Is income inequality improved by informal earnings and domestic activities?

Armagan T. Aktuna-Gunes[*] François Gardes[*] Okay Gunes[**] Christophe Starzec[***]

Preliminary Version

Abstract

The question posed by this work is two-fold: firstly, does domestic activity determine participation in informal activities to a proportionate degree? Secondly, does domestic activity and participation in informal activities result in better income distribution and a lowered level of poverty? In order to answer the first question, the size of the informal economy is first measured on the bases of declared monetary incomes and of full incomes (including monetary values of time use) through a model of a complete demand system for the self-employed and wage earners. The cross-sectional dataset used in these estimations is obtained by matching the Time Use Survey of 2006 with the Household Budget Survey over the years 2007 to 2011, inclusive. To attempt to answer the second question, we measure the Gini Index for extended income (informal earnings incorporated into declared income) and extended full income (informal earnings included in full incomes). We will also decompose income inequality using the Oaxaca-Blinder decomposition method for both working groups. The income inequality gap appears to favor self-employed workers who have informal earnings, while wage earners are worse off through their participation in informal activities. We adopt a concentration index to decompose the total effect of informal earnings on income inequality and to derive the elasticity of variables contributions to income inequality and the sensitivity of these contributions to informal earnings. According to our results, informal earnings contribute a great deal to inequality, expressed as 37.8% which decreases to 35.5% when time use values are added into the estimation; a conclusion which is consistent with the results of the informal economy estimation.

JEL Codes: E26, J22, D63, D31
Keywords: Informal economy, Time use, Income inequality, Decomposition

1. Introduction

It is crucial for governments to be specific in their policies and programmatic interventions in order to avoid the undesired economic and social costs of informality. Such costs extend to income inequality or to poverty which work to limit economic growth. It has been found that income distribution with full incomes (domestic production incorporated into monetary incomes) are more equally distributed than monetary incomes alone (see Bonke, 1992). These studies posit that there is a weak correlation between monetary income and domestic production. The underlying logic, as predicted by Gronau (1986), is that high level workers do less non-market work than low level workers. However, Franzis and Stewart (2011) also show that this weak correlation with the addition of a large value of domestic production to monetary income cannot explain this phenomenon. Furthermore, a strong relationship between these factors may reverse the effects of including household production in monetary incomes (Jenkins and O'Leary, 1996).

However, none of these works discuss the interaction between informal and domestic activities. In fact, participation in informal activities at a micro-decision level necessitates the consideration of domestic production since it is often used as a proxy for the household's informal activity decisions especially in developing economies (Aktuna-Gunes et al., 2014).

[*]Paris School of Economics, Université Paris I Panthéon-Sorbonne, Centre d'Economie de la Sorbonne, 106-112 Boulevard de l'Hôpital, 75647, Paris Cedex 13, France; tel : 01 44 07 82 87 ; *e-mails*: armagan.aktunagunes@gmail.com; gardes@univ-paris1.fr

[**] Université Paris I Panthéon-Sorbonne, Centre d'Economie de la Sorbonne, 106-112 Boulevard de l'Hôpital, 75647, Paris Cedex 13, France; tel : 01 44 07 82 87 ; *e-mails*: okaygunes@yahoo.com

[***] Paris School of Economics, CNRS-CES; tel: 01 44 07 82 59 *e-mail*: starzec@univ-paris1.fr

We can thereby expect that a shortage of resources, combined with a lower opportunity cost of time, will result in an increase in the rate of participation in informal activities to obtain necessary goods and services. This, in turn may modify income distribution, notably in developing economies. One could argue that ignoring domestic activities and under-reported incomes, both, may cause an underestimation of inequalities in income distribution. In this respect, does domestic activity determine participation in informal activities in the same manner? Do domestic activities and participation in informal earning activities result in better distribution of income and a lower level of poverty?

The aim of the present paper is precisely to answer these questions. We believe that the findings of this study lie in household informal earnings and domestic activities and the way in which they affect governmental policy related to income distribution and poverty in developing economies. The expected contribution of this paper is to decompose income distribution to show changes in poverty by incorporating the monetary time use values and unreported incomes into household income. To this end, we use cross-sectional data within the complete demand system framework and full prices with full expenditures (i.e. monetary expenditures and monetary time values of domestic activities combined) obtained through matching of the classic Family Budget and Time Use surveys for Turkey for the years 2007 to 2011. The details of the analysis are as follows:

1) The lack of reliable direct statistics on the informal economy requires both a specific methodological solution and appropriate databases to indirectly evaluate the size of unreported incomes. The most frequently used methods are based on a macroeconomic approach, very often giving disparate evaluations (Schneider and Enste, 2000)[1]. The background of these various macroeconomic methods is frequently discussed and criticized. For instance, Thomas (1999) points out that they are not based on any theory.

   a. In our study we use the complete demand system approach developed by Lyssiotou et al. (2004) (see also Fortin et al. (2009) and Aktuna-Gunes et al., 2014), for the estimation of the size of the black economy in Turkey. The model will be estimated on individual cross-section household data covering the period 2007-2011. The basic idea of this approach is to estimate the individual Engel curves and compare observed expenditure and income. Underreported income is recovered as the difference between the level of reported income and its theoretical level corresponding to the observed expenditure which is supposed to be exactly as reported. We propose a complete demand system approach for the estimation of the under-reported part of incomes both for self-employed and wage-earners[2]. This approach allows us to identify more accurate coefficients for under-reporting due to self-employment incomes and to wages by assuming that the consumption of each good, related to its marginal propensity of consumption, is the same as in the case of the revenue actually observed. Thus, it is

[1] The large differences between the estimates are due essentially to the method used. These differences prevent policy makers from evaluating the gravity of the problem to adopt appropriate policies. This is also the case of Turkey. Many methods have been used in the past such as *money demand* method by Ogunc and Yilmaz (2000) and also by Cetintas and Vergil (2003), the *tax collections* method by Ilgın (2002), the *electricity usage* method by Us (2004), *Dynamic Multiple Indicators Multiple Causes Method* (DYMIMIC) by Schneider and Savasan (2007) which rise a discussion about the reliability of the estimated size of the Turkish informal sector (see Ulgen and Ozturk (2006)). Indeed, these studies give very different estimations of informal economy in Turkey from 3.61% (for Temel et al. (1994)) to 139% (for Akalin and Kesikoglu (2007)) according to the method used for relatively recent and comparable periods.

[2] According to the research conducted by Republic of Turkey Social Security Institution in 2011, 75% of wage-earners declared the minimum wage lower than their real wage rate. The undeclared wage earners represent 45.63% of total wage earners.

possible to compute the size of the black economy on the basis of the information regarding the relative amount of self-employment and wage incomes in the GDP.

b. In this model we consider all goods and services and full income with full price values proposed by Gardes (2016) (see also Aktuna-Gunes et al., 2016) in a quadratic demand system in order to better identify the influences of domestic activities on informal earnings. "Full" values are obtained by integrating monetary time use values in income and in price.

2) We apply the model to Turkey, a developing country. The high level of domestic production in developing countries[3] increases the possibility of substitution between formal and informal incomes via, among others, domestic activities. Final goods via domestic production are produced by combining time use activities with market goods. In this respect, in the estimation, we combined Family Budget for the years from 2007 to 2011 with Time Use Survey for 2006. We use Rubin's (1986) matching specification in order to increase the quality of matching between time use data with household budget survey data (see also Moriarity and Scheuren, 2003). The main reason is that the "concatenation" method proposed by Rubin (1986, 1987) uses partial correlation between variables of interest for each dataset in a matching process which reveals uncertainty caused by the use of fabricated data.

3) We decompose income inequality into several *contributing factors*. The core idea, in this analysis, is to compare the distribution of the extended income (the monetary incomes with informal earnings of households) with the distribution of extended full income (monetary income and monetary time values including informal earnings) through a set of factors that vary systematically with informal activity decisions and socioeconomic status. This analysis allows us to identify the difference between the effects of domestic production and informal earnings on income inequality.

a. More precisely, we propose to decompose income inequality through the method proposed by Oaxaca and Blinder (1973). When we consider participation in informal activities, this allows us to analyze the ways that income inequality is affected, and possible explanations for these effects. However, the explained part of income inequality can be uncertain since heterogeneous time use patterns and restricted hours in work between self-employed workers and wage earners may result in a biased decomposition. In other words, informal earnings may favor or disfavor the income inequality according to time use constraints for households who have different types of occupations. Thus, we propose to decompose self-employed workers and wage earners separately in order to better identify the degree of inequality in income distribution with informal earnings for each group taken into consideration. Thus, we also decompose the effects of domestic production and informal earnings on the income gap for both working groups for each year of observation.

b. In this respect, the degree of inequality captured by the Oaxaca decomposition would be limited, given that group differences are examined here. That being the case, measurement and explanation of inequality in income across the entire distribution of informal earnings and socioeconomic status, would be preferable. We measure the Gini Index and adopt the Concentration Index proposed by Wagstaff et al. (2003) in order to demonstrate that informal earnings with other socioeconomic variables can be decomposed into contributions of the individual factors to income-related inequality.

[3]Domestic production takes the largest share in the daily life of Turkish households. According to Ilkkaracan and Gunduz (2009) this production represents values between 25% and 45% of GDP in 2006.

Thus, we can derive the total direction of the relationship between standard deviation of declared income for each income decile and the extended incomes (for monetary and full) through both workers' informal activity participation distribution. Yet its magnitude reflects both the strength of the relationship and the degree of variability in income inequality for Turkey from 2007 to 2011.

4) In this paper, new methods are proposed for the measurement of poverty which necessitate only aggregated information regarding changes between two periods, information which can be given by a pseudo-panel of repeated cross-sections and which do not necessitate panel data. In this estimation we use the Oaxaca decomposition equation for the poor households, for different periods, which provides the exogenous and endogenous components of poverty changes. Furthermore, we define the Multidimensional index of Poverty and Richness (MIPR) as proposed in Gardes, Gaubert and Langlois (2000) as crossing three dimensions of levels that are supposed to be independent of one another. We use individual informal earnings and monetary time use values in two dimensions of the MIPR, which enables to measure their effect on the change in poverty between two periods. These methods are based on pairing an anticipated probability of poverty computed by means of pseudo-panel data.

The remainder of the paper proceeds as follows: Section 1 presents the theoretical model of the complete demand system in the context of the under-reporting income from various sources with the model of full prices. Section 2 derives the econometric specification of the complete demand model, the decomposition methods and persistent poverty method. Section 3 introduces the combined Family Budget and Time Use surveys dataset used in estimations with a short description of the matching procedure. Section 4 reports the empirical results and section 5 concludes the paper.

2. The theoretical model

2.1 The Full Price Concept

Becker (1965) considers a set of *final goods* the quantities of which Z_i, i=1 to m, enter the direct utility of the consumer $u(Z_1, Z_2,\dots Z_m)$. In order to simplify the analysis, Becker states that a separate activity i produces the final good i in quantity Z_i using a unique market good in quantity x_i and unit time t_i per unit of activity i. Finally, time to produce activity i is supposed to be proportional the quantity of market factor: $t_i = \tau_i x_i$. Thus the final goods are produced by a set of domestic production functions f_i: $Z_i = f_i(x_i, \tau_i; W)$ with all other (socio-economic) characteristics of the household in vector W. This assumption allows him to write the consumer program: Max $u(Z_1, Z_2,\dots Z_m)$ such that $Z_i = f_i(x_i, \tau_i; W), \sum_i p_i x_i = y$ and $\sum_i \tau_i x_i + t_w = T$ with $y = wt_w + V$ the monetary income which sums labor and other incomes, t_w the labor time on the market and T total disposable time for one period. In case of multiple market goods used in activity i, a generalization to a bundle of market goods used to produce the activity can be performed by defining aggregate commodities of these market goods for i: the monetary price p_i can be defined as a price index for the bundle of corresponding goods coherent with the monetary budget constraint.

The sum of these three constraints gives the full budget constraint as depending on full income y^f defined as the maximum monetary income which could be earned working all disposable time T at the market wage rate net of taxes w: $y^f = wT$. The full price π_i for final goods i writes: $p_i x_i + \omega t_i$ with an opportunity cost for time ω which can eventually be taken

as the agent's market wage rate. If the agent's opportunity cost ω differs from net wage, the full budget constraint writes:

$$\sum_i (p_i x_i + \omega t_i) = y^f + (\omega - w)(T - t_w) = y^f + (\omega - w)\sum_i \tau_i x_i \quad (1)$$

In this case, the full income is corrected by means of a function of the domestic production time which represents the difference between the market and the personal valuation of that time: the agent substrates from her full income the transaction cost between her leisure and market labor opportunity cost for time (this correction applies whence the market labor supply t_w is predetermined, which defines the monetary income).

In case of complementary factors (market goods and time) used to domestic commodities (see Gardes, 2016), Becker's full price for commodity *i* can be written:

$$p_{ih}^f = p_i + \omega_h \tau_{ih} \quad (2)$$

with τ_{ih} the time use necessary to produce one unit of that activity and p_i the monetary price. Suppose that a Leontief technology allows the quantities of the two factors to be proportional to the activity:

$$\begin{aligned} x_{ih} &= \xi_{ih} z_{ih} \\ t_{ih} &= \theta_{ih} z_{ih} \end{aligned} \qquad \text{so that } t_{ih} = \tau_{ih} x_{ih} \text{ , yields } \tau_{ih} = \frac{\theta_{ih}}{\xi_{ih}}$$

This case corresponds to an assumption of complementarity between the two factors in the domestic technology[4], which allows calculating a proxy for the full price of activity *i* by the ratio of full expenditure (monetary expenditure and the value of time defined as time use per unit of the commodity multiplied by the opportunity cost of time ω) over its monetary component:

$$\pi_{ih} = \frac{(p_{it} + \omega_{ht}\tau_{ht})\, x_{ih}}{p_i x_{ih}} = \frac{p_{it} + \omega_{ht}\tau_{ht}}{p_i} = 1 + \frac{\omega_{ht}\tau_{ht}}{p_i} = \frac{1}{p_i} p_{ih}^f \quad (3)$$

Under the assumption of a common monetary price p_i for all households in a survey during the same period, this ratio contains all the information on the differences of full prices between households deriving from their opportunity cost for time ω_h and the coefficient of production τ_{ih}. If the monetary price changes between households or periods, the full price can be computed as the product of this proxy π_{ih} with p_{ih}: $p_{ih}^f = p_{iht}\pi_{ih}$. With these definitions, it is possible to measure the full prices, observing only monetary and full expenditures by equation (1). The market wage net of taxes have been used to calibrate the opportunity cost of time (see, e.g., Gardes and Starzec, 2015 for discussion on this subject).

2.2 Consumer expenditure system

Following Lyssiotou et al. (2004), Fortin et al. (2009) and Aktuna-Gunes et al.(2014) we consider households with separable preferences in durable and nondurable goods represented by a cost function: C(**p**, U) = *F*(*c*(**p**, U), *d*(**r**, U), U), where **p**, **r** and U correspond to the price vector of nondurable and durable goods, and to the household utility level. The *c*(.) and *d*(.) functions represent aggregate price indexes for nondurable and durable goods respectively. In

[4] An alternative hypothesis based on the substitutability between the two factors is discussed in Alpman and Gardes, 2016.

other words, they are the sub-cost functions which reflect the prices of unit costs paid by households for each type of good. Each of these functions increases in U and is linearly homogeneous in prices. This structure implies that household consumption decision can be decomposed into *two-stage budgeting*.

(a) The household begins with allocating its total revenue Y^* to the expenditure of durable and nondurable goods according to the cost minimizing rule (with the help of $c(.)$ and $d(.)$).

For example demand for the i^{th} good in the nondurable group writes:

$$q_i = \frac{\partial F(.)}{\partial c(.)} \frac{\partial c(.)}{\partial p_i} \tag{4}$$

So, we can aggregate the demand of q_i to obtain the household total expenditure of nondurable goods by using Shephard's lemma and the first degree homogeneity property on **p** of the $c(.)$ function.

$$y = \sum_i p_i q_i = \frac{\partial F(.)}{\partial c(.)} \sum_i p_i \frac{\partial c(.)}{\partial p_i} = \frac{\partial F(.)}{\partial c(.)} c(.) \tag{5}$$

(b) In the second step, the household chooses the part of the expenditure for each good which belongs to a given group (durable, nondurable) within the total expenditure of each group according to the price vector of this group and to the total utility level.

More precisely, the share of nondurable expenditures w_i within the total expenditure (y) is given by

$$w_i = \frac{p_i q_i}{y} = \frac{p_i \frac{\partial F(.)}{\partial c(.)} \frac{\partial c(.)}{\partial p_i}}{\frac{\partial F(.)}{\partial c(.)} c(.)} = \frac{p_i \frac{\partial c(.)}{\partial p_i}}{c(.)} = \frac{p_i}{\partial p_i} \frac{\partial c(.)}{c(.)} = \frac{\partial \ln c(.)}{\partial \ln p_i} \tag{6}$$

Following Banks et al. (1997), $c(.)$ and $d(.)$ are specified as Pig-log cost function, and equation (6) can thus be written as a Quadratic Almost Ideal Demand System (see the section *Econometric Model*).

3. Econometric model

3.1 Complete demand system

It can be assumed that the unit cost of goods, has the quadratic logarithmic form (Lewbel, 1990)

$$\ln c(\mathbf{p}, U) = a(\mathbf{p}) + b(\mathbf{p}) \left[\frac{U}{1 - g(\mathbf{p})\mathrm{U}} \right] \tag{7}$$

Where $a(\mathbf{p})$, $b(\mathbf{p})$ and $g(\mathbf{p})$ are some functions homogeneous in **p**. The Hicksian shares,

$$w_i = a(\mathbf{p}) + b(\mathbf{p}) \left[\frac{U}{1 - g(\mathbf{p})\mathrm{U}} \right] + \lambda_i(\mathbf{p}) \left[\frac{U}{1 - g(\mathbf{p})\mathrm{U}} \right]^2 \tag{8}$$

where $a_i(\mathbf{p})=\partial \ln a(\mathbf{p})/\partial \ln p_i$, $b_i(\mathbf{p})=\partial \ln b(\mathbf{p})/\partial \ln p_i$ and $\lambda_i(\mathbf{p})= b_i(\mathbf{p})\partial \ln g(\mathbf{p})/\partial \ln p_i$, and U is the households utility level. In order to calculate the budget share within the system of Engel Curves, the base period prices can be assumed to be one ($\mathbf{p} = \mathbf{r} = 1$ by introducing the h subscript which denotes the individual households:

$$w_{ih} = \alpha_i + \beta_i \left[\ln Y_h^* \right] + \delta_i \left[\ln Y_h^* \right]^2 \tag{9}$$

Where Y^* is the total (true) income and using the equation (8), $U/(1 - g_0 U)=(\ln Y^* - a_0)/ b_0$ and a_0, b_0 with g_0 are the values corresponding functions at $p_i = r_i = 1$. α, β, δ are the parameters. This equation represents the quadratic Engel curve derived from the Pig-log cost function.

We assume in our model that Y* is separated into three sources denoted a, s, r which respectively correspond to other income sources, wages, self-employment income. Thus, the total reported (true) income is supposed to be a weighted sum of these three sources.

$$Y_h^* = \sum_{m=a,s,r} \theta_m Y_{mh} \tag{10}$$

This equation implies that the true income must be equal to the sum of the observed incomes (Y_a, Y_s, Y_r) multiplied by their corresponding factors (θ_a, θ_s, θ_r), where we suppose θ_r, $\theta_s \geq 1$ (i.e., under reporting) and $\theta_a = 1$ (correct observation of the other incomes). It allows us to calculate the size of the underground economy and the saving tendencies with respect to the under reporting part of declared incomes by an estimation of θ_r and θ_s. In order to impose the constraints on the θ_r and θ_s parameter, Fortin et. al (2009) propose to express it by $(1+e^k)$ where k is a parameter estimated by the model. The true values of self-employment and wage income thus write $Y_r^*=(1+e^k)Y_r$ and $Y_s^*=(1+e^l)Y_s$.

Finally, the sum of each source of income can be determined as a ratio of the reported total income: $y_m = Y_m/Y$, where Y is the sum of other sources as fees, government transfers…etc. as well as wages and self-employment incomes. Following the model proposed by Aktuna-Gunes et al. (2014; based on Banks et al. 1997), we consider all goods and services with full price values in a quadratic demand system:

$$w_{ih} = \alpha_i + \sum_j \alpha_{ij} Z_{jh} + \beta_{1i} \left[\ln Y_h + \ln\left(\sum_{m=a,s,r} \theta_m y_m \right) \right] + \beta_{2i} \left[\ln Y_h + \ln\left(\sum_{m=a,s,r} \theta_m y_m \right) \right]^2 + \sum_j \gamma_{ij} \log \pi_{jh} + e_{ih} \tag{11}$$

where w, π, Z, represent respectively the budget share, the full prices and the household characteristics vector (which allows us to take into account the heterogeneity of preferences), and y_m the tree components of income. We cannot expect that the individuals from different social groups have the same reaction in consumption and saving choices with respect to the different types of incomes especially when there is uncertainty about these revenues.

3.2 Decomposition methods

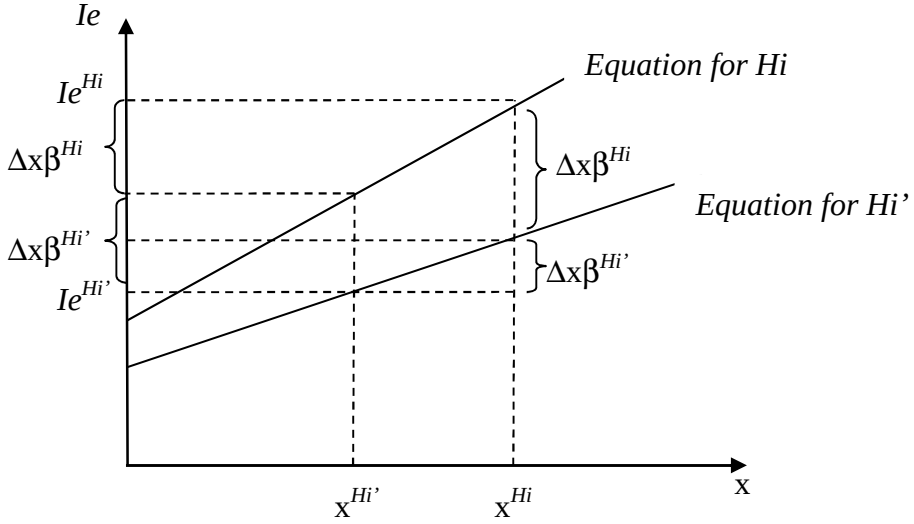
3.2.1 *Oaxaca's decomposition*

Suppose that our outcome variable of interest is income equality (*Ie*). Now, we have two groups, which we shall call the households having informal earnings (*Hi*) and households not having informal earnings (*Hi'*). We can assume that income inequality is explained by a vector of determinants, *x*, according to the regression model:

$$Ie_i = \begin{cases} \beta^{Hi'} x_i + \varepsilon_i^{Hi'} \; if \;\; Hi \\ \beta^{Hi} x_i + \varepsilon_i^{Hi} \; if \;\; Hi' \end{cases} \tag{12}$$

where the vectors of β parameters include intercepts. As it can be pointed in the **Figure 1**[5], In the case of one single regressor the households having informal earnings are assumed to have more advantageous regression line (as concerns the derivative if *Ie* over *x*) than households not having informal earnings.

Figure 1: Oaxaca decomposition for income equality and informal earnings



In the case of the *Hi'*, we read off the equation for the *Hi'* above $x^{Hi'}$, giving a value of *Ie* equal to $Ie^{Hi'}$. In the case of the *Hi*, we read off the equation for the *Hi* above x^{Hi}, giving a value of *Ie* equal to Ie^{Hi}. The gap between income inequality for $Ie^{Hi'}$ and for Ie^{Hi} could be expressed in either of two ways:

$$Ie^{Hi} - Ie^{Hi'} = \beta^{Hi} x^{Hi} - \beta^{Hi'} x^{Hi'} = \begin{cases} \Delta x \beta^{Hi'} + \Delta \beta x^{Hi} \\ \quad or \;\; as \\ \Delta x \beta^{Hi} + \Delta \beta x^{Hi'} \end{cases} \tag{13}$$

Where x^{Hi} and $x^{Hi'}$ are vectors of explanatory variables evaluated at the means for the households having informal earnings and the households not having informal earnings respectively where we assume that exogeneity corresponding to the conditional expectations of the error terms in equation (12), is zero. The two decompositions in equations (13) can be seen as special cases of a more general decomposition.

$$Ie^{Hi} - Ie^{Hi'} = \underbrace{\Delta x \beta^{Hi'}}_{=E} + \underbrace{\Delta \beta x^{Hi'}}_{=C} + \underbrace{\Delta x \Delta \beta}_{=EC} \tag{14}$$

so that the gap in mean outcomes can be thought of as deriving from a gap in endowments (E), a gap in coefficients (C), and a gap arising from the interaction of endowments and coefficients (CE). The two equations 13 are special cases in which

[5] We use the same specification of the Oaxaca decomposition presented by O'Donnel et al. (2008) for poor and nonpoor population.

$$Ie^{Hi} - Ie^{Hi'} = \Delta x\beta^{Hi'} + \Delta\beta x^{Hi} = E + (CE + C) \quad (15\text{-}1)$$

and

$$Ie^{Hi} - Ie^{Hi'} = \Delta x\beta^{Hi} + \Delta\beta x^{Hi'} = (E + CE) + C \quad (15\text{-}2)$$

So, in effect, the first decomposition places the interaction in the unexplained part, whereas the second places it in the explained part.

Oaxaca regression model

In our estimation setting, we use a standard deviation score for (Ie_z) measuring the difference between the declared monetary incomes for each individual and the median monetary incomes of each income quartiles. We classify the households according to whether they have informal earnings or not. Once equation (11) has been carried out, the estimated parameters of the Engel curves are used for the calculation of self employed workers and wage earners true income as $Y_r^* = \bar{\theta}_r Y_r$ and $Y_s^* = \bar{\theta}_s Y_s$ where Y_r^* and Y_s^* are the adjusted self-employers and wage earners incomes which are obtained by multiplying by their declared incomes Y_r and Y_s with $\bar{\theta}_r$ and $\bar{\theta}_s$ respectively. These parameters for each group *r* and *s* can simply be derived from equation (11) as a sum over consumption groups as follows:

$$\bar{\theta}_{r\,(or)\,s} = \frac{1}{n}\sum_{n}\left(\frac{10^{(\hat{R}_n)} - \sum_{m=a,s\ (or)\ r}\hat{\theta}_m y_m}{y_{r\,(or)\,s}}\right) \quad (16)$$

In this equation (16), the sum over households *h* corresponds to the individual values of *y* and θ for each household. θ_r and θ_s is computed for each consumption group by using the estimated parameters of the complete demand system given in the equation (11). $\bar{\theta}$ for *r* and *s* represents the arithmetic mean of the estimated θ_r and θ_s for *n* consumption groups[6]. Thus households having informal earnings (*Hi*=1) can easily be identified whether if households have $\bar{\theta}_r > 1$ and $\bar{\theta}_s > 1$.

3.2.2 Gini and Concentration Index

Measurement and explanation of income inequality across the entire distribution would be preferable than separately explaining informal income and socioeconomic-related income inequality for the self-employed and wage earners. Often inequality indices (Theil and Gini) can also be decomposed to measure sub-populations. Further, we also propose to measure how much income dispersion for each income quartile can be explained through the decomposition of the Concentration Index (CI). In fact, the CI has been used as a relative measure of inequality that indicates the extent to which a health indicator is concentrated

[6] $\hat{R}$ is defined in terms of the quadratic model as:

$\hat{R} = -\hat{\beta}_i - 2\hat{\delta}_i \pm \sqrt{((\hat{\beta}_i + 2\hat{\delta}_i)^2 - 4\hat{\delta}_i(\hat{\beta}_i \ln Y_h + \hat{\delta}_i(\ln Y_h)^2 - (\hat{\alpha}_i + \sum_j \hat{\alpha}_{ij} Z_{jh} + \sum_j \gamma_{ij} \log \pi_{jh} + \sum_t \hat{\alpha}_t A_t)}$. A_t represents the year dummies. $\hat{R}$ is a discriminant equation; hence only the positive roots have been chosen.

among the disadvantaged or the advantaged[7]. We propose to use CI as proposed by Wagstaff, van Doorslaer, and Watanabe (2003) since it enables us to measure the elasticity of the contributions of variables to income inequality and the sensitivity of contribution of the informal earnings and socioeconomic variables[8]. In our analysis, we determine CI as a means of quantifying the degree of informal earnings-related inequality in the dispersion of income among each income decile. To do so, we refer to an income inequality standard deviation score for (Ie_z) and to measure socioeconomic status and monetary and full income, including informal earnings. For any linear additive regression model of Ie_z, such as

$$Ie_z = \alpha + \sum_j \beta_j x_j + \varepsilon \tag{17}$$

the CI for Ie_z, C, can be written as follows:

$$C = \sum_j (\beta_j \bar{x}_j / \mu) C_j + GC_\varepsilon / \mu \tag{18}$$

where μ is the mean of the Ie_z, $\bar{x}_j$ is the mean of x_j, C_j is the CI for x_j (defined analogously to C), and GC_ε is the generalized CI for the error term (ε). Equation (18) shows that C is equal to a weighted sum of the CI of the j regressors, where the weight for x_j is the elasticity of Ie_z with respect to x_j $\left[\beta_j \frac{\bar{x}_j}{\mu}\right]$. The residual component—captured by the last term— reflects the informal income-related inequality as the dispersion of income that is not explained by systematic variation in the regressors by extended income, which should approach zero for a well-specified model.

3.3 Persistent Poverty

The persistence of poverty is the essential second dimension of this phenomenon, which adds to the number of households or individuals which are classified as poor by a poverty line. The information on the duration of this status can be recovered only by means of panel data. Duncan et al. (1995) show that one third of the poor families surveyed in the Panel Study of Income Dynamics are no more poor three years later. This figure is probably highly endogenous to the macroeconomic environment as well as to the households' characteristics. The panel, which is used to track poverty persistence, must be sufficiently long in order to predict the probability of each household type of remaining poor during a given period, since the decay of poverty along time is probably non-linear. Therefore, attrition may appear as a problem when recovering persistence, with a long panel.

In this section, new methods are proposed which necessitate only aggregate information on changes between two periods, information which can be given by a pseudo-panel of repeated cross-sections and which do not necessitate panel data. Once we compute the underreported parts of income, this application will allow us to obtain the estimates of

[7] The Concentration Index (Kakwani 1977, 1980), which is directly related to the concentration curve, does quantify the degree of socioeconomic-related inequality in a health variable (Kakwani, Wagstaff, and van Doorslaer 1997; Wagstaff, van Doorslaer, and Paci 1989, van Doorslaer et al. 1997, van Doorslaer et al. 2006). For the other usages of CI to explain income inequality, see Berrebi and Silber, (1987), Giorgi and Gigliarano, (2016).

[8] For the discussion of the model of estimation and decomposition and of other other indexes see O'Donnel and al.(2008).

persistent poverty using a pseudo-panel for the households participating in informal activities. The persistence of poverty can thus be recovered through four procedures:

1. A direct measure of panel data, as processed by Duncan et al. (1995) and Duncan (1999) on the PSID.
2. Estimating permanent income through the use of a dynamic model and defining a poverty line on monetary grounds by means of the households' permanent income, in turn to estimate permanent poverty (which may not be realized in each period, but which characterizes a large part of the household's life cycle. See Ben Rejeb, 2008 and Ben Rejeb et al, 2006 for examples).
3. Pairing an anticipated probability of poverty computed by means of pseudo-panel data.
4. Crossing the poor population and a grouping of the population in homogenous cells in order to recover poverty, for a later period, by the repartition of the same groups in the population.

This section presents the third method[9]. Poverty is defined by the usual indexes and by the Multidimensional Index of Poverty and Richness (MIPR)[10] as defined by Gardes and Langlois (Cardoso-Gardes, 1996, Gardes-Langlois, 2000). We also use extended monetary income and extended full income for the calculation of the indexes. The details of the two methods used to estimate persistent poverty is presented in Appendix: Two methods to estimate persistent poverty.

4. Micro Data, Matching Statistics

We use two household surveys: the Time use survey (TUS) in 2006 and the Household budget surveys (HBS) for the years between 2007 and 2011 inclusive from the Turkish Statistical Institute (TURKSTAT). The HBS have been conducted on a total of monthly 720 and annually 8640 households for each year. Three basic groups of variables have been obtained from these surveys: Variables of the socio-economic status of the households such as the status of property of house, living in village or in rural areas, etc; variables related to individuals (age, gender, academic background). Consumption expenditures variables (food and non alcoholic beverages, alcoholic beverages with cigarette and tobacco, clothing, health, transportation, education services, etc.) In the TUS in 2006, approximately 390 households were selected each month giving a total of 5070 households during the whole year. Within these households 11 815 members aged 15 years and over were interviewed and were asked to complete two diaries – one for a weekday and one for a weekend day – by recording all of their daily activities during 24 hours at ten-minute-slots. This survey on Time use in 2006 is matched independently on the Family budget survey realizing monetary and time expenditure data. In this application we do not take into account the possible spatial autocorrelation within regions.

We combine the monetary and time expenditures into a unique consumption activity at the individual level. We proceed with the matching of these surveys by using similar exogenous characteristics in both datasets as age, size of household using OECD equivalence scales, proportion of children in the households, matrimonial situation, home ownership, number of

[9] Method 4 is presented in Ben Rejeb, Salah-Matoussi and Gardes, 2009.
[10] This index has been recently included among the poverty indexes computed by Statistic Canada.

household members, geographical location separately for head of household and Female. The selection equation concerns the households which have a positive time use of their activities More precisely, we estimate 8 types of time use in the TUS which are also compatible with the available data from HBS as follows:

1. Food Time (TUS) - Food Expenditures (HBS)
2. Personal Care and Health Time (TUS) - Personal Care and Health Expenditures (HBS)
3. Housing Time (TUS) - Dwelling Expenditures (HBS)
4. Clothing Time (TUS) – Clothing Expenditures (HBS)
5. Education Time (TUS) - Education Expenditures (HBS)
6. Transport Time (TUS) - Transport Expenditures (HBS)
7. Leisure Time (TUS) - Leisure Expenditures (HBS)
8. Other Time (TUS) - Other Expenditures (HBS)

Food Time includes household and family care as the administration of food. Personal Care Time consists of personal care, commercial-managerial-personal services, helping sick or old household person. Housing Time corresponds to household-family care as home care, gardening and pet animal care, replacement of house-constructional work, repairing and administration of the household. Clothing Time consists of washing clothes and ironing. Education Time includes study (education) and childcare. Transport Time consists of travel and unspecified time use. Leisure Time corresponds to voluntary work and meetings, social life and entertainment as social life, entertainment-culture, and resting-holiday, sport activity as physical practice, hunting, fishing etc., sport, hobbies and games as art and hobbies, mass media as reading, TV/Video, radio and music. Other Time includes employment and labor searching times.

3.1. Matching Procedure

Rubin's (1986) matching approach is considered to be distinct from almost all other work on this topic (Moriarity and Scheuren, 2003). The matching procedure proposed by Rubin allows us to overcome two major problems relating to traditional matching methods: in order to impute the monetary expenditure allocated to activity i, denoted x_i, to the time use survey, traditional procedures use the regression coefficients of x_i for the whole dataset (where Z is a set of variables such as age and education common to both dataset). Traditional procedures assume that monetary and time expenditures t_i allocated to activity i are conditional independent given Z, disregarding as a consequence, the possible substitution between monetary and time inputs. Rubin (1986) show that this assumption may bias considerably the regression coefficients. Rubins' concatenation methodology allows to obtain the regression coefficients of x_i on $(1,Z, t_i)$ and t_i on $(1,Z, x_i)$ by assuming a partial correlation value between x_i and t_i given Z (where t_i is time allocated to activity i). Thus, x_i is predicted as a function of t_i and Z, while t_i is predicted as a function of x_i and Z for the whole dataset.

The second problem concerns the decrease in variance of the imputed values since traditional matching procedures smoothes the variation of the individual's expenditure data. As a consequence, inequalities in full income fall, which is a major concern when one seeks to address income inequalities. Rubins's approach matches each unit of the time survey to the observation with the closest predicted values of x_i in the consumer expenditure survey,

conditional on identical characteristics as informed by *Z*. It follows that the observed value of the match is imputed to the missing values.

In order to overcome two aforementioned problems we take into account the concatenation between imputed variables in the time dataset[11]. To summarize the concatenation methodology proposed by Rubin (1986, 1987), the variable *Y* in survey *A* is imputed in survey *B* and the variable *Z* in survey *B* is imputed in survey *A*. The software used for this matching was developed by Alpman (2016). The details of the matching procedure are as follows:

i. We consider three different kinds of variable sets: the first group of variables (*Y*) include the above-explained time use categories in the TUS. The second group (*Z*) represents the expenditure variables in the HBS corresponding to (*Y*) in the TUS. The third set is the common variables (*X*) such as sex, age, marital status, education level, geographic location, employment status, sector of work and type of firm in both surveys. The main hypothesis is that the partial correlation between *Y* and *Z* given *X* is supposed to be other than zero and is denoted thusly: $\rho_{Y,Z|X\neq 0}$.

ii. Thus, the partial variance of *Y* and *Z* given *X*, respectively $\rho_{Y|X}$ and $\rho_{Z|X}$, can be obtained by linear regressions of *Y* and *Z* on *X*. We begin with a linear regression model where *Y* and Z are successively regressed on *X*:

$$Y = a_0 + aX + \epsilon \quad (20\text{-}1)$$

$$Z = b_0 + bX + \mu \quad (20\text{-}2)$$

iii. The partial covariance of (*Y*, *Z*) given *X*, denoted $\sigma_{Y,Z|X}$, can be deduced from $\rho_{Y,Z|X}$ $(\rho_{Y|X} * \rho_{Z|X})^{1/2}$.

iv. Supposing that α and β are the column vectors of the regression coefficients of *Y* on (1,*X*) and *Z* on (1,*X*) respectively, *Y* and *Z* values may be generated for the dataset formed by *A* and *B* by using these regression coefficients. In this prediction, it is assumed that *Y* and *Z* values are conditionally independent for a given *X*. Rubin (1986) applies the sweep matrix operator: sweeping on *Y* gives the regression coefficients of *Z* on (1,*X*,*Y*) while sweeping on *Z* gives the regression coefficients of *Y* on (1,*X*,*Z*). The new regression coefficients are used to create new predicted *Y* and *Z* values for the dataset formed by *A* and *B*.

v. Thus, the predicted *Y* and *Z* are used in the prediction equation for *Y* given *X* and *Z* and in the prediction equation for *Z* given *X* and *Y*. These are the new prediction coefficients used to create new *Y* and *Z* values for the dataset formed by *A* and *B*: each missing unit of *Z* in *A* (and *Y* in *B*) is matched with the closest new predicted *Z* value in *B* (and *Y* in *A*), dependent on identical characteristics informed by *X*.

5. Empirical results

5.1.Informal economy

We estimate a complete demand expenditure system (equation 11) using Generalized Method of Moments (GMM) for both full expenditure (time plus money) and for monetary

[11] We would like to thank A. Alpman for his help in the application of this matching procedure. See a discussion of matching procedure Alpman and Gardes (2016).

expenditure alone. We integrate prices in the equation and the income variables are taken as endogenous. The same estimation is found in by Aktuna-Gunes et al. (2016) which measures the size of the informal economy for the years from 2003 to 2006 inclusive. The control variables included in the model are: the OECD equivalence scale, home ownership, males in white collar occupations, females in white collar occupations, the natural logarithm of the age of household members. We also control two interaction variables such as self-employed males and male wage workers, with male in white collar occupation, having permanent contract and fixed-term contract for male and female. We control the variables such as household type (classified as single, single with children, couples, couples with children and other family types), the dummies like education level of male and female and the durables goods dummies as computer owing, having a good heating system, we also control the number of rooms in the house and number of cell phones. After several experiments, so as to select the appropriate instruments, we choose the following variables: logarithm of the OECD equivalence scale, sex, the natural logarithm of the age of males and females, the square of the natural logarithm of the age of males and females, and the ratio of children to adults in the vector of household characteristics [12].

The estimation of the model for full expenditures and exclusively monetary expenditures from the pooled cross-sectional data covering the period of investigation 2007-2011, is presented in Table A2 and Table A3 respectively in the Appendix[13]. The size of the pooled sample increases to 33,765 households. Only the parameters of the estimates of seven budget share equations are reported in these tables since the parameters of the eighth equation (other goods/services) are redundant due to the adding up condition.

We obtain the size of the informal economy for each year (Table 1) by scaling up the under-reported parameter *k* and *l* (estimated by monetary and full expenditure) with the income part of self employers and wage earners in GDP (Table A4 in Appendix). The corresponding size of informal economy for self-employed workers varies between 32.12% and 29.52% and from 25.34% to 23.30% of GDP between 2007 and 2011, for the monetary and the full expenditure estimations. The size of the informal economy decreases on average by 6.53% (=30.96%-24.43%) owing to time use intensive domestic production of self-employed households.

Table 1: The size of informal economy in Turkey for the years between 2007 and 2011

(In %)

		2007	2008	2009	2010	2011	Average	Total
Monetary	Wage earners	39,21%	41,63%	39,64%	46,04%	46,61%	42,63%	73,59%
Expenditure	Self employed	32,12%	30,98%	31,61%	30,57%	29,52%	30,96%	
Full	Wage earners	42,74%	45,38%	43,21%	50,18%	50,81%	46,46%	70,89%
Expenditure*	Self employed	25,34%	24,45%	24,94%	24,12%	23,30%	24,43%	

** Full expenditure= monetary expenditures+monteary time use values*

Conversely, however, this estimation points out inverse results for wage earners. The corresponding size of the informal economy for wage earners varies between 46.61% and 39.21% and from 50.81% to 42.74% of GDP between 2007 and 2011 for the monetary and

[12] See also Lyssiotou et al. (2004), Aktuna-Gunes et al. (2014).

[13] Based on 2007 year variables, the over-identifying restriction in the estimation is 6.560. Chi-square p value for monetary estimations is 0.83 which is bigger than 0,05 where null hypotheses and the validity of the identifying instruments cannot be rejected for the chosen control variables. We keep the same control variables and don't add new ones in order to compare the results obtained from both estimations.

the full expenditure estimation, respectively. This indicates that the size of the informal economy increases on average by 3.83% (=42.63%-46.46%) due to the commodity intensive domestic production of wage earners' households. Taking into account that domestic activity leads to a decrease in the under-reporting-income ratio and thus in the size of the informal sector (-2.7% = 73.59% - 70.89%).

5.2.Oaxaca's decomposition

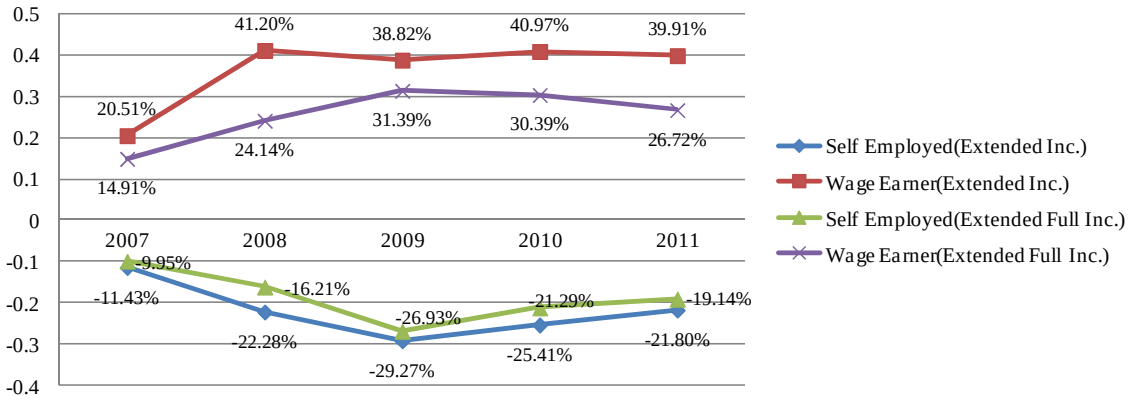
The bigger the informal earnings, the smaller the income inequality! In order to test this assumption we can study the determinants of the time allocation and informal earnings and look for the magnitude and significance of the income inequality coefficients. The results of the Oaxaca decomposition are shown in Table 2.

Table 2: Blinder-Oaxaca Decomposition

		Income Ineqality Gap	Explained Component (in %)	Unexplained Components (in %)
Extended income	Based on wage earners having informal earnings	0.366***	83.3***	16.7***
	Based on Self-Employed having informal earnings	-0.228***	58.4***	41.6***
Extended full income	Based on wage earners having informal earnings	0.262***	92.5***	7.5***
	Based on Self-Employed having informal earnings	-0.191***	125***	-25***

From the decomposition results, we observe that the income inequality gap for wage earners who have informal earnings is larger than the income inequality gap for self-employed workers. However, when we look at the effects of coefficients and endowments separately, we observe that the part of the income inequality gap that can be attributed to discrimination is less important than the endowments effect. The case of self-employed workers is of particular interest, since the income inequality gap seems to be in favor of self-employed populations with informal earnings. Informal earnings disfavor income inequality for wage earners. Therefore, extended full income results indicate that these inverse tendencies in income inequality for both working populations tends to decrease by increase-explained components. The income inequality gap using Oaxaca decomposition for both working groups covering the period 2007 to 2011 is represented in Graph 1.

Graph 1: Trend in the Income Inequality Gap for the Years between 2007 and 2011
(Based on Informal Earnings, and the Blinder-Oaxaca Decomposition)



The income inequality gap using extended monetary income reaches its peak for wage earners in the years 2010 and 2008. Therefore, the gap decreases significantly when we decompose the gap using extended full incomes. The gap took the largest value in 2009 for extended full incomes of wage earners. Generally speaking, the income inequality gap among wage earners increased dependent on the participation in informal activities for the years between 2007 and 2011 inclusive. The gap is decreased for the self-employed who participated in informal activities in this period. The lowest income inequality gap both for extended monetary and full income was in 2009, while a decreasing income inequality gap took the smallest value in 2007 for both extended income groups.

Another interesting finding, which is apparent from Graph 1, is that the differences between the income inequality gap measured through extended income and extended full income for wage earners are notably larger than for the self-employed for all years of observation. This finding clearly shows that the income inequality gap is improved by domestic production while informal earnings alone exacerbate the income inequality gap. Therefore, self-employed workers have the advantage of having flexible working hours which in turn enables them to produce time intensive final goods, yielding an improvement in the income inequality gap. At this point, it is crucial to underline that this result for self-employed workers and wage earners necessitates analysis of different categories of both working groups in order to better identify the different effects of informal earnings and domestic activities on the income inequality gap[14].

The Oaxaca decomposition also shows that, on average, the explained effect accounts for the proportion of the income inequality gap between self-employed workers with informal earnings. The fact is that the Oaxaca decomposition doesn't allow for the measurement of the total effect of informal earnings on income inequality since there are two inverse explained effects when we consider both groups in the estimation. We propose to use CI to overcome this problem.

5.3. The Gini and Concentration Indexes

[14] This is the subject of another work by the authors of this paper.

According to the average Gini coefficient over the 2007-2011 period given in Table 6A, we observe that the informal earnings increase income inequality in Turkey. The index rose to 41.5% when we include informal earnings estimated through declared income into monetary incomes. However, it is surprising to see that our extended full income Gini index is very close to that computed by World Bank (39.4% and 39.1% respectively).

The change and the differential between income inequalities for Turkish households may be driven by participating in informal activities. Therefore, the main advantage of our micro estimation is that we can observe the average share of changes in the income quartile of each household when we include informal earnings in their declared incomes and in full incomes. The details are given in Table 3.

Table 3: Share of changes in income quartiles over 2007 to 2011

	Extended Monetary Income			Extended Full Income		
	Wage Earners	Self-Emloyed	Total	Wage Earners	Self-Emloyed	Total
Remains same	20,49	18,12	19,73	23,90	21,55	23,15
Worse off	39,89	37,33	39,10	37,99	39,61	38,51
Better off	39,60	44,54	41,15	38,10	38,82	38,33
Total difference	*-0,29*	*7,21*	*2,05*	*0,11*	*-0,79*	*-0,18*

The households whose income quartiles remain constant after the inclusion of informal earnings are more important for the extended full income. Therefore, the position of self-employed households in income distribution is worse off for extended full income values. This is the inverse for wage earners with the exception that the change of the respective shares of wage earners becoming better and worse off between extended income and extended monetary income are both decreased.

We can also measure how much inequality as the dispersion of monetary income for each income deciles can be explained through the decomposition of the CI. The CI is an appropriate measure of socioeconomic and informal earnings-related income inequality when the income inequality standard deviation score is measured on a ratio scale with non-negative values. Using the estimation given in equation (17) and the concentration index in (18), the results are given in Table 4.

Table 4: Decomposition of the Concentration Index for the Income Inequality Standard Deviation Score of the Whole Population (2007-2011)

	Extended Monetary Income			Extended Full Income		
	Elasticity	Concentration Indices	Contributions	Elasticity	Concentration Indices	Contributions
Income	3,177	0,378	1,201	3,703	0,355	1,316
Sex	0,370	0,096	0,036	0,177	0,096	0,017
Ln(Education)	0,524	0,111	0,058	0,430	0,111	0,048
Ln(Age)	0,445	0,001	0,000	1,643	0,001	0,001
Urban	0,068	0,107	0,007	0,091	0,107	0,010
Male in white collar occupation	0,093	0,383	0,036	0,020	0,383	0,008
Female in white collar occupation	0,033	0,251	0,008	0,007	0,251	0,002
"Residuals"			-0,0193			-0,0969
Total		1,327	1,346		1,304	1,401

The sign of the concentration index indicates the direction of any relationship between the income inequality variable and having informal earnings. Its magnitude reflects both the strength of the relationship and the degree of variability in income inequality.

We decompose socio-economic and informal earning income-related income inequality for Turkey in the time period of 2007 to 2011 inclusive. The declared incomes' z-scores are multiplied by –1 such that a greater value indicates more inequality. Here we include commune-fixed effects to pick commune-level determinants of income inequality status. A summary of the decomposition of CI is presented in Table 4[15]. The (positive) concentration indices in the last row show that there was inequality in the income inequality standard deviation score to the advantage of households with informal earnings between the indicated years and that this inequality increased over time. The entries in each column are derived from equation (18) and give the elasticity of income inequality standard deviation score for each year, with respect to each factor, the Concentration Index for each factor, and the total contribution of each factor to the income inequality standard deviation score CI. For these years, most of the informal earnings related to inequality in the income inequality standard deviation score is explained by the direct effect of household earnings and by commune-level correlates of both income inequality and income. The large elasticities of the income inequality standard deviation score with respect to these factors are responsible for a large contribution to the income inequality standard deviation score CI. In contrast, there is a great deal of informal earning-related inequality in both with informal earnings (37.8%) and white collar occupation for males and females (38.3% and 25.1% respectively), but there is little sensitivity in the income inequality standard deviation score to variations in white collar occupations for males and females, and consequently they make little contribution to the income inequality standard deviation score concentration index. However, there is very large sensitivity in the income inequality standard deviation score to variation in extended income (120.1%). Education levels and living in a city are also contributors to the income inequality standard deviation score CI. This contribution is increased only by informal earnings for extended full incomes. Residuals are computed by the difference between the CI and the sum of the factor contributions which are close to zero indicating that we have a well-specified model.

5.4. Persistent Poverty

Work in Progress

6. Conclusion

In this paper, we show how the time use values of households may alter the size of the informal economy and which, in turn, determine the position of households in income distribution and poverty in Turkey for the years between 2007 and 2011 inclusive. To this end, we chose to observe two working occupation such -self-employed workers and wage earners- in order to better identify the size of the informal economy. At first, we use a new method to estimate the reporting part of household income on micro cross-sectional data within a complete demand expenditure system (equation 11) by using both the typical, purely

[15] For the estimation results see Table 7A in the Appendix.

monetary approach, and the full expenditures (money expenditures plus time) concept obtained by matching of the classic Family Budget and Time Use surveys. The concept of the full expenditure and the full income gives the opportunity to compute individual (full) prices and enables us to make a more adequate estimation of the Demand System, including all theoretical constraints.

The results show the importance of domestic activities in the estimation of the size of informal economy in a developing economy (Turkey). Our main findings are: (i) domestic activities determine the size of informal economy in two different ways: the domestic production of the households who have more flexible working times available, such as self-employed households, decreases the size of the informal economy. The main argument underpinning this behavior is that the time use substitution elasticity of final good production for self-employed workers would be elastic implying that they have more time intensive domestic production technology than other workers. Thus, the size of the informal economy for self-employed is decreased on average from 30.96% to 24.43% when we consider domestic production. (ii) However, the completely opposite result is apparent for wage earners, which confirm our argument. Inelastic working time hours necessitates wage earners to choose more commodity intensive final good production which causes high level of informal economy. The average size of informal economy among wage earners rose up from 42.63% to 46.46% in Turkey.

The model is well estimated with almost all significant parameters in place. We consider all goods taking into account the domestic production in a complete demand system framework by adding the valuated time use of various activities to the corresponding monetary expenditure. This decreases the average estimation of the size of informal economy in Turkey: in average 73.59% and 70.89% of GDP respectively for full expenditure and monetary ones over the years 2007-2011. Comparing our results for a developing country like Turkey, with Quebec (Fortin et al. (2009)) based on the same methodology shows a very large difference, as the informal sector is about 6% for Quebec in 2002 and 65.6% (using monetary income) and 79.28% (using full income) for Turkey (Aktuna-Gunes et al., 2016) over the years between 2003 and 2006.

In the second step, our interest was to find out whether informal earnings improve income distribution in Turkey. To test this, we first compute the standard deviation score of the self-employed and wage earners based on their declared monetary income derived from each income quartile and later decompose the gap in dispersion of their declared income z-scores according to the classification criterion in whether having informal earnings or not for each group of workers. It is quite interesting to find that the income inequality gap for the wage earners with informal earnings is more significant than the income inequality gap for a self-employed person who has informal earnings. However, we observe that the part of the income inequality gap that can be attributed to discrimination is less important than the endowments effect. In the case of self-employed workers is of particular interest, since the income inequality gap seems to be in favor of the self-employed population having informal earnings. Having informal earnings disfavors income inequality for wage earners. Therefore, extended full income results indicate that these inverse tendencies in income inequality for both working population tends to decrease by increasing explained components. In order to test the effect of domestic production on income inequality gap, we calculated the gap for each year

and for both working groups to see its evolution. The gap for wage earners took the largest value in 2009 for extended full incomes of wage earners. The gap is decreased for the self-employed who participated in informal activities in this period. The lowest income inequality gap for self-employed workers both for extended monetary and full income was in 2009, while the decreasing income inequality gap took the smallest value in 2007 for both extended income groups. However, our findings necessitate the analysis of different categories of both working groups in order to better identify the different effects of informal earnings and domestic activities on income inequality gap which is a question that remains for another research paper.

The fact is that the Oaxaca decomposition, doesn't allow for the measurement of the total effect of informal earnings on income inequality, since there are two inverse explained effects when we consider both working groups in the estimation. First we looked at the Gini Coefficient. It is surprising to see that our extended full income Gini index is very close to that computed by the World Bank (39.4% and 39.1% respectively). Later, we adopt CI to decompose the socio-economic and informal earning income-related income inequality for the time periods of 2007 to 2011 inclusive for Turkey. There is a great deal of informal earning-related inequality in both having informal earnings (37.8%) and white collar occupations for males and females (38.3% and 25.1% respectively), but there is a little sensitivity in the income inequality standard deviation score to variation in white collar occupations for males and females, and so they make little contribution to the income inequality standard deviation score concentration index. However, there is a very large sensitivity in income inequality standard deviation score to variation in extended income (120.1%). Education levels and living in the city also contribute to income inequality. This joint contribution is increased only by informal earnings for extended full incomes. Finally, the contribution of informal earnings on income inequality decreases for extended full income. This result is coherent with our findings regarding the size of informal economy decreased by domestic production.

Persistent poverty estimation is a work in progress.

References

Akalin G., Kesikoglu F. (2007), "Turkiye'de Kayitdisi Ekonomi ve Buyume Iliskisi"(The Relationship Between the Underground Economy and Economic Growth in Turkey), *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 3 (5), p. 71-87.

Aktuna-Gunes, A.T, Gardes F. and Starzec C. (2016), "Informal Markets, Domestic Production and Demand Elasticities: A Case Study for Turkey", under revision at *Economics Bulletin*.

Aktuna-Gunes, A.T, Starzec C. and Gardes F. (2014), "Une évaluation de la taille de l'économie informelle par un système complet de demande estimé sur données monétaires et temporelles", *Revue Economique*, 65(4), p. 567–590.

Alpman, A. (2016), "Implementing Rubin's alternative multiple-imputation method for statistical matching in Stata", *Stata Journal*, 6(3).

Alpman, A. and F. Gardes (2016), "Welfare Analysis of the Allocation of Time During the Great Recession", Working Paper, Centre d'Economie de la Sorbonne (CES), 2015.12- ISSN: p.1955-611X.

Banks J., Blundell R., Lewbel A. (1997), "Quadratic Engel Curves and Consumer Demand", *Review of Economic Studies*, 89 (4), p. 527-539.

Becker, G. (1965), “A Theory of the Allocation of Time”, *The Economic Journal*, 75, p.493–517.

Ben Rejeb, J ., Salah-Matoussi, M., Gardes, F. 2009, La mesure de la pauvreté persistante : de nouvelles méthodologies, *w.p. Université de Sfax, January.*

Bonke, J. (1992), “Distribution of economic resources: Implications of including household production.” *Review of Income and Wealth*, 38,p. 281-293.

Cetintas H., Vergil H. (2003), “Türkiye’de Kayıtdışı Ekonominin Tahmini”(Estimation of Underground Economy in Turkey), *Doğuş Üniversitesi Dergisi*, 4(1), p. 15-30.

Cardoso, F. Gardes, 1996, Caractérisation et Analyse des Comportements de Consommation des Ménages Pauvres sur Données Individuelles Françaises, *Revue Economique*, 687-698.

Duncan G.J.(1999), "The PSID and Me," IPR working papers 99-14, Institute for Policy Resarch at Northwestern University.

Duncan G., Boisjoly J. and Hofferth S. (1995), "L'accès à une première propriété : quels revenus considérer ? Analyse du "Panel Study of Income Dynamics"," *Économie et Prévision*, Programme National Persée, 121(5), p. 65-74.

Eckert-Jaffé O, Gardes F., Starzec C. (2000), The Cost of Children in Poland Using Panel Data, mimeo Ined-Insee, *Income and Wealth Conference*, Août.

Frazis, H. and J. Stewart (2010), “How Does Household Production Affect Measured Income Inequality?” *Journal of Population Economics,*24,p.3-22

Fortin, B., Lacroix, G. and Pinard, D. (2009), “Evaluation de l’économie souterraine au Québec: une approche micro-économétrique”, *Revue Economique*, 60 (5), p. 1257-1274.

Gardes, F. (2016), “The Estimation of Price Elacticity and the Value of Time in a Domestic Production Framework: An Application on to French Micro-data”, Under revision in *Annuals of Economics and Statistics*.

Gardes, F., Gaubert, P., Langlois, S., 2000, Pauvreté et Convergence des Structures de Consommation au Canada, *Canadian Review of Sociology and Anthropology*, n°1, 1-27.

Gardes, F., and Starzec, C. (2015), “Individual prices and household’s size: a restatement of equivalence scales using time and monetary expenditures combined.”, *Revue d’Economie Politique*, 125(3),p.317-474.

Gronau, R. (1986), Home production–a survey. In O. Ashenfelter & R. Layard (Eds.), Handbook of labor economics, Handbooks in economics series, No. 5 (Vol. 1, pp. 273–304). New York, Amsterdam and Oxford: North-Holland, distributed in North America by Elsevier Science.

Ilgin Y. (2002), “Kayıt Dışı Ekonomiyi Tahmin Yöntemleri ve Türkiyede Durum”(The informal economy estimation methods and the case of Turkey). *DPT Planlama Dergisi*, Özel Sayı.

Ilkkaracan A. I., Umut G. (2009), "Time-use, the Value of Non-Market Production and its Interactions with the Market Sector: The Case of Turkey", Paper presented at International Conference on Inequalities and Development in the Mediterranean Countries, *Mimeo*.

Jenkins, S. and O’Leary N. (1996), “Household Income Plus Household Production: The Distribution of Extended Income in the U.K.” *Review of Income and Wealth*, 42(4), p. 401-419.

Kakwani, N. C. (1977), “Measurement of Tax Progressivity: An International Comparison.” Economic Journal, 87(345)p.71–80.

Kakwani, N. C. (1980), Income Inequality and Poverty: Methods of Estimation and Policy Applications., New York: Oxford University Press.

Kakwani, N. C., Wagstaff A., and van Doorslaer E.. (1997), “Socioeconomic Inequalities in Health: Measurement, Computation and Statistical Inference.” Journal of Econometrics, 77(1)p. 87–104.

Lewbel, A. (1990), “Full rank demand systems”, *International Economic Review*, 31, p. 289–300.

Lyssiotou P., Pashardes P., Stengos T. (2004), “Estimates of the Black Economy based on Consumer Demand Approaches”¸ *The Economic Journal*, 114, p. 622-640.

Moriarity, C., and F. Scheuren (2003), “A Note on Rubin’s Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations”, *Journal of Business & Economic Statistics*, 21, p. 65−73.

Oaxaca, R. (1973), “Male-Female Wage Differentials in Urban Labor Markets.” *International Economic Review*, 14,p. 693–709.

O'Donnell, Owen; Van Doorsslaer, Eddy; Wagstaff, Adam; & Lindelöw, Magnus. (2008), *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation*. World Bank Publications.

Ogunc F., Yilmaz G. (2000), “Estimating Underground Economy in Turkey” *Discussion Paper*, Central Bank-Republic of Turkey.

Rubin, D. B. (1986), “Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations”, *Journal of Business & Economic Statistics*, 4, p.87−94.

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*, New York: John Wiley & Sons.Small, 1982.

Schneider F., Enste D. (2000), “Shadow Economies: size, causes and consequences” *Journal of Economic Literature*, 38, p.77-114.

Schneider F., Savaşan F. (2007), "Dymimic Estimates of the Size of Shadow Economies of Turkey and of Her Neighbouring Countries", *International Research Journal of Finance and Economics*, 9, p. 126-143.

Thomas, J. (1999), “Quantifying the black economy: ‘Measurement without theory yet again?”, *Economic Journal*, 109 (456), p.381-389.

Ulgen S, Ozturk U., (2006), *Kayitdisi Ekonomi ve Surdurulebilir Buyume: AB yolunda Degerlendirme ve Cozum Onerileri* (Informal Economy and Sustainable Growth: Solution Proposals on the Road to EU), TUSIAD Buyume Stratejileri Dizisi No: 8, T/2007-01/428.

Us V. (2004), “Kayitdisi Ekonomi Tahmini Yontem Onerisi : Turkiye Ornegi”(A proposal for the Informal Economy Estimation Method: The Case of Turkey), *Turkiye Ekonomi Kurumu*, 2004/17.

van Doorslaer, E., Masseria C., Koolman X., and the OECD Health Equity Research Group. (2006), “Inequalities in Access to Medical Care by Income in Developed Countries.” Canadian Medical Association Journal, 174p. 177–83.

van Doorslaer, E., Wagstaff A., Bleichrodt H., Calonge S., Gerdtham U. G., Gerfin M. , Geurts J., Gross L., Hakkinen U., Leu R. E., O’Donnell O., Propper C., Puffer F., Rodriguez M., Sundberg G., and Winkelhake O. (1997), “Income-Related Inequalities in Health: Some International Comparisons.” Journal Health Econ, 16(1)p. 93–112.

Wagstaff, A., van Doorslaer E., and Paci P. (1989), “Equity in the Finance and Delivery of Health Care: Some Tentative Cross-Country Comparisons.” Oxford Review of Economic Policy, 5(1)p. 89–112.

Wagstaff, A., van Doorslaer, E. (2003), “Catastrophe and Impoverishment in Paying for Health Care: with Applications to Vietnam 1993–98.” *Health Economics*, 12,p.921–34.

Appendix

Table 1A: Descriptive Statistics

Budget Shares	Variable	N	Moyenne	Ecart-type	Minimum	Maximum
Monetary Expenditures	Food	33765	0,2953	0,1481	0	0,9930
	Personal Care(with Health)	33765	0,0823	0,0835	0	1,0000
	Housing	33765	0,3995	0,1590	0	1,0000
	Clothing	33765	0,0529	0,0637	0	0,8424
	Education	33765	0,0157	0,0455	0	0,8726
	Transport	33765	0,1166	0,1347	0	0,9284
	Leisure	33765	0,0251	0,0497	0	0,7868
	Other	33765	0,0127	0,0370	0	0,7920
Full Expenditures	Food	33765	0,1328	0,0800	0	0,9264
	Personal Care(with Health)	33765	0,1509	0,0435	0	0,8018
	Housing	33765	0,1741	0,1077	0,01	1,0000
	Clothing	33765	0,0273	0,0328	0	0,5578
	Education	33765	0,0222	0,0288	0	0,8190
	Transport	33765	0,1110	0,0729	0	0,8460
	Leisure	33765	0,2210	0,0918	0	0,6910
	Other	33765	0,1608	0,1200	0	0,6697
Occupation dummies:		N	Mean	Std Dev	Minimum	Maximum
	Male in white collar occupation	33765	0,1459	0,3530	0	1
	Female in white collar occupation	33765	0,0463	0,2102	0	1
	Male wage worker	33765	0,3079	0,4616	0	1
	Female wage worker	33765	0,0703	0,2556	0	1
	Male self-employed	33765	0,1750	0,3800	0	1
	Female self-employed	33765	0,0378	0,1908	0	1
	Male having permanent contract	33765	0,3082	0,4617	0	1
	Female having permanent contract	33765	0,0707	0,2563	0	1
	Male having fixed-term contract	33765	0,0222	0,1475	0	1
	Female having fixed-term contract	33765	0,0150	0,1216	0	1
	Male not having a diploma	33765	0,1868	0,3897	0	1
	Male primary education	33765	0,1048	0,3063	0	1
	Male secondary education	33765	0,4745	0,4994	0	1
	Male superior education	33765	0,1383	0,3452	0	1
	Male other eduaction	33765	0,0957	0,2941	0	1
	Female not having a diploma	33765	0,8132	0,3897	0	1
	Female primary education	33765	0,0295	0,1693	0	1
	Female secondary education	33765	0,0917	0,2886	0	1
	Female superior education	33765	0,0307	0,1724	0	1
	Female other eduaction	33765	0,0349	0,1835	0	1
Household income share :		N	Mean	Std Dev	Minimum	Maximum
	ln(Total Income)	33765	6,8961	0,9378	0,6931	11,5179
	Other income / Total Income	33765	0,0752	0,1284	0	0,9747
	Self employment / Total Income	33765	0,3117	0,4632	0	1
	Extended (Self employment / Total Income)	33765	0,3937	0,5868	0	1,5940
	Full extended (Self employment / Total Income)	33765	0,3820	0,5721	0	1,6541
	Wage income / Total Income	33765	0,6131	0,4292	0	1
	Extended (Wage income/ Total Income)	33765	0,7423	0,5069	0	1,2918
	Full extended (Wage income/ Total Income)	33765	0,7460	0,5220	0	1,4812
	ln(Total Income) instrumented	33765	6,8423	0,6817	4	9,0783
	ln(Total full Income) instrumented	33765	6,7708	0,6786	4	8,8271
Demographic and regional characteristics:		N	Mean	Std Dev	Minimum	Maximum
	Ln(age)	33765	3,7933	0,2922	2,8904	4,5326
	Household type	33765	2,4743	1,2993	1	5
	OECD equivalence scale	33765	2,2141	0,7613	1	11,3
	City	33765	0,6946	0,4606	0	1
Durables and luxury goods :		N	Mean	Std Dev	Minimum	Maximum
	Home ownership	33765	0,6353	0,4814	0	1
	Number of rooms in the house	33765	3,4991	0,8181	1	10
	Computer	33765	0,3738	0,4838	0	1
	Good heating system	33765	0,2995	0,4581	0	1
	Number of cell phone	33765	2,0024	1,1325	0	9

Table 2A: Results for Monetary Expenditure Based on the Complete Demand System; All Population (GMM) 2007-2011

Variables	Food	*t* - ratio	Pc+Health	*t* - ratio	Housing	*t* - ratio	Clothing	*t* - ratio	Other	*t* - ratio	Transport	*t* - ratio	Leisure	*t* - ratio
Constant	0,921	1,690	2,621	2,680	-8,114	-4,490	0,030	0,030	0,119	0,120	3,606	5,200	0,628	1,210
2007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2008	0,005	2,190	-3,950	0,001	0,017	7,420	-0,006	-4,700	-0,003	-5,390	-0,002	-0,850	-0,002	-2,770
2009	0,000	0,020	4,080	0,001	0,013	5,430	-0,013	-10,220	-0,002	-3,140	-0,001	-0,450	0,002	2,810
2010	0,000	0,150	2,260	0,001	0,016	6,830	-0,013	-10,680	0,000	-0,320	-0,001	-0,470	0,001	1,350
2011	0,001	0,420	5,150	0,001	0,015	6,460	-0,016	-12,080	-0,001	-2,340	0,000	-0,060	0,000	0,020
OECD equivalence scale	0,019	14,500	-0,005	-7,890	-0,033	-27,020	0,011	14,510	0,000	0,420	0,003	2,660	-0,001	-1,780
Home ownership	0,012	7,410	-0,004	-5,110	0,008	4,900	-0,002	-1,930	-0,001	-2,380	-0,011	-7,810	-0,001	-2,660
Male in white collar occupation	-1,026	-2,960	-2,530	0,044	0,559	3,080	-0,030	-0,820	0,150	2,690	0,494	2,790	0,031	0,750
Female in white collar occupation	0,020	5,020	-0,003	-1,250	-0,019	-3,270	0,015	5,960	0,000	-0,130	-0,005	-1,040	-0,008	-4,550
Ln(age)	0,041	8,760	0,019	8,550	0,063	14,190	-0,033	-15,890	-0,008	-6,720	-0,034	-9,400	0,014	11,020
Male self employed x Male in white collar occupation	0,906	2,850	0,169	4,040	-0,499	-3,070	-0,072	-2,000	-0,117	-2,300	-0,413	-2,500	-0,007	-0,190
Male wage worker x Male in white collar occupation	1,132	3,030	1,570	0,047	-0,591	-2,990	0,093	2,330	-0,169	-2,810	-0,568	-3,000	-0,051	-1,140
Male having permenent contract	-0,113	-6,140	0,028	7,070	0,031	2,710	-0,032	-10,130	0,015	4,710	0,063	6,530	0,013	4,930
Female having permenent contract	0,002	0,370	0,006	2,170	-0,022	-4,910	0,009	4,600	-0,002	-1,450	0,011	2,680	-0,005	-3,860
Male having fixed-term contract	-0,023	-1,960	0,020	6,720	-0,013	-1,870	-0,011	-4,290	0,000	0,090	0,022	3,390	0,003	1,960
Female having fixed-term contract	0,005	1,090	-0,003	-1,130	-0,018	-3,580	0,013	5,830	0,000	-0,040	0,011	2,270	-0,006	-3,700
Male don't have education	-3,175	-1,870	-0,891	-1,510	5,816	2,770	-0,478	-0,470	-0,032	-0,020	-0,643	-0,600	0,000	0,000
Male having primary education	-0,027	-2,260	-0,017	-4,300	-0,090	-9,020	0,054	14,720	-0,002	-1,070	0,056	7,670	0,006	2,820
Male having secondary education	-0,019	-1,250	-0,005	-1,270	-0,073	-6,550	0,037	10,530	-0,004	-1,770	0,039	4,980	0,005	2,330
Male having superior education	-0,017	-1,450	0,000	-0,100	-0,044	-4,980	0,024	8,400	-0,006	-2,780	0,028	4,260	0,003	1,640
Female don't have education	-3,182	-1,880	-0,886	-1,500	5,874	2,790	-0,493	-0,490	-0,028	-0,020	-0,669	-0,620	-0,002	-0,730
Female having primary education	0,085	6,390	-0,105	-12,660	0,008	0,560	0,110	18,780	0,000	0,090	-0,071	-5,600	-0,026	-6,120
Female having secondary education	0,029	3,220	-0,069	-12,500	-0,040	-4,000	0,087	20,270	0,003	1,060	-0,008	-0,880	-0,015	-5,100
Female having superior education	-0,014	-2,600	-0,013	-4,040	-0,068	-9,750	0,035	12,100	-0,001	-0,310	0,043	6,660	0,004	1,700
Computer	-0,006	-3,200	-1,600	0,001	0,005	2,670	-0,003	-3,320	0,000	0,500	0,001	0,830	0,000	0,360
Good heating system	-0,009	-5,130	-7,750	0,001	0,043	24,740	-0,007	-6,930	0,000	0,770	-0,018	-10,690	-0,002	-4,480
Number of rooms in the house	-0,003	-3,710	-4,120	0,000	0,009	10,880	-0,002	-4,140	-0,001	-3,790	-0,001	-1,790	0,000	0,170
Urban	-0,021	-4,310	-4,040	0,001	0,069	19,070	0,001	1,110	-0,006	-6,900	-0,029	-10,330	-0,005	-6,480
Household type	-0,001	-1,230	1,810	0,000	0,003	5,440	-0,001	-3,390	0,000	0,150	-0,001	-2,140	0,000	1,860
Number of cell phone	-0,007	-7,820	0,001	3,130	0,004	4,420	0,002	3,410	0,001	2,860	-0,002	-3,110	0,000	-0,460
Full price-Food	-0,270	-55,410	0,017	26,010	0,186	38,880	0,024	19,390	0,007	13,340	0,025	28,860	0,006	15,060
Full price- Pc&Health	0,017	26,010	-0,060	-92,060	0,020	26,680	0,000	-0,380	0,002	13,650	0,013	30,320	0,004	22,110
Full price-Housing	0,186	38,880	0,020	26,680	-0,281	-50,070	0,016	12,680	0,009	14,760	0,036	40,800	0,007	13,900
Full price-Clothing	0,024	19,390	0,000	-0,380	0,016	12,680	-0,045	-52,840	0,000	1,140	0,005	12,620	-0,001	-5,760
Full price-Education	0,007	13,340	0,002	13,650	0,009	14,760	0,000	1,140	-0,001	-2,720	0,003	13,630	0,001	11,450
Full price-Transport	0,025	28,860	0,013	30,320	0,036	40,800	0,005	12,620	0,003	13,630	-0,092	-92,260	0,006	25,540
Full price-Leisure	0,006	15,060	0,004	22,110	0,007	13,900	-0,001	-5,760	0,001	11,450	0,006	25,540	-0,025	-84,360
Full price-Other	0,005	9,090	0,003	15,720	0,008	13,970	0,002	7,890	-0,023	-80,340	0,004	17,380	0,001	8,330
Y	0,740	9,810	-9,780	0,045	0,703	7,390	0,130	4,300	0,008	0,350	-0,847	-14,470	-0,166	-7,640
Y^2	-0,057	-10,670	9,250	0,003	-0,055	-8,090	-0,006	-2,940	0,000	-0,120	0,066	16,050	0,012	7,630

Under-reporting Self-employment (Yr) and Wage earners (Ys)	Parameter	*t* ratio
k (under reporting ratio for yr)	1,418	18,880
v (under reporting ratio for ys)	1,098	24,720

Stock-Yogo weak ID test (endogenous regressor: income)	(Critical values) 2SLS relative bias	>5%	>10%	>20%
Minimum eigenvalue statistic -F(5, 33732) = 17.94		18.37	10.83	6.77

Sargan statistic (overidentification test of all instruments): 6.560 Chi-sq(4) P-val = 0.8335

Table 3A: Results for Full Expenditure Based on the Complete Demand System; All Population (GMM) 2007-2011

Variables	Food	*t* - ratio	Pc+Health	*t* - ratio	Housing	*t* - ratio	Clothing	*t* - ratio	Other	*t* - ratio	Transport	*t* - ratio	Leisure	*t* - ratio
Constant	17,890	0,770	-15,190	-1,990	65,685	4,400	11,528	3,410	-27,908	-1,240	-15,540	-1,250	-18,910	-1,220
2007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2008	-0,006	-1,530	0,003	0,930	-0,010	-1,190	-0,005	-3,060	0,004	0,840	0,004	1,410	0,010	1,920
2009	-0,003	-0,760	0,007	2,620	-0,013	-1,630	-0,008	-5,030	-0,005	-1,010	0,005	1,810	0,014	2,840
2010	0,002	0,670	-0,002	-0,630	0,012	1,540	-0,003	-2,130	-0,008	-1,780	-0,001	-0,380	0,000	-0,090
2011	-0,015	-3,390	0,012	4,290	-0,035	-3,400	-0,013	-6,610	0,015	3,010	0,010	2,800	0,021	4,160
OECD equivalence scale	0,014	6,770	-0,006	-4,880	0,001	0,250	0,006	5,980	-0,001	-0,460	-0,001	-0,280	-0,011	-5,120
Home ownership	-0,012	-3,960	0,005	3,080	-0,016	-2,130	-0,005	-3,710	0,008	2,620	0,004	1,640	0,012	4,430
Male in white collar occupation	-0,080	-2,470	-0,619	-5,520	0,495	3,810	-0,296	-6,040	0,983	5,480	1,510	5,880	-1,985	-5,820
Female in white collar occupation	0,042	4,520	-0,026	-4,720	0,105	5,720	0,027	7,350	-0,043	-4,300	-0,033	-6,090	-0,062	-6,360
Ln(age)	0,002	0,230	0,031	5,100	-0,071	-2,970	-0,038	-8,200	-0,013	-1,120	0,001	0,100	0,121	12,040
Male self employed x Male in white collar occupation	0,108	2,680	0,511	4,600	-0,408	-2,630	0,245	5,110	-0,842	-4,910	-1,304	-5,330	1,689	5,100
Male wage worker x Male in white collar occupation	0,029	0,860	0,715	6,170	-0,631	-4,880	0,322	6,180	-1,019	-5,360	-1,656	-6,130	2,224	6,260
Male having permenent contract	0,001	0,070	-0,046	-7,590	0,053	2,430	-0,023	-4,820	0,055	4,280	0,097	7,720	-0,135	-9,940
Female having permenent contract	0,087	6,900	-0,043	-5,810	0,138	3,970	0,026	4,120	-0,075	-5,690	-0,013	-1,180	-0,107	-11,240
Male having fixed-term contract	-0,003	-0,420	0,006	0,850	0,001	0,080	0,000	0,100	-0,011	-1,070	-0,013	-0,930	0,021	1,110
Female having fixed-term contract	0,104	6,830	-0,051	-5,600	0,134	3,400	0,025	3,440	-0,065	-4,180	-0,005	-0,420	-0,127	-10,380
Male don't have education	0,123	0,010	4,931	0,350	-32,963	-1,420	-5,633	-1,080	9,549	1,640	10,100	0,780	0,000	0,000
Male having primary education	0,307	5,870	-0,164	-6,420	0,553	4,350	0,153	6,590	-0,384	-8,650	-0,100	-2,320	-0,294	-11,120
Male having secondary education	0,340	6,520	-0,179	-7,020	0,641	4,980	0,158	6,770	-0,433	-9,780	-0,153	-3,540	-0,300	-11,950
Male having superior education	0,248	6,630	-0,127	-6,830	0,468	5,040	0,113	6,750	-0,327	-10,190	-0,119	-3,850	-0,208	-11,030
Female don't have education	0,000	0,000	4,987	0,360	-33,133	-1,420	-5,680	-1,090	9,680	1,660	10,141	0,780	0,096	7,190
Female having primary education	-0,392	-5,430	0,231	5,970	-0,725	-4,180	-0,054	-1,750	0,371	5,140	0,151	3,010	0,385	7,120
Female having secondary education	-0,079	-2,330	0,038	1,790	-0,112	-1,340	0,043	2,800	0,026	0,620	0,052	2,210	0,044	1,300
Female having superior education	0,266	6,480	-0,159	-7,510	0,492	4,640	0,119	6,140	-0,311	-8,400	-0,062	-1,760	-0,295	-13,810
Computer	-0,001	-0,460	0,000	-0,290	-0,004	-0,610	-0,003	-1,890	0,005	1,340	0,006	1,790	-0,002	-8,660
Good heating system	-0,015	-5,260	0,003	1,580	0,003	0,380	-0,005	-3,420	0,005	1,640	0,001	0,190	0,007	2,230
Number of rooms in the house	0,000	-0,240	-0,001	-1,460	0,003	1,500	0,000	-0,800	-0,003	-2,400	0,000	-0,240	0,000	0,340
Urban	-0,028	-4,400	0,024	6,500	-0,034	-2,170	-0,011	-3,590	0,018	2,900	-0,031	-5,290	0,053	8,380
Household type	0,004	5,050	-0,002	-3,610	0,008	4,470	0,001	3,070	-0,005	-5,380	0,001	0,940	-0,006	-5,340
Number of cell phone	0,001	1,510	0,000	-0,740	0,007	3,810	0,002	5,410	-0,004	-3,760	-0,005	-5,260	-0,002	-1,190
Full price-Food	-0,043	-14,620	0,007	4,500	0,003	0,450	0,000	0,200	0,012	5,900	0,014	5,790	0,004	3,230
Full price- Pc&Health	0,007	4,500	-0,019	-21,560	0,016	3,890	0,001	1,950	-0,002	-1,450	-0,001	-0,480	-0,002	-3,830
Full price-Housing	0,003	0,450	0,016	3,890	-0,084	-3,810	0,003	0,780	0,012	1,920	0,029	4,170	0,013	5,820
Full price-Clothing	0,000	0,200	0,001	1,950	0,003	0,780	-0,012	-14,860	0,001	0,790	0,005	3,650	0,001	1,930
Full price-Education	0,012	5,900	-0,002	-1,450	0,012	1,920	0,001	0,790	-0,017	-6,260	0,001	0,490	-0,004	-4,090
Full price-Transport	0,014	5,790	-0,001	-0,480	0,029	4,170	0,005	3,650	0,001	0,490	-0,042	-18,440	-0,004	-4,410
Full price-Leisure	0,004	3,230	-0,002	-3,830	0,013	5,820	0,001	1,930	-0,004	-4,090	-0,004	-4,410	-0,005	-5,410
Full price-Other	0,003	3,650	-0,002	-4,680	0,008	3,870	0,001	3,840	-0,004	-5,880	-0,002	-3,670	-0,003	-10,600
Y	-5,333	-6,710	3,099	8,160	-9,707	-5,010	-1,754	-5,070	5,599	8,480	1,617	2,620	5,620	15,500
Y^2	0,389	6,700	-0,228	-8,220	0,711	5,000	0,131	5,150	-0,411	-8,560	-0,114	-2,510	-0,414	-16,020

Under-reporting Self-employment (Yr) and Wage earners (Ys)	Parameter	*t* ratio
k (under reporting ratio for *yr*)	1.184364	32.22
v (under reporting ratio for *ys*)	1.180907	29.73

Table 4A: The income part of wage earners and self employers between 2007 and 2011 (as % of GDP)

Years	Self-Employed	Wage Earners*
2007	0,211	0,355
2008	0,204	0,377
2009	0,208	0,359
2010	0,201	0,417
2011	0,194	0,423

Source : Republic of Turkey Social Security Institution

* Including regular employee

Table 5A: Estimation results of two step Heckman over 2007 to 2011 natural, logarithm of monthly wage

Variables	Male	Female
Ln(Age)	0.942***	1.025***
	(56.58)	(29.67)
Not having a diploma	-0.150***	-0.973***
	(-6.24)	(-19.19)
Primary education	0.263***	-0.689***
	(19.07)	(-20.61)
Secondary education	0.669***	0.204***
	(42.04)	(4.49)
Superior education	1.213***	1.224***
	(66.97)	(30.73)
City	0.174***	0.446***
	(18.17)	(17.66)
Couple	-0.077***	0.560***
	(-4.44)	(30.20)
N. of Child/ N. of Adults	-0.378***	0.441***
	(-11.33)	(11.09)
Ln(Size of hosehold)	0.620***	0.332***
	(37.78)	(17.66)
Constant	-2.657***	-3.477***
	(-47.40)	(-26.93)
λ	0.442***	0.663***
	(10.04)	(9.23)
rho	0.516	0.531
sigma	0.857	1.247
Total observation	45810	21090
Number of workers	34961	11072

Exponentiated coefficients; t statistics in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Table 6A: Average Gini Coefficients between 2007 and 2011 inclusive in Turkey

	Our Estimation	TURKSTAT*	World Bank*
Gini coefficient	-	0,386	0,391
Gini coefficient** (Based on total expenditure)	0,367	-	-
Gini coefficient** (Based on extended income)	0,415	-	-
Gini coefficient** (Based on extended full income)	0,394	-	-

*Including all population

**Computed only for self-employed and wage earners

Table 6A: Blinder-Oaxaca decomposition tables for wage earners and self-employed workers, extended income and extended full income (2007-2011)

6A-1) Blinder-Oaxaca decomposition for self-employed households, extended income (2007-2011)

H: Self-employed earner participated in informal activies			L: Self-employed earner not participated in informal activies		
Mean prediction high (H):	0.301	D is diagonal,			
Mean prediction low (L):	0.073	D:0 => E+(CE+C)			
Raw differential (R) {H-L}:	0.229	D:1 => (E+CE)+C.			
- due to endowments (E):	0.134	D:0.5 =>Cotton, (1998)			
- due to coefficients (C):	0.021	D:0.68 => Reimers, (1983)			
- due to interaction (CE):	0.074	D:* => Neumark,(1988)			
D:	0	1	0.5	0.688	*
Unexplained (U){C+(1-D)CE}:	0.095	0.021	0.058	0.044	0.039
Explained (V) {E+D*CE}:	0.134	0.208	0.171	0.185	0.190
% unexplained {U/R}:	41.6	9.0	25.3	19.2	17.1
% explained (V/R):	58.4	91.0	74.7	80.8	82.9

6A-2)Blinder-Oaxaca decomposition for wage earner households, extended income (2007-2011)

H: Wage earner participated in informal activies			L: Wage earner not participated in informal activies		
Mean prediction high (H):	0.418	D is diagonal,			
Mean prediction low (L):	0.053	D:0 => E+(CE+C)			
Raw differential (R) {H-L}:	0.366	D:1 => (E+CE)+C.			
- due to endowments (E):	0.304	D:0.5 =>Cotton, (1998)			
- due to coefficients (C):	-0.077	D:0.48 => Reimers, (1983)			
- due to interaction (CE):	0.138	D:* => Neumark,(1988)			
D:	0	1	0.5	0.485	*
Unexplained (U){C+(1-D)CE}:	0.061	-0.077	-0.008	-0.006	-0.002
Explained (V) {E+D*CE}:	0.304	0.443	0.374	0.371	0.368
% unexplained {U/R}:	16.7	-21.1	-2.2	-1.6	-0.6
% explained (V/R):	83.3	121.1	102.2	101.6	100.6

6A-3)Blinder-Oaxaca decomposition for wage earner households, full extended income (2007-2011)

H: Self-employed earner participated in informal activies			L: Self-employed earner not participated in informal activies		
Mean prediction high (H):	0.300	D is diagonal,			
Mean prediction low (L):	0.109	D:0 => E+(CE+C)			
Raw differential (R) {H-L}:	0.192	D:1 => (E+CE)+C.			
- due to endowments (E):	0.241	D:0.5 =>Cotton, (1998)			
- due to coefficients (C):	-0.079	D:0.65 => Reimers, (1983)			
- due to interaction (CE):	0.029	D:* => Neumark,(1988)			
D:	0	1	0.5	0.655	*
Unexplained (U){C+(1-D)CE}:	-0.050	-0.079	-0.064	-0.069	-0.043
Explained (V) {E+D*CE}:	0.241	0.270	0.256	0.260	0.235
% unexplained {U/R}:	-25.9	-41.1	-33.5	-35.8	-22.7
% explained (V/R):	125.9	141.1	133.5	135.8	122.7

6A-4)Blinder-Oaxaca decomposition for wage earner households, fullextended income (2007-2011)

H: Wage earner participated in informal activies			L: Wage earner not participated in informal activies		
Mean prediction high (H):	0.320	D is diagonal,			
Mean prediction low (L):	0.057	D:0 => E+(CE+C)			
Raw differential (R) {H-L}:	0.262	D:1 => (E+CE)+C.			
- due to endowments (E):	0.243	D:0.5 =>Cotton, (1998)			
- due to coefficients (C):	-0.045	D:0.67 => Reimers, (1983)			
- due to interaction (CE):	0.064	D:* => Neumark,(1988)			
D:	0	1	0.5	0.674	*
Unexplained (U){C+(1-D)CE}:	0.020	-0.045	-0.012	-0.024	-0.013
Explained (V) {E+D*CE}:	0.243	0.307	0.275	0.286	0.275
% unexplained {U/R}:	7.5	-17.0	-4.8	-9.0	-4.8
% explained (V/R):	92.5	117.0	104.8	109.0	104.8

Table 7A: Concentration Index Estimation Results

	Extended Monetary Income		Extended Full Income	
	Coef.	Std. err.	Coef.	Std. err.
Income	0.000***	0,000	0.000***	0,000
Sex	0.110***	0.005	0.112***	0.007
Ln(Education)	0.085***	0.004	0.111***	0.005
Ln(Age)	0.028***	0.008	0.028**	0.009
Urban	0.023***	0.004	0.038***	0.005
Male in white collar occupation	0.101***	0.005	0.116***	0.006
Female in white collar occupation	0.112***	0.010	0.116***	0.011
Constant	-0.872***	0.0317	-0.896***	0.0386
R-squared	0.874		0.887	
N. of cases	32014		22015	

* p<00.05, ** p<00.01, *** p<00.001

Persistent Poverty: Two methods to estimate persistent poverty

1. *Predicted probability*

Suppose only repeated cross-sections $\mathfrak{M}_t$ and $\mathfrak{M}_{t+1}$ are disposable to evaluate persistent poverty. Poverty, defined by some criterion (such that a poverty line defined by income or the multidimensional index of poverty and richness), separate the poor sub-population $\mathfrak{P}_t$ and the non-poor $\mathfrak{P}'_t$. An pairing between consecutive (non panelized) surveys, say between individuals *h* in *t* and *h'* in (t+1), cannot afford an efficient measure of the change in poverty at the individual level (i.e. the probable change in the status of *h*) because individual determinants are so important to explain the poverty status. Even if these determinants are close between *h* and *h'*, all latent influences (including individual capabilities) explain a large part of this status and impede to apply a pairing method except on an aggregate level. We propose a way to get rid of these latent individual determinants.

Consider equation (A1) where the probability to be poor $\pi_{h,t} = Prob(h \in \mathfrak{P}_t)$, is estimated by a qualitative response model over a set of explanatory variables Z observed at the individual level:

$$\pi_{h,t} = Z_{h,t}\beta_t + \varepsilon_t \qquad \text{(A1)}$$

We suppose that the explanatory variables in Z have been adjusted (for instance for inflation) in order to correspond to the same model of the poverty status for different periods.

The change of this estimate between periods *t* and (*t*+1) depends on the variation of the explanory variables in Z and on the change of the parameters β (which can be related to a changing environment, i.e. an endogeneity of the explanatory variables due a changing relation to latent variables such as expectations or macro variables-the unemployment rate, the individual effect of social policy…). This endogeneity can be taken into account by estimating equation (A1) over the two periods and measuring the change in the probability to be poor by the expression: $d\hat{\pi}_h = dZ_h\hat{\beta}$ but this method mixes the influence of individual

determinants with the change of the environment between the two periods, thus impeding to compare this change between *t* and (*t*+1) with the estimated change for two other periods[16].

In order to recover endogenous and exogenous changes, we apply an Oaxaca decomposition to equation (A1):

$$Z_{h,t+1}\hat{\beta}_{t+1} - Z_{h\prime,t}\hat{\beta}_t = [Z_{h,t+1} - Z_{h\prime,t}]\hat{\beta}_{t+1} + Z_{h,t}[\hat{\beta}_{t+1} - \hat{\beta}_t] \quad \text{(A2)}$$

The first term Δ_1 measures the *exogenous change* due to the difference between the observed individual determinants (conditional to the estimated behavior in the second period), while the second Δ_2 measures the *endogenous change* due to the environment (which modifies the model of poverty through a change in its coefficients β)[17]. The first term gives rise to an estimate of the probability change in *h*'s poverty status using an expectation of its (unobserved) determinants in (*t*+1)[18], $Z^e_{h,t+1}$, and conditional to an invariant model of poverty in the two periods ($\hat{\beta}_{t+1} = \hat{\beta}_t = \hat{\beta}$): $\mathrm{d}\hat{\pi} = [Z^e_{h,t+1} - Z_{h,t}]\hat{\beta} = \Delta_1$

This expectation $Z^e_{h,t+1}$can be made using semi-aggregate information of the changes between the two periods (for instance, it depends on the change for the sub-population $\mathcal{H}$ to which h pertains, this change being measured by means of a pseudo-panel based on the two repeated cross-sections)[19]. The second term measures the endogenous change in poverty due to a different influence of the observed determinants on the poverty status.

2. Application

The method is based on two principles: first, it uses aggregate information (in the computation of $Z^e_{h,t+1}$) in order to generate a dynamics in the static specification (A1). Second, the prediction is made on estimated probabilities and not on the discrete status of poverty. It can be applied by the following procedure:

(i) Estimate β on the union of the two (or more than two) consecutive surveys $\mathfrak{M}_t \cup \mathfrak{M}_{t+1}$. Equation (A1) is estimated with $\pi_{h,t} = 1$ for the poor (as defined by a given criteria) and 0 for the non-poor.

(ii) Suppose the estimated probability is $\hat{\pi}_{h,t} = Z_{h,t}\hat{\beta} = 0.8$ and $\hat{\pi}_{h,t+1} = Z^e_{h,t+1}\hat{\beta} = 0.6$ so that $\mathrm{d}\hat{\pi}_h = -0.2$. These figures indicate a tendency of h's poor status to be weaker in the future: thus this status is less persistent than for another individual who would have a greater estimated probability in (*t*+1). A classification of the population in *t* can be made according to the actual status of individual h in t, crossed with the information

[16] The estimation of β over a long period increases the effect of endogeneity.

[17] Alternative decompositions write: $\hat{\pi} = [Z_{h,t+1} - Z_{h',t}]\hat{\beta}_t + Z_{h',t+1}[\hat{\beta}_{t+1} - \hat{\beta}_t] = [Z_{h,t+1} - Z_{h',t}]\hat{\beta}_t + Z_{h',t}[\hat{\beta}_{t+1} - \hat{\beta}_t] + [Z_{h,t+1} - Z_{h',t}][\hat{\beta}_{t+1} - \hat{\beta}_t][Z_{h,t+1} - Z_{h',t}]\overline{\beta+Z}[\hat{\beta}_{t+1}\widehat{-\beta}_t]$ with $\bar{\beta}$ and $\overline{Z}$ the averages of $\hat{\beta}$ and Z over the two periods. These decompositions have obvious interpretations.

[18] Note that the determinants of probabilities in (*t*=1) are defined as relative to each survey, as they are used to computed the estimated probabilities at that period: for instance, income must be measured relatively to the income distribution for the period, so that it is deflated from one to period to another according to the increase of nominal income for the sub-population to which *h* pertains.

[19] Note that, computed with the average $\bar{\beta}$, formula (2) give the change in the probability to be poor conditional on the model of poverty in t (see the last decomposition in the previous note).

given by its estimated persistence: in the first round, the poor are defined by the MIPR index, which allows to estimate β. In the second round, the poor are defined as those individuals which are poor and which have an estimated probability greater than 0.7. Then β is estimated for this second sub-population till a convergence of the procedure in these parameters.

(iii) Similar differences can be computed for all couples of status i in t and t'. For Logit or Probit estimates, these differences sum to 0 since the estimated probabilities sum to 1 for all periods. The vector of these changes $(\widehat{\pi}_{i,h,t+1} - \hat{\pi}_{i,h,t}), i = 1 \; to \; I$ can be used to characterize the dynamic status of the individual.

(iv) Alternatively, as the anticipated vector $Z^{e}_{h,t+1}$ corresponds to a sub-population $\mathcal{H}$ to which *h* pertain in *t*, the changes in the probabilities can be easily computed for this sub-population. The average persistence of poverty in this sub-population can be attributed to all households in $\mathcal{H}$. These sub-populations (cells) can be defined, either by the crossing of characteristic variables such as age or family structure, or by a (non-linear) neuronal classification method (the classification can be operated over the union of the two surveys in *t* and (*t*+1), then each sub-population in some period is defined as the intersect of the cell defined over $\mathfrak{M}_t \cup \mathfrak{M}_{t+1}$ with each survey).

(v) These estimated differences in probabilities between two consecutive periods form a *transition matrix* which can be multiplied by itself in order to compute the expected change over several periods, which can be compared to average estimated probabilities for $\mathcal{H}$ in (*t*+*n*). Using panel data, the predicted status for an individual *h* in (*t*+*n*) can be compared to the actual measured in this period. Using repeated cross-sections, a similar test can be performed on each cell.

(vi) The procedure can be based on the Oaxaca decomposition of equation (A2) with separate estimation for the two periods. This provides the exogenous and endogenous components of poverty changes.