ECINEQ
Society for the Study of Economic Inequality

Working Paper Series

Measuring inequality of opportunity with imperfect data: The case of Turkey

Francisco H. G. Ferreira
Jérémie Gignoux
Meltem Aran

ECINEQ WP 2010 – 161

ECINEQ 2010-161
March 2010

www.ecineq.org

Measuring inequality of opportunity with imperfect data: The case of Turkey*

Francisco H. G. Ferreira†
The World Bank

Jérémie Gignoux
Paris School of Economics

Meltem Aran
Oxford University

Abstract

The measurement of inequality of opportunity has hitherto not been attempted in a number of countries because of data limitations. This paper proposes two alternative approaches to circumventing the missing data problems in countries where a demographic and health survey and an ancillary household expenditure survey are available. One method relies only on the DHS, and constructs a wealth index as a measure of economic advantage. The alternative method imputes consumption from the ancillary survey into the DHS. In both cases, the between-type share of overall inequality is computed as a lower bound estimator of inequality of opportunity. Parametric and non-parametric estimates are calculated for both methods, and the parametric approach is shown to yield preferable lower-bound measures. In an application to the sample of ever-married women aged 30-49 in Turkey, inequality of opportunity accounts for at least 26% (31%) of overall inequality in imputed consumption (the wealth index).

Keywords: Inequality of opportunity, wealth index, imputed consumption, Turkey
JEL classification: D31, D63, J62

* We are grateful, without implication, to Deon Filmer, Jesko Hentschel, Peter Lanjouw, David McKenzie and participants at the SPO-World Bank Social Policy Workshop in Ankara, on 22 October 2008, for comments on an earlier version of this paper. The views expressed in this paper are those of the authors, and should not be attributed to the World Bank, its Executive Directors, or the countries they represent.

† **Address of correspondence:** fferreira@worldbank.org

1. Introduction

A key development in modern thinking about social justice has been the theoretical incorporation of a central role for personal responsibility into the definition of fairness. Since Rawls's (1971) *A Theory of Justice*, and Sen's (1980) *Tanner Lectures*, political philosophers and economists have begun to ask what might be the right space in which equality should be promoted. A distinction began to be drawn between inequalities that are due to personal responsibility, and which may therefore be ethically acceptable, and those that are not, and which may therefore be classified as unjust.

An important strand of this thinking has argued that *equality of opportunity* provides the appropriate "currency of egalitarian justice" (Cohen, 1989). Society and the State, as its representative, should aim to provide a level playing field, eliminating, to the extent possible, inequalities due to morally irrelevant circumstances, whereas inequality reflecting differences in personal efforts might well be acceptable. Variants of this approach have been proposed by Dworkin (1981), Arneson (1989), and Roemer (1993, 1998). A recent overview of this literature can be found in Fleurbaey (2008).

Economists have also started considering the possibility that the distinction between inequality of opportunity and inequality in the space of outcomes may matter, not only normatively, but also positively. There is considerable evidence, for example, that attitudes to inequality affect attitudes to redistribution, and that the extent and nature of redistribution in turn affect both economic efficiency and equity.[2] And attitudes to inequality may differ depending on whether people perceive income differentials as arising from differences in effort, versus from differences in race, gender or family background. It has also been speculated that inequality of opportunity may be negatively associated with subsequent economic growth, whereas inequality that arises in response to efforts may actually provide useful incentives, and not be detrimental. See Bourguignon, Ferreira and Walton (2007), and Marrero and Rodríguez (2010).

In order to test these ideas empirically, a lively literature has developed on how inequality of opportunity – perforce a somewhat abstract concept – can be quantified and measured in practice. A number of approaches have recently been proposed, following the formal definitions in Roemer (1993, 1998) and van de Gaer (1993). These include

[2] On the first point, see e.g. Alesina et al. (2004). On the second, see, e.g. Bénabou and Tirole (2006).

Bourguignon, Ferreira and Menéndez (2007), Checchi and Peragine (2005), Ferreira and Gignoux (2008), and Lefranc, Pistolesi and Trannoy (2008, 2009).

Although these papers differ in important respects in how they propose to measure inequality of opportunity, they share some common features. In particular, they typically rely on individual- or household-level data on at least two sets of variables: an *advantage* (Roemer's term for an outcome that everyone can reasonably be presumed to value, such as income, wealth, educational achievement, or good health), and a number of *circumstances* (Roemer's term for variables that may be correlated with advantages, but over which individuals cannot exercise any control – such as race, gender or family background).

In practice, most studies have typically used some measure of economic well-being (such as earnings, income or consumption) as an advantage variable, and data on race, parental education and/or parental occupation as circumstances. For many countries, however, even such a limited set of variables is seldom available in the same data set. Specifically, most cross-sectional household income or expenditure surveys do not contain information on the education, occupation or socioeconomic status of the parents of today's adult earners. This limitation has prevented the application of existing techniques for measuring inequality of opportunity in a number of countries.[3]

This paper proposes two alternative methods for measuring inequality of opportunity in settings where a standard household survey does not contain information on the family background of today's adults, but where an alternative survey might. In particular, we explore the use of Demographic and Health Surveys (DHS), which are available for 83 countries around the world, and which contain relatively rich information on parental characteristics for a large subset of the adult population, namely all ever-married women. DHS surveys do not typically contain any estimate of household income or consumption expenditures, but they do include information on the ownership of an array of assets and durable goods, as well as on various indicators of housing quality and access to amenities. These variables have been used in the past to construct composite indicators of household wellbeing and we show how they can also be used to generate

[3] At the very least, these data limitations have sometimes caused researchers to use much older data sets that do contain the information on parents. An example is the use of PNAD 1996 data for Brazil in Ferreira and Gignoux (2008).

lower-bound estimates of inequality of opportunity, either on their own, or in combination with consumption data from a (separate) household expenditure survey.

The first proposed method relies on information from the DHS exclusively, and uses a “wealth index” – constructed as the first principal component of the asset and housing quality indicators – as a composite measure of socioeconomic status (following Filmer and Pritchett, 2001). The second proposed method relies on additional information from a household income or expenditure survey, from which the correlations between consumption and various covariates common to both surveys can be inferred. These correlations are then used to impute consumption expenditures onto the DHS sample, following McKenzie (2005).[4] By construction, each of these methods gives rise to distributions with very different properties, requiring different inequality indices for analysis. For each case we derive suitable measures of inequality of opportunity, and estimate them both parametrically and non-parametrically, along the lines of Ferreira and Gignoux (2008).

Although the two approaches are quite distinct, they do ultimately rely, at least in part, on the joint distribution of asset, housing and amenities indicators in the DHS, and part of our contribution is to compare the ways in which that same underlying information gives rise to different measures, as a result of incorporating data from other sources, or applying different statistical procedures in the analysis.

We compare the two approaches in the context of an assessment of the degree and nature of inequality of economic opportunity among Turkish women, using Turkey’s Demographic and Health Survey (TDHS) and Household Budget Survey (HBS). Our estimates suggest that between one quarter and one third of the observed inequality among women in Turkey is due to unequal opportunities, depending on which method is used. We also propose and describe an “opportunity profile”, which reveals that opportunity deprivation is particularly pronounced in rural areas of the Eastern provinces, and among families headed by people with mothers with no formal schooling.[5]

---

[4] This imputation method may be seen as a simplified version of the “poverty mapping” methodology of Elbers, Lanjouw and Lanjouw (2003).

[5] Turkey is an interesting application not only because of the data configuration and of its interesting geographical and ethnic disparities, but also because it is a country with middling levels of income inequality (Aran et al. 2008 report a Gini coefficient of 0.31 for consumption per equivalent adult), but where people appear to be highly averse to inequality, and to attribute it to “social injustice”. 85% of

The paper is organized as follows. Section 2 briefly summarizes our general approach to the measurement of inequality of opportunity, which is developed more fully in a companion paper (Ferreira and Gignoux, 2008). Section 3 describes the datasets and presents the two alternative indicators of economic advantage that we construct: the "wealth index" and an imputed measure of household per capita consumption expenditure. Section 4 adapts the measure of inequality of opportunity from Section 2 to these alternative indicators, and discusses alternative parametric and non-parametric estimation methods. Section 5 presents the results of the analysis for Turkey. Section 6 introduces the concept of opportunity profiles, and presents our estimates for Turkey. Section 7 concludes.

2. The measurement of inequality of opportunity[6]

Most empirical studies have followed Roemer (1998) in associating inequality of opportunity with that part of inequality which is due to morally irrelevant, pre-determined circumstances, over which individuals have no control, and for which they can therefore not take personal responsibility. Specifically, Roemer proposed that "leveling the playing field means guaranteeing that those who apply equal *degrees* of effort end up with equal achievement, regardless of their circumstances. The centile of the effort distribution of one's type provides a meaningful intertype comparison of the degree of effort expended in the sense that the *level* of effort does not" (1998, p.12, emphasis added).

To see what such a definition implies formally, consider a finite population of agents indexed by $i \in \{1,...N\}$, where each individual i is characterized exclusively by a set of attributes $\{y_i, C_i, e_i\}$, with y denoting an advantage, $\boldsymbol{C}$ denoting a vector of circumstance characteristics, and e denoting an effort level. Let us also follow Roemer (1998) in treating effort as a continuous variable, while the vector $\boldsymbol{C_i}$ consists of J

respondents to the Life in Transition Survey of 2006 felt that the "gap between the rich and the poor was too high" in Turkey, and when asked what was the "main reason why there are some people in need in our country today?" 63% choose "injustice in society" as their answer.

[6] This section is based on and summarizes the more comprehensive discussion in Ferreira and Gignoux (2008).

elements corresponding to each circumstance j (for individual i), with the typical entry being C_i^j. Furthermore, each element C_i^j takes a finite number of values, x_j, $\forall i$.

This permits us to partition the population into what Roemer calls *types*, i.e. population subgroups that are homogeneous in terms of circumstances. This partition is given by $\Pi = \{T_1, T_2, ..., T_K\}$, such that $T_1 \cup T_2 \cup ... \cup T_K = \{1, ..., N\}$, $T_l \cap T_k = \varnothing, \forall l, k$, and the vectors $C_i = C_j, \forall i, j | i \in T_k, j \in T_k, \forall k$. Let $G^k(e)$ denote the distribution of effort and $F^k(y)$ denote the distribution of advantage, each within type *k*. If we assume, as Roemer (1998) does, that advantage *y* is a monotonically non-decreasing function of effort *e*, it follows that the effort and advantage ranks must be the same within each type:

$$G^k(e) = \pi = F^k(y) \tag{1}$$

In this setting, Roemer's definition of equal opportunities as a situation in which levels of advantage are the same for each quantile of the effort distribution across all types (as implied in the earlier quote), can be written simply as:

$$F^k(y) = F^l(y), \forall l, k | T_k \in \Pi, T_l \in \Pi \tag{2}$$

This condition (2) has been presented as Roemer's "strong" definition of equality of opportunity in a number of recent papers, including Bourguignon, Ferreira and Walton (2007), Ferreira and Gignoux (2008) and Lefranc, Pistolesi and Trannoy (2008).[7] In this paper, we follow Ferreira and Gignoux (2008) in adopting a weaker criterion for the *empirical identification* of equality of opportunity, namely that mean advantage levels should be identical across types:[8]

$$\mu^k(y) = \mu^l(y), \forall l, k | T_k \in \Pi, T_l \in \Pi \tag{3}$$

Adopting this weaker empirical criterion for equal opportunities, it follows that the measurement of *inequality* of opportunity should seek to capture the extent to which $\mu^k(y) \neq \mu^l(y)$, for $k \neq l$. This would seem to call for an inequality index defined not on

[7] The definition in (2) is consistent with both the so-called *ex ante* and *ex post* approaches to measuring inequality of opportunity (Fleurbaey and Peragine, 2009). Differences between the two arise only outside equality. The approach we follow here falls within what those authors describe as *ex-ante*.

[8] Equation (3) is evidently much weaker than equation (2). It is not intended to replace (2) as a conceptual definition of equality of opportunity, but simply as an empirical criterion for identifying equality of opportunity in practice, when sample sizes cause the number of observations in each type to be too small to estimate full distributions for each type. See Ferreira and Gignoux (2008) for a discussion of the trade-offs involved in adopting this weaker criterion for empirical analysis.

the marginal distribution of advantages, $\boldsymbol{y} = (y_1,...,y_N)$, but on the corresponding *smoothed* distribution. A smoothed distribution, which we denote $\{\mu_i^k\}$, was originally defined by Foster and Shneyerov (2000), and is obtained from a distribution of advantages $\boldsymbol{y}$ and a partition Π by replacing each individual advantage y_i^k with the type-specific mean, $\mu^k(y)$. A natural scalar measure of inequality of opportunity is then simply given by the share of overall inequality in advantage which is accounted for by inequality in the smoothed distribution defined for a circumstance partition Π:

$$\theta_r = \frac{I(\{\mu_i^k\})}{I(y)} \tag{4}$$

Here, *I()* denotes a scalar inequality measure satisfying the axioms of symmetry, the transfer principle, scale invariance, population replication and additive decomposability.[9] Equation (4) then defines a measure of inequality of opportunity that is at once firmly rooted in Roemer's theory of inequality of opportunity, and quite intuitive. It is simply the between-group share of overall inequality in *y*, where the groups are given by a full partition of the population such that members of each group have identical circumstances: the "between-type inequality share".[10] In Ferreira and Gignoux (2008), where we formally derive this measure (and a closely related absolute index), we also note three of its properties, as follows.

(i) If we require the inequality index *I()* to further satisfy the axiom of path-independent decomposability, then the class of measures given by (4) collapses to a single measure:

$$\theta_r = \frac{E_0(\{\mu_i^k\})}{E_0(y)} \tag{5}$$

where E_0 denotes the mean logarithmic deviation.

---

[9] Formally, $\theta_r : \Omega \times \Lambda \rightarrow [0,1]$, where Ω denotes the space of joint distributions of *y* and *C*, and Λ denotes the space of possible partitions Π of such joint distributions.

[10] The between-group share defined by (4) corresponds to a standard decomposition of inequality by population subgroups, which uses overall inequality among individuals as the denominator. An alternative decomposition, proposed by Elbers et al. (2008), adjusts the reference inequality (the denominator) to take into account the number and relative sizes of groups in the partition. This alternative approach is specially well-suited to identifying the most salient cleavages in a particular society. While we find it less satisfactory as a lower-bound measure of inequality of opportunity – precisely because both the numerator and the denominator are sensitive to the design of the partition – future research should investigate its uses in describing the profile of opportunity.

(ii) θ_r itself satisfies the axioms of population replication, scale invariance, normalization, and within-type symmetry, where the latter two are defined in Ferreira and Gignoux (2008).

(iii) Given that not all relevant circumstances $\boldsymbol{C}$ are ever observed in the data, any empirical partition Π is an incomplete partition in terms of the theoretical full set of circumstances. There may well exist relevant circumstances that lie beyond an individual's own control and that affect their lifetime advantage, but which are not observed in the data. If we did observe them, and were able to further partition the population into groups defined by those variables, the between-group share of inequality might rise, but could certainly not fall. θ_r is therefore a *lower-bound* on the actual share of between-type inequality.

In the remainder of this paper, we apply this measure of inequality of opportunity to a situation where information on the advantage variable y and the circumstance vector $\boldsymbol{C}$ are not directly available in the same household survey, so that either y must be constructed as a composite aggregate of various underlying indicators (our "wealth index" method), or information on y from an ancillary survey must be used to impute it into the main survey containing information on $\boldsymbol{C}$ (our "imputed consumption" method). We compare the two methods in seeking to quantify inequality of opportunity in Turkey.

3. The data and two alternative indicators of economic advantage

In many countries, the analysis of inequality of opportunity is hampered by the fact that no single dataset contains information on both an adequate set of circumstance variables and on the desired advantage variable. This is the case in Turkey, for example. Whereas Turkey's Demographic and Health Survey (TDHS) provides detailed information on circumstances such as family background, place of birth and language/ethnicity, it contains no detailed data on earnings, income or consumption expenditures. The Turkish Household Budget Survey (HBS), on the other hand, provides detailed information on economic outcomes, but not on some of the most important candidate circumstance variables, such as the education of the parents of present-day workers.

We use the TDHS fielded between December 2003 and March 2004 by the Hacettepe Institute. The data were collected from a sample of 10,836 households, representative at the national level but also at the level of the five major regions of the country (the West, South, Central, North and East regions). Information on basic socio-economic characteristics of the population was collected for all household members, and all ever-married women between 15 and 49 years-old also answered a detailed questionnaire on family background, demography and health. 8075 women provided such information.

Although there is very limited information on earnings or consumption, the TDHS (like other DHS surveys elsewhere) collected reasonably detailed data on certain durable goods owned by households, on housing conditions, and on access to amenities. The TDHS survey also contains information on a set of circumstance variables for the sample of ever-married women, namely the region where they were born, the type of area of the place of birth (rural or urban), the levels of education of both the mother and father, the respondent's mother tongue, and the number of siblings[11].

A Household Budget Survey (HBS) was also collected in Turkey in 2003. This survey collected information on basic individual and household characteristics from a nationally representative sample of about 8,500 households. It is the staple survey for assessing the distribution of household consumption expenditures, and thus contains a reasonably detailed questionnaire on that topic, which provides the most reliable estimates of current living conditions for Turkish households. Although the 2003 HBS lacks information on a number of important circumstance variables, it does contain information on durables owned, housing conditions, and access to amenities, comparable to that available in the DHS.

This survey configuration permits two alternative methods to circumvent the missing data problem for measuring inequality of opportunity. The first method is to construct a household "wealth index" on the basis of information contained in the TDHS alone. Wealth indices constructed from DHS information on the ownership of durable

[11] Region was classified into three broad regions: West, Center, and East; the type of area of birth place into rural or urban according to whether the respondent considered it as a village or sub-district or not; parental education into four categories: no education or unknown level, primary, secondary, and higher education; mother tongue into Turkish or another language; and number of siblings into: less than 3, 4 to 5, 6 to 8, 9 or more.

goods (such as fridges, TV sets, cars, computers, etc.), on housing characteristics (such as the type of roof materials and floor cover), and on access to utilities (such as water and sanitation) have been widely used in estimating household welfare and in ranking households for targeting purposes.[12]

Following Filmer and Pritchett (2001), we define our "wealth index" as the first principal component of a vector of assets ***x*** (including durables, housing characteristics and utility access indicators) owned by households in the TDHS sample. In some cases, such as the floor material in the dwelling, or access to sanitation or water sources, there is arguably an ordinal nature to the alternative categories. In those cases it is statistically preferable to treat those variables explicitly as ordinal in the analysis (Kolenikov and Angeles, 2009). We therefore rank order the categories for those variables and aggregate categories for which there is ambiguity about the ranking, and in this regard our treatment differs slightly from the original Filmer-Pritchett method.

For each household *i*, the wealth index is given by:

$$y_i = \sum_p a_p \left(\frac{x_{pi} - \bar{x}_p}{s_p} \right) \qquad (6)$$

where the *p*-dimensional vector ***a*** is chosen so as to maximize the sample variance of *y*, subject to $\sum_p a_p^2 = 1$. *s* denotes a standard deviation, and the overbar denotes a mean.

Table 1 describes the elements underpinning Turkey's household wealth index, by listing each element of the vector ***x***, as well as its mean and standard deviation. In practice, we compute two (slightly different) wealth indexes: the main index uses the full set of asset variables available in the TDHS, and the subsidiary index uses only the asset variables that are also available, in an exactly comparable format, in the HBS (the "common set"). The subsidiary index is calculated to facilitate the comparison between the two methods being proposed. The last two columns of Table 1 present the scoring factors for each element of ***x*** in the TDHS sample (the vector ***a***), divided by the standard deviation, for the two asset indexes. The standard interpretation is that ***a*** yields the set of

[12] See Filmer and Scott (2008) for a recent (and sanguine) assessment of the robustness of household rankings based on asset indices originating from DHS information, when compared, inter alia, to detailed consumption expenditure data.

weights providing the maximum discrimination between households in the sample, in terms of their ownership of these particular assets ($\boldsymbol{x}$).[13]

McKenzie (2005) lists a number of reasons why an asset index such as this might in fact be preferable to consumption or income as a basis for inequality measurement, including the likelihood that recall bias might be smaller for asset ownership questions than for some income or expenditure questions. But he also highlights two potential pitfalls in using asset indices, namely the possibilities of truncation and clumping. Whereas truncation would most likely arise from not observing assets capable of distinguishing either the very poor from those just above them, or the very rich from those just below them, clumping might be caused by using too few assets, leading to "false modes" in the distribution, arising from insufficient discriminating power in the index. Figure 1 plots the superimposed histogram and kernel density estimate for our main asset index, revealing the absence of both truncation and clumping.

The second method we propose to circumvent the missing data problem relies on a simple statistical procedure for combining information on circumstances from the TDHS with information on consumption from the HBS. Ultimately, since the link between the two surveys is provided largely by components of the asset index (and a few additional covariates), this second exercise can be seen as an alternative way of using information on assets to measure inequality of opportunity in Turkey. Our approach here closely follows McKenzie (2005) in imputing consumption from the HBS into the TDHS, using a bootstrap prediction method.[14]

This procedure consists of combining a direct prediction based on a regression model, with a repeated draw of residuals comparable to a bootstrap. The relationship between wealth indicators X and per capita consumption c is estimated, on sample S_a (from the auxiliary HBS survey), using a log-linear regression model:

$$\ln(c) = X\beta + w\gamma + \varepsilon \tag{7}$$

[13] The TDHS data files contain a pre-constructed asset index, supposedly also given by (6). As the survey documentation does not describe the details of how that index is constructed, best research practice generally involves computing the index from the underlying data, as we have done here. The correlation coefficient between our main index and the TDHS index is 0.98, and the kernel density functions for both indices are very similar, although the kernel for our main index is considerably smoother.

[14] This approach is a simplified version of the consumption imputation procedures proposed by Elbers, Lanjouw and Lanjouw (2003).

where w are demographic controls. The estimation of (7) provides the fitted coefficients $\hat{\beta}$ and $\hat{\gamma}$ as well as estimated residuals $\hat{\varepsilon}$. In order to reproduce the observed levels of inequality, the imputation of per capita consumption into sample S_m (the "main" DHS survey) is constructed by adding the linear prediction, $X\hat{\beta} + w\hat{\gamma}$, and a prediction of the residual $\tilde{\varepsilon}$. The predicted residual $\tilde{\varepsilon}$ is drawn, for the sample S_m of the main survey, from the empirical distribution of residuals obtained when fitting (7) to the auxiliary sample S_a. The procedure allows for heteroskedasticity by drawing $\tilde{\varepsilon}$ from the distribution of residuals for households with similar assets[15]. This is done in six steps:

(1) The regression in (7) is estimated using the common set of wealth indicators, and the parameters $\hat{\beta}$, $\hat{\gamma}$ and residuals $\hat{\varepsilon}$ are obtained.

(2) The sample S_a of the HBS survey is divided into G = 10 groups, defined according to the deciles of the distribution of the first principal component (the wealth index) *y* for the set of wealth indicators common to the two surveys.[16] Separate distributions of the predicted residuals are identified for each of the 10 groups.

(3) The sample S_m of the DHS survey is then divided into the same 10 groups, using the same cut-off values for *y* as in the auxiliary sample.

(4) For each household *i* in group *g* in S_m, a residual $\tilde{\varepsilon}_i$ is drawn from the empirical distribution of residuals for households in group g in S_a. The imputed value of per capita consumption is given by:

$$c_i = \exp\left(\hat{\beta}' x_i + \hat{\gamma}' w_i + \tilde{\varepsilon}_i\right) \tag{8}$$

(5) Measures of inequality of opportunity are computed using the imputed distribution of per capita consumption.

(6) Following the bootstrap principle, steps (4) and (5) are repeated for a number *R* of drawn replicate distributions of the residuals, and the measures of inequality of opportunity are computed as the mean over the measures obtained for each replication. In our analysis, we use *R*=20 replications. This replication process allows averaging out the

[15] Heteroskedasticity might stem from a non-linear relationship between log consumption and wealth assets, or from the higher noise in this relationship for richer households than for poorer ones.

[16] We partition the sample into 10 groups in order to allow for a sufficiently high degree of heteroskedasticity, while keeping group sizes to the order of a few hundred observations.

bootstrap sampling error.

The set of wealth indicators common to the DHS and HBS surveys contains 14 variables for ownership of durable goods, and four variables for housing characteristics and access to utilities. A variable indicating the ownership of agricultural land, and nine variables for demographic controls and regional dummies are also included. Table 2 presents descriptive statistics for those variables in the two samples. The results for the regression of per capita household consumption on these variables in the HBS sample are then presented in Table 3. We use a log linear specification because of the likely nonlinear relationship between the ownership of assets and consumption.

Per capita consumption is then imputed using the fitted coefficients $\hat{\beta}$ and $\hat{\gamma}$ presented in Table 3 and the draws of the residuals. The descriptive statistics in Table 2 suggest that the set of regressors used for the imputation have similar distributions in the two samples.[17] Figure 2 depicts kernel density estimates of the distributions of total household consumption observed in the auxiliary HBS sample, and imputed in the main TDHS sample.[18] The two distributions have reasonably similar shapes, and the levels of inequality in actual consumption in the HBS and in imputed consumption in the TDHS are also close: for the sample of 30-49 year-old women, on which our analysis focuses below, the E(0)s are 0.337 and 0.360 respectively.

4. Estimating inequality of opportunity with missing data

We have now constructed two alternative economic advantage variables for each household in the TDHS sample. Both are based on information on "wealth" (as proxied by a vector of ownership indicators for assets and durable goods, housing quality, and access to amenities), although the second variable also uses information from an auxiliary survey on how those assets and a few other covariates correlate with measured consumption. (Crucially, this information includes the residuals of the consumption regression on the covariates common to both surveys.)

[17] Significant differences are found only for the share of urban residence because of the difference in the definitions of urban areas in the two surveys (agglomerations with 20,000 inhabitants for the HBS survey and 15,000 for the TDHS one), and access to piped water (the definition is more restrictive in the DHS).
[18] The distribution of imputed consumption in the TDHS that is shown corresponds to the first one of the R=20 draws.

In principle, we could now apply our scalar index for inequality of opportunity in equation (5) to the joint distribution of each of these variables (*y*), and the circumstance vector (***C***). However, the mean log deviation used in (5) is not suitable for measuring inequality in the distribution of the "wealth index" given by equation (6). By construction, this index is distributed with mean zero and a variance equal to the largest eigenvalue in the correlation matrix of ***x***. These properties mean that most standard inequality measures routinely used for income or consumption are unsuitable for the wealth index *y*. A zero mean impedes computation of most relative inequality measures (which generally divide by the mean), including the Gini coefficient and all members of the Generalized Entropy class. Negative values are problematic for logarithm-based measures (such as the mean log deviation, the Theil - T index, the variance of logarithms, and many others.)

For analyzing inequality in the wealth index, the simplest solution is to revert to the variance, which is straightforwardly decomposable and is also translation invariant. Since our general measure of inequality of opportunity (in equation 4) is by construction a *ratio* of inequality measures, the problem of scale dependence will vanish for the opportunity index, and the (related) issue of mean dependence would seem to be of no import for a variable that has mean zero *by construction*. We thus set $I(y) = Var(y)$ in (4), and our proposed measure of the "between-type inequality" in the "wealth index" is given by:

$$\hat{\theta}_r^N(\Pi) = \frac{\sum_k \frac{n_k}{n}(\bar{y}_k - \mu)^2}{Var(y)} \tag{9}$$

Since $Var(y) = \sum_k \frac{n_k}{n} \underset{i\in k}{Var}(y_i) + \sum_k \frac{n_k}{n}(\bar{y}_k - \mu)^2$, it is clear that (9) corresponds to the between-group share in a standard variance decomposition. Furthermore, since the weights in both the within-group and the between-group terms are simple population shares, and do not include income levels or shares, (9) describes a path-independent decomposition in the Foster-Shneyerov (2000) sense.

Equation (9) can obviously be computed non-parametrically from partition Π (hence the superscript N).[19] All that is required is the population share and mean wealth index for each cell of the partition, as well as the overall mean and variance for the complete sample. However, as the dimension of the circumstance vector $\boldsymbol{C}$ (J) and the number of discrete values that each element C_j can take (x_j), rise, the number of types in the partition ($K = \prod_{j=1}^{J} x_j$) increases geometrically. Naturally, for a given sample size, the precision of the estimates of group means will fall as J and x_j rise.

If the number of cells with fewer than 10 observations or so is non-trivial, it becomes worthwhile to estimate (9) parametrically. Following Ferreira and Gignoux (2008), this is done by estimating a linear regression of y on the circumstance vector $\boldsymbol{C}$:

$$y = C\psi + \varepsilon \tag{10}$$

Under the maintained functional form assumption in (10), a parametric estimate of the opportunity share of inequality $\hat{\theta}_r^P(\Pi)$ is given simply by the R^2 of (10), or:

$$\hat{\theta}_r^P(\Pi) = (\operatorname{var} y)^{-1}\left[\sum_k \psi_k^2 \operatorname{var} C_k + \frac{1}{2}\sum_k \sum_j \psi_k \psi_j \operatorname{cov}(C_k, C_j)\right] \tag{11}$$

Like most other parametric approaches in econometric estimation, this procedure economizes on data requirements, at the cost of making a functional form assumption. As discussed in Ferreira and Gignoux (2008), we see the parametric and non-parametric estimators as complementary: while the latter may suffer from imprecise estimation of mean advantage levels for types with low sample density, the former make functional form assumptions. The fact that they are empirically quite similar (as we will see in Section 5) provides some sense of methodological robustness. Just like its non-parametric counterpart in equation (9), $\hat{\theta}_r^P(\Pi)$ is a lower-bound estimate on the set of possible estimates for inequality of opportunity. If an additional element of C, which is presently omitted, were to become observable, the R^2 of (11) might rise, but it would not fall.

The parametric approach also allows for an additional decomposition: namely that of the total share of the variance due to the vector $\boldsymbol{C}$, into the components due to each element of the vector. These *partial shares* of inequality of opportunity, associated with

[19] The hat denotes that this is a sample estimate, and the subscript r distinguishes the relative measure from its absolute analogue, which is defined in Ferreira and Gignoux (2008).

each individual element C_j of the vector of circumstances, are computed using the regression coefficients from (10) and are defined as:[20]

$$\hat{\theta}_r^J(\Pi) = (\operatorname{var} y)^{-1}\left[\psi_J^2 \operatorname{var} C_J + \frac{1}{2}\sum_k \psi_k \psi_J \operatorname{cov}(C_k, C_J)\right] \tag{12}$$

Inspection of (12) immediately reveals that, for any given partition, these partial shares sum up to the overall parametric estimate of between-group inequality, given by (11). Besides this attractive additive decomposability property, this definition of circumstance-specific shares also satisfies a path-independence property. Although we have already noted that the overall *non-parametric* decomposition (9) is path-independent by construction, *parametric* estimation of the *partial shares* – based respectively on the smoothed and standardized distributions – are not the same.[21] However, as we show in the Appendix, equation (12) is the simple average between the direct and residual estimates of the partial shares, which correspond to the smoothed and standardized distributions, respectively. It is therefore a simple example of a Shapley decomposition, where averaging across alternative paths eliminates path-dependence. See Shorrocks (1999), and Foster and Shneyerov (2000).

Our second proposed advantage variable, namely imputed consumption, does not share the distributional peculiarities of the asset index. Imputed consumption c_i takes only positive values, so that equation (5) can be applied directly. The main advantage of using the mean log deviation (rather than the variance) in this case, is that the distributions of imputed consumption do not have mean zero by construction, so that mean- or scale-independence becomes, once again, a desirable property for *I()*. Moreover, unlike the variance, the mean log deviation also satisfies the principle of decreasing transfers, a

---

[20] Note that the estimates of the partial shares rely on the validity of the specific reduced-form coefficients ψ. They are not, therefore, lower-bound estimates like the measures in (9) or (11). They are meaningful estimates of the contribution of a particular circumstance to inequality of opportunities only under the much stronger assumption that those coefficients are unbiased, i.e. that any circumstance variables omitted from the reduced-form regression $y = C\psi + \varepsilon$ are orthogonal to $\boldsymbol{C}$. While we report some of the partial shares given by (12) in Section 5, we do not insist much on them, given this strong caveat.

[21] Just as a smoothed distribution is obtained from a vector y and a partition Π by replacing every y_i^k with the type-specific mean, $\mu^k(y)$, a standardized distribution is obtained by multiplying every y_i^k by $\mu \;/\mu^k(y)$.

possibly desirable property for a measure of economic inequality.[22] Using this index, our proposed measure of "between-type inequality" in imputed consumption is given by:

$$\hat{\theta}_r^N(\Pi) = \frac{\sum_k \frac{n_k}{n} E^0\left(c_i^k\right)}{E^0(c)} \tag{13}$$

which is simply a way of rewriting equation (5).

As in the case of the wealth index, we compute this share non-parametrically (using equation 13), as well as parametrically. In this case, given that the empirical distribution of residuals is approximately lognormal, the parametric estimate uses a log-linear specification of the relationship between circumstances and per capita consumption:

$$\ln c = C\varphi + \varepsilon \tag{14}$$

Just as the estimates of ψ from equation (10) could be used to implement a decomposition of overall inequality of opportunity into partial shares corresponding to individual circumstance variables, a similar procedure can be followed with estimates of φ (although these are not additive in the same way).[23] They are subject to the same caveats which applied to the partial shares for the wealth index, and are reported in the next section merely as a description of the data. Finally, in order to facilitate the comparison of results between the two methods (wealth index and imputed consumption), we also calculate equation (13) using the variance, as well as the mean log deviation. The results are discussed in the next section.

5. Results

This section presents our empirical estimates of inequality of opportunity as the between-type share of inequality in the "wealth index" and in imputed consumption, and compares the two sets of results. As discussed above, these estimates rely on statistical analysis of the joint distribution of each advantage variable with a comprehensive set of *circumstance* variables. To qualify as a circumstance in Roemer's sense, variables must

[22] This principle requires that, the lower the region of the distribution where a transfer occurs, the more it will reduce the level of inequality (Shorrocks and Foster, 1987).
[23] See our companion paper, Ferreira and Gignoux (2008), for the derivation of partial circumstance shares using a parametric estimation procedure and the mean log deviation as the inequality measure.

be impossible for the individual himself to affect by choice. Given the information available in the TDHS, our vector of circumstances consists of the type of area in which the woman was born, the region where she was born, her mother's and father's levels of education, her mother tongue, and the number of siblings the individual reports having. The discrete categories for each variable, as well as the distribution of the population across them, are presented in Table 4.

Table 5 reports the results of regressions (10) for the wealth index and (14) for log imputed consumption, on those circumstance variables. Regressions are reported for both the main wealth index (using the full set of asset variables) and the subsidiary wealth index (which uses the set of variables common to both surveys) described in Table 1. For the regressions in Table 5 and for all of the analysis that follows, the TDHS sample is restricted to ever-married women aged 30-49. Results for the full sample of every-married women (whose ages span 15-49 in the survey) are available from the authors on request, but are not reported here because early marriage is selective on circumstance variables.[24]

Since this is a reduced-form regression, coefficients should not be interpreted causally. They reflect partial correlations between individual circumstance variables and the household's wealth index (or imputed consumption), conflating both direct and indirect effects (e.g. through efforts). Nevertheless, the regression is informative. The share of explained variance, $\hat{\theta}_r^P(\Pi)$, is 27% for the main wealth index, 30% for the subsidiary index, and 26% for imputed consumption, suggesting broadly similar "between-type" shares of inequality, regardless of the aggregation method.

Being born in an urban area, having Turkish as mother tongue, and having more educated parents are all associated with higher adult levels of "wealth" and consumption. A greater number of siblings is associated with lower subsequent economic advantage. Perhaps most interestingly, once these circumstances are controlled for, there is only limited evidence of an association between birth region (at the three-region level) and

[24] In other words, the composition of the sample for younger women is particularly sensitive to whether they were born in the East or West, and to different kinds of families, leading to potential sample selection biases. This problem arises because detailed information on family background is collected in the TDHS only for women who are currently married or have been married in the past. Nevertheless, the results for the 15-49 age range are not very different from those reported here for the preferred sample.

economic advantage: only one of six possible regional coefficients is significant: the one for birth in the West region, in the imputed consumption regression.

Our measures of inequality of opportunity among ever-married Turkish women (aged 30-49) are presented in Table 6. This table summarizes results for both of our alternative methods ("wealth index" and imputed consumption), and presents both parametric and non-parametric estimates. In order to facilitate the comparison between the two methods, a number of "intermediate" alternatives are also presented. The first and second columns present the estimates for the main and subsidiary wealth indexes. The next four columns present estimates for imputed consumption, both with imputed residuals (using the bootstrap procedure described in Section 3) and without, and using both the variance and the mean log deviation (E_0 or MLD) as inequality aggregators. For each column, the first line simply reports the total inequality in the outcome variable. The second line reports the non-parametric estimate of between-group inequality, while the third line gives its parametric analogue.

As discussed in Section 4, our preferred estimates of inequality of opportunity are those given in the first and sixth columns. The first column uses the full wealth index as the advantage indicator, and the ratio of variances as the measure of inequality of opportunity (equations 9 or 11). The sixth column uses full predicted consumption (with imputed residuals) as the advantage indicator, and the ratio of mean log deviations as the measure of inequality of opportunity (equation 13). These are two alternative meaningful advantage indicators that one might construct, given the data available in a Demographic and Health Survey and an ancillary household survey (in this case the HBS), analyzed with the measures ideally suited for each. Parametric (non-parametric) estimates of inequality of opportunity are 31% (36%) for the wealth index, and 26% (32%) for imputed consumption.

However, examination of the full set of estimates sheds useful light on the implications of the various methodological choices: (a) the use of a wealth index or imputed consumption as the outcome variable, (b) the use of a full or reduced set of asset indicators to construct the wealth index, (c) the inclusion of draws for the residual term when imputed consumption in used, and (d) the reliance on the variance or mean log deviation in the estimates of inequality of opportunity in consumption.

For the wealth indexes in the first and second columns, the non-parametric estimates are consistently larger than the parametric ones, by about five percentage points in each case. These differences are consistent with the expected imprecision in the sample estimates of cell means in equation (9), owing to the fine partition of a finite sample. Since the exercise aims to derive lower-bound measures of inequality of opportunity as a share of observed wealth inequality, it seems preferable to rely on the parametric estimates in line 3 (from equation 11) as our benchmark result. This yields a tight range of 30% - 31% for the two variants of the wealth index.

Non-parametric estimates are also considerably larger than parametric ones for all four columns using imputed consumption as well, suggesting that one's choice of parametric estimation to generate lower-bound measures of inequality of opportunity is robust to the advantage indicator, at least in this application. Looking across the four consumption columns, it is clear that the opportunity shares are considerably higher (as high as 37%) when the residuals are not included in the consumption imputations. This was to be expected, since omitting the residuals excludes a large amount of heterogeneity which is uncorrelated with the observed covariates. Looking only at the parametric estimates for full imputed consumption (i.e. including residuals) in columns 4 and 6, we find shares of 20% using the variance and 26% using the MLD. As discussed in the previous section, an estimate based on the scale-invariant MLD measure seems superior to one based on the variance, for this advantage indicator.

Setting aside the differences due to the inequality aggregator (variance versus MLD), it would appear that the gap between our preferred measures of inequality of opportunity for ever-married women in Turkey, namely 31% for the wealth index and 26% for imputed consumption, is driven, at least in part, by differences in the information used to generate the two advantage indicators. The difference between a quarter and (almost) a third is not trivial, to be sure. But neither is it worrying large, once one acknowledges that the advantage concepts are actually intrinsically distinct: the wealth index relies exclusively on more permanent indicators, such as assets and durable goods owned, housing characteristics, and access to amenities like running water and sanitation.

There is very little transitory consumption in the building blocks of this index, whereas there is much more in the imputed consumption indicator, particularly when the residuals are included. This is very clear from a comparison of columns 1, 2, and 3: when the residuals are not imputed, and the same inequality measure is used, the opportunity shares are very similar: 30%, 31% and 33%. It is the inclusion of the residuals that drives most of the difference between our preferred estimates in columns 1 and 6. While this surely reflects, at least in part, differences in the methodological and statistical procedures employed, such as principal components analysis and two-sample regression-based imputation, a plausible claim can be made that it also reflects, at least in part, a real difference in the nature of the advantage variable being investigated, with a greater weight for transitory components in the imputed consumption variable.

The bottom panel of Table 6 reports the partial shares of overall inequality associated with each individual circumstance included in the partition. These shares are computed using (12) for the variance, and an analogous procedure described in Ferreira and Gignoux (2008) for the mean log deviation As noted earlier, these shares are included here purely for descriptive purposes, and should not be interpreted causally in any way.

Although there are differences in the absolute numbers, both the broad orders of magnitude and the relative importance of each circumstance are fairly similar between columns 1 and 6. Whether a Turkish woman is born in an urban or rural area appears to be a powerfully associated with her economic advantage as an adult. More than a third of the overall (lower-bound) opportunity share of wealth inequality is accounted for by this circumstance alone. Parental education follows, both for the wealth index and for imputed consumption, although the order between the two is reversed in the two cases. Taken together, they are more important than rural/urban birth in accounting for the overall share.

Mother tongue and number of siblings follow. The number of siblings result, with roughly 10% (20%) of the share of overall wealth (consumption) inequality accounted for by circumstances is not trivial, particularly when considering that this is after controlling for the education of both parents, as well as the geography of birth. As before, and despite the salience of regional differences in the literature on Turkey, the three-way (East, Center, West) partition of the country has only a limited importance in accounting for inequality in opportunity for economic advantage, once a few other basic circumstances are controlled for.

6. Opportunity profiles: identifying the least advantaged groups

The partition of the population into *types* (circumstance-homogeneous groups), that was used above to compute lower-bound measures of inequality of opportunity, can also be used to shed light on the distribution of opportunities among Turkish women in a more direct and disaggregated manner. We know from equation (2) that equality of opportunity requires that advantage distributions be identical across types. Differences in wealth or consumption distributions among types, therefore, are taken to reveal (or arise from) inequality of opportunity.

The cardinal measures presented in the previous section rely fundamentally on differences across conditional *means*. Because of sample size restrictions, it is impossible to estimate density or distribution functions for all 768 types used in our decomposition. But it is still informative to look at more aggregated conditional distributions, where the population is partitioned into groups by one specific circumstance at a time. Figure 3 plots kernel density estimates for the “wealth index” distribution for various such “aggregated types”: women born in rural versus urban areas in panel (a), women born in each of the three main regions in panel (b), women with parents with different educational backgrounds in panels (c) and (d); women with different mother tongues in (e), and women with different numbers of siblings in panel (f).

These conditional wealth distributions differ markedly across these social groups, and not only in means, but in other moments and in general shape as well. Women born in the East, or in rural areas, are evidently at a considerable disadvantage. Those whose mothers and fathers had achieved secondary education or higher, conversely, tend to enjoy much higher levels of wealth in adult life, as do native Turkish speakers. Such pronounced disparities across advantage distributions that are conditional on exogenous, pre-determined circumstances, is prima-facie evidence of the inequality of opportunity for which we estimated (lower-bound) scalar measures in the previous section.

At least conceptually, it is not unreasonable to see the support of such conditional distributions as an individual *i*'s ($i \in k$) *opportunity set* for outcome *y*, and $F^k(y)$ as the probability distribution associated with the opportunity set. After all, given *i*'s circumstances, only *i*'s own efforts and luck will determine his final position $\pi_i = F^k(y_i)$.[25] If it were possible, therefore, to rank conditional distributions $F^k(y)$ across *k* in a meaningful way, we would obtain a ranking of opportunity sets across types.

At the level of disaggregation implicit in Figure 3, one could of course look for robust rankings across conditional distributions by means of stochastic dominance relationships (see Lefranc et al., 2008). However, such broad groupings may be less useful for policymakers interested in identifying pockets of exclusion than a more detailed profile, that exploits the full K = 768 cells in the fine partition of the population analyzed in Section 5. Although the corresponding conditional distributions cannot be plotted and stochastic dominance relationships cannot be established given the sample size, the types can still be ranked by a particular moment of their conditional distributions. While this is certainly less robust than a dominance-based ranking, there are offsetting gains in terms of the ability to generate a complete ranking of types by their opportunity sets, and in terms of a much sharper description of the disadvantaged groups.

Following Ferreira and Gignoux (2008), we rank each type in our fine partition by the mean of its conditional advantage distribution, $\mu^k(y)$. This is consistent with our criterion for the empirical identification of equality of opportunity, given in (3). Once types are so ordered, the circumstances which define them constitute an *opportunity profile*. A little more formally, we define an *opportunity profile* as the ordered partition $\Pi^* = \{T_1, T_2, ..., T_K\} | \ \mu^1 \leq \mu^2 \leq ... \leq \mu^K$, corresponding to any original partition *Π*. This is simply an ordered set of types, ranked by their mean level of advantage.

To focus on the worst-off types, we further define an *opportunity-deprivation profile* as a subset of *Π** that includes only a certain fraction *π* of the population that belongs to the lowest-ranked types. Formally:

$$\Pi^*_\pi = \{T_1, T_2, ..., T_j, ..., T_J\} | \mu^1 \leq \mu^2 \leq ... \leq \mu^J \ ; \ \mu^J < \mu^k, \forall k > J \ ; \text{ and } \sum_{j=1}^{J-1} N_j \leq \pi N \leq \sum_{j=1}^{J} N_j$$

---

[25] Luck is absent from Roemer's (1998) conceptualization of equality of opportunity, which we summarized very briefly in Section 2. However, see Lefranc et al. (2009) for an illuminating discussion of luck and inequality of opportunity.

If, for example π=0.1, then $\Pi^*_{0.1}$ is simply the ordered set of types, ranked by mean advantage, up until the type that brings the population share of the set over 10%. Table 7 lists the circumstances that define the types in $\Pi^*_{0.1}$ for our sample of 30-49 ever-married women in Turkey, when imputed consumption is chosen as the relevant advantage. Such a detailed profile might permit identifying those groups most deserving of policy support, from the perspective of a Rawlsian social planner who adopted an equality of opportunities perspective to define social groups. See Roemer (2006), and Bourguignon, Ferreira and Walton (2007).

While Table 7 describes each type in the opportunity-deprivation profile individually, Table 8 better summarizes the composition of the bottom and top tenths of the (consumption) opportunity profile in Turkey. This table reveals that 99% of those women in the most advantaged group were born in urban areas, while 88% of the bottom tenth was born in rural areas. 95% of the bottom tenth of the opportunity profile was born in Eastern provinces, and 97% had mothers with no formal education whatever. A similar proportion was born in households where Turkish was not the primary language spoken, and over 70% had six or more siblings. The contrast between the two columns in Table 8 is stark: when Turkish women are ranked by the mean imputed consumption of their types, and we look at the bottom and top tenths of the ensuing distribution, they come from strikingly different backgrounds, geographically, educationally and ethnically.

7. Conclusion

Rising interest in inequality of opportunity among both normative and positive economists has led to various recent attempts to measure it empirically. However, because the measurement of inequality of opportunity generally requires reasonably detailed data on both a measure of advantage (such as income or consumption) and on a set of pre-determined background circumstances (such as parental education, wealth or occupation), these attempts have run afoul of data limitations in a number of countries. The most common problem has been the absence of information on the parents of today's adults in the same surveys that document the incomes or consumption expenditures of those adults.

This paper proposes two alternative statistical approaches to circumvent this missing data problem, for those cases where a Demographic and Health Survey (DHS) is available. The first approach relies on the DHS alone, and uses a "wealth index" as the Roemer *advantage* variable. This index is computed as a principal component of a vector of assets and durable goods owned, housing characteristics and access to amenity indicators. The second approach relies on an additional, ancillary survey, and imputes a measure of consumption from that survey into the DHS.

Once these advantage variables are constructed, we apply an intuitive measure of inequality of opportunity developed in a companion paper (Ferreira and Gignoux, 2008) to their distributions: the "between-type inequality share". The measure relies on a partition of the population by a small set of observed circumstances which can be confidently interpreted as completely independent of individual choices: region and area of birth, the educational attainment of both parents, mother tongue, and the number of siblings a person grew up with. Because this is an incomplete set of circumstances, the inequality share is interpreted as a lower bound on inequality of opportunity.

Since the wealth index and the imputed consumption distributions are rather different statistical constructs, different versions of the between-type inequality share are calculated for each indicator. A ratio of variances is used for the zero-mean wealth-index distribution, while a ratio of mean-log deviations is used for the distribution of imputed consumption. These measures are estimated both parametrically and non-parametrically, but the parametric approach yields preferable lower-bound estimators, given sample-size restrictions.

In an application of these methods to the sample of ever-married women aged 30-49 in Turkey, we found that inequality of opportunity accounts for at least 26% of total inequality in predicted consumption, and 31% of total inequality in the wealth index. We attribute the difference between these two numbers primarily to the greater transitory or unexplained heterogeneity that is present in the consumption, but not in the wealth measure. This is consistent with the fact that the between-type inequality share is much higher for imputed *predicted* consumption (i.e. without imputed residuals). Non-parametric estimates are higher for both advantage indicators.

Partial circumstance shares are also computed for each method, and are interpreted purely as descriptions of the data. Rural versus urban birth, and parental education appear to be the main correlates of future economic advantage, both when measured in terms of a wealth index and of imputed consumption. The language spoken at home and sibship size are also important. Interestingly, once the aforementioned circumstances are controlled for, the broad geographical region in which a woman was born (Eastern, Central or Western) appears less important. Since wealth distributions do differ substantially across these regions (as do consumption and education levels), this finding suggests that such differences are due to heterogeneity in the composition of the population across regions, in terms of the other circumstances, rather than to any intrinsic regional effects.

The paper also explores the *opportunity profile* for Turkey, constructed by ranking household *types* by their mean level of imputed consumption. Once households are so ranked, the bottom 10% of the distribution is 88% rural and 96% Eastern (by birth). 97% of them hail from non-Turkish speaking households, and the same share had mothers with no formal education. 84% had fathers with no formal schooling, and 70% had six or more siblings. The contrast with the top tenth of the opportunity distribution was striking along every dimension.

Such marked differences in economic opportunity across groups defined by morally irrelevant and pre-determined characteristics might explain, at least in part, why Turks appear relatively inequality averse, despite a middling position in the world's ranking of consumption inequality. Perhaps more importantly, the opportunity profile of social groups, constructed on the basis of these pre-determined circumstances, might be useful to Turkish policymakers as they seek to target scarce resources and policy attention with the aim of fostering a more inclusive growth process.

Appendix

Table 6 reports partial shares of inequality of opportunity, associated with each individual element C_j of the vector of circumstances $\boldsymbol{C}$. These partial shares, which in the variance decompositions are computed through equation (12), using the regression coefficients from (10), have the attractive property that they sum up to the total share of inequality of opportunity computed through equation (11), using the same regression coefficients.

This appendix shows that (12) is a simple average of the two alternative paths of the variance decomposition. It therefore corresponds to the Shapley value decomposition proposed by Shorrocks (1999). This explains its additive decomposability.

Recall that $y = C\psi + \varepsilon$ (10').

Therefore $$\operatorname{var}(y) = \sum_j \psi_j^2 \operatorname{var} C_j + \frac{1}{2}\sum_k \sum_j \psi_k \psi_j \operatorname{cov}(C_j, C_k) + \operatorname{var} e \qquad \text{(A1)}$$

The partial contribution of a particular circumstance C_J to var (y) can be calculated in two alternative ways. Both focus on the first two terms in (A1), i.e. set var (e) = 0. The *direct* estimate holds all $C_j, \forall j \neq J$ constant in (A1), and computes the remaining variance as a share of the total:

$$\hat{\theta}_{dir}^J = \frac{\psi_J^2 \operatorname{var} C_J}{\operatorname{var} y} \qquad \text{(A2)}$$

The indirect, or *residual*, estimate takes holds C_J itself constant, and takes the difference between var (y) and the ensuing variance:

$$\hat{\theta}_{res}^J = \frac{\operatorname{var} y - \operatorname{var} \tilde{y}^J}{\operatorname{var} y} = \frac{\operatorname{var} y - \left[\sum_{j \neq J} \psi_j^2 \operatorname{var} C_j + \frac{1}{2}\sum_{j \neq J}\sum_{k \neq J} \psi_j \psi_k \operatorname{cov}(C_j, C_k) + \operatorname{var} e\right]}{\operatorname{var} y} = \qquad \text{(A3)}$$
$$= \frac{\psi_J^2 \operatorname{var} C_J + \sum_k \psi_k \psi_J \operatorname{cov}(C_k, C_J)}{\operatorname{var} y}$$

Taking the average between (A2) and (A3) yields (12):

$$\frac{1}{2}\left(\theta_d^{PJ} + \theta_r^{PJ}\right) = \frac{\psi_J^2 \operatorname{var} C_j + \frac{1}{2}\sum_k \psi_k \psi_J \operatorname{cov}(C_k, C_J)}{\operatorname{var} y} = \hat{\theta}_r^J(\Pi) \ \blacksquare$$

References

Alesina, Alberto, Rafael Di Tella and Robert MacCulloch (2004): "Inequality and happiness: are Europeans and Americans different?", *Journal of Public Economics*, **88**: 2009-2042.

Aran, Meltem, Sirma Demir, Ozlem Sarica and Hakan Yazici (2008): “Poverty and Welfare Changes in Turkey 2003-2006”, The World Bank Turkey Office, mimeo.

Arneson, Richard (1989): “Equality of Opportunity for Welfare”, *Philosophical Studies,* **56**: 77-93.

Bénabou, Roland and Jean Tirole (2006): “Belief in a Just World and Redistributive Politics”, *Quarterly Journal of Economics,* **121** (2): 699-746.

Bourguignon, François, Francisco H.G. Ferreira and Marta Menéndez (2007): “Inequality of Opportunity in Brazil”, *Review of Income Wealth,* **53** (4): 585-618.

Bourguignon, François, Francisco H.G. Ferreira and Michael Walton (2007): “Equity, Efficiency and Inequality Traps: A research agenda”, *Journal of Economic Inequality* **5**: 235-256.

Checchi, Daniele and Vitoroco Peragine (2005): “Regional Disparities and Inequality of Opportunity: the Case of Italy”, *IZA* Discussion Paper, 1874/2005.

Cohen, Gerry A. (1989): “On the Currency of Egalitarian Justice”, *Ethics,* **99**: 906-944.

Dworkin, Ronald (1981): “What is Equality? Part 2: Equality of Resources What is Equality? Part 2: Equality of Resources”, *Philosophy and Public Affairs,* **10** (4): 283-345.

Elbers, Chris, Jean O. Lanjouw and Peter Lanjouw (2003): “Micro-level Estimation of Poverty and Inequality”, *Econometrica,* **71** (1): 355-364.

Elbers, Chris, Peter Lanjouw, Johan Mistiaen and Berk Özler (2008): “Reinterpreting between-group inequality”, *Journal of Economic Inequality,* **6** (3): 231-245.

Ferreira, Francisco H. G. and Jeremie Gignoux (2008): "The measurement of inequality of opportunity: theory and an application to Latin America", *Policy Research Working Paper Series,* 4659, The World Bank.

Filmer, Deon and Lant Pritchett (2001): “Estimating Wealth Effects Without Expenditure Data – or Tears: An application to educational enrollments in states of India”, *Demography,* **38** (1): 115-132.

Filmer, Deon and Kinnon Scott (2008): "Assessing asset indices", *Policy Research Working Paper Series,* 4605, The World Bank.

Fleurbaey, Marc (2008): *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press.

Fleurbaey, Marc and Vito Peragine (2009): "Ex ante versus ex post equality of opportunity", ECINEQ Working Paper 2009-141.

Foster, James and Artyom Shneyerov (2000): "Path Independent Inequality Measures", *Journal of Economic Theory,* **91**: 199-222.

Kolenikov, Stanislav and Gustavo Angeles (2009): "Status Measurement with Discrete Proxy Variables: Is principal component analysis a reliable answer?", *Review of Income and Wealth*, **55** (1): 128-165.

Lefranc, Arnaud, Nicolas Pistolesi and Alain Trannoy (2008): "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?", *Review of Income and Wealth,* **54** (4): 513-546.

Lefranc, Arnaud, Nicolas Pistolesi and Alain Trannoy (2009): "Equality of Opportunity and Luck: Definitions and testable conditions, with an application to income in France", *Journal of Public Economics* **93**: 1189-1207.

Marrero, Gustavo A. and Juan G. Rodríguez (2010): "Inequality of Opportunity and Growth". Universidad Rey Juan Carlos, Spain. Mimeo.

McKenzie, David (2005): "Measuring Inequality with Asset Indicators", *Journal of Population Economics,* **18**: 229-260.

Rawls, John (1971): *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Roemer, John E. (1993): "A pragmatic theory of responsibility for the egalitarian planner", *Philosophy and Public Affairs* **22**: 146-166.

Roemer, John E. (1998): *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Roemer, John E. (2006) "The 2006 World Development Report: Equity and Development – A Review", *Journal of Economic Inequality*.

Sen, Amartya (1980): "Equality of What?", in S. McMurrin (ed.) *Tanner Lectures on Human Values*. Cambridge: Cambridge University Press.

Shorrocks, Anthony (1999): "Decomposition Procedures for Distributional Analysis: A unified framework based on the Shapley Value", University of Essex, mimeo.

Shorrocks, Anthony and James E. Foster (1987): "Transfer Sensitive Inequality Measures", *Review of Economic Studies,* **54**: 485-97.

Van de Gaer, Dirk (1993): *Equality of Opportunity and Investment in Human Capital*. PhD dissertation. Catholic University of Leuven, Belgium.

Figure 1: The Main Household Asset Index for Turkey: density

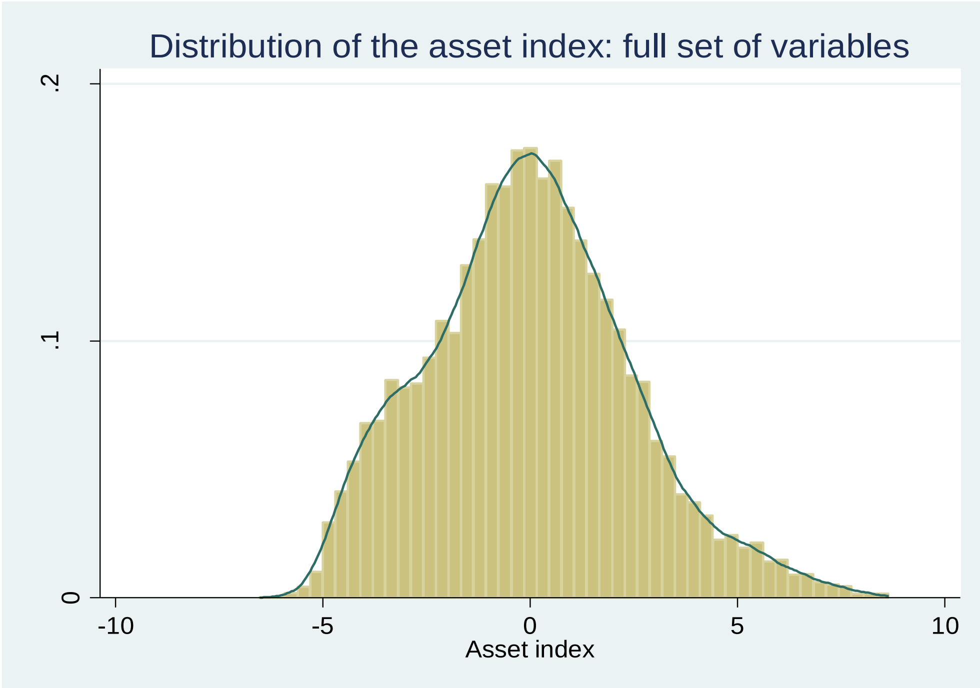

Figure 2: Distribution of annual household consumption expenditures: observed in HBS 2003 and imputed in TDHS 2003

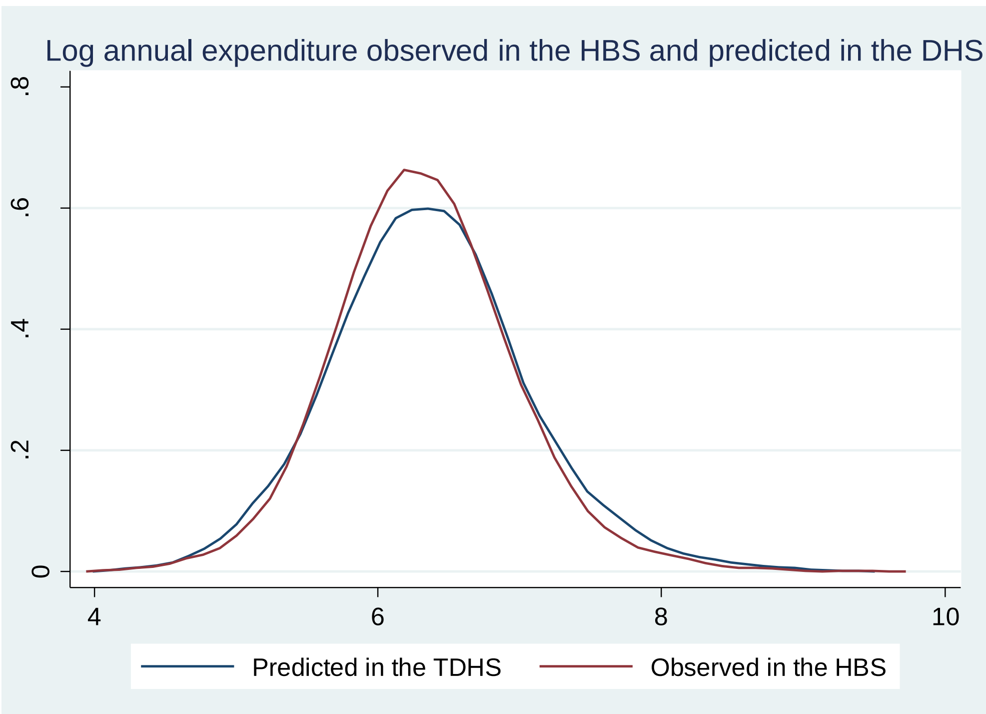

Figure 3: Household Wealth Distributions for Different Circumstance Groups in Turkey:

Kernel Density Estimates

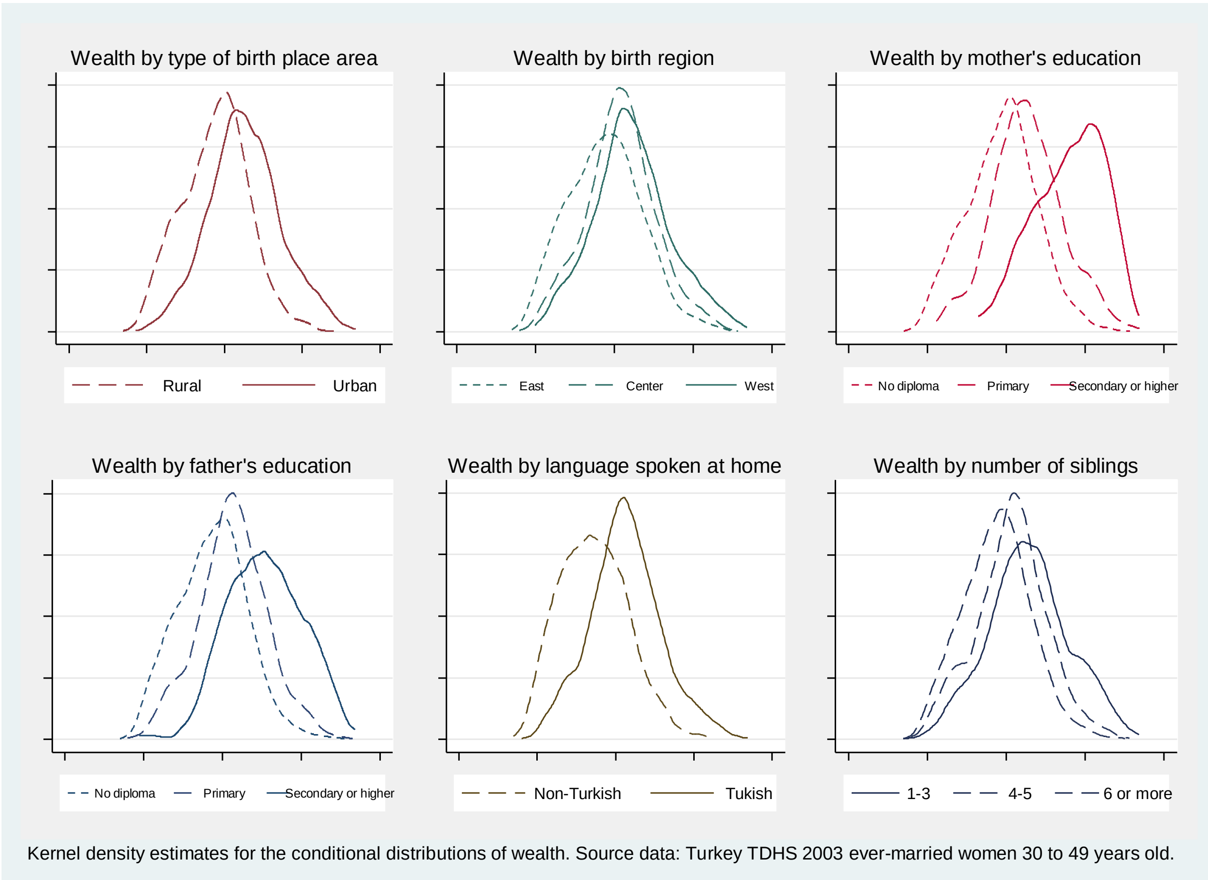

Table 1: The Household wealth index

Principal components and summary statistics for asset indicators

Variable	Mean	Std. Dev.	Scoring factor (/sd): full set of variables	Scoring factor (/sd): common set of variables
Has gas or electric oven	0.712	0.453	0.234	
Has microwave oven	0.072	0.259	0.138	0.191
Has dishwasher	0.221	0.415	0.257	0.331
Has blender/mixer	0.392	0.488	0.269	
Has DVD/VCD player	0.317	0.465	0.218	0.258
Has washing machine	0.783	0.412	0.243	0.265
Has video camera	0.035	0.184	0.140	0.208
Has iron	0.851	0.356	0.221	
Has satellite antenna	0.143	0.350	0.106	
Has vacuum cleaner	0.756	0.429	0.263	
Has air conditioner	0.047	0.212	0.140	0.205
Has television	0.947	0.223	0.128	0.138
Has video	0.073	0.259	0.153	0.212
Has cable TV	0.062	0.240	0.164	0.240
Has camera	0.339	0.473	0.249	
Has CD player	0.182	0.386	0.205	
Has cellular phone	0.671	0.470	0.223	0.252
Has computer	0.116	0.320	0.222	0.316
Has internet	0.063	0.242	0.196	0.295
Has private car	0.258	0.437	0.195	0.251
Has motorcycle	0.045	0.208	-0.009	-0.026
Has bicycle	0.193	0.394	0.116	
Works own or family's agricultural land	0.137	0.344	-0.136	-0.182
Source of water for drinking (ordered variable)	0.501	0.861	0.105	
Piped water inside dwelling	0.742	0.437		0.244
Type of toilet (ordered variable)	0.675	1.946	0.224	
Toilet inside dwelling	0.782	0.413		0.266
Type of floor material in dwelling (ordered variable)	0.041	0.520	0.219	
Dwelling is owned by a household member	0.620	0.485	-0.043	-0.047
Dwelling is rented	0.248	0.432	0.062	
Dwelling is a lodging	0.014	0.118	0.039	0.043
No rent paid for dwelling	0.116	0.321	-0.031	
Other type of dwelling	0.002	0.040	-0.017	-0.043
Number of members per sleeping room	2.412	1.223	-0.133	
				-0.165
Number of members per room	1.325	0.872		
Observations	10,836			

Notes: mean and standard deviation of the ownership, access to amenities and dwelling characteristics (full and reduced) set of variables, and scoring factors for the first principal components, divided by the standard deviation.

Table 2: Descriptive statistics for the asset indicators and demographic variables common to the HBS and DHS samples

	DHS 2003			HBS 2003		
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Oven	10,836	0.072	0.259	25,764	0.063	0.242
Dishwasher	10,836	0.221	0.415	25,764	0.229	0.420
Dvd player	10,836	0.317	0.465	25,764	0.200	0.400
Washing machine	10,836	0.783	0.412	25,764	0.801	0.399
Video camera	10,836	0.035	0.184	25,764	0.022	0.146
Air conditioner	10,836	0.047	0.212	25,764	0.028	0.166
TV	10,836	0.947	0.223	25,764	0.971	0.168
Video	10,836	0.073	0.259	25,764	0.068	0.251
Cable TV	10,836	0.062	0.240	25,764	0.060	0.237
Cellular	10,836	0.671	0.470	25,764	0.545	0.498
Computer	10,836	0.116	0.320	25,764	0.093	0.291
Internet	10,836	0.063	0.242	25,764	0.036	0.187
Car	10,836	0.258	0.437	25,764	0.241	0.428
Moto	10,836	0.045	0.208	25,764	0.026	0.158
Agricultural land	10,836	0.137	0.344	25,764	0.131	0.338
Piper water	10,836	0.742	0.437	25,764	0.932	0.251
Toilets inside	10,836	0.782	0.413	25,764	0.884	0.321
House owned	10,836	0.620	0.485	25,764	0.719	0.449
House lodge	10,836	0.014	0.118	25,764	0.013	0.115
House other	10,836	0.118	0.323	25,764	0.051	0.221
Household members per room	10,836	1.325	0.872	25,764	1.271	0.699
Log household size	10,836	1.301	0.538	25,764	1.308	0.485
Number of children 0 to 4	10,836	0.382	0.686	25,764	0.344	0.626
Number of children 5 to 14	10,836	0.802	1.134	25,764	0.845	1.106
Female household head	10,836	0.125	0.331	25,764	0.096	0.295
Age of head	10,836	47.218	15.071	25,764	46.841	13.658
Squared age of head (/10)	10,836	24.566	15.525	25,764	23.806	13.867
Years of education of head	10834	6.518	4.541	25,764	6.662	3.474
Squares years of education of head (/10)	10834	6.310	7.615	25,764	5.645	5.227
Urban area	10,836	0.705	0.456	25,764	0.638	0.481
Region 2: West Marmara	10,836	0.051	0.220	25,764	0.050	0.218
Region 3: Aegean	10,836	0.153	0.360	25,764	0.153	0.360
Region 4: East Marmara	10,836	0.089	0.285	25,764	0.088	0.284
Region 5: West Anatolia	10,836	0.099	0.299	25,764	0.097	0.296
Region 6: Mediterranean	10,836	0.128	0.334	25,764	0.134	0.341
Region 7: Central Anatolia	10,836	0.056	0.229	25,764	0.062	0.240
Region 8: West Black Sea	10,836	0.065	0.247	25,764	0.066	0.248
Region 9: East Black Sea	10,836	0.038	0.191	25,764	0.044	0.205
Region 10: Northeast Anatolia	10,836	0.028	0.165	25,764	0.028	0.165
Region 11: Central east Anatolia	10,836	0.041	0.199	25,764	0.041	0.198
Region 12: Southeast Anatolia	10,836	0.071	0.257	25,764	0.067	0.251

Notes: Statistics given for the full samples of households of each survey.

Table 3: Regression of annual household consumption on covariates in the HBS

Coefficient	Log annual expenditure	(cont'd)	
Oven	0.08***	Number of children 0 to 4	-0.03***
	[0.02]		[0.01]
Dishwasher	0.19***	Number of children 5 to 14	-0.05***
	[0.010]		[0.00]
Dvd player	0.08***	Female household head	0.02
	[0.01]		[0.01]
Washing machine	0.22***	Age of head	0.02***
	[0.01]		[0.00]
Video camera	0.27***	Squared age of head (/10)	-0.01***
	[0.03]		[0.00]
Air conditioner	0.24***	Years of education of head	0.01***
	[0.03]		[0.00]
TV	0.17***	Squared head educ. (/10)	0.01***
	[0.02]		[0.00]
Video	0.05***	Urban area	0.07***
	[0.02]		[0.01]
Cable TV	0.27***	Istanbul	Ref.
	[0.02]		
Cellular	0.19***	West Marmara	-0.16***
	[0.01]		[0.01]
Computer	0.12***	Aegean	-0.23***
	[0.02]		[0.01]
Internet	0.14***	East Marmara	-0.19***
	[0.03]		[0.02]
Car	0.22***	West Anatolia	-0.18***
	[0.01]		[0.01]
Moto	0.08***	Mediterranean	-0.22***
	[0.02]		[0.01]
Agricultural land	0.06***	Central Anatolia	-0.23***
	[0.01]		[0.02]
Piper water	0.08***	West Black Sea	-0.31***
	[0.02]		[0.01]
Toilets inside	0.11***	East Black Sea	-0.20***
	[0.01]		[0.02]
House owned	0.06***	Northeast Anatolia	-0.20***
	[0.01]		[0.02]
House rented	Ref.	Central east Anatolia	-0.24***
			[0.02]
House lodge	0.07**	Southeast Anatolia	-0.20***
	[0.03]		[0.02]
House other	0.04**	Constant	4.62***
	[0.02]		[0.06]
Members per room	-0.06***		
	[0.01]	Observations	25,764
Log hh size	0.41***	R-squared	0.539
	[0.02]		

Sample of households in the HBS 2003. Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

Table 4: Partition of the population by circumstances

Characteristics	Pop. Share
	Percent
Type of area	
Rural area	57.0
Urban area	43.0
Birth region	
East	22.3
Central	45.7
West	32.0
Mother's education	
no education or unknown	68.8
primary education	28.3
Secondary education	2.5
higher education	0.3
Father's education	
no education or unknown	43.0
primary education	48.5
Secondary education	6.8
higher education	1.7
Mother tongue	
Other language	14.9
Turkish	85.1
Number of siblings	
less than 3	21.7
4 to 5	48.8
6 to 8	25.0
9 or more	4.5

Sample of ever-married women aged 30-49
Source: TDHS 2003

Table 5: Reduced-form regression of "wealth indices" and imputed per capita consumption on circumstances

	Main wealth index	Subsidiary wealth index	Imputed per capita consumption
Birth in a urban area	1.05***	1.15***	0.35***
	[0.05]	[0.06]	[0.02]
Birth in the Central region	0.09	0.07	0.03
	[0.06]	[0.07]	[0.03]
Birth in the West region	0.08	0.10	0.15***
	[0.07]	[0.09]	[0.04]
Mother's primary education	0.35***	0.45***	0.16***
	[0.05]	[0.07]	[0.03]
Mother's secondary education	1.10***	1.84***	0.60***
	[0.19]	[0.25]	[0.09]
Mother's higher education	0.79*	0.99	0.71***
	[0.40]	[0.60]	[0.18]
Father's primary education	0.26***	0.38***	0.12***
	[0.05]	[0.06]	[0.03]
Father's secondary education	0.88***	1.15***	0.33***
	[0.10]	[0.14]	[0.05]
Father's higher education	1.89***	2.01***	0.62***
	[0.23]	[0.32]	[0.11]
Turkish mother tongue	0.66***	0.54***	0.37***
	[0.07]	[0.09]	[0.04]
4 to 5 siblings	-0.28***	-0.31***	-0.07**
	[0.06]	[0.08]	[0.03]
6 to 8 siblings	-0.43***	-0.45***	-0.17***
	[0.07]	[0.09]	[0.04]
9 or more siblings	-0.71***	-0.73***	-0.34***
	[0.10]	[0.14]	[0.06]
Constant	-1.20***	-1.03***	18.18***
	[0.08]	[0.10]	[0.04]
Observations	8074	5229	5229
R-squared	0.274	0.302	0.256

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1
Source: TDHS 2003 with consumption imputed from HBS 2003
Sample of ever-married women, aged 30-49.

Table 6: Measures of Inequality of Opportunity for ever-married Turkish women

Economic outcomes measure	Asset index (main)	Asset index (subsidiary)	Imputed consumption (no residuals)	Imputed consumption	Imputed consumption (no residuals)	Imputed consumption
Inequality measure	Variance	Variance	Variance	Variance	MLD	MLD
Parametric decomposition					Log	Log
Overall inequality	6.01	4.14	26985	57512	0.255	0.362
	[0.13]	[0.12]	[1895]		[0.007]	
Total share of inequality of opportunity						
Non parametric	0.358	0.357	0.407	0.264	0.426	0.323
	[0.012]	[0.014]	[0.023]		[0.015]	
Parametric	0.311	0.302	0.334	0.195	0.374	0.262
	[0.012]	[0.014]	[0.021]		[0.016]	
Partial shares associated with circumstances						
Type of area	0.116	0.114	0.073	0.042	0.154	0.107
	[0.009]	[0.009]	[0.007]		[0.011]	
Birth region	0.004	0.003	0.016	0.008	0.039	0.025
	[0.004]	[0.004]	[0.005]		[0.009]	
Mother's education	0.058	0.064	0.113	0.069	0.136	0.097
	[0.008]	[0.009]	[0.018]		[0.016]	
Father's education	0.074	0.074	0.085	0.048	0.128	0.090
	[0.009]	[0.011]	[0.015]		[0.015]	
Mother tongue	0.030	0.022	0.016	0.010	0.073	0.051
	[0.005]	[0.005]	[0.004]		[0.008]	
Number of siblings	0.029	0.025	0.031	0.018	0.070	0.049
	[0.006]	[0.005]	[0.005]		[0.010]	
Observations	5229					

Bootstrap standard errors in brackets. Bootstrap S.E.s not reported for imputed consumption with residuals, given double boostrapping.
Source: TDHS 2003 with consumption imputed from HBS 2003. Sample: ever-married women, aged 30-49

Table 7: The (Consumption) Opportunity-Deprivation Profile for Turkey

Birth Area	Birth Region	Mother's Education	Father's Education	Mother Tongue Spoken at home	Number of Siblings	Population share (percent)	Group mean per capita annual consumption
urban area	East	primary education	no education or unknown	other language	9 or more	0.01	29.0
rural area	East	no education or unknown	secondary education	other language	4 to 5	0.03	40.6
rural area	West	primary education	primary education	Turkish	9 or more	0.07	44.7
urban area	East	primary education	no education or unknown	other language	4 to 5	0.02	50.4
rural area	East	primary education	primary education	other language	6 to 8	0.01	50.8
rural area	East	primary education	no education or unknown	other language	9 or more	0.01	52.3
urban area	West	primary education	no education or unknown	Turkish	6 to 8	0.03	56.3
rural area	West	no education or unknown	secondary education	Turkish	4 to 5	0.05	57.0
rural area	East	primary education	no education or unknown	other language	4 to 5	0.02	67.9
rural area	East	primary education	secondary education	other language	6 to 8	0.02	69.8
rural area	Central	primary education	primary education	other language	9 or more	0.02	70.6
urban area	East	no education or unknown	primary education	other language	9 or more	0.04	71.7
rural area	East	no education or unknown	primary education	other language	less than 3	0.16	72.3
rural area	East	no education or unknown	no education or unknown	other language	9 or more	1.21	72.7
rural area	East	no education or unknown	primary education	other language	9 or more	0.36	73.4
rural area	West	no education or unknown	primary education	Turkish	9 or more	0.04	76.4
rural area	West	no education or unknown	no education or unknown	Turkish	9 or more	0.02	79.6
rural area	East	no education or unknown	no education or unknown	other language	less than 3	0.44	80.0
rural area	East	no education or unknown	no education or unknown	other language	4 to 5	2.21	80.8
rural area	East	no education or unknown	no education or unknown	other language	6 to 8	3.31	82.3
rural area	Central	no education or unknown	secondary education	other language	4 to 5	0.01	82.5
urban area	Central	primary education	no education or unknown	other language	4 to 5	0.02	86.9
rural area	Central	primary education	primary education	other language	6 to 8	0.02	90.2
rural area	East	no education or unknown	secondary education	Turkish	4 to 5	0.03	90.2
rural area	East	no education or unknown	primary education	other language	6 to 8	0.69	91.4
urban area	East	no education or unknown	no education or unknown	other language	6 to 8	0.95	92.4
urban area	West	primary education	secondary education	Turkish	9 or more	0.04	93.1

Source: TDHS 2003. Sample of ever married women aged 30-49. Annual per capita consumption is imputed from the HBS using the assets common to the TDHS. Consumption figures in millions of 2003 Turkish Lira (substituted in December 2003 by the second Turkish Lira (TRY) at the rate of 1TRY=1E6 TRL).

Table 8: The opportunity-Deprived and the Opportunity-Hoarders: Characteristics of the bottom and top tenths of the opportunity profile

Percentage of the advantaged and disadvantaged groups of women that fall into each category of circumstances	Advantaged 10%	Disadvantaged 10%
Birth Area		
Rural	0.6	88.4
Urban	99.4	11.6
Birth Region		
East	5.3	95.5
Central	35.9	2.0
West	58.8	2.5
Mother's Education		
No Diploma/Illiterate	3.6	97.3
Primary School	69.4	2.7
Secondary School	23.9	0.0
Higher Education	3.1	0.0
Father's Education		
No Diploma/Illiterate	3.2	84.3
Primary School	42.6	14.0
Secondary School	38.5	1.7
Higher Education	15.8	0.0
Mother Tongue		
Non-Turkish	1.0	97.3
Turkish	99.0	2.7
Number of Siblings		
Less than 3	86.0	6.0
3 to 5	12.7	23.8
6 to 8	1.3	50.3
More than 9	0.0	19.9

Source: TDHS 2003. Sample includes only ever-married women ages 30-49.
Consumption per capita is imputed from the HBS using the assets common to the TDHS data.