\mathbf{X}_i) = k \;\; \forall$ i = 1,…,m.
From (1), if we define $\mathbf{Y}_a = \sum_i \mathbf{Y}_i$ we can write the *derived aggregate* model H_d (sometimes improperly referred to as the "disaggregate model") :

$$H_d : \mathbf{Y}_a = \sum_i \mathbf{X}_i\,\boldsymbol{\beta}_i + \sum_i \mathbf{u}_i \tag{2}$$

Alternatively, defining $\mathbf{X}_a = \sum_i \mathbf{X}_i$, we can write the *aggregate* model H_a as:

$$H_a : \mathbf{Y}_a = \mathbf{X}_a\mathbf{b} + \mathbf{v}_a \tag{3}$$

where $\mathbf{v}_a \equiv \sum_i \mathbf{V}_i\,\boldsymbol{\beta}_i + \sum_i \mathbf{u}_i$ and $\sum_i V_i\,\boldsymbol{\beta}_i \equiv \sum_i \mathbf{X}_i\,\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b}$ [2].
We denote by k_a the rank of $\mathbf{X}_a$.

Following Theil, we define perfect aggregation as non-contradiction between the derived aggregate model H_d and the aggregate model H_a.

Model H_d states that $\mathbf{Y}_a$ belongs to the sum of the subspaces spanned by the columns of the $\mathbf{X}_i$ matrices (plus a random disturbance of null expected value). Vice versa, the aggregate model states that $\mathbf{Y}_a$ belongs to the subspace spanned by the columns of the $\mathbf{X}_a$ matrix (plus a random disturbance).

Let us call S_i the subspace spanned by $\mathbf{X}_i$, and define $T = S_1 + \ldots + S_m$. Moreover let us call S_a the subspace spanned by the columns of $\mathbf{X}_a$. Of course S_a is also – by definition – a subspace of T.

Non-contradiction between the two models requires:

$$dim(T) = dim(S_a)$$

in other words it requires the subspaces T and S_a to be isomorphic (See for example Hefferson, 2008, for a definition).

---

[2] The definition of $\mathbf{v}_a$ stems from Theil's' auxiliary equation: : $\mathbf{X}_i = \mathbf{X}_a\,\Gamma_i + \mathbf{V}_i$, with: $\sum_i \mathbf{V}_i = \mathbf{0}$ and $\sum_i \boldsymbol{\Gamma}_i = I$.

First of all note that $dim(\mathrm{T}) = rank\,(\mathbf{X}_1 \vdots \ldots\ldots \vdots \mathbf{X}_m)$. Thus a necessary condition for perfect aggregation to hold is:

$$rank\,(\mathbf{X}_1 \vdots \ldots\ldots \vdots \mathbf{X}_m) = k \qquad \text{(c.1)}$$

Obviously, given that $rank(\mathbf{X}_i) = k$, this condition is met if and only if any one of the $\mathbf{X}_i$ matrices spans all the subspaces $\mathrm{S}_{i'}$, $\forall\ i' = 1,\ldots,m$. In order to derive an equivalent condition, we can resort to a decomposition of the $\mathbf{X}_i$ matrices (the so-called auxiliary equations). In fact, since the $\mathbf{X}_i$ matrices all belong to $\mathbb{R}^n$, and since they all span k-dimensional subspaces of $\mathbb{R}^n$, for a given i and for $\forall\ i'=1,\ldots,m$, $\mathbf{X}_{i'}$ can be decomposed in the sum of two matrices, one belonging to S_i, the other belonging to the orthogonal subspace $\mathrm{S}_i^{\,c}$. Thus we can write:

$$\mathbf{X}_{i'} = \mathbf{X}_i\,\mathbf{C}_{i'/i} + \mathbf{F}_{i'/i} \qquad (4)$$

where $\mathbf{X}_i'\,\mathbf{F}_{i'/i} = \mathbf{0}$.

Since $rank\,(\mathbf{X}_{i'}) = rank\,(\mathbf{C}_{i'/\mathrm{i}}) + rank(\mathbf{F}_{i'/i})$, in order for $(\mathbf{X}_i \vdots \mathbf{X}_{i'})$ to have rank k a necessary condition is $rank(\mathbf{F}_{i'/i}) = 0$ (condition that implies $\mathbf{F}_{i'/i} = \mathbf{0}$ and $rank\,(\mathbf{C}_{i'/\mathrm{i}}) = k$) . Thus, a necessary condition for $dim(\mathrm{T}) = k$ is:

$$rank(\mathbf{F}_{i'/i}) = 0 \qquad \forall\ i'=1,\ldots,m \qquad \text{(c.2)}$$

When this last condition is met (together with the rank condition for the $\mathbf{X}_i$ matrices) $\mathbf{C}_{i'/i}$ is non singular $\forall\ i'$. Thus (c.1) and (c.2) are equivalent.

From auxiliary equation (4) we can also derive:

$$\mathbf{X}_a = \mathbf{X}_i \sum_{i'} \mathbf{C}_{i'/i} + \sum_{i'} \mathbf{F}_{i'/i} = \mathbf{X}_i\,\mathbf{C}_i + \mathbf{F}_i \qquad (5)$$

thus implying that, whenever conditions (c.1) or (c.2) hold:

$$\mathbf{X}_a = \mathbf{X}_i\,\mathbf{C}_i \qquad \text{(c.3)}$$

Condition (c.3) can also be stated as exact linear relations among independent variables. It is easy to see that (c.3) ↔(c.1) or (c.2). Thus conditions (c.1), (c.2) and (c.3) are all equivalent.

However, since $rank(\mathbf{X}_a) = rank(\mathbf{C}_i) + rank(\mathbf{F}_i)$, condition (c.3) does not itself guarantee that $rank(\mathbf{X}_a) = k$; in fact $\mathbf{C}_i$ is the sum of square full-rank matrices $\mathbf{C}_{i'/i}$, but this does not guarantee its non-singularity; moreover there are no "obvious" conditions to be posed in order for this requirement to be fulfilled. Thus in the context of linear prediction, perfect aggregation requires:

$$rank\,(\mathbf{X}_1 \vdots \ldots\ldots \vdots \mathbf{X}_m) = rank(\mathbf{X}_a) = k \qquad \text{(c.4)}$$

or, in other words: $dim(\mathrm{T}) = dim(\mathrm{S}_a)$.
This condition is both necessary and sufficient for perfect aggregation.

Of course, whenever the (correctly specified) micro-relations state:

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{u}_i \qquad i = 1,2,\ldots,m$$

the derived aggregate model is equivalent to the aggregate model without further assumptions. Note that in this case condition (c.4) holds without further assumptions. In fact the derived aggregate model is given by:

$$\mathrm{H}_d : \mathbf{Y}_a = \textstyle\sum_i \mathbf{X}_i\,\mathbf{b} + \sum_i \mathbf{u}_i = \mathbf{X}_a\mathbf{b} + \sum_i \mathbf{u}_i$$

thus stating that that $\mathrm{E}(\mathbf{Y}_a)$ belongs to the S_a subspace, spanned by the columns of the $\mathbf{X}_a$ matrix .

Let us now turn our attention to the estimation context. The conditions stated by Theil are: $\mathbf{X}_i = \mathbf{X}_a\,\mathbf{\Gamma}_i,\ \forall i = 1,2,\ldots,m$ or $\boldsymbol{\beta}_i = \mathbf{b}\ \forall i = 1,2,\ldots,m$ (equality of the micro coefficients). It is easy to see that the first of these conditions is equivalent to condition (c.4), where as the assumption of equality of the micro coefficients although sufficient, seems unduly restrictive.

Obviously, in order for perfect aggregation to hold condition (c.4) must be satisfied in and out of sample. This extended condition (also known as *compositional stability*) is usually stated as:

$$\mathbf{X}_a = \mathbf{X}_i\, \mathbf{C}_i \qquad \forall i, \quad \forall h=1,\dots,n,\dots,$$

3. Prediction and within sample goodness of fit

The question is whether to predict the aggregate variable $\mathbf{Y}_a$ using model H_d or model H_a. In the first case we predict the aggregate variable $\mathbf{Y}_a$ aggregating the predicted values of the micro-dependent variables $\mathbf{Y}_i$, $i = 1,...,m$; in the second case we predict $\mathbf{Y}_a$ by means of the aggregate independent variables.

Assume we use LS. We can thus define the two predictors:

$$\hat{\mathbf{y}}_d = \sum_i \mathbf{X}_i \hat{\boldsymbol{\beta}}_i = \sum_i \mathbf{X}_i \boldsymbol{\beta}_i + \sum_i \mathbf{A}_i \mathbf{u}_i \text{ and } \hat{\mathbf{y}}_a = \mathbf{X}_a \hat{\mathbf{b}} = \mathbf{A}_a \sum_i \mathbf{y}_i \ ;$$

and the two residuals

$$\mathbf{e}_d = \mathbf{y}_a - \hat{\mathbf{y}}_d = \sum_i \mathbf{M}_i \mathbf{u}_i,\ \mathbf{e}_a = \mathbf{y}_a - \hat{\mathbf{y}}_a = \sum_i V_i \boldsymbol{\beta}_i + \mathbf{M}_a \sum_i \mathbf{u}_i$$

where

$$\mathbf{A}_i = \mathbf{X}_i \left(\mathbf{X}_i' \mathbf{X}_i\right)^{-1} \mathbf{X}_i',\ \mathbf{A}_a = \mathbf{X}_\mathbf{a} \left(\mathbf{X}_\mathbf{a}' \mathbf{X}_\mathbf{a}\right)^{-1} \mathbf{X}_a',\ \mathbf{M}_i = \mathbf{I} - \mathbf{A}_i,\ \mathbf{M}_a = \mathbf{I} - \mathbf{A}_a$$ [3]

It is well known that the predictor $\hat{\mathbf{y}}_a$ is the orthogonal projection of the aggregate dependent variable on the subspace spanned by the columns of the $\mathbf{X}_a$ matrix. Thus it belongs to the k_a dimensional subspace S_a.

Vice versa, $\hat{\mathbf{y}}_d$ is the sum of the orthogonal projections of the micro dependent variables on the S_i subspaces, and it belongs to a subspace T. It should be noted that, in general, $\hat{\mathbf{y}}_d$ is not an orthogonal projection of $\mathbf{y}_a$ on the subspace T.
In order to give a graphical representation of the different subspaces, assume we have observed two variables $\boldsymbol{x}, \boldsymbol{y}$, on two micro-units (i=1,2)

[3] In order for X_a to be invertible we will assume either: *rank* $X_a = k$, or that X_a is an nxk_a full rank matrix.

in 3 different occasions (h=1,2,3). Let us consider the three-dimensional space $\mathbb{R}^3$ where the axes are associated with occasions and each point-vector represents the time series of observations on a single variable for one single unit i. The disaggregate model projects $\boldsymbol{y}_i$ (separately for each micro-relation), on the subspace S_i spanned by the vectors $\boldsymbol{x}_i$ and the unit vector $\mathbf{1}$=(1, 1, 1). We can represent the two subspaces S_1 and S_2 in figure 1:

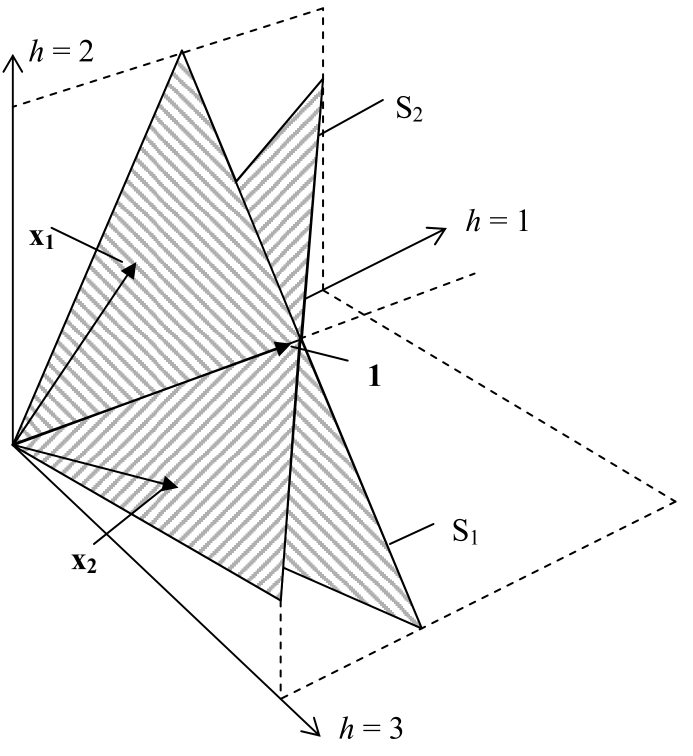

Figure 1

Now let us give a geometrical representation of the two predictors $\hat{\boldsymbol{y}}_d$, $\hat{\mathbf{y}}_a$. The derived aggregate predictor $\hat{\boldsymbol{y}}_d$ is the sum of the disaggregate predictors $\hat{\boldsymbol{y}}_1$ $\hat{\boldsymbol{y}}_2$, that lie on different planes, thus $\hat{\boldsymbol{y}}_d$ lies on a subspace T which is the sum of the two subspaces S_1 and S_2, and – in principle – could be a three-dimensional subspace. The aggregate predictor $\hat{\boldsymbol{y}}_a$ lies on the subspace S_a spanned by the vector $\boldsymbol{x}_a = \boldsymbol{x}_1 + \boldsymbol{x}_2$ and the unit vector **1**, so it lies on a two-dimensional subspace of $\mathbb{R}^3$.

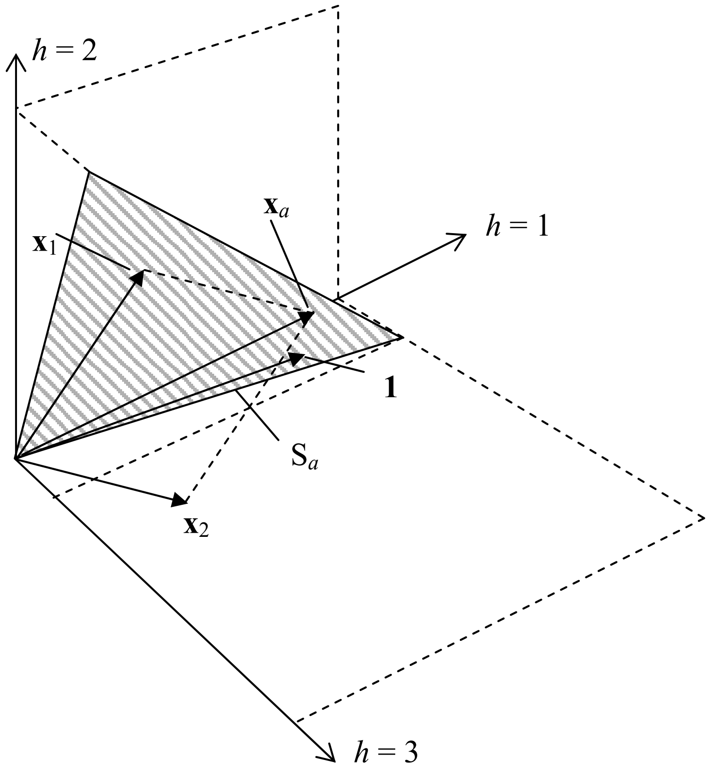

Figure 2

Since $dim(T) \geq dim(S_a)$ we should expect $\hat{\mathbf{y}}_d$ to have a better fit than $\hat{\mathbf{y}}_a$; however a goodness of fit measure is needed. Since $\hat{\mathbf{y}}_d$ is not an orthogonal projection of $\mathbf{y}_a$, for the disaggregate model we cannot resort to the usual definition of R^2 as:

$$R^2 = \frac{dev(\hat{\mathbf{y}}_d)}{dev(\mathbf{y}_a)} = 1 - \frac{\mathbf{e}_d{}'\mathbf{e}_d}{\mathbf{y}_a{}'\mathbf{y}_a}$$

since the equality does no longer hold. But, instead, we have to choose an appropriate definition among the most frequently used in literature. We could define, as GG do, $\mathrm{R}^2 = 1 - \frac{e'e}{y'y}$. However, since we are interested in a projection problem, the most appropriate goodness of fit criterion seems to be the closeness between predicted and observed values, as measured by the square of the cosine of their angle. We thus define:

$$\tilde{R}^2 = \cos^2(\hat{\mathbf{y}}, \mathbf{y}) = \frac{(\mathbf{y}'\hat{\mathbf{y}})^2}{(\hat{\mathbf{y}}'\hat{\mathbf{y}})(\mathbf{y}'\mathbf{y})}$$

It can be easily seen that for the aggregate model $\mathbf{y}_a'\hat{\mathbf{y}}_a = \hat{\mathbf{y}}_a'\hat{\mathbf{y}}_a$, thus:

$$\tilde{R}_a^2 = \cos^2(\hat{\mathbf{y}}_a, \mathbf{y}_a) = \frac{(\hat{\mathbf{y}}_a'\hat{\mathbf{y}}_a)}{(\mathbf{y}_a'\mathbf{y}_a)} = R_a^2$$

where as for the derived model H$_d$:

$$\tilde{R}_d^2 = \cos^2(\hat{\mathbf{y}}_d, \mathbf{y}_a) = \frac{(\mathbf{y}_a'\hat{\mathbf{y}}_d)^2}{(\hat{\mathbf{y}}_d'\hat{\mathbf{y}}_d)(\mathbf{y}_a'\mathbf{y}_a)} = \frac{(\mathbf{e}_d'\hat{\mathbf{y}}_d)^2}{(\mathbf{y}_a'\mathbf{y}_a)(\hat{\mathbf{y}}_d'\hat{\mathbf{y}}_d)} + R_d^2$$

Thus, in general $\tilde{R}_d^2 \geq R_d^2$. Moreover $\tilde{R}_d^2 = R_d^2$ if and only if $\mathbf{e}_d$ and $\hat{\mathbf{y}}_d$ are orthogonal.

We now want to show that defining goodness of fit as $\tilde{R}^2$ leads, on average, to select the model H$_d$ unless perfect aggregation holds (in which case the two models are equivalent).

In other words the selection criterion we introduce – unlike the criterion based on $\left(R_a^2 - R_d^2\right)$ – is unbiased.

It can be easily seen that:

$$\left(\widetilde{R}_d^2 - \widetilde{R}_a^2\right)\mathbf{y}_a'\mathbf{y}_a = \mathbf{e}_a'\mathbf{e}_a + \left(\frac{\left(\mathbf{e}_d'\hat{\mathbf{y}}_d\right)^2}{\hat{\mathbf{y}}_d'\hat{\mathbf{y}}_d} - \mathbf{e}_d'\mathbf{e}_d\right)$$

All terms that appear in this expression are ≥0; however $\frac{\left(\mathbf{e}_d'\hat{\mathbf{y}}_d\right)^2}{\hat{\mathbf{y}}_d'\hat{\mathbf{y}}_d} \leq \mathbf{e}_d'\mathbf{e}_d$. Thus $\left(\widetilde{R}_d^2 - \widetilde{R}_a^2\right)$ attains its minimum when $\frac{\left(\mathbf{e}_d'\hat{\mathbf{y}}_d\right)^2}{\hat{\mathbf{y}}_d'\hat{\mathbf{y}}_d} = 0$, in other words when $\hat{\mathbf{y}}_d$ is an orthogonal projection of $\mathbf{y}_a$.[4]

However, when $\mathbf{e}_d'\hat{\mathbf{y}}_d = 0$,

$$\mathbf{e}_d'\mathbf{e}_d = \mathbf{e}_d'\mathbf{y}_a = \sum_i\sum_j \mathbf{y}_i'\mathbf{M}_i\mathbf{y}_j = \sum_i\sum_j \boldsymbol{\beta}_i'\mathbf{X}_i'\mathbf{M}_j\mathbf{u}_j + \sum_i\sum_j \mathbf{u}_i'\mathbf{M}_j\mathbf{u}_j\,^{5}.$$

Moreover, since it is always $\mathbf{V}_i'\mathbf{X}_a = \mathbf{0}$[6]**:**

$$\mathbf{e}_a'\mathbf{e}_a = \sum_i\sum_j \mathbf{u}_i'\mathbf{M}_a\mathbf{u}_j + \sum_i\sum_j \boldsymbol{\beta}_i'\mathbf{V}_i'\mathbf{M}_a\mathbf{V}_j\boldsymbol{\beta}_j = \sum_i\sum_j \mathbf{u}_i'\mathbf{M}_a\mathbf{u}_j + \sum_i\sum_j \boldsymbol{\beta}_i'\mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta}_j$$

Thus:

$$\mathbf{e}_a'\mathbf{e}_a - \mathbf{e}_d'\mathbf{e}_d =$$
$$\sum_i\sum_j \mathbf{u}_i'\left(\mathbf{M}_a - \mathbf{M}_j\right)\mathbf{u}_j + \sum_i\sum_j \boldsymbol{\beta}_i'\mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta}_j - \sum_i\sum_j \boldsymbol{\beta}_i'\mathbf{X}_i'\mathbf{M}_j\mathbf{u}_j$$

and:

---

[4] Of course perfect aggregation is a sufficient condition for $\hat{\mathbf{y}}_d$ to be the orthogonal projection of y_a.

[5] Recall that $\mathbf{M}_j = \mathbf{I} - \mathbf{X}_j\left(\mathbf{X}_j'\mathbf{X}_j\right)^{-1}\mathbf{X}_j'$

[6] Like the other auxiliary equations we introduced, the auxiliary equation: $\mathbf{X}_i = \mathbf{X}_a\Gamma_i + \mathbf{V}_i$, assumes $\mathbf{V}_i \in S_a^c$ and thus: $\mathbf{V}_i'\mathbf{X}_a = \mathbf{0}$.

$$E\left(\mathbf{e}_a^{'}\mathbf{e}_a - \mathbf{e}_d^{'}\mathbf{e}_d\right) = \left(k - k_a\right)\sum_i \sum_j \sigma_{ij} + \sum_i \sum_j \boldsymbol{\beta}_i^{'}\mathbf{V}_i^{'}\mathbf{V}_j\boldsymbol{\beta}_j \geq 0$$

so that our selection criterion is unbiased.
It can easily be seen that, whenever condition (c.4) holds $\mathbf{e}_d^{'}\hat{\mathbf{y}}_d = 0$, k_a=k and $\mathbf{V}_i = \mathbf{0}\ \forall i$. Thus, in this case $\widetilde{R}_d^2 = R_a^2$.

4. Concluding remarks

Although Theil's findings clearly show that prediction via micro equations would yield more precise estimates of the aggregate dependent variable than the corresponding macro equation, some authors still claim the opposite could well be true (see Stoker, 1993 for a survey of the subsequent econometric literature, and Barreto and Howland, 1998, for a survey of the issues concerning the aggregation problem and how they have sometimes been disregarded in economic theory).

In their paper Grunfeld and Griliches (1960) wonder: "Is aggregation necessarily bad?" and conclude that under certain circumstances it can give rise to a better fit. They consequently suggest a selection criterion based on $E\left(\mathbf{e}_a^{'}\mathbf{e}_a - \mathbf{e}_d^{'}\mathbf{e}_d\right)$ and define perfect aggregation as $\sum_i \sum_j \boldsymbol{\beta}_i^{'}\mathbf{V}_i^{'}\mathbf{V}_j\boldsymbol{\beta}_j = 0$. It is easy to see that $\sum_i \sum_j \boldsymbol{\beta}_i^{'}\mathbf{V}_i^{'}\mathbf{V}_j\boldsymbol{\beta}_j = 0$ if

a) $\boldsymbol{\beta}_i = \mathbf{b} \quad \forall\ i = 1,\ldots,m$

or:

b) $\mathbf{V}_i = \mathbf{0} \quad \forall\ i = 1,\ldots,m \Leftrightarrow \begin{cases} \mathbf{F}_{i'7i} = \mathbf{0} \forall i' = 1,\ldots,m \\ rank(\mathbf{C}_i) = k, \forall i = 1,\ldots,m \end{cases}$

These properties suggest that a perfect aggregation test could be based on the statistic $\hat{\xi} = \sum_i \mathbf{X}_i\hat{\boldsymbol{\beta}}_i - \mathbf{X}_a\hat{\mathbf{b}}$. In fact this is the suggestion from Pesaran *et al.* (1989).

On the contrary we have argued, in line with Pieraccini (2005), that there has been much misunderstanding on the performances of aggregate models, misunderstanding that can sometimes be reconduced to a biased selection criterion, and that has led to a less restrictive definition of perfect aggregation, in contrast to Theil's findings. If perfect aggregation within the prediction framework is to be defined as non contradiction between the two models with respect to some goodness of fit criterion, it is necessary to resort to an appropriate, unbiased goodness of fit measure; for example the $\widetilde{R}^2$ that we have suggested. However it would also be of great use to have a test for perfect aggregation. Unfortunately the test introduced by Pesaran *et al.* is not implementable. In fact they show that, under the assumption that $\boldsymbol{u}$ is normally distributed with mean zero and known covariance matrix, when $\sum_i \sum_j \boldsymbol{\beta}_i' \mathbf{V}_i' \mathbf{V}_j \boldsymbol{\beta}_j = 0$:

$$m^{-1}\left(\hat{\xi}' \boldsymbol{\psi}_m^{-1} \hat{\xi} \right) \sim \chi^2_{\nu}$$

where:

$$\boldsymbol{\psi}_m = m^{-1} \sum_{i,j=1}^{m} \sigma_{ij} \mathbf{H}_i \mathbf{H}_j;$$

$$\mathbf{H}_i = (\mathbf{A}_i - \mathbf{A}_a)$$

$$\nu = rank(\boldsymbol{\psi}_m)$$

They also derive a sufficient (but not necessary) condition for $\boldsymbol{\psi}_m$ to have full rank.

It is easy to see that when condition (c.2) holds:

$$\mathbf{M}_a - \mathbf{M}_j = \mathbf{A}_j - \mathbf{A}_a = \mathbf{X}_j \left(\mathbf{X}_j' \mathbf{X}_j\right)^{-1} \mathbf{X}_j' - \mathbf{X}_j \mathbf{C}_j \left(\mathbf{X}_a' \mathbf{X}_a\right)^{-1} \mathbf{C}_j' \mathbf{X}_j'$$

is a symmetric and idempotent matrix; thus it is positive semi definite with rank k-$rank(\mathbf{C}_j)$. However when $\mathbf{V}_i = \mathbf{0}$, $rank(\mathbf{C}_i) = k \; \forall i$, so that, in fact, under perfect aggregation the test for perfect aggregation is not implementable.

Acknowledgments: The author wishes to thank Professor Luciano Pieraccini for helpful, stimulating discussions and for his encouragement in writing this article.

References

Barreto H., Howland F.M. (1998), The treatment of aggregation in modern economic analysis. *HES Conference Proceedings*, Montreal.

Grunfeld Y., Griliches Z. (1960), Is aggregation necessarily bad? *Review of Economics and Statistics* , 42, 1-13.

Hefferson J. (2008) *Linear Algebra*, GNU Free Documentation.

Pesaran M.H, Pierse R.G., Kumar M.S. (1989), Econometric analysis of aggregation in the context of linear prediction models. *Econometrica*, 57, 861-888.

Pieraccini L. (2005), Is aggregation ever necessary? *Quaderni di Statistica*, 7, 1-15.

Sasaki K. (1978), An empirical analysis of linear aggregation problems. The case of investment behavior in Japanese firms, *Journal of Econometrics*, 7, 313-331.

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.