BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

Measurement errors in consumption surveys and the estimation of poverty and inequality indices

by Giovanni D'Alessio

Number 1116

June 2017

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Measurement errors in consumption surveys and the estimation of poverty and inequality indices

by Giovanni D'Alessio

Number 1116 - June 2017

ISSN 1594-7939 (print)
ISSN 2281-3950 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

MEASUREMENT ERRORS IN CONSUMPTION SURVEYS AND THE ESTIMATION OF POVERTY AND INEQUALITY INDICES

by Giovanni D'Alessio*

Abstract

This paper firstly aims to evaluate the incidence of measurement error affecting the main variables collected in surveys on consumption. The assessment is carried out on two Tanzania surveys which provide both diary and panel data. Diary data can be employed to obtain reliability coefficients for time-invariant variables. When variables vary over time, as in the case of panel data, an estimation of the incidence of measurement error on the total variance can be obtained by applying models which allow the decomposition of observed variability into true dynamics and noise (e.g. the Heise model and the latent Markov model). Some evaluations of the reliability of the data are also conducted on the basis of the internal consistency criterion, an approach that does not require panel data. On the basis of the reliability estimates obtained, examples of possible impacts of measurement errors on poverty analysis are briefly discussed. These experiments clearly show the importance of the topic in poverty and inequality data analysis.

JEL Classification: D31, I32.
Keywords: inequality, poverty, survey data, measurement errors, reliability.

Contents

---

* Bank of Italy, Directorate General for Economics, Statistics and Research.

1. Introduction[1]

Among the possible sources of error affecting survey data, the discrepancy between recorded and 'true' data, which originates from the response or from oversights in the processing phase prior to estimation, is of particular interest although seldom analysed.

The survey design, in all its parts, may have an impact on survey responses. In some cases questions can be asked ambiguously or face limitations due to the cognitive processes of the respondent: people may not actually know the exact answer to the questions they are asked, especially in cases where response by proxy is allowed, and answers on quantities tend to be rounded up or down. Moreover, retrospective questions mean recalling events of the past while hypothetical ones require some abstract reasoning which may generate uncertain answers. All the above aspects may interact among themselves and with other factors affecting the quality of survey data. Interviewer behaviour, for example, can be very important: there are a number of ways of asking the same question in a face-to-face setting, and each can induce a different psychological reaction, ultimately affecting the answer. More general aspects such as the motivation of respondents and their willingness to give their time and effort to a survey should also be assumed to influence data quality. Finally, the use of the tools employed in the CAPI programmes for improving the consistency among the answers and outlier detection may be relevant (Converse and Presser, 1986, Tourangeau et al., 2000, Grosh and Glewwe, 2000). As to consumption surveys, specific factors may be at work (diary/recall, usual/specific month, 7/14 days of recall, long/short item list, and so on) and may determine significant variations in survey results (Friedman et al. 2016).[2]

In what follows, we will not deal with the various causes of error, taken one by one. We will focus instead on the tools enabling the assessment of the magnitude of measurement errors in the main variables collected in consumption survey data. This perspective offers useful information to data users, whose estimates can be severely biased when measurement errors are not taken into account, and to data producers, who may find a tool for discussing and improving the data collection process in these quantifications.

In regression analysis it is well known that the presence of classical measurement errors[3] in the explanatory variables leads to biased and inconsistent OLS estimators; in simple regression and correlation analysis the bias assumes the form of attenuation bias, i.e. a tendency towards zero. In such cases, instrumental variables are a common tool for obtaining unbiased and consistent estimates (Chen et al., 2007).

Measurement errors also affect the estimates of mobility in panel data. Either one looks at mobility tables describing the transitions from one state to another of a sample

[1] This paper was prepared during my stay as visiting scholar at the Survey and Methods Unit (DECSM), Development Data Group, the World Bank. I would like to thank G. Carletto, A. Zezza, D. Jolliffe, L.Cannari and two anonymous referees for their useful insights. The opinions expressed in this article are only mine and do not necessarily represent the views of the World Bank or its affiliated organizations or those of the Bank of Italy.

[2] In particular Beegle et al. (2012) conduct a field experiment in Tanzania to test various alternative methods of measuring household consumption, finding significant differences among the consumption values reported.

[3] A measurement error is defined as classical when the error term has a zero mean, finite variance and is uncorrelated with other variables.

observed in two consecutive waves or one tries to estimate the growth of a variable observed against the initial value, (classical) measurement errors tend to overstate the actual mobility.

Measurement errors are also supposed to have an impact on inequality and poverty measures. According to classical hypotheses, errors tend to inflate the variance and the tails of a distribution,[4] thus boosting inequality and poverty indices.

In what follows we will show how measures of reliability can be estimated using survey data, and will discuss some typical drivers of measurement errors in consumption surveys. The reliability of a measure denotes the variability of the estimates over repeated trials and in the same approximate conditions. It is different from the accuracy of a measure, which implies both a small variability of estimates and a closeness to the true value (Hand et al., 2001).

This approach does not require the availability of true data to make a comparison with survey data[5] although it confines the analysis to those cases where a hypothesis of classical measurement errors holds, at least approximately. In this regard it is important to consider that voluntary under-reporting, which is one of the main drivers of non-classical measurement errors, is usually much less significant for consumption than for income and wealth. Cannari and D'Alessio (1993) and Gottschalk and Huynh (2010) evaluate the effects of measurement errors on the distribution of earnings and financial wealth respectively by comparing survey data with a benchmark containing 'true data' and find that survey data actually underestimate inequality. Similar results are obtained by D'Alessio and Neri (2015) who adopt a completely different approach, based on calibration techniques. Cifaldi and Neri (2013), on analysing the reporting behaviour of Italian households, find that misreporting of consumption has a different association with the reported amounts than with income; while under-reporting increases with declared income, there is no similar evidence for consumption. The explanation may be twofold: on the one hand, consumption is a less sensitive topic than income, because fiscal authorities are not interested in such amounts, on the other hand, consumption is more difficult to hide from an interviewer in a face-to-face interview.

For the users of survey data, there are many reasons to worry about the amount, the origin and the expected effects on estimates of measurement errors affecting data, as well as to look for techniques able to take them into account. On the other hand, the relevance of the topic should also push data providers to minimize as much as possible the measurement errors they can control, by complementing microdata with more information on the magnitude and expected effects of such errors on the estimates.[6]

The possible impacts of measurement errors on poverty and inequality analyses are briefly discussed. The exercise is conducted using data from two surveys conducted in Tanzania, but the tools proposed are quite general and the discussion could easily be extended to countries conducting similar surveys.

---

[4] Classical measurement errors also affect the standard errors of estimators, widening their intervals of confidence.

[5] In practice, the availability of true data on income and wealth is a rare event. Administrative data can also be considered contaminated by some form of measurement error, once tax evasion and elusion have been considered.

[6] The interest of data producers to contain measurement errors is of course part of a more general strategy for lowering all potential errors affecting estimates as much as possible.

The paper is structured as follows. Section 2 briefly describes the data used in the paper, with a special focus on the aspects relating to data quality. Section 3 shows the statistical tools that we use for evaluating the degree of reliability of collected data. Section 4 presents some descriptive statistics on measurement error for the two surveys. Section 5 briefly concludes.

2. Statistical tools for estimating reliability in consumption surveys

2.1 What is reliability?

Let X be a continuous variable measured with an error, so that the measure Y differs from the true value X by a random component: Y = X + e. If the disturbance has the following properties (called *homoscedastic, uncorrelated error*):

$E(e) = 0$; $E(X, e) = \sigma_{X,e} = 0$; $E(e, e) = \sigma^2_e$

the variance of Y may be written as $\sigma^2_Y = \sigma^2_X + \sigma^2_e = \sigma^2_X / \lambda^2$ where $\lambda = \dfrac{\sigma_X}{\sigma_Y}$.

Under the above conditions, the coefficient λ is known as the *reliability index*; it expresses the share of variability in Y which belongs to the true phenomenon X (Lord and Novick, 1968), as opposed to the part due to the factors that contaminate the measurement process.[7]

The estimation of λ, as previously defined, would require knowledge of Y, which can seldom be assumed. An easy way of obtaining an estimate of the reliability index is to resort to the test-retest, which involves a double measurement of X in the same (approximate) survey conditions.

Let Y_1 and Y_2 be such measurements, $Y_1 = X + e_1$ and $Y_2 = X + e_2$ for which we also assume that e_1, and e_2 are uncorrelated between themselves and with X, i.e. $E(e_1,e_2) = \sigma_{e1,e2} = 0$ and $E(X_t, e_{t'}) = \sigma_{Xt,et'} = 0\ \forall t,t'$. Under these conditions, the correlation coefficient between the two measurements Y_1 and Y_2 equals the square of the reliability index:

$\rho_{y1,y2} = \sigma_{y1,y2} / \sigma_{y1}\sigma_{y2} = \sigma^2_x / (\sigma^2_x + \sigma^2_e) = \sigma^2_x / \sigma^2_y = \lambda^2$

It is worth noting that if the assumption $E(e) = 0$ does not hold, where $E(e) = \delta$, as may be the case in a particular survey design (i.e. the choice of the usual month consumption), the index λ only captures the variability of the two repeated Ys and not its closeness to X (which is unknown). This means that the reliability index measured in this way evaluates the degree to which an instrument provides consistent measures; it does not indicate the instrument's truthfulness as, at least generally speaking, it cannot assess the distance between collected and true data.

If we are dealing with categorical variables, the test-retest model needs to be revised (Biemer and Trewin, 1997). Let X be a categorical variable (with K categories) and Y its measurement. A reliability index for categorical features measured twice (Y_1 and Y_2) on the same set of n units is the fraction of units that are classified consistently: $\lambda^* = tr(F)/n = \Sigma_i f_{ii}/n$ where F is the cross tabulation of Y_1 and Y_2 whose generic element is f_{ij}.

[7] For a review of reliability analysis, see Webb, Shavelson and Haertel (2006).

The index λ^*, however, does not take into account the fact that consistent answers could be partly random: if the two measures Y_1 and Y_2 are independent random variables, the expected share of consistent units is $\Sigma_I f_{i.} f_{.i}/n$. It is thus preferable to resort to Cohen's kappa coefficient κ, a reliability index which controls for this effect by normalizing the share of observed matching cases with respect to their expected incidence if the two measurements of Y_1 and Y_2 are independent: $\kappa = (\lambda^* - \Sigma_i f_{i.} f_{.i}/n) / (1- \Sigma_i f_{i.} f_{.i}/n)$.[8]

2.2 Reliability within a single survey: diary data as repeated measures

The estimation of a reliability coefficient using data captured in a single survey is not an easy task. It may be unpleasant to ask a question more than once in the same survey (e.g. 'How much rent did you pay last month?') and even if one did it is likely that respondents tend to provide coherent answers, leading to an overestimation of reliability. In fact, the test-retest formula of λ relies on the assumption of uncorrelated errors; this assumption may be violated if the respondents realize that they have already answered the same question. Only a few studies provide reliability coefficients based on a double measurement in a single survey (e.g. Crossley and Kennedy, 2002).

In consumption surveys, however, the collection of data is often done by means of a diary, which can be often organized in a way that allows the computation of the reliability coefficient on repeated measures over time. For example, if the diary is held by households over two weeks, it is possible to assume that, for the i-th generic household, the consumption of a good over a certain period Y_1 (say those occurring during the first week, or on odd-numbered days) is a random variable with the same mean and variance of the consumption measured in a second period Y_2 (over the second week, or on even-numbered days); most of the time the assumption of uncorrelated errors may reasonably hold.[9]

In such a case, where $Y_{i1}, = X_{i1} + e_{i1}$; $Y_{i2}, = X_i + e_{i2}$; with $E(e_i)=0$; $E(X_i,e_i)=0$; $E(e_i,e_i)=\sigma^2_e$, one can simply estimate the reliability of the weekly consumption as the correlation coefficient between the two measures. Moreover, according to the hypotheses described above, any other rearrangements of the daily measures into two halves may be used to compute the correlation coefficient (split-half approach) expression of reliability.[10]

In order to obtain an estimate of the reliability of the sum of the two weeks' consumption (Y_1+Y_2) starting from a measure of consistency between the two halves, we use the Spearman-Brown formula:[11] $\lambda = [2\ (\rho_{y1,y2})] / [1 + (\rho_{y1,y2})]$. If there are more than 2 periods, the generalized formula applies: $\lambda = [n\ (\rho_{y1,y2})] / [1 + (n-1)\ (\rho_{y1,y2})]$.

Table 1 shows that a reliability index of 0.6 obtained for the weekly consumption corresponds to a reliability of 0.75 for the 2-week estimate and of 0.86 for the 4-week estimate. As the weekly reliability increases, the gain obtained by extending the period for

---

[8] Both the indices λ^* and κ can be adopted to assess the reliability of single categories of qualitative variables, computing them on the dummy variables by opposing each category to all the others (Biancotti et al., 2008). This can help in understanding where the main classification problems lie.

[9] We will show below that also where the assumption does not hold, information on individual variations over time can be very important in the poverty and inequality analysis.

[10] It is worth noting that the reliability coefficient measured in this way is not affected by a change over time in the average value of Y, as may happen in the case of a uniform fatigue effect across units. In such a case, the reliability index cannot account for the bias but still measures the variance across units correctly.

[11] Brown, 1910 and Spearman, 1910. Alternative estimators of reliability in the split-half scheme are found in Rulon (1939) and Guttman (1945).

which the diary is kept reduces. This kind of information can help in evaluating the trade-off between a more stable estimation due to a longer diary data collection and the higher costs associated with such a choice.

It is important to bear in mind that the estimation of reliability as described above implies the independence of measurements over time (i.e. between the two periods considered). If this is not the case, as for example when the purchasing frequency is low and one purchase in a day implies a reduced or even a zero value in the contiguous days, the reliability coefficients are underestimated by the exposed procedure, because of the presence of correlated errors.

Table 1 - Reliability of repeated measures (Spearman-Brown formula)

Number of repeated measures						
1	2	3	4	12	26	52
0.10	0.18	0.25	0.31	0.57	0.74	0.85
0.20	0.33	0.43	0.50	0.75	0.87	0.93
0.30	0.46	0.56	0.63	0.84	0.92	0.96
0.40	0.57	0.67	0.73	0.89	0.94	0.97
0.50	0.67	0.75	0.80	0.92	0.96	0.98
0.60	0.75	0.82	0.86	0.95	0.97	0.98
0.70	0.82	0.88	0.90	0.97	0.98	0.99
0.80	0.89	0.92	0.94	0.98	0.99	0.99
0.90	0.95	0.96	0.97	0.99	0.99	0.99
1.00	1.00	1.00	1.00	1.00	1.000	1.000

In general, the intertemporal variance of consumption survey data may be influenced by a purchasing frequency inadequate to the length of the observation window (see Gibson and Kim, 2011). For example, in a 2-week diary, households found with zero expenditure on clothes are presumably those who will have some clothing expenses during the rest of the year (and these amounts will probably be offset on average by other households for which positive expenses have been found during the observation window). If the window is too brief (compared with the purchasing frequency) the individual data will provide an unbiased estimate of the individual means but with a consistent standard error. Strictly speaking, this is not a matter of reliability, although it has similar effects on estimates.

Consumption surveys usually have a yearly reference period (S) while the estimates are often obtained on the basis of shorter periods of observation/recall (w), i.e. a 1- or 2-week diary or 1-3 months of information (i.e. for household bills), which are scaled up by means of an expansion factor (S/w). This practice does not take into account that while the average is not affected by the length of the collection period (w) over a homogeneous period, the variance of the reconstructed yearly consumption may depend on it (Deaton and Grosh, 1997).[12] With regard to this, the practice of spreading the full sample over the entire year with independent subsamples may be useful in the estimation of the mean (if it cannot be assumed to be constant over time) but does not help in any way in the estimation of the variance (or of the poverty and inequality indices).

Let us consider the case of a consumption of a good observed over a semester Y_1, whose amount is doubled to obtain a yearly estimate $Y'=2*Y_1$. In such a case the variance of this estimator Y' is simply $Var(Y') = 4\ \sigma_1^2$. If we observe both the semesters and derive the

[12] Clarke et al. (2008) show how an optimal trade-off between the higher precision characterizing short periods of recall and the greater stability affecting a wider range of observation can be determined.

yearly estimate by summing up these two components, $Y'' = Y_1 + Y_2$, under the hypothesis of an equal variance of the two semesters (i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$), we can derive the variance of Y'' as: $Var(Y'') = 2\,\sigma^2\,(1+\rho)$.

By comparing the two expressions we have that $Var(Y') \geq Var(Y'')$, where the equality holds only if the two components are perfectly correlated (i.e. $\rho=1$). Given the constraint $\sigma_1^2=\sigma_2^2=\sigma^2$, a perfect correlation implies the equality of the two components $Y_1 = Y_2$. In other words, the variance of the extrapolated estimate Y' is always greater than that which would be obtained by collecting data over the whole period, and will be equal only if there is no variation over time (i.e. $Y_1 = Y_2$).

The above result obviously holds even in the case of monthly (or weekly) estimates. For example, if one considers the usual estimator $Y'=12*Y_1$ which multiplies the amount observed over a month by 12, under the assumption of equal monthly variance, the variance is $Var\ (Y')=12^2\ \sigma^2$. The variance of the sum of the 12 monthly components $Y''=\Sigma_t Y_t$ ($t=1,\ldots,12$), can be written instead as: $Var\ (Y'')=12\ \sigma^2+2\sigma^2\ \Sigma_{j<k}\ \rho_{jk}$ where ρ_{jk} are all the $T(T-1)/2=66$ different pairwise correlation coefficients between the monthly measures. Again we conclude that the variance of the yearly estimator Y', obtained by extrapolating a single observed monthly (or weekly) measure, is always greater than or equal to the variance of the estimator Y'' obtained by adding up all the components, being equal only if the measures are all equal among themselves (Gibson, Huang, Rozelle, 2003; Gibson 2016).[13]

The conclusion is important for poverty and inequality analysis: if we can assume a stability over time in the variance of a phenomenon, an assumption that is reasonable most of the time, all other things being equal, inequality and poverty measures tend to have an upward bias both when the reliability of the measures is not perfect and when we use reduced observation/recall windows, in the presence of intertemporal variability.[14] [15] It is particularly important that even if the variations over time are not due to measurement errors, they produce the same effect when the extrapolation strategy is applied, thus inflating the variance (and the poverty rates and inequality measures) of the phenomenon.

A corollary of the above statement is that, except for the unrealistic case of completely reliable and stable variables, inequality or poverty indices derived from consumption surveys with different observed periods are not immediately comparable, and would require some adjustment to take into account the abovementioned bias.

2.3 Reliability within a single survey: internal consistency

Apart from in diary data, it is rare to find repeated measures in the same consumption survey. Nonetheless it is sometimes possible to assume that a set of variables is the expression of a unique latent variable. One could assume, for example, that the components

---

[13] On the relationship between poverty measurement and the variability of economic outcomes within a year, see also Jolliffe and Serajuddin (2015).

[14] Following Istat (2016) which describes in detail the new methodology employed in the Italian Household Budget Survey and measures the impact of the changes on the estimates, the widening of the reference period for the consumption data collected in diaries has significantly contributed to lowering the relative poverty ratio.

[15] As we have already said, we have not considered other possible effects on estimates attributable to the length of the time period, such as the decrease in reporting due to a fatigue effect which affects average values.

of household consumption describe behaviour which should have some internal coherence; marked deviations from this scheme could be an indicator of potential problems in the data.[16]

Under this assumption, some descriptive statistics drawn from this field of analysis can be useful to derive information on the reliability of consumption variables. In the following we will discuss some indicators, such as the correlation coefficients between each consumption components and the sum of all the other components, or the correlation coefficient between the consumption components and the values predicted by all the other components.

A different implementation of the same rationale is based on the use of Principal Component Analysis (PCA), a tool widely employed in denoising data. Following the singular value decomposition (Eckart and Young, 1936), we know that every matrix X with n observations and p variables can be fully decomposed on the base of eigenvalues φ_m and eigenvectors u_m and v_m of the corresponding quadratic forms X'X and XX' respectively (with common non-zero eigenvalues φ_m but different eigenvectors u_m and v_m):

$$X = \Sigma_m \varphi_m u_m v_m' \qquad (m=1,\ldots,p)$$

Considering just the first k principal components as relevant (i.e. those corresponding to the highest eigenvalues) leads to a decomposition of X in one matrix of signal (X*) and one of noise (E), whose information can be discarded:

$$X = \Sigma_j \varphi_j u_j v_j' + E = X^* + E \qquad (j=1,\ldots,k<p)$$

The first k principal components are the linear combinations of the original variables maximizing the variance, under the orthogonality constraint; they account for the maximum share of the global variance, expressed by the ratio $\tau=\Sigma_j \varphi_j /\Sigma_m \mathrm{Var}(x_m)$. However, τ is an average measure, as not all the variables are approximated in the same way. In this framework, the ratio of the standard deviation of each variable as approximated (by means of a linear prediction) by the first k principal components to its original standard deviation may be assumed as a measure of reliability. In geometrical terms, the reliability of each variable can be seen as the ratio of the length of the vector projected onto the optimal (in terms of explained variance) subspace of the first k principal components and the length of the same vector in the full p-dimensional space.

The choice of the number of principal components to retain is crucial. Sometimes, the 'eigenvalue one' rule of thumb is applied, which implies the retention of all the principal components whose variance is higher than that of the original (standardized) variables. In other cases, an analysis of the plot of the eigenvalues can help, for example when it shows a clear drop in the explanatory power of the principal components. In some cases, information on the possible magnitude attributable to noise obtained by means of methods like those shown in this paragraph can help with this task. From a practical point of view, as there is not always a unique and clear solution, a sensitivity analysis with various numbers of principal components is advised.[17]

[16] This is the general framework in which Cronbach's alpha coefficient of internal consistency is usually computed (Cronbach, 1951 and 2004).

[17] For a discussion on this topic see Gavish and Donoho, 2014.

2.4 Reliability of panel data

In panel surveys households are generally interviewed with a sufficient time lag to avoid any contamination of the first interview on the subsequent answers; for all the variables common to the waves, for which no changes may reasonably have occurred from one wave to another (i.e. time-invariant), a quantification of measurement error can be obtained by applying the test-retest formula.[18]

For time-varying variables, which are the majority of variables collected in consumption surveys, the analysis of measurement errors requires more sophisticated instruments because the quantities vary with time, and it is necessary to define models to distinguish actual change from movements induced by wrong measurements.

A method for estimating reliability indexes using longitudinal data is provided by the simplex model (Heise, 1969; Alwin, 2007), within the more general framework of Structural Equation Modelling (SEM). The reliability of data on time-varying quantities can be assessed by means of the simplex model, provided that at least three separate measurements of the variable on the same panel units are available; the separation of real dynamics from measurement error is obtained under mild regularity conditions (Biemer et al.,2009).

A special case is the Heise method (Heise, 1969), which hypothesizes that the 3 variables X_1, X_2 and X_3 are measured by Y_1, Y_2, and Y_3 respectively, $Y_t = X_t + e_t \; \forall t$, with homoscedastic, uncorrelated error.

X_1, X_2 and X_3 are assumed to be pairwise related through independent, first-order autoregressive models, which do not need to be stationary:

$X_1 = \delta_1$; $\qquad X_2 = \beta_{21} X_1 + \delta_2$; $\qquad X_3 = \beta_{32} X_2 + \delta_3$

where $\beta_{t+1,t}$ is the autoregressive coefficient and δ_t is the process innovation. Innovations are uncorrelated pairwise.

Assuming a constant reliability across the measures, the correlation coefficient between the observed values Y_t and Y_{t+1} can be written as $\rho_{Yt,Yt+1} = \lambda_Y{}^2 \rho_{Xt,Xt+1}$, that is the correlation between X_t and X_{t+1} is attenuated by measurement errors both on Y_t and Y_{t+1}.

In such a case, the estimation of λ - assumed to be constant over the 3 waves - is obtained by means of the ratio of simple correlation coefficients:[19]

$$\lambda_Y = \sqrt{\frac{\rho_{Yt-1,Yt}\,\rho_{Yt,Yt+1}}{\rho_{Yt-1,Yt+1}}}$$

Under a first-order autoregressive assumption AR1, the above ratio should be equal to one if the variables are perfectly measured; when measurement errors are present the ratio tends to decrease correspondingly.[20]

---

[18] Biancotti et al. (2008) use Italian data to estimate the reliability of the variable measuring the floor area of residential dwellings, having selected the subsample of those households who didn't move or incur extraordinary renovation expenses between the two survey waves. The reliability coefficient is λ=0.80.

[19] In the example provided by Biemer et al. (2009), the Heise measure is approximately the average of the measures obtained over the single waves with alternative stationarity assumptions needed to identify the model.

[20] As observed by Biancotti et al. (2008), the Heise index measured under the AR1 hypothesis tends to be a downwards-biased estimate of the reliability value if data actually follow an AR2 process.

It is worth noting that the parameters of the autoregressive model do not need to be stationary, i.e. they may vary from one change to the next. What is supposed to be constant is the amount of measurement errors, an assumption that may be reasonably made in surveys conducted on a regular basis with unchanged collection procedures.

3. The data

Two main data sources have been considered in the paper: the Tanzania National Household Budget Survey (TNHBS) and the Tanzania National Panel Survey (TNPS).

As to TNHBS, in the paper we use the 28-day diary data from the 2011-2012 wave, conducted on a sample of 10,186 households with completed interviews drawn from the 2002 Population and Housing Census frame. A stratified multi-stage sample design was used for this survey. At the first stage the primary sampling units (PSUs) selected 400 enumeration areas (EAs). At the second stage the EAs had an average of 133 households each, (155 for rural EAs and 94 for urban EAs). As some households were observed for longer than a month, only information concerning the first 4 weeks was retained in the analysis. In the paper estimates, sampling weights are used.

The Tanzania National Panel Survey (TNPS) is a survey conducted on a regular basis by the National Bureau of Statistics and the Ministry of Finance. The original sample, designed to be representative of the national, urban/rural, and major agro-ecological zones, consisted of about 3,200 households in the first 2008-2009 wave. The sample households were clustered in 409 EAs across Tanzania and Zanzibar.

In the second wave (2010-2011) the sample included the originally sampled households plus split-off households, while in the third wave (2012-2013) all the households interviewed during the previous two waves were contacted for the interview. Thus the total sample of the last two waves is greater than that of the first wave (almost 4,000 units).

As the purpose of our analysis is to estimate the reliability of consumption measures, we have built our models only considering the approximately 1,000 households who did not change their composition across the 3 waves. In this way, the models accounting for changes over time can remain simple and deviations from the model can be attributed to measurement errors.

The attrition rate between the 2010/2011 and 2012/2013 waves was quite low, at around 3.5 percent for households and 7.5 percent for individuals.

4. The reliability of Tanzania consumption data

4.1 Reliability of diary data (TNHBS)

In order to assess the reliability of diary data collected by the TNHBS we have grouped household expenses according to the week in which they occurred (1 to 4) and to the COICOP (Classification Of Individual COnsumption by Purpose) codes.

For every group of goods and services, the correlation of weekly and bi-weekly household expenses has been computed (Table 2). As the diaries include 4 weeks, the average of the 6 weekly and the 3 bi-weekly correlations have been computed in order to summarize the results. Moreover, following the Spearman-Brown formula described above, the estimated reliability of the 4-week amount is presented.

In the data analysis, it is important to take into account that diary data are only a part of household consumption/expenditure, and that the share accounted for by the diary data may vary with the type of goods and services considered. For example, while food consumption items are fully collected in the diary, housing diary expenditures do not include the monthly (actual or imputed) rents for the house of residence and for other houses held as well as many other housing expenses collected by the questionnaire with reference to the last month (expenses for electricity, water and sewage services, waste collection and so on) or to the last three months (gas in cylinders, charcoal, kerosene, coal and firewood). Analogously, fixed and mobile telephone bill and Internet subscriptions are not included in the communication expenses of the diary nor is the TV licence included in the recreation and culture expenses.

Conscious of these limitations, in the following we will discuss the reliability of the diary data only, which do not fully represent the entire category except for food and beverages expenses.

Food and non-alcoholic beverages have an average weekly correlation of 0.5, which according to Spearman-Brown formula implies an estimated reliability for the corresponding 4-week diary figures of around 0.8. Both transport and communication have a slightly lower 4-week reliability, of around 0.75, while alcoholic beverages show a reliability of around 0.7 and clothing and footwear a reliability of around 0.6. All the other figures are lower, in some cases as a clear effect of a typically low purchasing frequency (i.e. furnishings). For these latter items, the estimates of reliability – intended as the closeness of collected data to real values - are likely to be biased downwards; nonetheless, the low correlations signal instability over time which may add undue variance to consumption estimates.[21]

It is worth noting that even if most of the expense items are complemented with other components from outside the diary, the limited reliability of some shares implies that additional variance is added to final estimates. Moreover, the collection of data for components outside the diary, as for example the expenses for electricity, may also add variance to the total expenditure estimate, as they are collected on a last-month (or 3-month) basis and expanded to the year without taking measurement errors into account the.

Values for the average bi-weekly correlations that are consistently higher than that of the average weekly correlations signal a tendency to obtain more stable estimates as the diary period is extended.

If we look at data collected using diaries as panel data, we can also estimate reliability indexes following the Heise model, allowing for some true variation over time on the base of an AR1 model. The estimates obtained on the basis of the correlations observed between the expenditures over both the first 3 weeks and those of the last 3 weeks largely agree (Table 2).

[21] The computation of average correlations is based on the assumption of an equal reliability of weekly expenses. In our data, some descriptive analyses seem to suggest that this might not be entirely the case. For example, in 7 out of 11 types of goods and services considered, the correlation coefficients between the weekly expenses tend to increase, moving from the first to the second week and decreasing thereafter. The deviations are often not so important as to seriously affect our discussion based on an average measure; however they could reflect both an initial learning effect in compiling the diary and a subsequent fatigue effect.

Table 2 - Reliability of diary aggregates

Consumption aggregates	Average weekly correlation	Average bi-weekly correlation	Estimate of reliability of 4 weeks' expenditures *	Estimate of reliability of 4 weeks' expenditures **	Heise model reliability coefficients – (weeks 1 to 3)	Heise model reliability coefficients – (weeks 2 to 4)
Food and non-alcoholic beverages	0.500	0.614	0.800	0.761	0.867	0.812
Alcoholic beverages	0.374	0.564	0.705	0.721	0.698	0.585
Clothing and footwear	0.292	0.459	0.622	0.629	0.708	0.478
Housing, water, electricity, gas	0.236	0.370	0.553	0.540	0.874	0.760
Furnishings	0.078	0.146	0.252	0.256	0.214	0.304
Health	0.100	0.183	0.308	0.309	0.367	0.590
Transport	0.480	0.638	0.787	0.779	0.798	0.758
Communication	0.455	0.623	0.770	0.768	0.705	0.673
Recreation and culture	0.049	0.082	0.172	0.152	0.514	0.343
Education	0.004	0.006	0.016	0.011	0.383	0.015
Other goods and services	0.233	0.381	0.549	0.552	0.423	0.434
Total expenditures	0.532	0.671	0.820	0.803	0.896	0.828

* Obtained by applying the Spearman-Brown formula shown in the text to the average weekly correlation. ** Obtained by applying the Spearman-Brown formula shown in the text to the average bi-weekly correlation.

4.2 Reliability of panel data (TNP)

Table 3 shows the reliability coefficients computed in different ways for twelve main components of total household consumption collected in the TNP.

The first column refers to the coefficient obtained by applying the Heise model to household consumption components. As these estimates, as well as the other estimates considered in the table, are computed by evaluating the heterogeneity of the answers provided by panel households over time, only the approximately 1,000 households who did not change their composition were considered. As a robustness check, the second column shows the reliability coefficients computed on the ranks, which are less influenced by outliers. The following two columns refer to the corresponding estimates obtained for the equivalent household consumption, while the remaining two are the averages of the yearly coefficients of the SEM coefficients linking the observed to latent variables (computed for household consumption only) under the hypotheses of equal variance of error components and of variable components respectively.

The results show quite a satisfactory reliability of total consumption, with an average of the various estimates just below 0.9 both for nominal and real figures. In other words, 90 per cent of the variability of these indicators is consistent between the measures while the remaining part is attributable to measurement error.

However, reliability is not constant across the consumption components. It is higher both for utilities and education, which account for a few expenses on a more regular basis. On the other hand, the lowest reliability is observed for recreational consumption, which is more difficult to capture due to its lower regularity over time and higher granularity across household members. A modest reliability also characterizes transportation and (in most estimates) health consumption.

The reliability of food consumption consumed both at home and away from home (excluding alcoholic beverages), is around 0.8, quite similar to the estimates of the previous paragraph.

Table 3 – Reliability coefficients for some expenditure aggregates

Consumption aggregates (a)	Heise model coefficients - Household consumption		Heise model coefficients - Equivalent household consumption		Average of annual SEM coefficients - Household consumption	
	Value	Ranks	Value	Ranks	Model 1	Model 2
1. Food and non-alcoholic beverages: at home and away from home	0.780	0.810	0.820	0.757	0.871	0.998
2. Alcohol and tobacco: at home and away from home	0.764	0.787	0.812	0.793	0.586	0.736
3. Food, beverages, alcohol and tobacco: at home	0.719	0.813	0.650	0.695	0.819	0.659
4. Food, beverages, alcohol and tobacco: away from home	0.877	0.656	0.747	0.661	0.843	0.709
5. Utilities: water, kerosene, lighting	0.935	0.893	0.905	0.905	0.962	0.849
6. Furnishings and household expenses	0.870	0.671	0.756	0.666	0.663	0.896
7. Health	0.854	0.576	0.461	0.532	0.664	0.622
8. Transportation	0.622	0.665	0.531	0.660	0.715	0.524
9. Communications	0.762	0.876	0.714	0.880	0.788	0.742
10. Recreation	0.232	0.318	0.378	0.319	0.372	0.520
11. Education	0.996	0.968	0.793	0.968	0.964	0.666
12. Other consumption	0.654	0.833	0.748	0.847	0.826	0.724
13. Total consumption - nominal	0.905	0.882	0.919	0.842	0.990	0.779
14. Total consumption – real	0.884	0.867	0.899	0.826	0.977	0.749

(a) Aggregates 1+2 = 3+4 = Total food consumption.

The reliability of total consumption does not grow when the least reliable variables are excluded from the sum. For example, the sum of all consumption items excluding recreation provides an aggregate whose reliability is just a little lower than that of the complete aggregate; also excluding transportation or health consumption from the total slightly decreases reliability. In general, adding up items improves the reliability of aggregates.

As already observed for health, the estimates of reliability coefficients do not always display stable behaviour. Heise coefficients computed on consumption values and on ranks only show a moderate agreement; some differences between these estimates are quite large (e.g. 0.877 and 0.656 for food consumption away from home, or 0.854 and 0.576 for health consumption). Moreover, the coefficients derived from the SEM are sometimes divergent from the other estimates. On the whole, the analysis of the different coefficients does not always provide a clear picture or eliminate any doubts as to the real situation.

A greater instability characterizes the reliability coefficients computed on more detailed food consumption items (Table 1A in the appendix A). In fact, correlation coefficients of specific food consumption between two consecutive waves – always computed only on households who did not change their composition - are often quite low, around 0.2 on average and only in a few cases are they significantly higher (never greater than 0.625).[22]

[22] As a comparison, the average of one-lag correlation coefficients computed on the main consumption aggregates is around 0.55.

By collecting specific consumption items over a single week, the survey captures a behaviour of households that is only weakly confirmed in the second wave. As we discussed earlier, although the short reference period reduces the memory biases and other forms of contamination, and they are presumably near to the actual data, the collected data are not a good picture of the consumption behaviour of that household over the entire year. In any case, low correlations over the waves imply a greater instability in the estimates of Heise coefficients, which may even go outside the range 0-1 (as happens in almost a quarter of the cases).[23]

As we have already said, a different approach to dealing with the reliability of answers relies on the analysis of internal consistency. In consumption surveys, several instruments developed for this kind of data can be fruitfully used.

Table 4 shows several indexes that can help in understanding the reliability of the collected data. The first column shows the correlation between the specific variable in row and the sum of all the other components of total household consumption. The higher the value, the more the data contained in the variable is coherent with the sum of all the other components. The second column shows the correlation of the component in the row and the predicted values of the multiple regression with all the other components. As the least squares solution is the linear combination maximizing the predictability of the dependent component, this measure is always higher than that computed on the sum of the components. The third and the fourth columns of the table show the share of standard deviation of the component which is accounted for by the first and the first three principal components respectively. As the first principal component is a linear combination of all the components, including that on the row, it tends to be higher than the previous two measures, although this is not always the case (for example, see recreation consumption). The same four measures are then computed for the equivalent consumption.

On the whole, the picture drawn by these indicators is coherent with that described above, mainly when considering the Heise indices computed on the ranks rather than on the values.

Recreation and health consumption show low indices of internal consistency, confirming the results obtained with the Heise model. A low internal consistency index also characterizes ‘Alcohol and tobacco: at home and away from home’ which, on the contrary, showed a good performance in terms of coherence over time. On the other hand, a good performance is found for utilities, communication, other consumption and food (excluding alcohol and tobacco), largely confirming the previous results.

---

[23] Although the above reasons advise caution, some results will be commented on below. On average, the variables indicating the consumption during the week (Yes/No), the quantity and the value have almost the same reliability. The items with the largest reliability indices (around 0.7-0.8) are sugar, coconuts, rice (husked), fresh fish and seafood, peas, beans, lentils and other pulses, dry tea, salt, bread and onions, tomatoes, carrots, green pepper, and other viungo. Low reliability coefficients are found for the following items: eggs, other domestic/wild meat products, other starches, milk products (like cream, cheese, yoghurt etc.), chicken and other poultry, millet and sorghum (grain), coffee and cocoa, canned milk/milk powder and sweet potatoes. The coefficients for wine and spirits, bottled beer, prepared tea, coffee, honey, syrups, jams, marmalade, jellies, canned fruits, seeds, products from nuts/seeds (excl. cooking oil), packaged/canned fish and other raw materials for drinks were always outside the range 0-1.

Table 4 – Internal consistency of 2012 consumption aggregates

Consumption aggregates (a)	Household consumption				Equivalent consumption			
	Correlation with the sum of all the other components	Multiple correlation with all other components	Correlation with the first principal component	Multiple correlation with the first three principal components	Correlation with the sum of all the other components	Multiple correlation with all other components	Correlation with the first principal component	Multiple correlation with the first three principal components
1. Food and non-alcoholic beverages: at home and away from home	0.630	0.648	0.833	0.923	0.613	0.681	0.842	0.909
2. Alcohol and tobacco: at home and away from home	0.234	0.238	0.323	0.832	0.205	0.196	0.282	0.744
3. Food, beverages, alcohol and tobacco: at home	0.410	0.517	0.663	0.832	0.235	0.529	0.560	0.909
4. Food, beverages, alcohol and tobacco: away from home	0.383	0.498	0.623	0.813	0.291	0.495	0.614	0.934
5. Utilities: water, kerosene, lighting	0.631	0.723	0.760	0.796	0.648	0.732	0.772	0.826
6. Furnishings and household expenses	0.489	0.553	0.613	0.682	0.484	0.533	0.604	0.647
7. Health	0.296	0.316	0.357	0.552	0.237	0.398	0.349	0.668
8. Transportation	0.549	0.617	0.682	0.700	0.517	0.609	0.659	0.679
9. Communications	0.639	0.694	0.757	0.767	0.681	0.710	0.773	0.776
10. Recreation	0.190	0.293	0.282	0.756	0.195	0.266	0.268	0.485
11. Education	0.517	0.568	0.633	0.657	0.274	0.329	0.370	0.516
12. Other consumption	0.679	0.747	0.794	0.814	0.676	0.725	0.771	0.805

(a) Aggregates 1+2 = 3+4 = Total food consumption. The two pairs of food aggregates have been used alternately in the computation of all the indices of the table.

In the comparative analysis of these results, it is worth taking into account that the reliability measures have been computed in the two frameworks under different assumptions. Clearly, random errors affecting indicators imply both a reduced ability of an AR1 model to account for data, and a lower internal consistency of data. A different performance of an indicator under the two frameworks could signal some deviation from the hypotheses on which the models are built.

This could be the case of 'alcohol and tobacco', for which the indices based on the models give a picture of satisfying reliability, which is not confirmed on looking at the coherence with other consumption items. A similar result is found for health consumption. This suggests that these kinds of consumption do not share the same latent variable, as the conditions at their base (the need to smoke or drink, or poor health conditions) are only partially related to the consumption behaviour.

Furthermore, the reliability indicators of food items based on the internal consistency largely confirm the results obtained with the above models (Table 2A in Appendix A).[24]

5. The impact on poverty and inequality measures

The impact of measurement errors on poverty and inequality measures has been studied quite extensively over the years.

As already mentioned in the introduction, many applications have been conducted using a case-by-case approach, by comparing survey data with administrative or other approximations of 'true data' (Cannari and D'Alessio, 1993; Gottschalk and Huynh, 2010) whose conclusions cannot easily be extended to different contexts.

Much attention has been devoted to the estimation of poverty dynamics, which is greatly affected by measurement errors. Methods for obtaining mobility estimates accounting for the upward bias induced by measurement errors have been proposed by many authors (Neri, 2009, Luttmer, 2002; Glewwe, 2012; Burger et al., 2016; Lee et al., 2017).

Widespread attention has also been paid to the impact of outliers on poverty and inequality measures. Such studies have produced a much deeper knowledge of the sensitivity of various poverty and inequality measures to data contamination (Cowell and Flachaire, 2007).[25] For example, these studies have made it clear that, generally speaking, inequality measures are more sensitive to extreme values than poverty measures. This is particularly true if poverty lines are exogenous (i.e. \$1.25 per day) or they are built on more stable in-sample statistics (i.e. median rather than mean) (Cowell and Victoria-Feser, 1996a; Cowell and Victoria-Feser, 1996b). Most of the time, the proposed estimators are obtained through the use of parametric models or by combining a parametric robust estimation of the upper tail of the distribution with the empirical data (the semi-parametric approach) (Victoria-Feser, 2000; Cowell and Victoria-Feser, 2007).

Less numerous are the studies dealing with the impact of measurement errors - not limited to extreme values - on poverty and inequality measures, maybe because a precise estimation of this effect would require knowledge of the joint distribution of true income (or consumption) measures and the corresponding errors.[26] Scott (1992) shows a method that can be used to reduce the variance of extrapolated annual expenditures. A similar adjustment is applied by Gibson et al. (2003) for the treatment of measurement errors on Chinese

---

[24] We observe quite good reliability for rice (husked), sugar, bread, onions, tomatoes, carrots, green pepper, other viungo, cooking oil, beef including minced sausage, spinach, cabbage and other green vegetables, ripe bananas, peas, beans, lentils and other pulses, bottled/canned soft drinks (soda, juice, water), coconuts (mature/immature), Irish potatoes, dry tea, fresh milk, mangoes, avocadoes and other fruits, other spices, fresh fish and seafood (including dagaa). A bad performance is instead observed for prepared tea, coffee, packaged/canned fish, wild birds and insects, other domestic/wild meat products, other raw materials for drinks, spinach, cabbage and other green vegetables, cashews, almonds and other nuts, other starches, seeds, products from nuts/seeds (excl. cooking oil), local brews, wine and spirits, sugarcane, pork including sausages and bacon, dried/salted fish and seafood (incl. dagaa), yams/cocoyams, canned milk/milk powder, milk products (like cream, cheese, yoghurt and so on), groundnuts in shell/shelled, coffee and cocoa.

[25] Of course extreme values are not necessarily measurement errors, although a certain degree of overlapping between the two phenomena exists. Robust estimators for outliers are also a useful tool for dealing with measurement errors.

[26] A different field of analysis analyses the impact of measurement errors on the estimation of the equality of opportunities (see Ferreira and Gignoux (2011).

consumption data. Chesher and Schluter (2002) propose an approximated method for the estimation of such impacts that can be estimated on error-contaminated data only.

As the classical measurement errors we have dealt with in this paper add noise to variables, it is expected that poverty and inequality measures are correspondingly affected. We have also seen that the extrapolation of data observed over a short period of time to the whole period of interest may produce the same outcomes. Having estimated the magnitude of these effects, it is of great interest to analyse the impact of these issues on a specific case.

Adopting a standard definition of the relative poverty rate, i.e. the count of households whose equivalent (nominal) consumption falls below half the median, we find a share of 15.8 of households in the TNP.

As we have shown in Table 3, the total consumption collected in the TNP has a reliability of around 0.9, that is to say that 10 per cent of the standard deviation is due to measurement errors and should be accounted for.

Of course, having an estimate of the magnitude of measurement errors does not tell us which part of the variability we have to discard. We only know that, if the assumptions of the classical measurement errors hold, they should be spread across the units quite uniformly.

A simple solution can be obtained by means of the method proposed by Scott (1992), which defines a transformation of collected data Y_i in a way that – preserving the mean - the standard deviation of the new variable X_i is 0.9 times that of the old variable: $X_i = M + (Y_i - M)*0.9$, where M is the mean of the consumption variable. This transformation, which implies greater corrections in the tails of the distribution and lower corrections for values near to the mean, can help give an idea of the possible impact of measurement errors on poverty and inequality estimates. By adopting this criterion, the poverty rate obtained on these transformed variables would be much lower than that obtained with the original data (8.2 per cent).[27] The Gini concentration index would also be greatly reduced from 0.436 to 0.393.

If we extend Scott's approach, we can adopt a different transformation of the data such that only the unexplained variance of a model is compressed in order to obtain the desired variability. So for example, by using panel data we can regress the 2008 and 2010 data on the 2012 expenditures and use the predicted values instead of the unconditional mean in the above formula; the coefficient for the compression of the residuals is correspondingly adjusted. In such a case, the poverty rates for the 1,000 homogeneous panel sample units decline by 2.7 percentage points. The change from original to adjusted estimates is smaller than that obtained in the previous adjustment but is still considerable. The Gini concentration index is reduced to 0.393, as in the previous experiment.

A further experiment consisted in finding an approximation of the components of household equivalent expenditures by means of a Principal Component Analysis. Table 5 shows, for various possible approximations (with 1, 2, …, k principal components), the share of the standard deviation accounted for by the principal components for each variable, and the poverty rate obtained on the total equivalent expenditures derived by adding up the values predicted by the k principal components.

[27] In general terms, the headcount poverty ratio defined in absolute terms can be even more affected by such a transformation, because the poverty line does not shift with the distribution. The impact depends on the mass of the distribution around the poverty line.

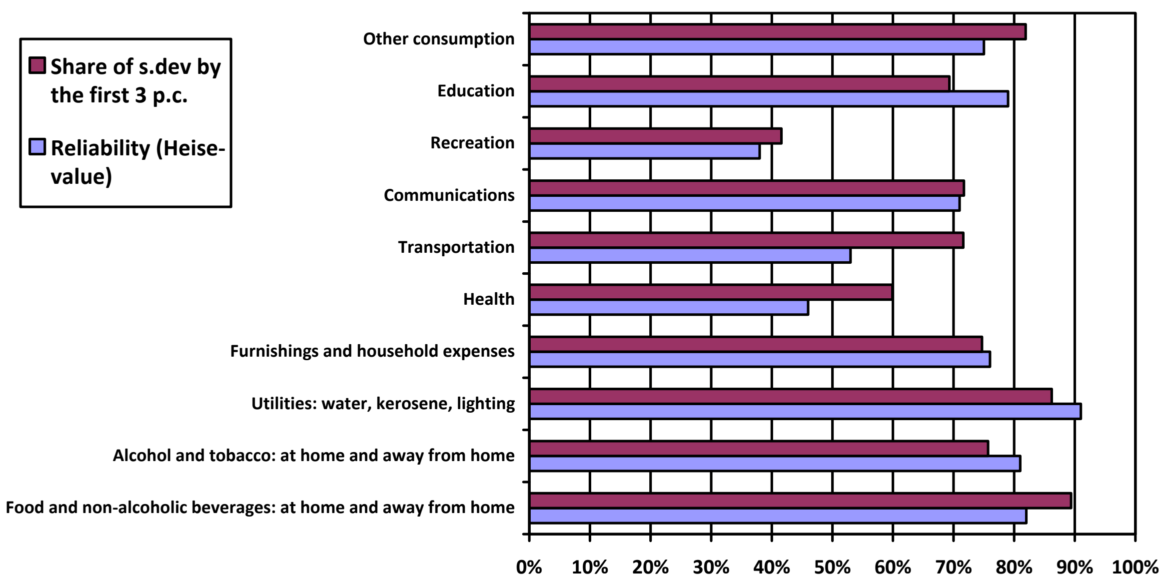
The poverty rate is around 10 per cent when just a few principal components are considered (up to 3); it grows to 12 per cent with 4 or 5 principal components and then slowly goes up to 15.8 per cent when all the possible components are considered (i.e. no errors are considered). As a possible criterion for selecting the number of principal components to retain, we observe that with the first 3 principal components we obtain an approximation of the original components of the total expenditures that is quite close to that derived by means of the Heise model (Figure 1). In other words, the two methods seem to converge independently towards similar results. As in the previous examples, there are clear indicators that poverty rates could be significantly overestimated when using standard estimators.

Table 5 – Principal Component Analysis of 2012 main expenditures items

	Share of the standard deviation accounted for by the first k principal components								
	1	2	3	4	5	...	8	...	12
Food and non-alcoholic beverages	84.1	87.7	89.4	89.4	89.2	...	98.0	...	100.0
Alcohol and tobacco	27.4	75.4	75.7	76.2	77.7	...	99.5	...	100.0
Utilities: water, kerosene, lighting	82.7	86.5	86.2	86.1	85.9	...	87.8	...	100.0
Furnishings and household expenses	67.6	71.9	74.7	83.4	83.0	...	99.5	...	100.0
Health	26.4	37.8	59.9	69.9	79.5	...	99.1	...	100.0
Transportation	66.7	66.8	71.6	72.0	78.9	...	91.2	...	100.0
Communications	71.5	71.8	71.7	71.8	75.2	...	78.0	...	100.0
Recreation	20.2	20.9	41.6	89.0	89.3	...	99.7	...	100.0
Education	43.7	53.0	69.3	79.3	99.9	...	99.7	...	100.0
Other consumption	76.4	79.8	81.9	85.4	88.4	...	88.7	...	100.0
Poverty rate *	10.0	9.4	9.1	12.5	12.6	...	13.4	...	15.8
Gini index **	0.414	0.416	0.415	0.415	0.416	...	0.423	...	0.436

* Household poverty rate computed on the total equivalent expenditures obtained as the sum of the estimated components predicted by the k principal components. ** Gini index computed on the total equivalent expenditures obtained as the sum of the estimated components predicted by the k principal components.

Figure 1 – Share of standard deviation accounted for by the first 3 principal components and the reliability coefficients obtained by means of the Heise model



As to the Gini concentration index, when considering up to 5 principal components it is always around 0.414, a little higher than in the previous two experiments but markedly lower than the original value (0.436).

Although the only purpose of the experiments is to provide an indication of the possible impacts of measurement errors in poverty and inequality measures, the results converge towards the conclusion that the estimates can be notably upward-biased.

6. Conclusions

The paper analysed the measurement errors affecting the most important variables collected in two consumption surveys carried out in recent years in Tanzania: the Tanzania National Household Budget Survey (TNHBS) and the Tanzania National Panel Survey (TNPS). These surveys gave us the chance to study measurement errors using both diary and panel data, and to address general issues regarding the relationships between the quality of data and the estimation of poverty and inequality measures.

The main results can be summarized as follows:

- according to our estimates and models, all the variables collected in these surveys are affected by a share of measurement errors which tend to inflate their variance; the common practice of extrapolating data observed over a short period of time (i.e. one month) to the whole reference period (i.e. the year) can be seen as a measurement error; it tends to inflate the variance of indicators.[28] Other things being equal, the longer the length of the observation period (i.e. the period for which diary data are collected) the lower the variance of collected data;
- the researchers who use consumption micro-data should properly take into account that – other things being equal - the higher the reliability the lower the poverty and inequality measures. In the paper some methodological hints for more robust estimates are provided, but more research in this field is needed;
- given that in sample surveys a certain degree of measurement errors is unavoidable, all the above considerations suggest both the adoption of best practices in the collection of survey data, in order to improve the reliability of data, and a move towards a standardization of the collection methods employed, which reduces the risks of contaminating the comparisons with spurious effects. In particular, in sample surveys conducted on a regular basis, the improvements that usually occur in the data collection procedures could reduce measurement errors over time, and thus produce a bias in the trend of poverty and inequality measures;
- in the diary data collected in the TNHBS, food, transport and communication expenditures show quite good reliability (around 0.8); instead lower reliability characterizes furnishings and - for the share collected by diary - education expenses;

[28] This result is presumably dominant in consumption surveys, in particular in those conducted in developing countries. It is worth noting however that in income and wealth surveys, it is possible that under-reporting behaviour also leads to further measurement errors, pushing poverty and inequality measures downwards.

- the reliability of single food consumption items is instead generally lower, and presumably affected by the consumption/purchasing frequency. Quantities and values have similar reliability;
- for TNPS data, the reliability is quite high for total consumption (around 0.9); food and non-alcoholic beverages have reliability values of around 0.8; a lower reliability is found for some components, such as expenditure on recreation, for which t a lower regularity over time and more expenses spread across individuals can be presumed;
- low reliability is also found for some estimates concerning expenditure on health, mainly when the internal consistency approach is adopted. The result underlines that these expenses are not fully driven by the same latent variable of other expenditures.

Appendix A – Reliability of food consumption items

Table 1A – Reliability coefficients of food consumption items (Heise model), 2008-2013

Item Code	Food consumption item	Yes/No	Quantities		Values	
			Value	Ranks	Value	Ranks
101	Rice (paddy)	0.544	a	0.551	a	a
102	Rice (husked)	0.665	0.798	0.771	0.820	0.796
103	Maize (green, cob)	0.716	0.539	0.726	0.308	0.269
104	Maize (grain)	a	a	a	a	0.468
105	Maize (flour)	0.622	0.646	0.657	0.627	0.714
106	Millet and sorghum (grain)	0.293	0.229	0.296	a	a
107	Millet and sorghum (flour)	0.379	0.377	0.383	0.325	0.356
108	Wheat, barley grain and other cereals	0.575	0.082	0.567	0.638	0.643
109	Bread	0.699	0.684	0.725	0.624	0.741
110	Buns, cakes and biscuits	0.763	0.491	0.736	0.564	0.723
111	Macaroni, spaghetti	0.393	0.286	0.395	0.297	0.403
112	Other cereal products	0.490	a	0.467	0.414	0.322
201	Cassava fresh	0.596	a	0.628	0.508	0.475
202	Cassava dry/flour	0.750	0.625	0.765	0.503	0.734
203	Sweet potatoes	0.761	0.234	0.735	0.539	0.784
204	Yams/cocoyams	0.619	0.539	0.626	a	a
205	Irish potatoes	0.649	0.417	0.670	0.536	0.661
206	Cooking bananas, plantains	0.630	a	0.665	0.528	0.532
207	Other starches	a	0.278	a	a	a
301	Sugar	0.752	0.826	0.820	0.767	0.794
302	Sweet potatoes	a	0.084	a	0.288	a
303	Honey, syrups, jams, marmalade, jellies, canned fruits	a	a	a	a	a
401	Peas, beans, lentils and other pulses	0.783	0.735	0.788	0.600	0.775
501	Groundnuts in shell/shelled	0.573	0.478	0.588	0.606	0.351
502	Coconuts (mature/immature)	0.765	0.767	0.816	0.809	0.788
503	Cashew, almonds and other nuts	0.535	0.250	0.538	a	a
504	Seeds, products from nuts/seeds (excl. cooking oil)	a	a	a	a	a
601	Onions, tomatoes, carrots, green pepper, other viungo	0.723	0.596	0.724	0.654	0.764
602	Spinach, cabbage and other green vegetables	0.609	0.501	0.655	0.664	0.684
603	Canned, dried and wild vegetables	0.530	0.559	0.526	a	a
701	Ripe bananas	0.560	0.749	0.577	0.344	0.537
702	Citrus fruits (oranges, lemons, tangerines, etc.)	0.413	0.289	0.449	0.372	0.428
703	Mangoes, avocadoes and other fruits	0.591	0.527	0.626	0.908	0.615
704	Sugarcane	0.609	0.626	0.626	0.641	0.578
801	Goat meat	0.424	0.621	0.448	0.319	0.393
802	Beef including minced sausage	0.587	0.541	0.612	0.623	0.642
803	Pork including sausages and bacon	0.541	0.547	0.545	0.414	0.469
804	Chicken and other poultry	0.308	a	0.322	0.185	0.281
805	Wild birds and insects	0.637	0.617	0.639	a	a
806	Other domestic/wild meat products	0.325	0.239	0.327	a	a
807	Eggs	0.417	0.036	0.424	0.288	0.418
808	Fresh fish and seafood (including dagaa)	0.690	0.769	0.759	0.782	0.734
809	Dried/salted fish and seafood (incl. dagaa)	0.654	0.599	0.677	0.586	0.567
810	Packaged/Canned fish	a	a	a	a	a
901	Fresh milk	0.554	0.549	0.572	0.426	0.491
902	Milk products (like cream, cheese, yoghurt etc)	0.307	0.511	0.321	0.055	0.182
903	Canned milk/milk powder	0.251	0.102	0.252	0.138	0.251
1001	Cooking oil	0.818	0.177	0.764	0.670	0.780
1002	Butter, margarine, ghee and other fatty products	0.471	0.360	0.473	0.412	0.498
1003	Salt	0.855	0.561	0.886	0.463	0.825
1004	Other spices	0.437	0.163	0.421	0.364	0.445
1101	Tea dry	0.802	a	0.739	0.599	0.779
1102	Coffee and cocoa	0.365	0.055	0.365	a	0.152
1103	Other raw materials for drinks	a	a	a	a	a
1104	Bottled/canned soft drinks (soda, juice, water)	0.518	a	0.533	0.498	0.538
1105	Prepared tea, coffee	a	a	a	a	a
1106	Bottled beer	a	a	a	a	a
1107	Local brews	0.552	0.582	0.560	0.540	0.503
1108	Wine and spirits	a	a	a	a	a

(a) Coefficients outside the range (0-1).

Table 2A –Internal consistency of 2012 food consumption items

Item Code	Food consumption item	Correlation with the sum of all the other components	Multiple correlation with all other components	Correlation with the first principal component	Multiple correlation with the first three principal components
101	Rice (paddy)	0.003	0.161	0.012	0.012
102	Rice (husked)	0.581	0.695	0.536	0.536
103	Maize (green, cob)	0.162	0.298	0.169	0.169
104	Maize (grain)	0.035	0.215	0.056	0.056
105	Maize (flour)	0.187	0.373	0.185	0.185
106	Millet and sorghum (grain)	0.028	0.122	0.036	0.036
107	Millet and sorghum (flour)	0.121	0.259	0.143	0.143
108	Wheat, barley grain and other cereals	0.268	0.602	0.245	0.245
109	Bread	0.429	0.578	0.393	0.393
110	Buns, cakes and biscuits	0.284	0.436	0.236	0.236
111	Macaroni, spaghetti	0.291	0.469	0.312	0.312
112	Other cereal products	0.130	0.242	0.121	0.121
201	Cassava fresh	0.185	0.411	0.181	0.181
202	Cassava dry/flour	-0.006	0.192	0.011	0.011
203	Sweet potatoes	0.172	0.294	0.180	0.180
204	Yams/cocoyams	0.143	0.313	0.148	0.148
205	Irish potatoes	0.478	0.557	0.461	0.461
206	Cooking bananas, plantains	0.343	0.409	0.347	0.347
207	Other starches	0.030	0.150	0.032	0.032
301	Sugar	0.468	0.694	0.448	0.448
302	Sweet potatoes	0.289	0.440	0.293	0.293
303	Honey, syrups, jams, marmalade, jellies, canned fruits	0.320	0.487	0.360	0.360
401	Peas, beans, lentils and other pulses	0.488	0.568	0.454	0.454
501	Groundnuts in shell/shelled	0.171	0.310	0.197	0.197
502	Coconuts (mature/immature)	0.474	0.673	0.407	0.407
503	Cashew, almonds and other nuts	0.032	0.118	0.035	0.035
504	Seeds, products from nuts/seeds (excl. cooking oil)	0.022	0.170	0.031	0.031
601	Onions, tomatoes, carrots, green pepper, other viungo	0.649	0.727	0.642	0.642
602	Spinach, cabbage and other green vegetables	0.517	0.585	0.492	0.492
603	Canned, dried and wild vegetables	0.016	0.137	0.019	0.019
701	Ripe bananas	0.470	0.573	0.485	0.485
702	Citrus fruits (oranges, lemons, tangerines, etc.)	0.385	0.496	0.404	0.404
703	Mangoes, avocadoes and other fruits	0.403	0.530	0.426	0.426
704	Sugarcane	0.060	0.226	0.082	0.082
801	Goat meat	0.172	0.496	0.180	0.180
802	Beef including minced sausage	0.536	0.626	0.552	0.552
803	Pork including sausages and bacon	0.080	0.224	0.092	0.092
804	Chicken and other poultry	0.318	0.473	0.338	0.338
805	Wild birds and insects	0.012	0.105	0.004	0.004
806	Other domestic/wild meat products	0.016	0.098	0.009	0.009
807	Eggs	0.373	0.524	0.388	0.388
808	Fresh fish and seafood (including dagaa)	0.393	0.567	0.382	0.382
809	Dried/salted fish and seafood (incl. dagaa)	0.136	0.310	0.148	0.148
810	Packaged/Canned fish	0.001	0.067	-0.001	-0.001
901	Fresh milk	0.418	0.517	0.440	0.440
902	Milk products (like cream, cheese, yoghurt etc)	0.174	0.296	0.202	0.202
903	Canned milk/milk powder	0.160	0.335	0.178	0.178
1001	Cooking oil	0.613	0.734	0.599	0.599
1002	Butter, margarine, ghee and other fatty products	0.306	0.484	0.356	0.356
1003	Salt	0.194	0.465	0.188	0.188
1004	Other spices	0.420	0.542	0.409	0.409
1101	Tea dry	0.476	0.605	0.426	0.426
1102	Coffee and cocoa	0.177	0.341	0.197	0.197
1103	Other raw materials for drinks	0.014	0.102	0.014	0.014
1104	Bottled/canned soft drinks (soda, juice, water)	0.442	0.645	0.438	0.438
1105	Prepared tea, coffee	-0.004	0.049	-0.005	-0.005
1106	Bottled beer	0.213	0.610	0.223	0.223
1107	Local brews	0.016	0.282	0.021	0.021
1108	Wine and spirits	0.040	0.223	0.044	0.044

References

Beegle K., J. De Weerdt, J. Friedman, J. Gibson (2012), Methods of household consumption measurement through surveys: Experimental results from Tanzania, Journal of Development Economics 98 pp. 3–18.

Biancotti C., G. D'Alessio, A. Neri, 2008. Measurement Error in The Bank of Italy's Survey of Household Income and Wealth, *Review of Income and Wealth*, International Association for Research in Income and Wealth, vol. 54(3), pp. 466-493, 09.

Biemer P.P., S.L. Christ, C. A Wiesen (2009), A general approach for estimating scale score reliability for panel survey data, Psychological Methods, Vol 14(4), pp. 400-412. http://dx.doi.org/10.1037/a0016618

Biemer, P.P., D. Trewin (1997), A Review of Measurement Error Effects on the Analysis of Survey Data, L. Lyberg et al. eds., Survey Measurement and Process Quality, Wiley, pp. 603-633.

Brown W. (1910), Some experimental results in the correlation of mental abilities, British Journal of Psychology, 3, pp. 296–322.

Burger R.P., S. Klasen, A. Zoch (2016), Estimating income mobility when income is measured with error: the case of South Africa, REDI3x3 Working paper, No.14, April.

Cannari L., G. D'Alessio (1993), Non-reporting and Under-reporting Behavior in the Bank of Italy's Survey of Household Income and Wealth, in "Bulletin of the International Statistical Institute", vol. LV, No. 3, Pavia, pp. 395-412.

Chen X., H. Hong, D. Nekipelov (2007), Measurement Error Models, http://web.stanford.edu/~doubleh/eco273B/survey-jan27chenhandenis-07.pdf

Chesher A., C. Schluter (2002), Welfare Measurement and Measurement Error, The Review of Economic Studies, Vol. 69, No. 2 (Apr., 2002), pp. 357-378.

Cifaldi G, A. Neri (2013), Asking income and consumption questions in the same survey: what are the risks?, Temi di discussione (Working papers), No. 908, Banca d'Italia.

Clarke P.M., D.G. Fiebig, U.G. Gerdtham (2008), Optimal recall length in survey design, Journal of Health Economics, 27, pp. 1275–1284

Converse J.M., S. Presser (1986), Survey Questions: Handcrafting the Standardized Questionnaire. Beverly Hills, California, Sage Publications.

Cowell F.A., E. Flachaire (2007), Income distribution and inequality measurement: The problem of extreme values, Journal of Econometrics, No. 141, pp. 1044–1072.

Cowell F.A., M.P. Victoria-Feser (1996a), Poverty measurement with contaminated data: a robust approach, European Economic Review, No. 40, pp. 1761-1771.

Cowell F.A., M.P. Victoria-Feser (1996b), Robustness properties of inequality measures, Econometrica, Vol. 64, No. 1, pp. 77-101.

Cowell F.A., M.P. Victoria-Feser (2007), Robust stochastic dominance: A semi-parametric approach, Journal of Economic Inequality, No. 5, pp. 21-37.

Cronbach L.J. (1951), Coefficient alpha and the internal structure of tests, Psychometrika 16, 297–334.

Cronbach L.J. (2004), My current thoughts on coefficient alpha and successor procedures, Educational and Psychological Measurement 64, pp. 391–418.

Crossley T.F., S. Kennedy (2002), The reliability of self-assessed health status, Journal of Health Economics, Vol. 21, No. 4, pp. 643-58.

D'Alessio G., A. Neri (2015), Income and wealth sample estimates consistent with macro aggregates: some experiments, Questioni di Economia e Finanza (Occasional Papers), No. 272, Banca d'Italia.

Deaton A., M. Grosh (1997), Designing Household Survey Questionnaires for Developing Countries: Lessons from 10 Years of the Living Standards Measurement Study, Edited by M. Grosh and P. Glewwe, The World Bank.

Eckart G., G. Young (1936), The approximation of one matrix by another of lower rank, Psychometrika, 1, pp. 211-218.

Ferreira F.H.G., J. Gignoux (2011), The measurement of inequality of opportunity: theory and an application to Latin America, Review of Income and Wealth, 57: pp. 622-657.

Friedman J., K. Beegle, J. De Weerdt, J. Gibson (2016), Decomposing Response Errors in Food Consumption Measurement: Implications for Survey Design from a Survey Experiment in Tanzania, World Bank Policy Research Working Paper 7646, April.

Gavish M., D.L. Donoho (2014), The Optimal Hard Threshold for Singular Values is 4/sqrt (3), arXiv:1305.5870v3 [stat.ME], 4 June, http://arxiv.org/pdf/1305.5870.pdf

Gibson J. (2016), Measuring Chronic Hunger from Diet Snapshots: Why 'Bottom up' Survey Counts and 'Top down' FAO Estimates Will Never Meet, Working Papers in Economics, University of Waikato, Department of Economics.

Gibson J., B. Kim (2011), How reliable are household expenditures as a proxy for permanent income? Implications for the income–nutrition relationship, Department of Economics Working Paper Series, Number 03/11, Hamilton, New Zealand: University of Waikato.

Gibson J., J. Huang, S. Rozelle (2003), Improving Estimates of Inequality and Poverty from Urban China's Household Income and Expenditure Survey, Review of Income and Wealth, Series 49, Number 1, March 2003.

Glewwe P. (2012), How Much of Observed Economic Mobility is Measurement Error? IV Methods to Reduce Measurement Error Bias, with an Application to Vietnam, World Bank Economic Review, Volume 26, Issue 2: pp. 236-264, June

Gottschalk P., M. Huynh (2010), Are Earnings Inequality and Mobility Overstated? The Impact of Non-Classical Measurement Error, The Review of Economics and Statistics, Vol. 92, No. 2, pp. 302-315.

Grosh M., P. Glewwe (2000), Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study, Volumes 1, 2, and 3, Oxford University Press (for the World Bank).

Guttman L. (1945), A basis for analyzing test-retest reliability. Psychometrika, 10 (4), pp. 255-282.

Hand D., H. Mannila, P. Smyth (2001), Principles of Data Mining, MIT Press, Cambridge, MA.

Heise D. (1969), Separating Reliability and Stability in Test-Retest Correlation, *American Sociological Review*, Vol. 34, No. (1), pp. 93-101.

Istat (2016), La nuova indagine sulle spese per consumi in Italia, (edited by D. Grassi, N. Pannuzi and C. Freguja), Metodi – Letture Statistiche.

Jolliffe D., U. Serajuddin (2015), Estimating poverty with panel data, comparably: an example from Jordan, Policy Research Working Paper, No. 7373.

Lee N., G. Ridder, J. Strauss (2017), Estimation of poverty transition matrices with noisy data, Journal of Applied Econometrics, No. 32, pp. 37–55.

Lord F.M., M.R. Novick (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, MA.

Luttmer E.F.P. (2002), Measuring Economic Mobility and Inequality: Disentangling Real Events from Noisy Data, Harris School of Public Policy, Univ. of Chicago.

Neri A. (2009), Measuring wealth mobility, Temi di discussione (Working papers), No. 703, Banca d'Italia.

Rulon, P.J. (1939), A simplified procedure for determining the reliability of a test by split-halves, Harvard Educational Review, 9, pp. 99-103.

Scott C. (1992), Estimation of Annual Expenditure from One-month Cross-sectional Data in a Household Survey, Inter-Stat, 8, pp. 57-65.

Spearman C. (1910), Correlation calculated from faulty data, British Journal of Psychology, 3, pp. 271–295.

Tourangeau R., L.J. Rips, K. Rasinski (2000), The Psychology of Survey Response, Cambridge University Press.

Victoria-Feser (2000), A General Robust Approach to the Analysis of Income Distribution, Inequality and Poverty, International Statistical Review, Vol. 68, pp. 277-293.

Webb N.M., R.J Shavelson, E.H. Haertel (2006), Reliability Coefficients and Generalizability Theory, Handbook of Statistics, Vol. 26, Elsevier.

RECENTLY PUBLISHED "TEMI" (*)

N. 1092 – *Copula-based random effects models for clustered data*, by Santiago Pereda Fernández (December 2016).

N. 1093 – *Structural transformation and allocation efficiency in China and India*, by Enrica Di Stefano and Daniela Marconi (December 2016).

N. 1094 – *The bank lending channel of conventional and unconventional monetary policy*, by Ugo Albertazzi, Andrea Nobili and Federico M. Signoretti (December 2016).

N. 1095 – *Household debt and income inequality: evidence from Italian survey data*, by David Loschiavo (December 2016).

N. 1096 – *A goodness-of-fit test for Generalized Error Distribution*, by Daniele Coin (February 2017).

N. 1097 – *Banks, firms, and jobs*, by Fabio Berton, Sauro Mocetti, Andrea Presbitero and Matteo Richiardi (February 2017).

N. 1098 – *Using the payment system data to forecast the Italian GDP*, by Valentina Aprigliano, Guerino Ardizzi and Libero Monteforte (February 2017).

N. 1099 – *Informal loans, liquidity constraints and local credit supply: evidence from Italy*, by Michele Benvenuti, Luca Casolaro and Emanuele Ciani (February 2017).

N. 1100 – *Why did sponsor banks rescue their SIVs?*, by Anatoli Segura (February 2017).

N. 1101 – *The effects of tax on bank liability structure*, by Leonardo Gambacorta, Giacomo Ricotti, Suresh Sundaresan and Zhenyu Wang (February 2017).

N. 1102 – *Monetary policy surprises over time*, by Marcello Pericoli and Giovanni Veronese (February 2017).

N. 1103 – *An indicator of inflation expectations anchoring*, by Filippo Natoli and Laura Sigalotti (February 2017).

N. 1104 – *A tale of fragmentation: corporate funding in the euro-area bond market*, by Andrea Zaghini (February 2017).

N. 1105 – *Taxation and housing markets with search frictions*, by Danilo Liberati and Michele Loberto (March 2017).

N. 1106 – *I will survive. Pricing strategies of financially distressed firms*, by Ioana A. Duca, José M. Montero, Marianna Riggi and Roberta Zizza (March 2017).

N. 1107 – *STEM graduates and secondary school curriculum: does early exposure to science matter?*, by Marta De Philippis (March 2017).

N. 1108 – *Lending organization and credit supply during the 2008-09 crisis*, by Silvia del Prete, Marcello Pagnini, Paola Rossi and Valerio Vacca (April 2017).

N. 1109 – *Bank lending in uncertain times*, by Piergiorgio Alessandri and Margherita Bottero (April 2017).

N. 1110 – *Services trade and credit frictions: evidence from matched bank-firm data*, by Francesco Bripi, David Loschiavo and Davide Revelli (April 2017).

N. 1111 – *Public guarantees on loans to SMEs: an RDD evaluation*, by Guido de Blasio, Stefania De Mitri, Alessio D'Ignazio, Paolo Finaldi Russo and Lavina Stoppani (April 2017).

N. 1112 – *Local labour market heterogeneity in Italy: estimates and simulations using responses to labour demand shocks*, by Emanuele Ciani, Francesco David and Guido de Blasio (April 2017).

N. 1113 – *Liquidity transformation and financial stability: evidence from the cash management of open-end Italian mutual funds*, by Nicola Branzoli and Giovanni Guazzarotti (April 2017).

N. 1114 – *Assessing the risks of asset overvaluation: models and challenges*, by Sara Cecchetti and Marco Taboga (April 2017).

(*) 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.

2015

AABERGE R. and A. BRANDOLINI, *Multidimensional poverty and inequality,* in A. B. Atkinson and F. Bourguignon (eds.), Handbook of Income Distribution, Volume 2A, Amsterdam, Elsevier, **TD No. 976 (October 2014).**

ALBERTAZZI U., G. ERAMO, L. GAMBACORTA and C. SALLEO, *Asymmetric information in securitization: an empirical assessment*, Journal of Monetary Economics, v. 71, pp. 33-49, **TD No. 796 (February 2011).**

ALESSANDRI P. and B. NELSON, *Simple banking: profitability and the yield curve,* Journal of Money, Credit and Banking, v. 47, 1, pp. 143-175, **TD No. 945 (January 2014).**

ANTONIETTI R., R. BRONZINI and G. CAINELLI, *Inward greenfield FDI and innovation,* Economia e Politica Industriale, v. 42, 1, pp. 93-116, **TD No. 1006 (March 2015).**

BARONE G. and G. NARCISO, *Organized crime and business subsidies: Where does the money go?,* Journal of Urban Economics, v. 86, pp. 98-110, **TD No. 916 (June 2013).**

BRONZINI R., *The effects of extensive and intensive margins of FDI on domestic employment: microeconomic evidence from Italy,* B.E. Journal of Economic Analysis & Policy, v. 15, 4, pp. 2079-2109, **TD No. 769 (July 2010).**

BUGAMELLI M., S. FABIANI and E. SETTE, *The age of the dragon: the effect of imports from China on firm-level prices*, Journal of Money, Credit and Banking, v. 47, 6, pp. 1091-1118, **TD No. 737 (January 2010).**

BULLIGAN G., M. MARCELLINO and F. VENDITTI, *Forecasting economic activity with targeted predictors,* International Journal of Forecasting, v. 31, 1, pp. 188-206, **TD No. 847 (February 2012).**

BUSETTI F., *On detecting end-of-sample instabilities,* in S.J. Koopman, N. Shepard (eds.), Unobserved Components and Time Series Econometrics, Oxford, Oxford University Press, **TD No. 881 (September 2012).**

CESARONI T., *Procyclicality of credit rating systems: how to manage it,* Journal of Economics and Business, v. 82. pp. 62-83, **TD No. 1034 (October 2015).**

CIARLONE A., *House price cycles in emerging economies*, Studies in Economics and Finance, v. 32, 1, **TD No. 863 (May 2012).**

CUCINIELLO V. and F. M. SIGNORETTI, *Large banks,loan rate markup and monetary policy,* International Journal of Central Banking, v. 11, 3, pp. 141-177, **TD No. 987 (November 2014).**

DE BLASIO G., D. FANTINO and G. PELLEGRINI, *Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds,* Industrial and Corporate Change, v. 24, 6, pp. 1285-1314, **TD No. 792 (February 2011).**

DEPALO D., R. GIORDANO and E. PAPAPETROU, *Public-private wage differentials in euro area countries: evidence from quantile decomposition analysis,* Empirical Economics, v. 49, 3, pp. 985-1115, **TD No. 907 (April 2013).**

DI CESARE A., A. P. STORK and C. DE VRIES, *Risk measures for autocorrelated hedge fund returns,* Journal of Financial Econometrics, v. 13, 4, pp. 868-895, **TD No. 831 (October 2011).**

FANTINO D., A. MORI and D. SCALISE, *Collaboration between firms and universities in Italy: the role of a firm's proximity to top-rated departments*, Rivista Italiana degli economisti, v. 1, 2, pp. 219-251, **TD No. 884 (October 2012).**

FRATZSCHER M., D. RIMEC, L. SARNOB and G. ZINNA, *The scapegoat theory of exchange rates: the first tests,* Journal of Monetary Economics, v. 70, 1, pp. 1-21, **TD No. 991 (November 2014).**

NOTARPIETRO A. and S. SIVIERO, *Optimal monetary policy rules and house prices: the role of financial frictions,* Journal of Money, Credit and Banking, v. 47, S1, pp. 383-410, **TD No. 993 (November 2014).**

RIGGI M. and F. VENDITTI, *The time varying effect of oil price shocks on euro-area exports,* Journal of Economic Dynamics and Control, v. 59, pp. 75-94, **TD No. 1035 (October 2015).**

TANELI M. and B. OHL, *Information acquisition and learning from prices over the business cycle,* Journal of Economic Theory, 158 B, pp. 585–633, **TD No. 946 (January 2014).**

2016

ALBANESE G., G. DE BLASIO and P. SESTITO, *My parents taught me. evidence on the family transmission of values,* Journal of Population Economics, v. 29, 2, pp. 571-592, **TD No. 955 (March 2014).**

ANDINI M. and G. DE BLASIO, *Local development that money cannot buy: Italy's Contratti di Programma,* Journal of Economic Geography, v. 16, 2, pp. 365-393, **TD No. 915 (June 2013).**

BARONE G. and S. MOCETTI, *Inequality and trust: new evidence from panel data,* Economic Inquiry, v. 54, pp. 794-809, **TD No. 973 (October 2014).**

BELTRATTI A., B. BORTOLOTTI and M. CACCAVAIO, *Stock market efficiency in China: evidence from the split-share reform,* Quarterly Review of Economics and Finance, v. 60, pp. 125-137, **TD No. 969 (October 2014).**

BOLATTO S. and M. SBRACIA, *Deconstructing the gains from trade: selection of industries vs reallocation of workers,* Review of International Economics, v. 24, 2, pp. 344-363, **TD No. 1037 (November 2015).**

BOLTON P., X. FREIXAS, L. GAMBACORTA and P. E. MISTRULLI, *Relationship and transaction lending in a crisis,* Review of Financial Studies, v. 29, 10, pp. 2643-2676, **TD No. 917 (July 2013).**

BONACCORSI DI PATTI E. and E. SETTE, *Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register,* Journal of Financial Intermediation , v. 25, 1, pp. 54-76, **TD No. 848 (February 2012).**

BORIN A. and M. MANCINI, *Foreign direct investment and firm performance: an empirical analysis of Italian firms,* Review of World Economics, v. 152, 4, pp. 705-732, **TD No. 1011 (June 2015).**

BRAGOLI D., M. RIGON and F. ZANETTI, *Optimal inflation weights in the euro area,* International Journal of Central Banking, v. 12, 2, pp. 357-383, **TD No. 1045 (January 2016).**

BRANDOLINI A. and E. VIVIANO, *Behind and beyond the (headcount) employment rate,* Journal of the Royal Statistical Society: Series A, v. 179, 3, pp. 657-681, **TD No. 965 (July 2015).**

BRIPI F., *The role of regulation on entry: evidence from the Italian provinces,* World Bank Economic Review, v. 30, 2, pp. 383-411, **TD No. 932 (September 2013).**

BRONZINI R. and P. PISELLI, *The impact of R&D subsidies on firm innovation,* Research Policy, v. 45, 2, pp. 442-457, **TD No. 960 (April 2014).**

BURLON L. and M. VILALTA-BUFI, *A new look at technical progress and early retirement,* IZA Journal of Labor Policy, v. 5, **TD No. 963 (June 2014).**

BUSETTI F. and M. CAIVANO, *The trend–cycle decomposition of output and the Phillips Curve: bayesian estimates for Italy and the Euro Area,* Empirical Economics, V. 50, 4, pp. 1565-1587, **TD No. 941 (November 2013).**

CAIVANO M. and A. HARVEY, *Time-series models with an EGB2 conditional distribution,* Journal of Time Series Analysis, v. 35, 6, pp. 558-571, **TD No. 947 (January 2014).**

CALZA A. and A. ZAGHINI, *Shoe-leather costs in the euro area and the foreign demand for euro banknotes,* International Journal of Central Banking, v. 12, 1, pp. 231-246, **TD No. 1039 (December 2015).**

CIANI E., *Retirement, Pension eligibility and home production,* Labour Economics, v. 38, pp. 106-120, **TD No. 1056 (March 2016).**

CIARLONE A. and V. MICELI, *Escaping financial crises? Macro evidence from sovereign wealth funds' investment behaviour,* Emerging Markets Review, v. 27, 2, pp. 169-196, **TD No. 972 (October 2014).**

CORNELI F. and E. TARANTINO, *Sovereign debt and reserves with liquidity and productivity crises,* Journal of International Money and Finance, v. 65, pp. 166-194, **TD No. 1012 (June 2015).**

D'AURIZIO L. and D. DEPALO, *An evaluation of the policies on repayment of government's trade debt in Italy,* Italian Economic Journal, v. 2, 2, pp. 167-196, **TD No. 1061 (April 2016).**

DE BLASIO G., G. MAGIO and C. MENON, *Down and out in Italian towns: measuring the impact of economic downturns on crime,* Economics Letters, 146, pp. 99-102, **TD No. 925 (July 2013).**

DOTTORI D. and M. MANNA, *Strategy and tactics in public debt management,* Journal of Policy Modeling, v. 38, 1, pp. 1-25, **TD No. 1005 (March 2015).**

ESPOSITO L., A. NOBILI and T. ROPELE, *The management of interest rate risk during the crisis: evidence from Italian banks,* Journal of Banking & Finance, v. 59, pp. 486-504, **TD No. 933 (September 2013).**

MARCELLINO M., M. PORQUEDDU and F. VENDITTI, *Short-Term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility,* Journal of Business & Economic Statistics , v. 34, 1, pp. 118-127, **TD No. 896 (January 2013).**

RODANO G., N. SERRANO-VELARDE and E. TARANTINO, *Bankruptcy law and bank financing,* Journal of Financial Economics, v. 120, 2, pp. 363-382, **TD No. 1013 (June 2015).**

ZINNA G., *Price pressures on UK real rates: an empirical investigation,* Review of Finance,v. 20, 4, pp. 1587-1630, **TD No. 968 (July 2014).**

2017

ADAMOPOULOU A. and G.M. TANZI, *Academic dropout and the great recession,* Journal of Human Capital, V. 11, 1, pp. 35–71, **TD No. 970 (October 2014).**

ALBERTAZZI U., M. BOTTERO and G. SENE, *Information externalities in the credit market and the spell of credit rationing,* Journal of Financial Intermediation, v. 30, pp. 61–70, **TD No. 980 (November 2014).**

ALESSANDRI P. and H. MUMTAZ, *Financial indicators and density forecasts for US output and inflation,* Review of Economic Dynamics, v. 24, pp. 66-78, **TD No. 977 (November 2014).**

BRUCHE M. and A. SEGURA, *Debt maturity and the liquidity of secondary debt markets,* Journal of Financial Economics, v. 124, 3, pp. 599-613, **TD No. 1049 (January 2016).**

DE BLASIO G. and S. POY, *The impact of local minimum wages on employment: evidence from Italy in the 1950s,* Journal of Regional Science, v. 57, 1, pp. 48-74, **TD No. 953 (March 2014).**

LOBERTO M. and C. PERRICONE, *Does trend inflation make a difference?,* Economic Modelling, v. 61, pp. 351–375, **TD No. 1033 (October 2015).**

MOCETTI S., M. PAGNINI and E. SETTE, *Information technology and banking organization,* Journal of Journal of Financial Services Research, v. 51, pp. 313-338, **TD No. 752 (March 2010).**

MOCETTI S. and E. VIVIANO, *Looking behind mortgage delinquencies,* Journal of Banking & Finance, v. 75, pp. 53-63, **TD No. 999 (January 2015).**

PALAZZO F., *Search costs and the severity of adverse selection,* Research in Economics, v. 71, 1, pp. 171-197, **TD No. 1073 (July 2016).**

PATACCHINI E., E. RAINONE and Y. ZENOU, *Heterogeneous peer effects in education,* Journal of Economic Behavior & Organization, v. 134, pp. 190–227, **TD No. 1048 (January 2016).**

FORTHCOMING

ADAMOPOULOU A. and E. KAYA, *Young Adults living with their parents and the influence of peers,* Oxford Bulletin of Economics and Statistics, **TD No. 1038 (November 2015).**

BOFONDI M., L. CARPINELLI and E. SETTE, *Credit supply during a sovereign debt crisis,* Journal of the European Economic Association, **TD No. 909 (April 2013).**

BRONZINI R. and A. D'IGNAZIO, *Bank internationalisation and firm exports: evidence from matched firm-bank data,* Review of International Economics, **TD No. 1055 (March 2016).**

BURLON L., *Public expenditure distribution, voting, and growth,* Journal of Public Economic Theory, **TD No. 961 (April 2014).**

BUSETTI F., *Quantile aggregation of density forecasts,* Oxford Bulletin of Economics and Statistics, **TD No. 979 (November 2014).**

CESARONI T. and R. DE SANTIS, *Current account "core-periphery dualism" in the EMU,* World Economy, **TD No. 996 (December 2014).**

CESARONI T. and S. IEZZI, *The predictive content of business survey indicators: evidence from SIGE,* Journal of Business Cycle Research, **TD No. 1031 (October 2015).**

CONTI P., D. MARELLA and A. NERI, *Statistical matching and uncertainty analysis in combining household income and expenditure data,* Statistical Methods & Applications, **TD No. 1018 (July 2015).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, **TD No. 787 (January 2011).**

D'AMURI F. and J. MARCUCCI, *The predictive power of google searches in forecasting unemployment,* International Journal of Forecasting, **TD No. 891 (November 2012).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, **TD No. 877 (September 2012).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court,* Journal of Economic Geography, **TD No. 898 (January 2013).**

MANCINI A.L., C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading,* Review of Economics of the Household, **D No. 958 (April 2014).**

MEEKS R., B. NELSON and P. ALESSANDRI, *Shadow banks and macroeconomic instability,* Journal of Money, Credit and Banking, **TD No. 939 (November 2013).**

MICUCCI G. and P. ROSSI, *Debt restructuring and the role of banks' organizational structure and lending technologies*, Journal of Financial Services Research, **TD No. 763 (June 2010).**

NATOLI F. and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, **TD No. 1025 (July 2015).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment,* Macroeconomic Dynamics, **TD No. 871 July 2012).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, **TD No. 1100 (February 2017).**

SEGURA A. and J. SUAREZ, *How excessive is banks' maturity transformation?,* Review of Financial Studies, **TD No. 1065 (April 2016).**