BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

The use of survey weights in regression analysis

by Ivan Faiella

January 2010

number 739

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

The use of survey weights in regression analysis

by Ivan Faiella

Number 739 - January 2010

THE USE OF SURVEY WEIGHTS IN REGRESSION ANALYSIS

by Ivan Faiella*

Abstract

While there is a wide consensus in using survey weights when estimating population parameters, it is not clear what to do when using survey data for analytic purposes (i.e. with the objective of making inference about model parameters). In the model-based framework (MB), under the hypothesis that the underlying model is correctly specified, using survey weights in regression analysis potentially involves a loss of efficiency. In a design-based perspective (DB), weighted estimates are both design consistent and can provide robustness to model mis-specification. In this paper, I suggest that the choice of using survey weights can be seen in a regression diagnostic set. The survey data analyst should check if the design information included in survey weights has some explanatory power in describing the model outcome. To accomplish this task a set of econometric tests is suggested, that could be supplemented by the analysis of model features under the two strategies.

JEL Classification: C42, C52.
Keywords: survey methods, model evaluation and testing.

Contents

---

* Bank of Italy, Structural Economic Analysis Department. I wish to thank Gianni Betti, Laura Neri and Vijay Verma for the stimulating discussions we had during my visiting period at the University of Siena. I am also grateful to Stefano Iezzi and Giuseppe Ilardi for their useful insights. The work also benefited from the comments of Leandro D'Aurizio, Romina Gambacorta and Andrea Venturini and of two anonymous referees. The views expressed are those of the author and do not necessarily reflect those of the Bank of Italy. E-mail: ivan.faiella@bancaditalia.it.

All models are wrong; some models are useful.
George Box, 1979.

1 Introduction

Microdata are often collected using sample surveys and their design typically involves specific techniques such as clustering and stratification which, if ignored, generally lead to an inaccurate estimation of the variance. Furthermore, when the process of sample selection and the response mechanism is *non-ignorable*,[1] disregarding survey weights can result in biased estimators. Incorporating all the design features requires the use of survey weights and a strategy to estimate the sampling variance that includes information about the sampling process.

While there is a far-reaching agreement about using survey weights in descriptive inference, it is less clear cut if their use must be automatically extended when studying relationships among survey variables. This case is referred to in the literature as the "analytic use of sample surveys" (Skinner et al., 1989). The objective of analytic inference is to draw conclusions about a super-population assumed to have generated the actual population (Särndal, Swensson, and Wretman 1992). Generally, analytic inference relies on a regression model linking the study variables with a set of explanatory variables (covariates). The optimal estimators (e.g. OLS) for this class of models relies on assumptions usually not met by complex survey data.

First, sample observations typically have different selection probabilities. Nathan and Smith (1989) show that, unless the selection of the sampling units is ignorable subject to the covariates of the model, OLS estimates are biased and inconsistent. Note that this selection pattern depends both on the actual sampling scheme (i.e. how the population elements are included in the sample) and on the response process.[2] When this information is not relevant for the model a condition of design ignorability is met and it is possible to make recourse to more efficient estimators.

Secondly, the usual standard errors formula for LS and ML estimates are not appropriate because the sampling units are not identically and independently distributed across all possible samples. The observed sample is in fact

[1]The design is *ignorable* when the selection of the theoretical sample and the response mechanism that leads to the actual sample depend only on the observed data. More on this in the next sections.

[2]It is common practice to incorporate non-response adjustments in survey weights.

the output of a selection on a stratified population and often the sampling elements are "clusters" of the statistical units, suggesting that variance estimators could be negatively biased unless they account for the similarity of the units pertaining to the same cluster (*intra-cluster correlation*).

The present study is focused on the first point only. All the variance estimates subsequently used will adopt the randomization level determined by sampling design. It will later be shown that this stance can be seen as a natural extension of the "sandwich" estimator routinely used by practitioners to derive standard errors robust to model variance mis-specification; it is also the same as the procedures adopted in the econometrics of cluster samples (see for example Wooldridge, 2006).

Fundamentally, the questions we want to answer in this study are three:

1. What are the pros and cons of using survey weights when modelling sample survey data?

2. Is there a simple way to test if survey weights provide the model with additional information?

3. Can the choice of using survey weights be incorporated among the model diagnostic tools?

The paper is structured as follows: in the next section I briefly recall the differences between the design and the model-based approach to inference with particular reference to the use of survey weights in regression analysis (an excellent reference on the topic is Binder and Roberts, 2003); I will then briefly touch on the fact that model builders often implement randomization based estimators to correct variance-covariance model matrices for heteroskedasticity and cluster samples; in section 3, I briefly evaluate if there is an alternative to using survey weights augmenting a model with survey information; in section 4 I present a set of procedures that can help the data analyst decide whether to use survey weights; in section 5 all the previous findings are put into action using microdata from the two surveys conducted by the Bank of Italy; finally the main conclusions are formulated.

2 The theoretical debate

2.1 The design-based approach

The foundation of design-based (henceforth DB) inference lies on the concept of *randomization.* Consider a finite survey population U as a collection of N

elements: $U = \{1, 2, 3, ..., i, ..., N\}$. To select a sample we need to define a *sampling design* that establishes all the samples - the set S - that it is possible to draw from this population: $S = \{s_1, s_2, .., s_r\}$. Given a sampling design *p(s)*, it is possible to associate all the population elements with an indicator variable I_i equal to 1 if the *i-th* element of the population is included in the sample and zero otherwise: $I = \{I_1, I_2, .., I_i, ..., I_N\}$. When an actual sample s is drawn from the set S of the possible samples, the indicator variable is conditioned to this sample: $i_s = \{i_{1s}, i_{2s}, .., i_{is}, ..., i_{Ns}\}$. Then a formal definition of sampling design will be $p(s) = P(S = s) = P(I = i_s)$ and each $i - th$ unit will be included in a sample s with probability $\pi_i = \sum_{s \ni i} p(s)$ (the *inclusion probability* of element i).

For each sample element it is possible to measure (for hypothesis without error) one or more characteristics (for example a study variable y_i and a vector of auxiliary variables x_i^T). **These elements are fixed both in the sample and in the population**. What governs the randomness of the process is the sampling distribution; this in turn depends on the inclusion probabilities of the sample elements. The inference process is founded on the variability across all the possible samples (determined, as seen, by $p(s)$).

In the DB context, the analyst uses models as a statistical tool to study the correlation structure between the dependent variable and a set of predictors.[3] Model estimators are usually seen as the combination of a class of design-unbiased estimators known in the survey literature as the *Horvitz-Thompson*-like (HT) estimator of the form $\hat{Y}_{HT} = f(y, x^T, \pi)$, also known as the π-estimator (Särndal, Swensson, and Wretman 1992).[4] The rationale of this approach is to inflate each sample observation y_i, x_i^T dividing it by its inclusion probability π_i. In the literature, the resulting estimators are termed as *Census parameters* (Chambers and Skinner 2003) or *Descriptive population quantities* (Pfeffermann 1993).

Consider for example the linear regression model. We have a sample with n observations. For each *i-th* observation we can observe the study variable y_i, a vector of k predictors x_i^T and the survey weight w_i computed by the survey organization. The well known formula for the least square solution if we plug-in survey weight is

$$\hat{\beta}_w = \left(X^T W X\right)^{-1} X^T W Y \qquad (1)$$

where W is a $n \times n$ diagonal matrix with the survey weights in the diagonal, Y is a $n \times 1$ vector and X is a $n \times k$ matrix. Note that the vector

[3] He focuses more on "estimable" than on "structural" models (Wooldridge 2002)

[4] The HT estimator is also referred to as the Horvitz-Thompson-Narain estimator, as Narain independently presented a similar general theory in 1951.

$\hat{\beta}_w$ is the ratio of two HT estimators. The resulting estimator is biased but its bias is of the order of $n-1$, negligible in large samples; furthermore its relative bias is bounded by the coefficient of variation of the denominator $(X'WX)$, usually very small for large samples (Kish,1965). This approach can be extended to Generalized Linear Models. In this case the score function is rewritten to resemble an HT estimator (see for example Binder, 1983 and Nordberg, 1989).[5]

A potential shortcoming of the DB estimator is related to its potential inefficiency. If the sampling design conveys no additional information into the model (the design is *ignorable*) survey weights pointlessly risk inflating the variance of the estimators. In presence of large sample sizes this problem can be overstated. Moreover, in univariate context, Little and Vartivarian (2005) show that calibration, in the presence of the right choice of post-strata, i.e. strata formed at the estimation stage, can actually decrease the variance of the weighted estimator. Other scholars show that the efficiency of the DB estimator can be improved by smoothing survey weights with an appropriate model (Beaumont 2008).

Briefly, what are the benefits and the drawbacks of the DB approach?

Advantages of the DB approach:

1. Using DB inference, no assumption is necessary regarding the distribution of the residuals.

2. In a DB perspective, π-weighted estimators are both design consistent and provide robustness to model mis-specification (e.g. they are robust to the problem of omitted variables). The advocates of this method underline that the parameters estimated using survey weights are more robust because they are model unbiased if the model is true and design consistent if it is not (Kott 1991).

Disadvantages of the DB approach:

1. Within the DB framework, in building the model the analyst does not have a clear rationale on how to choose among competing estimators (Little 1981).

2. DB models are not always useful for prediction: in some cases, the reference population can be misleading in generalizing the results to other possible populations.[6]

[5]For GLM, a more sophisticated approach that also leads to a HT-like estimator is described in Pfeffermann and Sverchkov, 2003.

[6]Kalton (1989), presents an example regarding the extrapolation of weighted proportions to a general population in a simple Markov chain model.

3. The properties of the $\hat{\beta}_w$ in small samples are unknown (Pfeffermann and Sverchkov 1999).

2.2 The model-based approach

In a model-based framework (henceforth MB) the focus is on the data generating process. In a finite population context, the actual population can be seen as a realization of the infinite possible ones generated by a superpopulation mechanism: a population model is specified and a sample is drawn - using Simple Random Sampling (SRS) with replacement - from the population so that, in case of a linear specification,

$$y_i = \beta x_i^T + e_i \tag{2}$$

where y, x^T are observed on the i-th unit while e is an error term (unobservable) assumed orthogonal to the covariates. β is the parameter vector (constant in a frequentist framework) to be estimated. MB-inference studies the sampling distribution of the statistics over repeated realizations generated by the model: **the selected sample is held fixed**.

When using MB tools, the researcher is usually interested not in the particular population observed, but rather in the causal process linking the predictors and the response variable (econometricians call these "structural" models).

The benefits and drawbacks of the MB are the following:

Advantages of the MB approach:

1. If the model is correctly specified the unweighted estimator performs better than any competing estimator in terms of variance (it is BLUE).

2. Analysts can rely on a huge literature covering model building and diagnostics.

Disadvantages of the MB approach:

1. If the model is mis-specified and the predictors correlated to the response variable are omitted, MB estimates might be biased and inconsistent.

2. MB variance estimators usually rest on tight assumptions on the distribution of the unobserved errors, thus underestimating the actual variance.

2.3 A note on estimating model variance

While the use of survey weights is rarely dealt with in econometric textbooks (with the exception of Wooldridge, 2002 and Cameron and Trivedi, 2005), there is in general a wide acceptance that MB standard errors are downward biased thus impairing the validity of the computed confidence intervals.

In the MB framework the usual hypothesis about the distribution of errors is that they are independently and identically distributed (i.e. they follow an *iid* process). This means that data are sampled from the population using SRS with replacement and that the size of the errors is the same across different observations.

To deal with the problem of a non-identical distribution of the residuals across the sample, an asymptotic variance-covariance matrix robust to misspecification can be adopted (Greene 2002).

An equivalent procedures used by survey statisticians is to define a score such as $\hat{z}_i = x_i^T(y_i - x_i^T\hat{\beta})$ and compute the deviance as $\hat{z}^T\hat{z}$, where $\hat{z}$ is the score vector. [7][8]

Likewise, the assumption of error independence across the sample is usually not met when using sample survey data. The sampled populations are finite and sampling is without replacement.[9] Moreover sample design typically involves specific techniques such as clustering and stratification.

Because in stratified-clustered samples observations within a stratum are correlated, central limit theory does not hold (Wooldridge 2002). Ignoring these sampling features generally leads to an inaccurate estimation of the variance. Hence, in the presence of cluster samples, the assumption of independence must be relaxed at cluster level (i.e. model errors are independent between clusters).

[7] Binder (1983) devises a general estimation procedure to estimate the variance of the parameters of general linear models. This method requires that the variance of the score is "sandwiched" between the first derivatives of the score function.

[8] A formal argument to justify why a randomization-based estimator is consistent for the "correct" MB estimator relies on the concept of *anticipated variance*: this is defined as the variance of the estimator with respect to the sampling design and the superpopulation model (Isaki and Fuller 1982).

[9] This feature is not relevant if the sampling fraction (the share of the population sampled) is negligible (a common circumstance in household sample surveys).

3 Handling design information when modelling sample survey data

Some scholars tried to find a bridge between DB robustness and MB completeness. Pfeffermann (1993) and more recently other authors (e.g. Gelman 2007, Little 2004), propose a sort of "third way" to take the best from the DB and the MB approaches. This strategy relies on specifying a model using MB tools for inference, but focusing on estimators that are design consistent for a given census parameter (i.e. they consistently estimate a corresponding parameter in the population). This means relaxing some optimality rules in exchange for design consistency protection from model failure. A seminal work of Scott and Smith (1969), then developed by Pfeffermann and Lavange (1989), suggested setting up a multi-level model, exploiting the hierarchical structure of survey data. Strata are treated as fixed effects (population effect) and clusters as random effects (sample effect).[10]

Some studies advocate taking full advantage of the hierarchical modelling: if sample design and population information is available it is possible to build models that account at the same time for the factors underlying the analysed phenomenon and for the sample selection process (clusters and strata information and associated covariates), survey units participation (non response rates, as in Yuan and Little, 2007a, 2007b), population information in some relevant dimensions (thus including post-stratification as in Little, 2004 and Gelman, 2007).[11]

What all these approaches have in common is that they structure the model in order to control for *design ignorability* (henceforth DI). We have DI whenever the information on how the population elements are included in the sample is not relevant in explaining the modelled outcome. The practical limit of this approach arises form the consideration that design variables are not always available for the analyst. More formally, given the definition of a ξ model to estimate a parameter β, the concept of design ignorability implies that, under ξ model validity, the data collection process and response mechanism do not provide any additional information to estimate β (for an analytical description see Chapter 7 of Gelman et al. 2003).

It is often the case that cluster and stratum information is not dissemi-

[10] If a more parsimonious approach is followed, the information on clusters can be collapsed to the strata then estimated as random effects (Pfeffermann and Lavange 1989). This idea can be explored to solve the problem of data confidentiality.

[11] Other scholars (using a more "econometric" approach), rely on simultaneous procedures where the phenomenon and the selection of the sampling units are modelled jointly (see for example Magee et al., 1998 and De Luca and Peracchi, 2007).

nated due to confidentiality protection. Geographical information (an ordinary choice for stratification) is disseminated at the aggregate level.[12] The proposal of those scholars pushing for an MB that "augments" model information with survey design variables is thus limited by this practice of the survey organizations. Those that can exploit this wealth of information, are then a limited number of researchers or the officers within the survey organizations.[13] In this study I take the perspective of data users (and not that of data producer) and therefore the possibility of implementing such a complex model will not be explored.

4 Testing for design ignorability

If the analyst does not have access to survey design information he/she has two alternatives:

1. disregard all the survey design information taking a standard MB stance incurring the risk of relying on inconsistent estimators;

2. adhere to the DB approach thus accepting some inefficiency as a price for protection against model mis-specification.

But there is another option: he/she can use the information embodied in survey weights to establish what the consequences are of excluding it from the model. In practice this means testing to see if the design is ignorable.

In the literature, DuMouchel and Duncan (1983) first proposed testing for the difference between weighted and unweighted estimators. This can be accomplished by various strategies.

4.1 Formal testing

Consider model (2) and add survey weights and their cross-products to form the augmented model $y_i = \gamma z_i^T + u_i$. It is straightforward to perform a Wald to evaluate if the coefficients of survey weights and their cross-products with the predictors are statistically different from zero. If $Rb = r$ denotes the set of q linear hypotheses to be jointly tested, then the Wald test statistic is:

[12] For example, the Survey on Household Income and Wealth SHIW, releases the nuts-2 (region) information; in the microdata of Eurostat coordinated surveys on living conditions (EU-SILC) and of the Federal Reserve Survey of Consumer Finances, only the nuts-1 variables (geographical area) are supplied in the dataset.

[13] A recent survey of the possible use of MB approach by data producers is in van den Brakel and Bethlehem, 2008.

$$W = (Rb - r)(RVR')^{-1}(Rb - r) \sim {\chi^2}_q; W/q \sim F(q, df) \tag{3}$$

The number of degrees of freedom (df) in the presence of a complex survey should reflect the randomization level. For example, in the case of a multi-stage, stratified design, they should be computed as *n° of clusters - n° of strata - n° of predictors*. When using replication-based variance estimates, the degrees of freedom are given by the number of replications (Faiella 2008). Such a test can be easily implemented with the functions usually embedded in the statistical software packages (e.g. `regTermTest` in the R survey package, `test` in the `svy:` Stata environment).

A Hausmann test is suggested by Pfefferman (1993) . Define $\hat{\beta_w}$ as the weighted least squares estimator, $\hat{\beta}$ as the standard LS estimator, and let $var(\hat{\beta_w} - \hat{\beta})$ be some robust measure of the variance of the difference in the two estimators (estimated using replication techniques). Then

$$\left(\hat{\beta_w} - \hat{\beta}\right)' \left[var\left(\hat{\beta_w} - \hat{\beta}\right)\right]^{-1} \left(\hat{\beta_w} - \hat{\beta}\right) \tag{4}$$

is a statistic asymptotically distributed as a χ^2_p where $p = dim(\hat{\beta})$. This is a test of DI in the sense that it verifies if excluding survey weights has a significant effect on the consistency of $\hat{\beta}$. In fact, under the null both $\hat{\beta}$ and $\hat{\beta}_w$ are consistent, while under the alternative only $\hat{\beta}_w$ is consistent.

When models are non-linear in the parameters it is better to use a statistic whose specification is invariant to non-linear transformations of the parameters. This property is violated by the Wald statistic but satisfied by the Lagrange Multiplier (LM) statistics (Kleibergen 2008).

An LM-score test can be derived as follows. Regress $y_i = (x_i^T \beta)$ obtain the residuals $\hat{u}_i = y_i - (x_i^T \hat{\beta})$ and run a second regression of the residuals on $x_i^T * (1 + w_i)$. The LM statistic is computed as the sample size times the coefficient of determination of this regression and it is compared with a χ_q, where q is the number of restrictions on the previous equation (in this case the number of predictors). This test can be extended to non-linear regressions comparing the ratio of the squares of efficient scores to model variance with a χ_q (Greene 2002).

This version of the LM is biased towards type I error. Kiviet (1986) proposes an F-test form of the LM test statistic with improved performances, defined as $LMF = \frac{n-k}{q} \frac{R^2}{1-R^2}$. Under the null $LMF \sim F(q, df)$.

4.2 Parameter exploration and other heuristics

A complement of formal testing consists in plotting the residuals (or a transformation of the residuals) of the unweighted regression against survey weights or design variables (if available) to look for correlation patterns that, if present, would suggest that survey weights have some role in predicting the outcome variable, even after controlling for a group of predictors.

Another useful check relies on the graphical representation of the unweighted and weighted parameters distribution. Given $\hat{\beta}$ and $\hat{\beta_w}$ and the associate standard errors, draw m random variates $\hat{\beta}_{sim} \sim N(\hat{\beta}, \hat{\sigma}_{\hat{\beta}})$ and $\hat{\beta}_{w\,sim} \sim N(\hat{\beta_w}, \hat{\sigma}_{\hat{\beta_w}})$. Then compare the MB and the DB estimators, looking at the distribution of these variates (graphical inspection such as density estimation or boxplots can help in spotting differences between the two).

5 Hands-on survey data

In this section I will look how to implement in practice the tests and the other diagnostic tools presented to help the researcher to choose between an MB and a DB estimator. As an example, I will make use of two surveys conducted by the Bank of Italy.

The first is the Survey on Household Income and Wealth (SHIW), conducted to collect information on the economic behaviour of Italian households. The sample comprises about 8,000 households and is drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and demographic size. Microdata, documentation and publications (in Italian and English) can be downloaded free of charge from the Bank of Italy's website.[14]

The second is the Survey of Industrial and Service Firms (SISF), that collects information on about 4,000 non-financial private service firms with 20 or more employees. The survey adopts a one-stage stratified sample design. The strata are combinations of the branch of activity, size class and regional location of the firm's head office.[15] Microdata can be elaborated using the Bank of Italy's Remote access to micro Data (BIRD) (Bruno, D'Aurizio, and Tartaglia-Polcini 2008).

The first is a complex survey (involving stratification, multiple stages of sampling, probability proportional to size selection methods and a split-panel

[14] `www.bancaditalia.it/statistiche/indcamp/bilfait`.

[15] Further details are available in the SISF report freely downloadable from the Bank website (`http://www.bancaditalia.it`)

design) with a rather low response rate (40 per cent) and this complexity is reflected in an elaborate multi-step construction of the survey weights (Faiella and Gambacorta 2007). SISF sample is instead a one stage stratified sample with a good response rate (75 per cent) and a more standard weighting set-up.

In the next section I estimate a linear model on these survey data with and without survey weights and I run the battery of tests and the heuristic procedures previously described to check if MB estimates capture the same information of the estimators that incorporate survey weights (DB).

Following the indications given in section 2.3, the variance of the estimators is computed using a randomization-based method. In practice a *replication-based method* known as *Jackknife Repeated Replications* (JRR) is adopted (details on the properties of this method are provided in Faiella, 2008).

5.1 A model for household expenditures

As a first example, I make use of a linear model of household expenditures. The analysis is based on SHIW 2006 data (7,768 households). The outcome variable is the log of household expenditures. The predictors, listed in Table A.1, are grouped in three categories: attributes of the head of household (defined as the main income earner within the household) such as age, job status etc.; characteristics of the household (household size, number of earners; etc.); indicators of the household economic situation such as household income, presence of liabilities, etc.

Table A.5 presents model results without survey weights (MB estimates), while Table A.6 shows the weighted (DB) estimates.

Table A.3 reports the results of the 3 tests previously presented: all the tests show that at 1 per cent confidence level the null hypothesis that design is ignorable is rejected. Hansen, Madow and Tepping (1983) and Lohr (1999) suggest that the decision to include survey weights in regression models implies a trade-off between bias and variance of the estimators; then a rule of thumb can be to include them when sample size is large and the sample size helps to mitigate the possible loss of efficiency. To test how sample size influences results, I perform the tests on a SHIW subsample that excludes the panel component (about 50 per cent of the full sample, about 3,900 observations). The results, in the bottom part of Table A.3, confirm the full-sample outcome.

I then check the difference in model features exploring the distribution of the parameters in the DB and the MB context. The relevant moments of the parameters distribution are computed (Table A.7) and a panel containing 4 figures is graphed: the first two report the density estimation of the MB and DB parameter. The third plots the MB simulated parameter against the DB one: if they are equal they should lie on the bisecting line. If points lie below (above) the bisecting line, it means that the simulated MB parameters are systematically lower (higher) then the DB parameters. Finally a boxplot of the MB and DB parameters summarizes the information on their distribution.[16]

Exploring the table and the panels we can conclude that:

1. it is not always the case that the DB estimator presents more variability: looking at the columns of the coefficient of variation (CV) of Table A.7 for 13 predictors out of 17, DB parameters are more volatile than MB parameters, but for the other 3 the reverse happens;

2. while for the majority of the parameters DB and MB estimates produce pretty similar results, the correlation between household expenditure and geographical information regarding household residence and household size appears to be quite different (see Figure 3-5).

It is apparent that, in the case under examination, both test results and the diagnostics exercise suggest that the DB estimator should be preferred over the MB estimator. In particular, the difference in the parameters related to the geographical location of the household (a typical piece of information used in designing the sample) seems to indicate that the sample is somehow "unbalanced" if compared with the distribution in the population. Note also that the deviance of the residuals of the DB model is (slightly) smaller, thus indicating that DB performs better in terms of explained variance (DB R-squared is 0.632, MB R-squared is 0.626).

5.2 A model to explain firms' turnover

In the second example, I model firms' turnover using SISF data. The analysis is based on 2008 data (about 4,000 firms). The outcome variable is the log of the turnover per employee. The predictors, listed in Table A.2, cover firm characteristics (age, sector, location and size), overseas sales, the intensity of activity during the year and the investment level in the previous year.

[16]In order not to burden the reader, I present the diagnostic plots of selected covariates only. Complete results and the R code to generate this diagnostic is available from the author.

Table A.8 presents model results without survey weights (MB estimates), while Table A.9 shows the weighted (DB) estimates.

Table A.4 reports that as with SHIW data all the tests reject the null hypothesis that design is ignorable. To test how sample size influences tests results, I perform the same tests on a subsample that randomly excludes about 50 per cent of the observation. The results, in the bottom part of Table A.4, confirm the full-sample outcome.

Exploring the table and the panels with the distribution of the parameters we can conclude that:

1. using SISF data the DB estimator is always more variable: looking at the columns of the CV of Table A.10 for 7 predictors out of 10, DB parameters are at least twice more variable than MB parameters;

2. while for the majority of the parameters DB and MB estimates produce pretty similar results, the association with firm location and with investment shows important discrepancies (see Figure 6-7).

SISF analysis confirms that the DB estimator should be preferred over the MB estimator and it suggests that the sample distribution is "unbalanced" if compared with the distribution in the population.

6 Conclusions

In this paper, I have presented the benefits and the costs of using MB or DB estimators. What I pointed out is that:

1. in estimating the variance of the parameters randomization-based methods are robust to mis-specification thus suggesting that this should be the preferred strategy by the researcher;

2. instead of deciding what approach to use on the basis of devotion to a theory, the survey data analyst should look at the differences in DB and MB estimators;

3. to accomplish this task a set of econometric tests is suggested. These tests are somewhat modified to be sure that the underlying variance and degrees of freedom measures account for the randomization process;

4. the result of the tests should be supplemented by the analysis of model features. For this reason a set of diagnostic tools (heuristics) is suggested, simulating DB and MB parameters and looking (also graphically) at their distribution.

I applied these principles to a linear model using a survey whose weights reflect a sophisticated procedure (SHIW) and a survey with a more standard weighting process (SISF). The results indicate that in both cases it is safer to use survey weights, because MB specification seems to fail in capturing the information incorporated in the survey weights.

The alternative approach to set-up a multilevel model is not explored because its application is constrained by the limited design information that the majority of researchers are provided with. In fact for reasons of confidentiality protection, strata and cluster information are usually not disseminated in sample survey micro-data.

I would like to conclude with this 1987 ASA communication from Alexander: "[...] the proponents of weighting (such as the author) would assert that no model will include all the relevant variables, and that few analysts will wish to include in their model all the geographic and operational variables which determine sampling rates. It is difficult to object in principle with the goal of correctly modelling all relevant variables, including the variables relating to sampling. However, the theoretical and empirical tasks of deriving, fitting, and validating such models seem formidable for many complex national demographic surveys."(Alexander 1987)

Modern PC's computational power and the availability of statistical software (in the case of R, even in the public domain) *it is as revolutionary for research in the behavioural sciences as the microscope in biology* (Hiaschi and Selvin, 1967, cited in Skinner et al., 1989).

Giving the increasing opportunity to explore microdata to fully account for the heterogeneity in the individual behaviour, the modellers should check if the information about the population that survey practitioners adopt in building survey weights is relevant in explaining the object of the analysis (i.e design ignorability). If this condition is not met, the model should incorporate design information.

References

Alexander, C. H. (1987): "A Model-Based Justification for Survey Weights," in *Proceedings of the Survey Research Methods Section American Statistical Association*.

Beaumont, J.-F. (2008): "A new approach to weighting and inference in sample surveys," *Biometrika*, 95(3), 539–553.

Binder, D. (1983): "On the Variance of Asimptotically Normal Estimators from Complex Surveys," *International Statistical Review*, 51, 279–272.

Binder, D., and G. Roberts (2003): "Design-based and Model-based Methods for Estimating," in *Analysis of survey data*, pp. 29–33. Wiley.

Bruno, G., L. D'Aurizio, and R. Tartaglia-Polcini (2008): "Remote processing of firm microdata at the Bank of Italy Giuseppe Bruno," Discussion paper, Bank of Italy - Occasional Papers.

Cameron, A. C., and P. K. Trivedi (2005): *MICROECONOMETRICS: Methods and Applications*. Cambridge University Press.

Chambers, R. L., and C. J. Skinner (eds.) (2003): *Analysis of survey data*. Wiley, New York.

DuMouchel, W. H., and G. J. Duncan (1983): "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples," *Journal of the American Statistical Association*, 78(383), 535–543.

Faiella, I. (2008): "Accounting for sampling design in the SHIW," *Temi di Discussione del Servizio Studi*, 662.

Faiella, I., and R. Gambacorta (2007): "The weighting process in the SHIW," *Temi di Discussione del Servizio Studi*, 636.

Gelman, A. (2007): "Struggles with Survey Weighting and Regression Modeling," *Statistical Science*, 22, 153–173.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003): *Bayesian Data Analysis*, Chapman & Hall Texts in Statistical Science. Chapman & Hall, 2 edn.

Greene, W. H. (2002): *Econometric Analysis*. Prentice Hall.

HANSEN, M., W. MADOW, AND B. TEPPING (1983): “An evaluation of model-dependent and probability sampling inferences in sample surveys,” *Journal of the American Statistical Association*, 78, 776–793.

ISAKI, C. T., AND W. A. FULLER (1982): “Survey Design Under the Regression Superpopulation Model,” *Journal of the American Statistical Association*, 77, 89–96.

KALTON, G. (1989): “Modeling Considerations: Discussion from a Survey Sampling Perspectives,” in *Panel Surveys*, pp. 575–585. Wiley.

KISH, L. (1965): *Survey Sampling*. Wiley, New York.

KIVIET, J. (1986): “On the rigour of some misspecification tests for modelling dynamic relationships,” *Review of Economic Studies*, 53, 241–261.

KLEIBERGEN, F. (2008): “testing,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.

KOTT, P. (1991): “A Model-Based Look at Linear Regression with Survey Data,” *The American Statistician*, 45, 107–112.

LITTLE, R. (1981): “Robust model-based inference for a finite population mean from unequally weighted samples,” in *Proceedings of the Survey Research Methods Section American Statistical Association.*

LITTLE, R. J., AND S. VARTIVARIAN (2005): “Does Weighting for Nonresponse Increase the Variance of Survey Means?,” *Survey Methodology*, 31, 161–168.

LITTLE, R. J. A. (2004): “To Model or Not To Model? Competing Modes of Inference for Finite Population Sampling,” *Journal of the American Statistical Association*, 11, 546–556.

LOHR, S. (1999): *Sampling: Design and Analysis*. Duxbury Press.

LUCA, G. D., AND F. PERACCHI (2007): “A sample selection model for unit and item nonresponse in cross-sectional surveys,” *CEIS Tor Vergata RESEARCH PAPER SERIES*, 99.

MAGEE, L., A. ROBB, AND J. BURBIDGE (1998): “On the use of sampling weights when estimating regression models with survey data,” *Journal of Econometrics*, 84, 251–271.

Narain, R. (1951): "On sampling without replacement with varying probabilities," *Journal of the Indian Society of Agricultural Statistics*, 3, 169–174.

Nathan, G., and T. Smith (1989): "The Effect of Selection in Regression Analysis," in *Analysis of Complex Surveys*, pp. 149–163. Wiley.

Nordberg, L. (1989): "Generalized linear modeling of sample survey data," *Journal of Official Statistics*, 5, 223–239.

Pfeffermann, D. (1993): "The role of sampling weights when modeling survey data," *International Statistical Review*, 61, 317–337.

Pfeffermann, D., and L. Lavange (1989): "Regression Models for Stratified Multistage Cluster Samples," in *Analysis of Complex Surveys*, pp. 237–260. Wiley.

Pfeffermann, D., and M. Sverchkov (1999): "Parametric and semiparametric estimation of regression models fitted to survey data," *SANKHYA*, 61, 166–186.

Pfeffermann, D., and M. Y. Sverchkov (2003): "Fitting Generalized Linear Model under Informative Sampling," in *Analysis of survey data*, pp. 175–195. Wiley.

Scott, A., and M. Smith (1969): "Estimation in multi-stage surveys," *Journal of the American Statistical Association*, 64, 830–840.

Skinner, C., D. Holt, and T. Smith (eds.) (1989): *Analysis of Complex Surveys*. Wiley, New York.

Särndal, C., B. Swensson, and J. Wretman (1992): *Model Assisted Survey Sampling*. Springer-Verlag.

van den Brakel, J., and J. Bethlehem (2008): "Model-Based Estimation for Official Statistics," *Statistics Netherlands Discussion papers*, 08002, http://www.cbs.nl/nl-NL/menu/methoden/research/discussionpapers/archief/2008/2008-02-x10-pub.htm.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

——— (2006): "Cluster-Sample Methods In Applied Econometrics: An Extended Analysis," *Mimeo*.

Yuan, Y., and R. Little (2007a): "Model-Based Estimates of the Finite Population Mean for Two-Stage Cluster Samples with Unit Non-response," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56, 79–97.

——— (2007b): "Parametric and Semiparametric Model-based Estimates of the Finite Population Mean for Two-Stage Cluster Samples with Item Nonresponse," *Biometrics*, 63, 1172–1180.

APPENDIX: Tables and Figures

Table A.1. Predictors for the (log) consumption equation

Name of the variable	Description
Head of Household	
$I(SEX == 2)$	Female=1; 0 otherwise
ETA	Age
$I(ETA^2)$	Age squared
$SICK$	Household head sick=1 (self reported); 0 otherwise
$I(STUDIO > 3)$	Holds at least a secondary school diploma=1; 0 otherwise
$I(CIT == 1)$	Italian citizen=1; 0 otherwise
$I(STACIV == 1)$	Married=1; 0 otherwise
$I(Q == 2)$	Self-employed=1; 0 otherwise
Household	
$I(NCOMP > 2)$	Size (n° of members) >2=1; 0 otherwise
$I(NPERC > 1)$	Income earners>1=1; 0 otherwise
$I(AREA3 == 3)$	Residing in the South and Islands=1; 0 otherwise
$I(ACOM4C == 3)$	Municipality with 500k+ inhabitants=1; 0 otherwise
Household economic condition	
y	Log of household income
$I(CLW < 3)$	Net wealth less than the 30th percentile=1; 0 otherwise
$I(CLW > 8)$	Net wealth more than the 80th percentile=1; 0 otherwise
$I(PF > 0)$	Debt ownership=1; 0 otherwise

Table A.2. Predictors for the (log) turnover per employee

Name of the variable	Description
$I(SETTOR3! = MANIFATT.)$	1=Service sector; 0=Manufacturing sector
$fattest$	Foreign turnover (log)
$ladd$	Employees (log)
$ladd2$	Employees (log) squared
$orelav$	Number of hours worked in the year (log)
$orestra$	Number of hours worked in the year – overtime (log)
$linv0$	Previous year investments (log)
$I(AREAG4 == 4)$	1=South and Islands; 0 otherwise
eta	Age of the firm

Table A.3. Test results for Design Ignorability (SHIW data)

Full sample (7768 obs.)		
	Distribution under H0	P-value
Wald	$\chi^2_{df=q}$	0.000
Hausmann	$\chi^2_{df=dim(\beta)}$	0.004
LM score [LMF version]	$\chi_q[F(q,df)]$	0.000[0.000]
Excluding the panel component (about half of the sample=3881 obs.)		
	Distribution under H0	P-value
Wald	$\chi^2_{df=q}$	0.001
Hausmann	$\chi^2_{df=dim(\beta)}$	0.001
LM score [LMF version]	$\chi_q[F(q,df)]$	0.011[0.010]

Table A.4. Test results for Design Ignorability (SISF data)

Full sample (3848 obs.)		
	Distribution under H0	P-value
Wald	$\chi^2_{df=q}$	0.000
Hausmann	$\chi^2_{df=dim(\beta)}$	0.000
LM score [LMF version]	$\chi_q[F(q,df)]$	0.000[0.000]
Excluding randomly about half of the sample=1,899 obs.)		
	Distribution under H0	P-value
Wald	$\chi^2_{df=q}$	0.000
Hausmann	$\chi^2_{df=dim(\beta)}$	0.000
LM score [LMF version]	$\chi_q[F(q,df)]$	0.000[0.000]

Table A.5. Expenditure equation: unweighted (MB) estimates

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.988	0.365	16.401	$<$ 2e-16	***
y	0.343	0.039	8.715	$<$ 2e-16	***
$I(SEX == 2)$	-0.014	0.012	-1.148	0.252	
ETA	0.007	0.002	3.961	0.000	***
$I(ETA^2)$	0.000	0.000	-4.986	0.000	***
$SICK$	-0.068	0.034	-2.012	0.045	*
$I(STUDIO > 3)$	0.124	0.014	8.992	$<$ 2e-16	***
$I(NCOMP > 2)$	0.084	0.011	7.373	0.000	***
$I(NPERC > 1)$	0.020	0.018	1.106	0.269	
$I(CIT == 1)$	0.151	0.028	5.383	0.000	***
$I(STACIV == 1)$	0.115	0.011	10.709	$<$ 2e-16	***
$I(Q == 2)$	-0.003	0.015	-0.192	0.847	
$I(CLW < 3)$	-0.072	0.016	-4.398	0.000	***
$I(CLW > 8)$	0.194	0.017	11.367	$<$ 2e-16	***
$I(PF > 0)$	0.091	0.010	8.733	$<$ 2e-16	***
$I(AREA3 == 3)$	-0.127	0.012	-10.187	$<$ 2e-16	***
$I(ACOM4C == 3)$	0.085	0.019	4.376	0.000	***

Signif. codes: *** 0.001 ** 0.01 * 0.05 . 0.1
n=7768, degrees of freedom=329, Resid.Dev=0.10366

Table A.6. Expenditure equation: weighted (DB) estimates

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.949	0.363	16.384	$<$ 2e-16	***
y	0.343	0.039	8.686	$<$ 2e-16	***
$I(SEX == 2)$	-0.006	0.013	-0.446	0.656	
ETA	0.007	0.002	3.335	0.001	***
$I(ETA^2)$	0.000	0.000	-4.009	0.000	***
$SICK$	-0.084	0.047	-1.771	0.077	*
$I(STUDIO > 3)$	0.121	0.017	7.306	0.000	***
$I(NCOMP > 2)$	0.112	0.013	8.801	$<$ 2e-16	***
$I(NPERC > 1)$	0.005	0.019	0.244	0.808	
$I(CIT == 1)$	0.179	0.032	5.529	0.000	***
$I(STACIV == 1)$	0.106	0.013	8.081	0.000	***
$I(Q == 2)$	0.002	0.018	0.106	0.915	
$I(CLW < 3)$	-0.064	0.017	-3.786	0.000	***
$I(CLW > 8)$	0.196	0.019	10.459	$<$ 2e-16	***
$I(PF > 0)$	0.107	0.012	8.626	0.000	***
$I(AREA3 == 3)$	-0.157	0.018	-8.812	$<$ 2e-16	***
$I(ACOM4C == 3)$	0.104	0.015	7.065	0.000	***

Signif. codes: *** 0.001 ** 0.01 * 0.05 . 0.1
n=7768, degrees of freedom=329, Resid.Dev=0.10351

Table A.7. Statistics on the parameter distribution (SHIW data)

	Mean	CV	P0	P25	P50	P75	P100
Parameter 1 *Intercept*							
MB estimates	5.98	6.07	4.76	5.73	5.98	6.23	7.11
DB estimates	5.94	6.08	4.73	5.70	5.94	6.19	7.07
Parameter 2 y							
MB estimates	0.34	11.45	0.21	0.32	0.34	0.37	0.46
DB estimates	0.34	11.48	0.21	0.32	0.34	0.37	0.46
Parameter 3 $I(SEX == 2)$							
MB estimates	-0.01	-84.84	-0.05	-0.02	-0.01	-0.01	0.02
DB estimates	-0.01	-211.66	-0.05	-0.01	-0.01	0.00	0.03
Parameter 4 ETA							
MB estimates	0.01	25.27	0.00	0.01	0.01	0.01	0.01
DB estimates	0.01	30.04	0.00	0.01	0.01	0.01	0.01
Parameter 5 $I(ETA^2)$							
MB estimates	0.00	-19.85	0.00	0.00	0.00	0.00	0.00
DB estimates	0.00	-24.66	0.00	0.00	0.00	0.00	0.00
Parameter 6 $MALATO$							
MB estimates	-0.07	-48.84	-0.18	-0.09	-0.07	-0.05	0.04
DB estimates	-0.08	-55.40	-0.24	-0.12	-0.08	-0.05	0.06
Parameter 7 $I(STUDIO > 3)$							
MB estimates	0.12	11.09	0.08	0.11	0.12	0.13	0.17
DB estimates	0.12	13.66	0.07	0.11	0.12	0.13	0.17
Parameter 8 $I(NCOMP > 2)$							
MB estimates	0.08	13.53	0.05	0.08	0.08	0.09	0.12
DB estimates	0.11	11.33	0.07	0.10	0.11	0.12	0.15
Parameter 9 $I(NPERC > 1)$							
MB estimates	0.02	91.92	-0.04	0.01	0.02	0.03	0.07
DB estimates	0.00	452.97	-0.06	-0.01	0.00	0.02	0.06
Parameter 10 $I(CIT == 1)$							
MB estimates	0.15	18.56	0.06	0.13	0.15	0.17	0.24
DB estimates	0.18	18.07	0.07	0.16	0.18	0.20	0.28
Parameter 11 $I(STACIV == 1)$							
MB estimates	0.12	9.31	0.08	0.11	0.12	0.12	0.15
DB estimates	0.11	12.35	0.06	0.10	0.11	0.11	0.15
Parameter 12 $I(Q == 2)$							
MB estimates	0.00	-458.87	-0.05	-0.01	0.00	0.01	0.04
DB estimates	0.00	1212.18	-0.06	-0.01	0.00	0.01	0.06
Parameter 13 $I(CLW < 3)$							
MB estimates	-0.07	-22.49	-0.13	-0.08	-0.07	-0.06	-0.02
DB estimates	-0.06	-26.10	-0.12	-0.08	-0.06	-0.05	-0.01
Parameter 14 $I(CLW > 8)$							
MB estimates	0.19	8.77	0.14	0.18	0.19	0.21	0.25
DB estimates	0.20	9.53	0.13	0.18	0.20	0.21	0.25
Parameter 15 $I(PF > 0)$							
MB estimates	0.09	11.42	0.06	0.08	0.09	0.10	0.12
DB estimates	0.11	11.56	0.07	0.10	0.11	0.12	0.15
Parameter 16 $I(AREA3 == 3)$							
MB estimates	-0.13	-9.74	-0.17	-0.14	-0.13	-0.12	-0.09
DB estimates	-0.16	-11.26	-0.22	-0.17	-0.16	-0.15	-0.10
Parameter 17 $I(ACOM4C == 3)$							
MB estimates	0.08	22.85	0.02	0.07	0.08	0.10	0.14
DB estimates	0.10	14.13	0.05	0.09	0.10	0.11	0.15

Table A.8. Turnover equation: unweighted (MB) estimates

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.6323392	0.6636729	-0.953	0.34076	
$I(SETTOR3! = MANIFATT.)$	0.3036911	0.0329328	9.222	$<$ 2e-16	***
$fattest$	0.0771898	0.0045432	16.990	$<$ 2e-16	***
$ladd$	-1.1162046	0.1058763	-10.543	$<$ 2e-16	***
$ladd2$	0.0065252	0.0051963	1.256	0.20929	
$orelav$	0.8021574	0.0876852	9.148	$<$ 2e-16	***
$orestra$	-0.0459207	0.0151399	-3.033	0.00244	**
$linv0$	0.1479054	0.0093250	15.861	$<$ 2e-16	***
$I(AREAG4 == 4)$	-0.1581184	0.0313940	-5.037	4.96e-07	***
eta	0.0010174	0.0005546	1.835	0.06663	
Signif. codes: *** 0.001 ** 0.01 * 0.05 . 0.1					
n=3848, degrees of freedom=3780, Resid.Dev=2506					

Table A.9. Turnover equation: weighted (DB) estimates

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.738289	1.352248	0.546	0.585118	
$I(SETTOR3! = MANIFATT.)$	0.334894	0.048262	6.939	4.62e-12	***
$fattest$	0.077259	0.007828	9.870	$<$ 2e-16	***
$ladd$	-0.756831	0.227485	-3.327	0.000886	***
$ladd2$	-0.002938	0.012520	-0.235	0.814483	
$orelav$	0.587234	0.178975	3.281	0.001043	**
$orestra$	-0.064673	0.024685	-2.620	0.008830	**
$linv0$	0.112803	0.014619	7.716	1.52e-14	***
$I(AREAG4 == 4)$	-0.250276	0.054579	-4.586	4.67e-06	***
eta	-0.001326	0.001274	-1.040	0.298262	
Signif. codes: *** 0.001 ** 0.01 * 0.05 . 0.1					
n=3848, degrees of freedom=3780, Resid.Dev=2784					

Table A.10. Statistics on the parameter distribution (SISF data)

	Mean	CV	P0	P25	P50	P75	P100
			Parameter n° 1 Intercept				
MB estimates	-0.6484	-101.798	-2.8671	-1.0963	-0.6452	-0.198	1.4117
DB estimates	0.7055	190.6469	-3.815	-0.207	0.7121	1.6233	4.9031
			Parameter n° 2 I(d$SETTOR3!=MANIFATT.)				
MB estimates	0.3029	10.8143	0.1928	0.2807	0.303	0.3252	0.4051
DB estimates	0.3337	14.3839	0.1724	0.3012	0.334	0.3665	0.4835
			Parameter n° 3 fattest				
MB estimates	0.07708	5.86239	0.06189	0.07401	0.0771	0.08016	0.09118
DB estimates	0.07707	10.10184	0.0509	0.07179	0.07711	0.08238	0.10137
			Parameter n° 4 ladd				
MB estimates	-1.1188	-9.4127	-1.4727	-1.1902	-1.1182	-1.0469	-0.7901
DB estimates	-0.7623	-29.6794	-1.5228	-0.9159	-0.7612	-0.6079	-0.0562
			Parameter n° 5 ladd2				
MB estimates	0.0064	80.76622	-0.01097	0.00289	0.00642	0.00993	0.02253
DB estimates	-0.00324	-384.129	-0.0451	-0.01169	-0.00318	0.00526	0.03562
			Parameter n° 6 orelav				
MB estimates	0.8	10.9012	0.5069	0.7409	0.8005	0.8596	1.0722
DB estimates	0.58289	30.53952	-0.01542	0.46212	0.58377	0.70437	1.13846
			Parameter n° 7 orestra				
MB estimates	-0.04629	-32.5319	-0.0969	-0.0565	-0.04621	-0.03601	0.00071
DB estimates	-0.06527	-37.6152	-0.14779	-0.08193	-0.06515	-0.04852	0.01135
			Parameter n° 8 linv0				
MB estimates	0.1477	6.2804	0.1165	0.1414	0.1477	0.1540	0.1766
DB estimates	0.11245	12.93049	0.06358	0.10258	0.11252	0.12237	0.15783
			Parameter n° 9 I(AREAG4==4)				
MB estimates	-0.15888	-19.6532	-0.26383	-0.18006	-0.15873	-0.13757	-0.06143
DB estimates	-0.2516	-21.576	-0.43406	-0.28843	-0.25133	-0.21455	-0.08218
			Parameter n° 10 eta				
MB estimates	0.001	54.93891	-0.00085	0.00063	0.00101	0.00138	0.00273
DB estimates	-0.00136	-93.4286	-0.00562	-0.00222	-0.00135	-0.00049	0.0026

Figure 1. Distribution of the parameter: (log) household income (y)

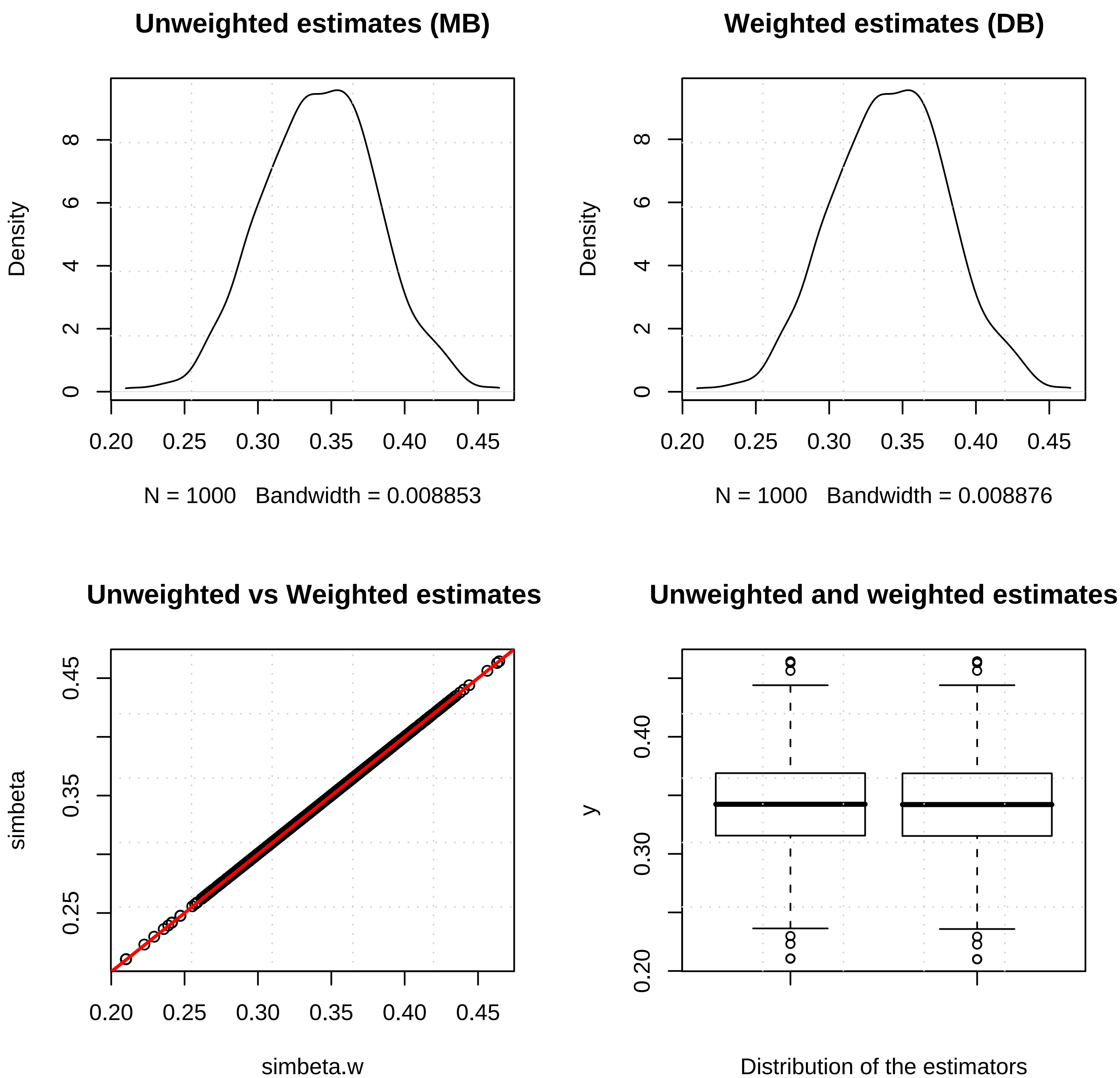


Figure 2. Distribution of the parameter: head of household age (ETA)

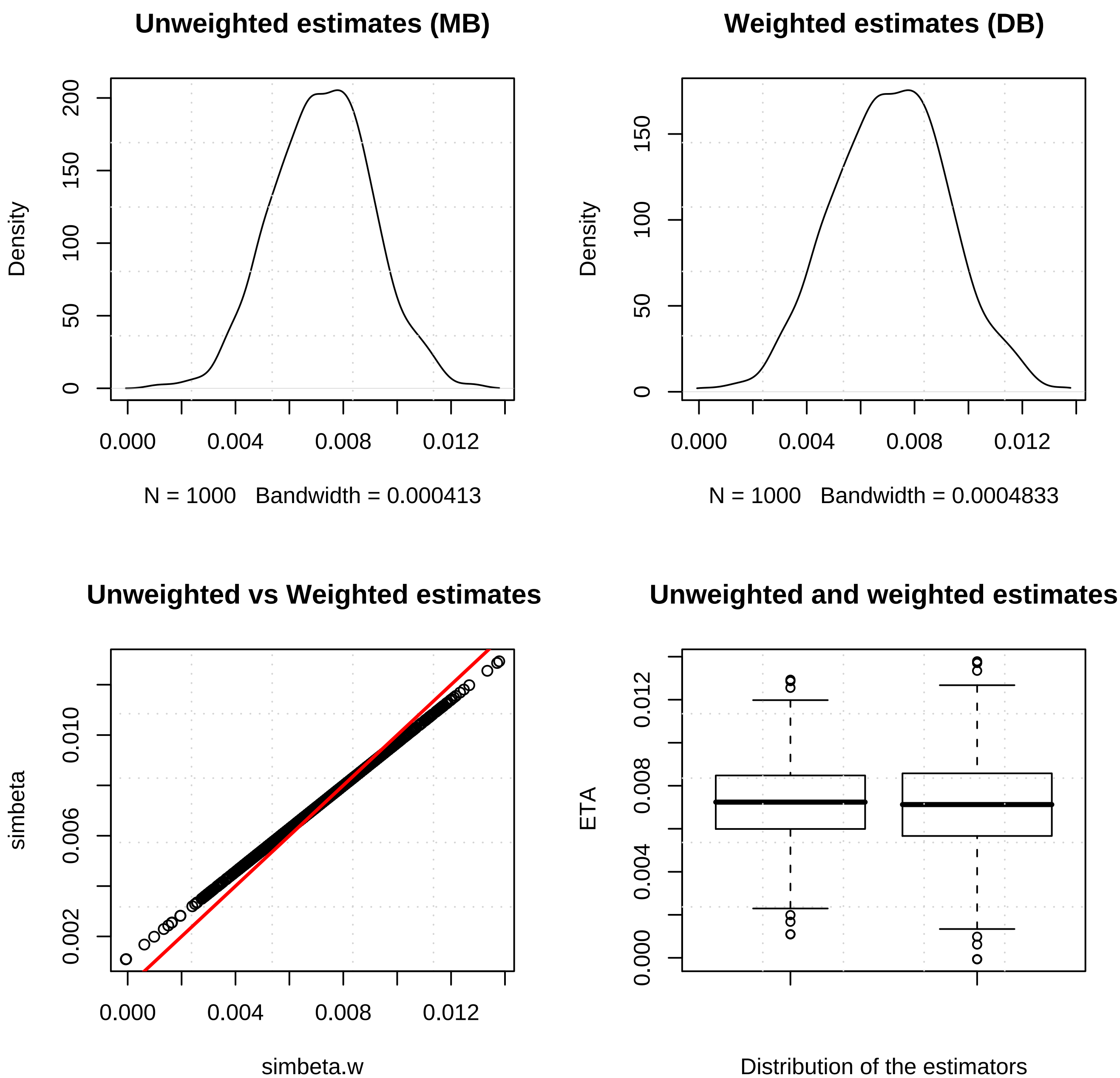


Figure 3. Distribution of the parameter: number of household members greater than 2 ($I(NCOMP > 2)$)

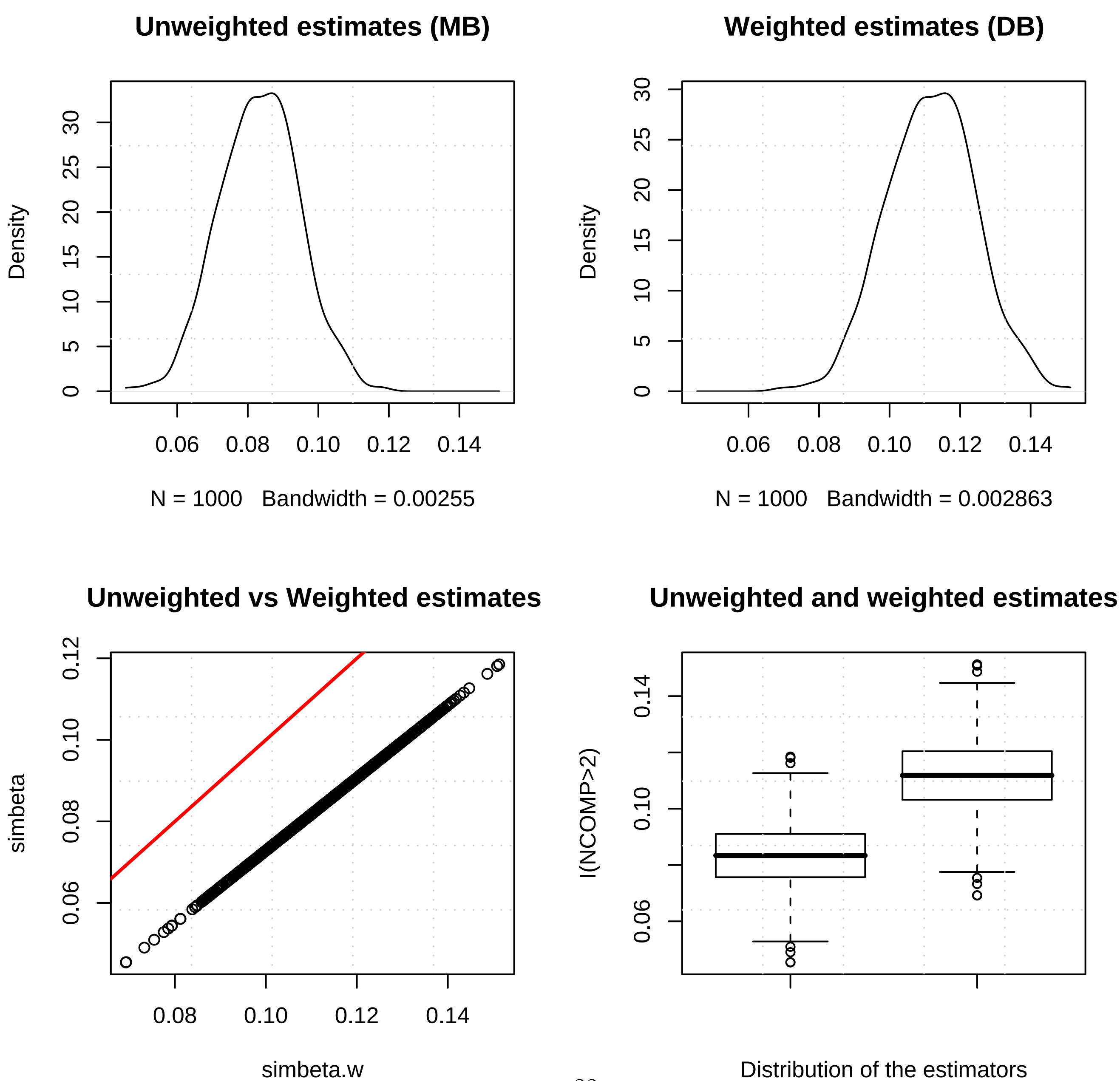


Figure 4. Distribution of the parameter: household residing in the South and Islands ($I(AREA3 == 3)$)

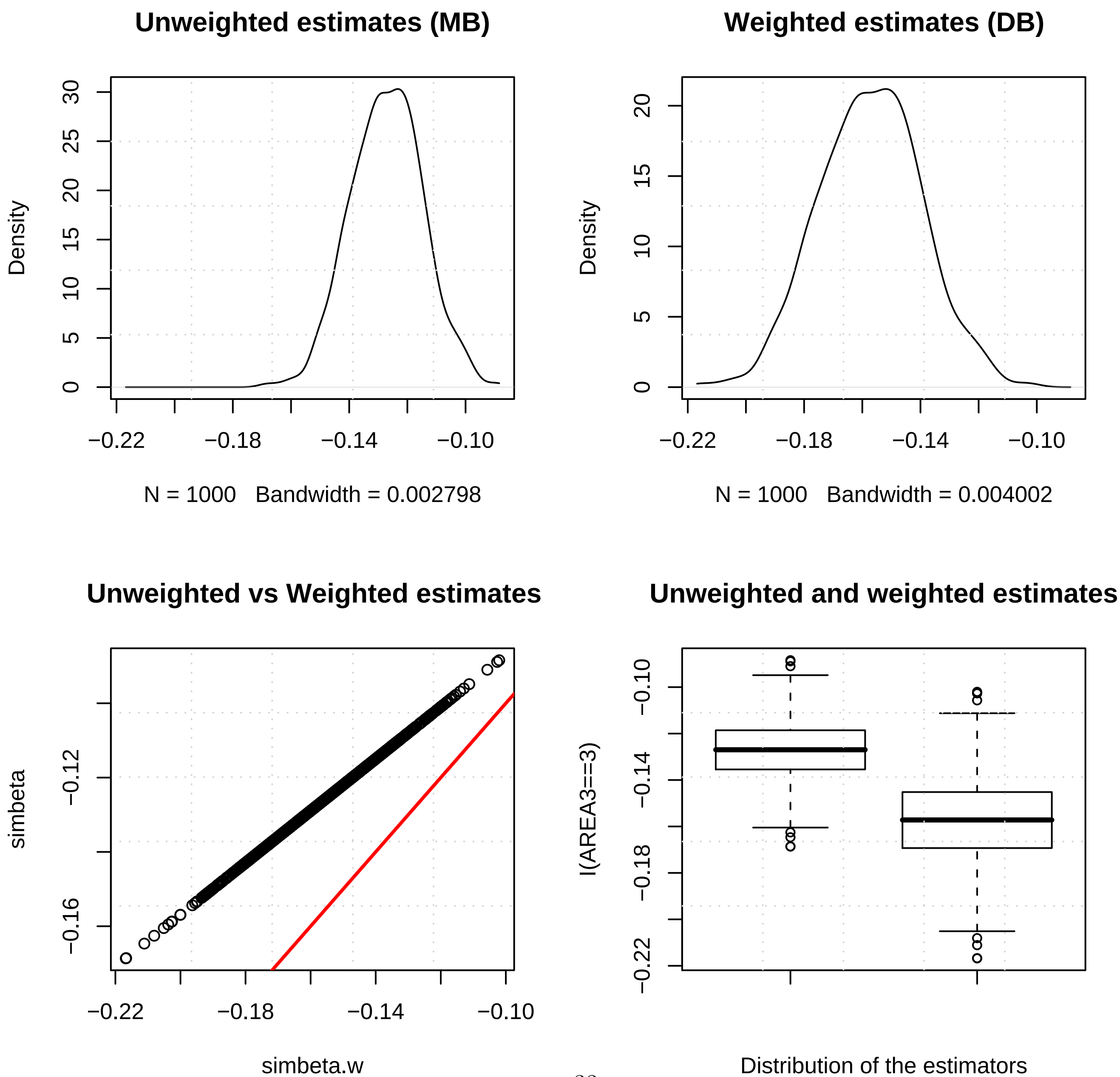


Figure 5. Distribution of the parameter: household residing in municipalities with more than 500k inhabitants ($I(ACOM4C == 3)$)

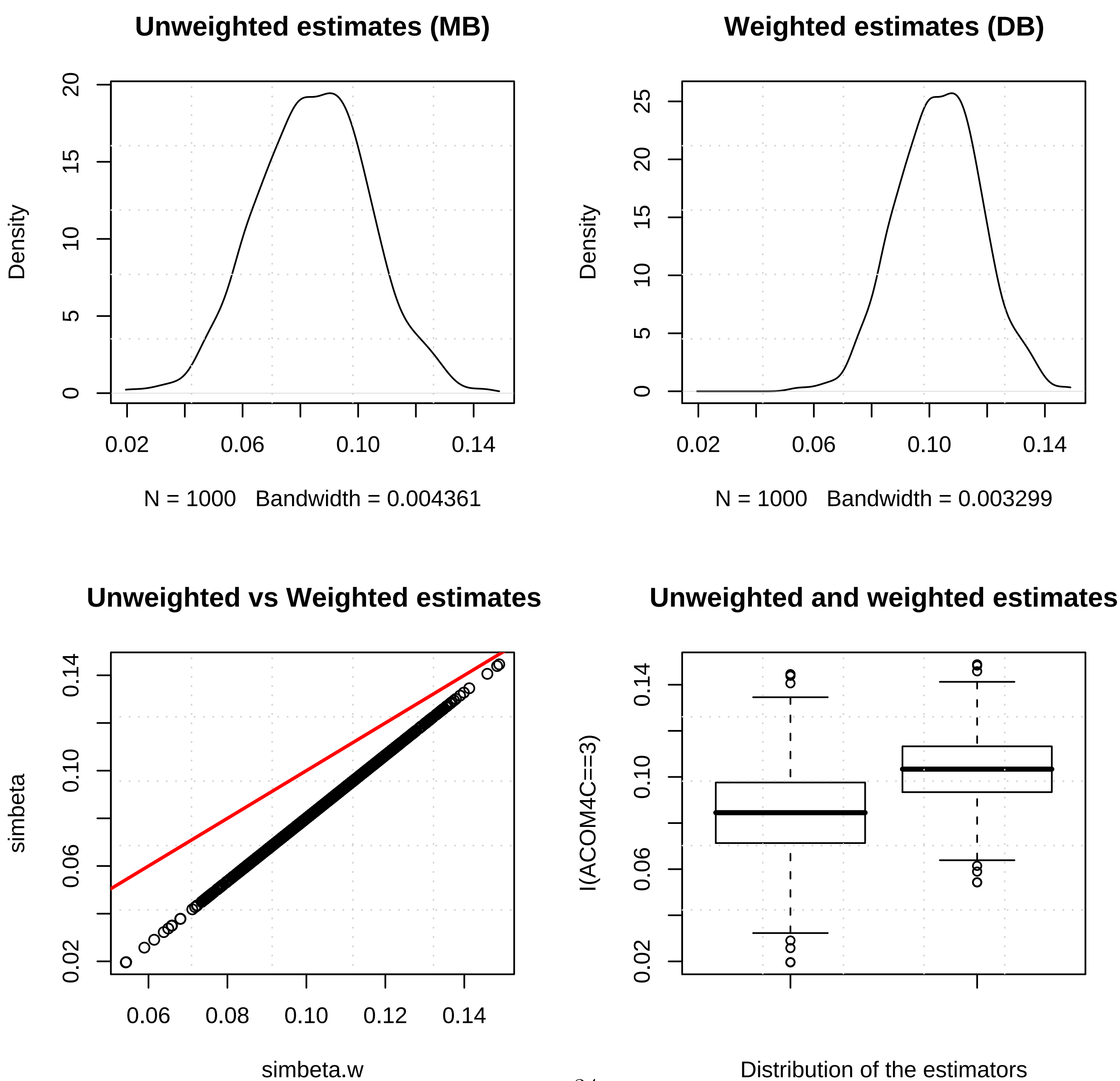


Figure 6. Distribution of the parameter: previous year investments ($linv0$)

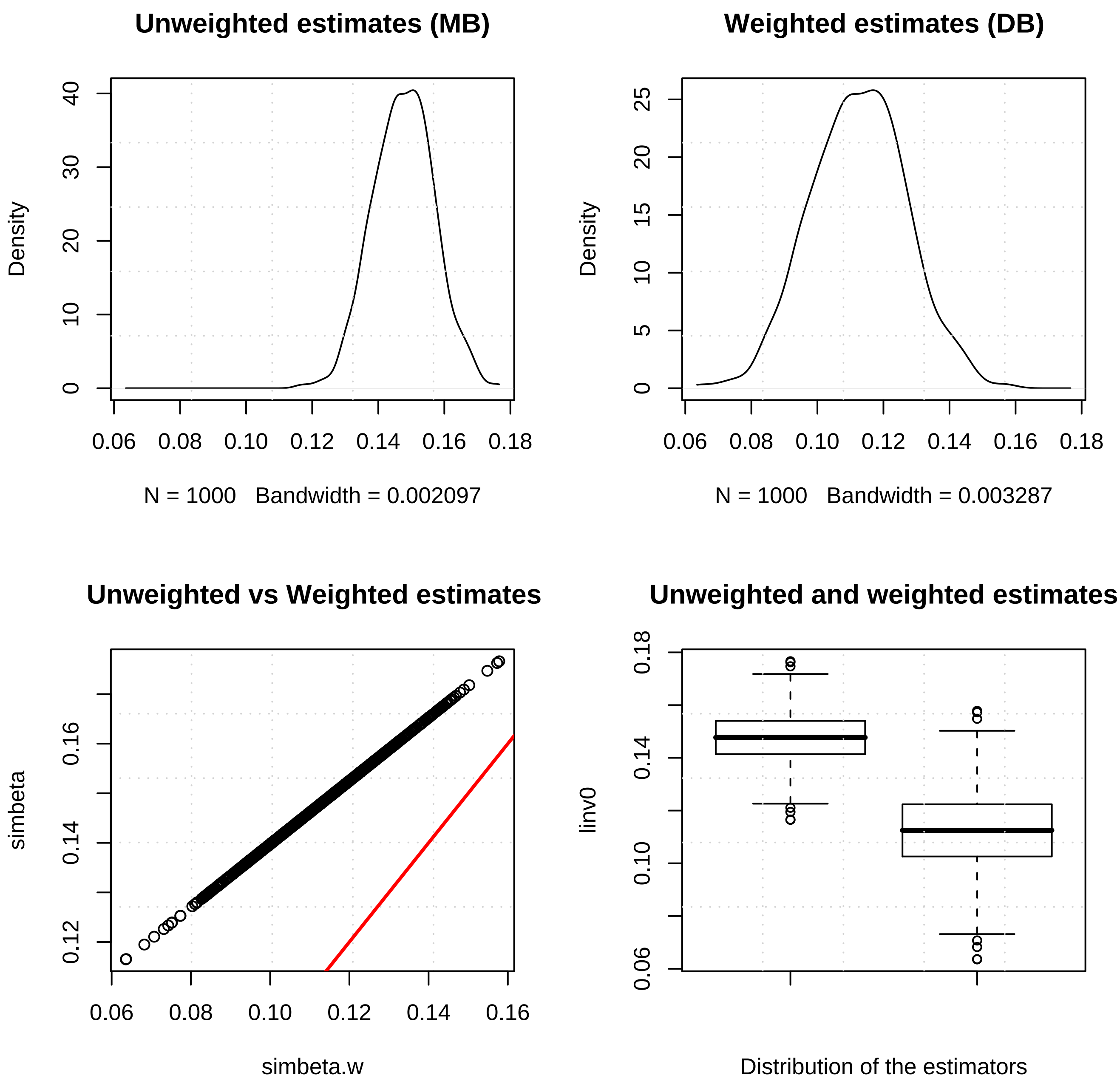
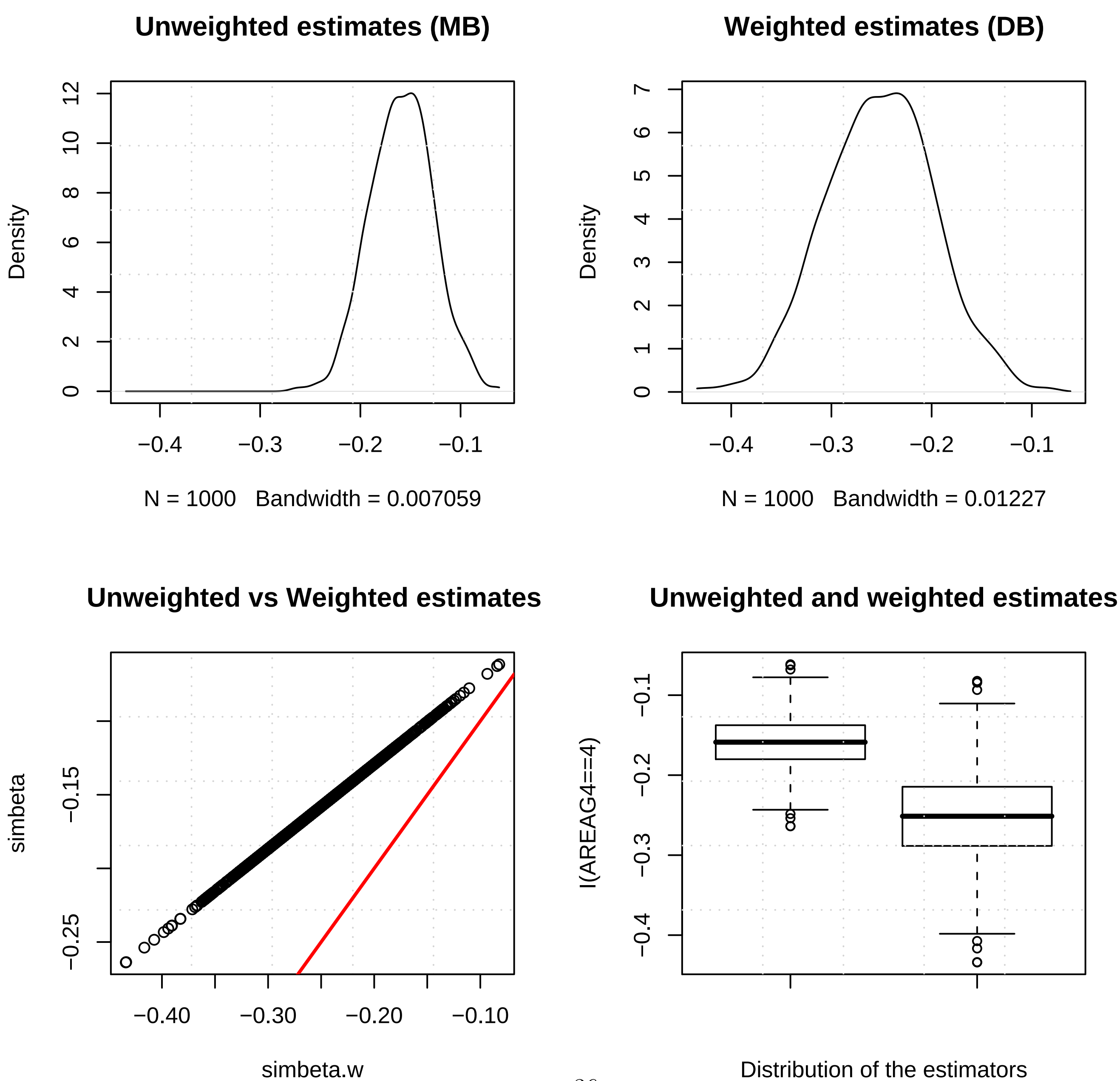


Figure 7. Distribution of the parameter: firms located in the South and Islands ($I(AREAG4 == 4)$)



RECENTLY PUBLISHED “TEMI” (*)

N. 714 – *L'attività retail delle banche estere in Italia: effetti sull'offerta di credito alle famiglie e alle imprese,* by Luigi Infante and Paola Rossi (June 2009)

N. 715 – *Firm heterogeneity and comparative advantage: the response of French firms to Turkey's entry in the European Customs Union,* by Ines Buono (June 2009).

N. 716 – *The euro and firm restructuring,* by Matteo Bugamelli, Fabiano Schivardi and Roberta Zizza (June 2009).

N. 717 – *When the highest bidder loses the auction: theory and evidence from public procurement,* by Francesco Decarolis (June 2009).

N. 718 – *Innovation and productivity in SMEs. Empirical evidence for Italy,* by Bronwyn H. Hall, Francesca Lotti and Jacques Mairesse (June 2009).

N. 719 – *Household wealth and entrepreneurship: is there a link?,* by Silvia Magri (June 2009).

N. 720 – *The announcement of monetary policy intentions*, by Giuseppe Ferrero and Alessandro Secchi (September 2009).

N. 721 – *Trust and regulation: addressing a cultural bias*, by Paolo Pinotti (September 2009).

N. 722 – *The effects of privatization and consolidation on bank productivity: comparative evidence from Italy and Germany*, by E. Fiorentino, A. De Vincenzo, F. Heid, A. Karmann and M. Koetter (September 2009).

N. 723 – *Comparing forecast accuracy: a Monte Carlo investigation*, by Fabio Busetti, Juri Marcucci and Giovanni Veronese (September 2009).

N. 724 – *Nonlinear dynamics in welfare and the evolution of world inequality*, by Davide Fiaschi and Marzia Romanelli (October 2009).

N. 725 – *How are firms' wages and prices linked: survey evidence in Europe* , by Martine Druant, Silvia Fabiani, Gabor Kezdi, Ana Lamo, Fernando Martins and Roberto Sabbatini (October 2009).

N. 726 – *Low skilled immigration and the expansion of private schools*, by Davide Dottori and I-Ling Shen (October 2009).

N. 727 – *Sorting, reputation and entry in a market for experts*, by Enrico Sette (October 2009).

N. 728 – *Ricardian selection*, by Andrea Finicelli, Patrizio Pagano and Massimo Sbracia (October 2009).

N. 729 – *Trade-revealed TFP*, by Andrea Finicelli, Patrizio Pagano and Massimo Sbracia (October 2009).

N. 730 – *The riskiness of corporate bonds*, by Marco Taboga (October 2009).

N. 731 – *The interbank market after august 2007: what has changed and why?*, by Paolo Angelini, Andrea Nobili and Maria Cristina Picillo (October 2009).

N. 733 – *Dynamic macroeconomic effects of public capital: evidence from regional Italian data*, by Valter Di Giacinto, Giacinto Micucci and Pasqualino Montanaro (November 2009).

N. 734 – *Networks with decreasing returns to linking* , by Filippo Vergara Caffarelli (November 2009).

N. 735 – *Mutual guarantee institutions and small business finance*, by Francesco Columba, Leonardo Gambacorta and Paolo Emilio Mistrulli (November 2009).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2006

F. BUSETTI, *Tests of seasonal integration and cointegration in multivariate unobserved component models,* Journal of Applied Econometrics, Vol. 21, 4, pp. 419-438, **TD No. 476 (June 2003).**

C. BIANCOTTI, *A polarization of inequality? The distribution of national Gini coefficients 1970-1996*, Journal of Economic Inequality, Vol. 4, 1, pp. 1-32, **TD No. 487 (March 2004).**

L. CANNARI and S. CHIRI, *La bilancia dei pagamenti di parte corrente Nord-Sud (1998-2000),* in L. Cannari, F. Panetta (a cura di), Il sistema finanziario e il Mezzogiorno: squilibri strutturali e divari finanziari, Bari, Cacucci, **TD No. 490 (March 2004).**

M. BOFONDI and G. GOBBI, *Information barriers to entry into credit markets,* Review of Finance, Vol. 10, 1, pp. 39-67, **TD No. 509 (July 2004).**

W. FUCHS and LIPPI F., *Monetary union with voluntary participation,* Review of Economic Studies, Vol. 73, pp. 437-457 **TD No. 512 (July 2004).**

E. GAIOTTI and A. SECCHI, *Is there a cost channel of monetary transmission? An investigation into the pricing behaviour of 2000 firms*, Journal of Money, Credit and Banking, Vol. 38, 8, pp. 2013-2038 **TD No. 525 (December 2004).**

A. BRANDOLINI, P. CIPOLLONE and E. VIVIANO, *Does the ILO definition capture all unemployment?,* Journal of the European Economic Association, Vol. 4, 1, pp. 153-179, **TD No. 529 (December 2004).**

A. BRANDOLINI, L. CANNARI, G. D'ALESSIO and I. FAIELLA, *Household wealth distribution in Italy in the 1990s,* in E. N. Wolff (ed.) International Perspectives on Household Wealth, Cheltenham, Edward Elgar, **TD No. 530 (December 2004).**

P. DEL GIOVANE and R. SABBATINI, *Perceived and measured inflation after the launch of the Euro: Explaining the gap in Italy,* Giornale degli economisti e annali di economia, Vol. 65, 2 , pp. 155-192, **TD No. 532 (December 2004).**

M. CARUSO, *Monetary policy impulses, local output and the transmission mechanism*, Giornale degli economisti e annali di economia, Vol. 65, 1, pp. 1-30, **TD No. 537 (December 2004).**

L. GUISO and M. PAIELLA, *The role of risk aversion in predicting individual behavior, In P. A. Chiappori e C. Gollier (eds.) Competitive Failures in Insurance Markets: Theory and Policy Implications, Monaco, CESifo,* **TD No. 546 (February 2005).**

G. M. TOMAT, *Prices product differentiation and quality measurement: A comparison between hedonic and matched model methods*, Research in Economics, Vol. 60, 1, pp. 54-68, **TD No. 547 (February 2005).**

L. GUISO, M. PAIELLA and I. VISCO, *Do capital gains affect consumption? Estimates of wealth effects from Italian household's behavior*, in L. Klein (ed), Long Run Growth and Short Run Stabilization: Essays in Memory of Albert Ando (1929-2002), Cheltenham, Elgar, **TD No. 555 (June 2005).**

F. BUSETTI, S. FABIANI and A. HARVEY, *Convergence of prices and rates of inflation,* Oxford Bulletin of Economics and Statistics, Vol. 68, 1, pp. 863-878, **TD No. 575 (February 2006).**

M. CARUSO, *Stock market fluctuations and money demand in Italy, 1913 - 2003,* Economic Notes, Vol. 35, 1, pp. 1-47, **TD No. 576 (February 2006).**

R. BRONZINI and G. DE BLASIO, *Evaluating the impact of investment incentives: The case of Italy's Law 488/92.* Journal of Urban Economics, Vol. 60, 2, pp. 327-349, **TD No. 582 (March 2006).**

R. BRONZINI and G. DE BLASIO, *Una valutazione degli incentivi pubblici agli investimenti,* Rivista Italiana degli Economisti , Vol. 11, 3, pp. 331-362, **TD No. 582 (March 2006).**

A. DI CESARE, *Do market-based indicators anticipate rating agencies? Evidence for international banks,* Economic Notes, Vol. 35, pp. 121-150, **TD No. 593 (May 2006).**

R. GOLINELLI and S. MOMIGLIANO, *Real-time determinants of fiscal policies in the euro area*, Journal of Policy Modeling, Vol. 28, 9, pp. 943-964, **TD No. 609 (December 2006).**

2007

S. SIVIERO and D. TERLIZZESE, *Macroeconomic forecasting: Debunking a few old wives' tales*, Journal of Business Cycle Measurement and Analysis , v. 3, 3, pp. 287-316, **TD No. 395 (February 2001).**

S. MAGRI, *Italian households' debt: The participation to the debt market and the size of the loan*, Empirical Economics, v. 33, 3, pp. 401-426, **TD No. 454 (October 2002).**

L. CASOLARO. and G. GOBBI, *Information technology and productivity changes in the banking industry*, Economic Notes, Vol. 36, 1, pp. 43-76, **TD No. 489 (March 2004).**

G. FERRERO, *Monetary policy, learning and the speed of convergence*, Journal of Economic Dynamics and Control, v. 31, 9, pp. 3006-3041, **TD No. 499 (June 2004).**

M. PAIELLA, *Does wealth affect consumption? Evidence for Italy*, Journal of Macroeconomics, Vol. 29, 1, pp. 189-205, **TD No. 510 (July 2004).**

F. LIPPI. and S. NERI, *Information variables for monetary policy in a small structural model of the euro area*, Journal of Monetary Economics, Vol. 54, 4, pp. 1256-1270, **TD No. 511 (July 2004).**

A. ANZUINI and A. LEVY, *Monetary policy shocks in the new EU members: A VAR approach*, Applied Economics, Vol. 39, 9, pp. 1147-1161, **TD No. 514 (July 2004).**

D. JR. MARCHETTI and F. Nucci, *Pricing behavior and the response of hours to productivity shocks*, Journal of Money Credit and Banking, v. 39, 7, pp. 1587-1611, **TD No. 524 (December 2004).**

R. BRONZINI, *FDI Inflows, agglomeration and host country firms' size: Evidence from Italy*, Regional Studies, Vol. 41, 7, pp. 963-978, **TD No. 526 (December 2004).**

L. MONTEFORTE, *Aggregation bias in macro models: Does it matter for the euro area?*, Economic Modelling, 24, pp. 236-261, **TD No. 534 (December 2004).**

A. NOBILI, *Assessing the predictive power of financial spreads in the euro area: does parameters instability matter?*, Empirical Economics, Vol. 31, 1, pp. 177-195, **TD No. 544 (February 2005).**

A. DALMAZZO and G. DE BLASIO, *Production and consumption externalities of human capital: An empirical study for Italy*, Journal of Population Economics, Vol. 20, 2, pp. 359-382, **TD No. 554 (June 2005).**

M. BUGAMELLI and R. TEDESCHI, *Le strategie di prezzo delle imprese esportatrici italiane*, Politica Economica, v. 23, 3, pp. 321-350, **TD No. 563 (November 2005).**

L. GAMBACORTA and S. IANNOTTI, *Are there asymmetries in the response of bank interest rates to monetary shocks?*, Applied Economics, v. 39, 19, pp. 2503-2517, **TD No. 566 (November 2005).**

P. ANGELINI and F. LIPPI, *Did prices really soar after the euro cash changeover? Evidence from ATM withdrawals*, International Journal of Central Banking, Vol. 3, 4, pp. 1-22, **TD No. 581 (March 2006).**

A. LOCARNO, *Imperfect knowledge, adaptive learning and the bias against activist monetary policies*, International Journal of Central Banking, v. 3, 3, pp. 47-85, **TD No. 590 (May 2006).**

F. LOTTI and J. MARCUCCI, *Revisiting the empirical evidence on firms' money demand*, Journal of Economics and Business, Vol. 59, 1, pp. 51-73, **TD No. 595 (May 2006).**

P. CIPOLLONE and A. ROSOLIA, *Social interactions in high school: Lessons from an earthquake*, American Economic Review, Vol. 97, 3, pp. 948-965, **TD No. 596 (September 2006).**

L. DEDOLA and S. NERI, *What does a technology shock do? A VAR analysis with model-based sign restrictions*, Journal of Monetary Economics, Vol. 54, 2, pp. 512-549, **TD No. 607 (December 2006).**

F. VERGARA CAFFARELLI, *Merge and compete: strategic incentives for vertical integration*, Rivista di politica economica, v. 97, 9-10, serie 3, pp. 203-243, **TD No. 608 (December 2006).**

A. BRANDOLINI, *Measurement of income distribution in supranational entities: The case of the European Union*, in S. P. Jenkins e J. Micklewright (eds.), Inequality and Poverty Re-examined, Oxford, Oxford University Press, **TD No. 623 (April 2007).**

M. PAIELLA, *The foregone gains of incomplete portfolios*, Review of Financial Studies, Vol. 20, 5, pp. 1623-1646, **TD No. 625 (April 2007).**

K. BEHRENS, A. R. LAMORGESE, G.I.P. OTTAVIANO and T. TABUCHI, *Changes in transport and non transport costs: local vs. global impacts in a spatial network*, Regional Science and Urban Economics, Vol. 37, 6, pp. 625-648, **TD No. 628 (April 2007).**

M. BUGAMELLI, *Prezzi delle esportazioni, qualità dei prodotti e caratteristiche di impresa: analisi su un campione di imprese italiane*, v. 34, 3, pp. 71-103, Economia e Politica Industriale, **TD No. 634 (June 2007).**

G. ASCARI and T. ROPELE, *Optimal monetary policy under low trend inflation*, Journal of Monetary Economics, v. 54, 8, pp. 2568-2583, **TD No. 647 (November 2007).**

R. GIORDANO, S. MOMIGLIANO, S. NERI and R. PEROTTI, *The Effects of Fiscal Policy in Italy: Evidence from a VAR Model*, European Journal of Political Economy, Vol. 23, 3, pp. 707-733, **TD No. 656 (January 2008).**

B. ROFFIA and A. ZAGHINI, *Excess money growth and inflation dynamics*, International Finance, v. 10, 3, pp. 241-280, **TD No. 657 (January 2008).**

G. BARBIERI, P. CIPOLLONE and P. SESTITO, *Labour market for teachers: demographic characteristics and allocative mechanisms*, Giornale degli economisti e annali di economia, v. 66, 3, pp. 335-373, **TD No. 672 (June 2008).**

E. BREDA, R. CAPPARIELLO and R. ZIZZA, *Vertical specialisation in Europe: evidence from the import content of exports*, Rivista di politica economica, numero monografico,**TD No. 682 (August 2008).**

2008

P. ANGELINI, *Liquidity and announcement effects in the euro area*, Giornale degli Economisti e Annali di Economia, v. 67, 1, pp. 1-20, **TD No. 451 (October 2002).**

P. ANGELINI, P. DEL GIOVANE, S. SIVIERO and D. TERLIZZESE, *Monetary policy in a monetary union: What role for regional information?*, International Journal of Central Banking, v. 4, 3, pp. 1-28, **TD No. 457 (December 2002).**

F. SCHIVARDI and R. TORRINI, *Identifying the effects of firing restrictions through size-contingent Differences in regulation*, Labour Economics, v. 15, 3, pp. 482-511, **TD No. 504 (June 2004).**

L. GUISO and M. PAIELLA,, *Risk aversion, wealth and background risk*, Journal of the European Economic Association, v. 6, 6, pp. 1109-1150, **TD No. 483 (September 2003).**

C. BIANCOTTI, G. D'ALESSIO and A. NERI, *Measurement errors in the Bank of Italy's survey of household income and wealth*, Review of Income and Wealth, v. 54, 3, pp. 466-493, **TD No. 520 (October 2004).**

S. MOMIGLIANO, J. HENRY and P. HERNÁNDEZ DE COS, *The impact of government budget on prices: Evidence from macroeconometric models*, Journal of Policy Modelling, v. 30, 1, pp. 123-143 **TD No. 523 (October 2004).**

L. GAMBACORTA, *How do banks set interest rates?,* European Economic Review, v. 52, 5, pp. 792-819, **TD No. 542 (February 2005).**

P. ANGELINI and A. GENERALE, *On the evolution of firm size distributions*, American Economic Review, v. 98, 1, pp. 426-438, **TD No. 549 (June 2005).**

R. FELICI and M. PAGNINI, *Distance, bank heterogeneity and entry in local banking markets*, The Journal of Industrial Economics, v. 56, 3, pp. 500-534, **No. 557 (June 2005).**

S. DI ADDARIO and E. PATACCHINI, *Wages and the city. Evidence from Italy*, Labour Economics, v.15, 5, pp. 1040-1061, **TD No. 570 (January 2006).**

M. PERICOLI and M. TABOGA, *Canonical term-structure models with observable factors and the dynamics of bond risk premia,* Journal of Money, Credit and Banking, v. 40, 7, pp. 1471-88, **TD No. 580 (February 2006).**

E. VIVIANO, *Entry regulations and labour market outcomes. Evidence from the Italian retail trade sector*, Labour Economics, v. 15, 6, pp. 1200-1222, **TD No. 594 (May 2006).**

S. FEDERICO and G. A. MINERVA, *Outward FDI and local employment growth in Italy*, Review of World Economics, Journal of Money, Credit and Banking, v. 144, 2, pp. 295-324, **TD No. 613 (February 2007).**

F. BUSETTI and A. HARVEY, *Testing for trend*, Econometric Theory, v. 24, 1, pp. 72-87, **TD No. 614 (February 2007).**

V. CESTARI, P. DEL GIOVANE and C. ROSSI-ARNAUD, *Memory for prices and the Euro cash changeover: an analysis for cinema prices in Italy*, In P. Del Giovane e R. Sabbatini (eds.), The Euro Inflation and Consumers' Perceptions. Lessons from Italy, Berlin-Heidelberg, Springer, **TD No. 619 (February 2007).**

B. H. HALL, F. LOTTI and J. MAIRESSE, *Employment, innovation and productivity: evidence from Italian manufacturing microdata*, Industrial and Corporate Change, v. 17, 4, pp. 813-839, **TD No. 622 (April 2007).**

J. Sousa and A. Zaghini, *Monetary policy shocks in the Euro Area and global liquidity spillovers,* International Journal of Finance and Economics, v.13, 3, pp. 205-218, **TD No. 629 (June 2007).**

M. Del Gatto, Gianmarco I. P. Ottaviano and M. Pagnini, *Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms,* Journal of Regional Science, v. 48, 1, pp. 97-129, **TD No. 635 (June 2007).**

P. Del Giovane, S. Fabiani and R. Sabbatini, *What's behind "inflation perceptions"? A survey-based analysis of Italian consumers,* in P. Del Giovane e R. Sabbatini (eds.), *The Euro Inflation and Consumers' Perceptions. Lessons from Italy,* Berlin-Heidelberg, Springer, **TD No. 655 (January 2008).**

B. Bortolotti, and P. Pinotti, *Delayed privatization,* Public Choice, v. 136, 3-4, pp. 331-351, **TD No. 663 (April 2008).**

R. Bonci and F. Columba, *Monetary policy effects: New evidence from the Italian flow of funds,* Applied Economics , v. 40, 21, pp. 2803-2818, **TD No. 678 (June 2008).**

M. Cucculelli, and G. Micucci, *Family Succession and firm performance: evidence from Italian family firms,* Journal of Corporate Finance, v. 14, 1, pp. 17-31, **TD No. 680 (June 2008).**

A. Silvestrini and D. Veredas, *Temporal aggregation of univariate and multivariate time series models: a survey,* Journal of Economic Surveys, v. 22, 3, pp. 458-497, **TD No. 685 (August 2008).**

2009

F. Panetta, F. Schivardi and M. Shum, *Do mergers improve information? Evidence from the loan market*, Journal of Money, Credit, and Banking, v. 41, 4, pp. 673-709, **TD No. 521 (October 2004).**

P. Pagano and M. Pisani, *Risk-adjusted forecasts of oil prices*, The B.E. Journal of Macroeconomics, v. 9, 1, Article 24, **TD No. 585 (March 2006).**

M. Pericoli and M. Sbracia, *The CAPM and the risk appetite index: theoretical differences, empirical similarities, and implementation problems*, International Finance, v. 12, 2, pp. 123-150, **TD No. 586 (March 2006).**

S. Magri, *The financing of small innovative firms: the Italian case*, Economics of Innovation and New Technology, v. 18, 2, pp. 181-204, **TD No. 640 (September 2007).**

S. Magri, *The financing of small entrepreneurs in Italy*, Annals of Finance, v. 5, 3-4, pp. 397-419, **TD No. 640 (September 2007).**

F. Lorenzo, L. Monteforte and L. Sessa, *The general equilibrium effects of fiscal policy: estimates for the euro area*, Journal of Public Economics, v. 93, 3-4, pp. 559-585, **TD No. 652 (November 2007).**

R. Golinelli and S. Momigliano, *The Cyclical Reaction of Fiscal Policies in the Euro Area. A Critical Survey of Empirical Research*, Fiscal Studies, v. 30, 1, pp. 39-72, **TD No. 654 (January 2008).**

P. Del Giovane, S. Fabiani and R. Sabbatini, *What's behind "Inflation Perceptions"? A survey-based analysis of Italian consumers*, Giornale degli Economisti e Annali di Economia, v. 68, 1, pp. 25-52, **TD No. 655 (January 2008).**

F. Maccheroni, M. Marinacci, A. Rustichini and M. Taboga, *Portfolio selection with monotone mean-variance preferences,* Mathematical Finance, v. 19, 3, pp. 487-521, **TD No. 664 (April 2008).**

M. Affinito and M. Piazza, *What are borders made of? An analysis of barriers to European banking integration*, in P. Alessandrini, M. Fratianni and A. Zazzaro (eds.): The Changing Geography of Banking and Finance, Dordrecht Heidelberg London New York, Springer, **TD No. 666 (April 2008).**

L. Arciero, C. Biancotti, L. D'Aurizio and C. Impenna, *Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG*, Journal of Artificial Societies and Social Simulation, v. 12, 1, **TD No. 686 (August 2008).**

A. Calza and A. Zaghini, *Nonlinearities in the dynamics of the euro area demand for M1*, Macroeconomic Dynamics, v. 13, 1, pp. 1-19, **TD No. 690 (September 2008).**

L. Francesco and A. Secchi, *Technological change and the households' demand for currency*, Journal of Monetary Economics, v. 56, 2, pp. 222-230, **TD No. 697 (December 2008).**

M. Bugamelli, F. Schivardi and R. Zizza, *The euro and firm restructuring*, in A. Alesina e F. Giavazzi (eds): Europe and the Euro, Chicago, University of Chicago Press, **TD No. 716 (June 2009).**

B. Hall, F. Lotti and J. Mairesse, *Innovation and productivity in SMEs: empirical evidence for Italy*, Small Business Economics, v. 33, 1, pp. 13-33, **TD No. 718 (June 2009).**

FORTHCOMING

L. MONTEFORTE and S. SIVIERO, *The Economic Consequences of Euro Area Modelling Shortcuts*, Applied Economics, **TD No. 458 (December 2002).**

M. BUGAMELLI and A. ROSOLIA, *Produttività e concorrenza estera*, Rivista di politica economica, **TD No. 578 (February 2006).**

G. DE BLASIO and G. NUZZO, *Historical traditions of civicness and local economic development*, Journal of Regional Science, **TD No. 591 (May 2006).**

R. BRONZINI and P. PISELLI, *Determinants of long-run regional productivity with geographical spillovers: the role of R&D, human capital and public infrastructure,* Regional Science and Urban Economics, **TD No. 597 (September 2006).**

E. IOSSA and G. PALUMBO, *Over-optimism and lender liability in the consumer credit market,* Oxford Economic Papers, **TD No. 598 (September 2006).**

U. ALBERTAZZI and L. GAMBACORTA, *Bank profitability and the business cycle,* Journal of Financial Stability, **TD No. 601 (September 2006).**

A. CIARLONE, P. PISELLI and G. TREBESCHI, *Emerging Markets' Spreads and Global Financial Conditions*, Journal of International Financial Markets, Institutions & Money, **TD No. 637 (June 2007).**

V. DI GIACINTO and G. MICUCCI, *The producer service sector in Italy: long-term growth and its local determinants,* Spatial Economic Analysis, **TD No. 643 (September 2007).**

Y. ALTUNBAS, L. GAMBACORTA and D. MARQUÉS, *Securitisation and the bank lending channel,* European Economic Review, **TD No. 653 (November 2007).**

F. BALASSONE, F. MAURA and S. ZOTTERI, *Cyclical asymmetry in fiscal variables in the EU*, Empirica, **TD No. 671 (June 2008).**

M. BUGAMELLI and F. PATERNÒ, *Output growth volatility and remittances*, Economica, **TD No. 673 (June 2008).**

M. IACOVIELLO and S. NERI, *Housing market spillovers: evidence from an estimated DSGE model,* American Economic Journal: Macroeconomics, **TD No. 659 (January 2008).**

A. ACCETTURO, *Agglomeration and growth: the effects of commuting costs,* Papers in Regional Science, **TD No. 688 (September 2008).**

L. FORNI, A. GERALI and M. PISANI, *Macroeconomic effects of greater competition in the service sector: the case of Italy,* Macroeconomic Dynamics, **TD No. 706 (March 2009).**

Y. ALTUNBAS, L. GAMBACORTA, and D. MARQUÉS-IBÁÑEZ, *Bank risk and monetary policy,* Journal of Financial Stability, **TD No. 712 (May 2009).**