

THE MEASUREMENT OF INEQUALITY OF OPPORTUNITY:  
THEORY AND AN APPLICATION TO LATIN AMERICA

BY FRANCISCO H. G. FERREIRA\*

*The World Bank and IZA*

AND

JÉRÉMIE GIGNOUX

*Paris School of Economics*

Building on the existing literature, this paper constructs a simple scalar measure of inequality of opportunity and applies it to six Latin American countries. The measure—which captures between-group inequality when groups are defined exclusively on the basis of predetermined circumstances—is shown to yield a lower-bound estimate of true inequality of opportunity. Absolute and relative versions of the index are defined, and alternative parametric and non-parametric methods are employed to generate robust estimates. In the application to Latin America, we find inequality of opportunity shares ranging from one quarter to one half of total consumption inequality. An opportunity-deprivation profile that identifies the worst-off *types* in each society is also formally defined, and described for the same six countries. In three of them, 100 percent of the opportunity-deprived were found to be indigenous or Afro-descendants.

**JEL Codes:** D31, D63, J62

**Keywords:** inequality of opportunity, Latin America

1. INTRODUCTION

Economic inequality—usually measured in terms of income or consumption—is neither all bad nor all good. Most people view income gaps that arise from the application of different levels of effort as less objectionable than those that are due, say, to racial discrimination. Indeed, the distinction between inequalities due to the exercise of individual responsibility on the one hand, and those due to predetermined circumstances on the other, has become central to the

*Note:* We are grateful to Caridad Araujo, Pranab Bardhan, Ricardo Paes de Barros, Marc Fleurbaey, James Foster, Markus Goldstein, Stephen Jenkins, Peter Lanjouw, Marta Menéndez, Vito Peragine, John Roemer, Jaime Saavedra, and three anonymous referees for helpful comments on earlier drafts. Insightful comments were also received at conferences or seminars at the World Bank, the IDB, the Brookings Institution, Cornell University, the Catholic University of Milan, Universidad de los Andes in Bogotá, Colégio de México, Universidad Torcuato di Tella in Buenos Aires, and the Universities of Essex, London, Lund, Manchester, and Oxford. We also thank Carlos Becerra, Jofre Calderón, and Leo Gasparini for kindly providing us with access to data. The views expressed in the paper are those of the authors, and should not be attributed to the World Bank, their Executive Directors, or the countries they represent.

\*Correspondence to: Francisco H. G. Ferreira, The World Bank, 1818 H Street NW, Washington, DC 20433, USA (fferreira@worldbank.org).

© 2011 The Authors

Review of Income and Wealth © 2011 International Association for Research in Income and Wealth  
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St,  
Malden, MA, 02148, USA.

literature on social justice in political philosophy, social choice, and, increasingly, in mainstream economics.

Dworkin (1981), Arneson (1989), Cohen (1989), and, to some extent, Sen (1985) are among a number of influential authors who have argued that inequality in the distribution of particular *outcomes*—such as incomes—is not the appropriate yardstick for assessing the fairness of a given allocation or social system. Despite important differences in nuance, these authors have all suggested that some outcome differences—which are attributable to differences in choices for which individuals can be held responsible—may be ethically acceptable. In this view, unacceptable inequalities reside in a logically prior space—of resources, capabilities, opportunities—for which individuals cannot be held responsible.<sup>1</sup>

John Roemer (1998), for instance, calls those factors over which individuals have a measure of control, “efforts” (e.g. how long one studies, or how hard one works), while those for which they cannot reasonably be held to have any responsibility are referred to as “circumstances” (e.g. race, gender, or family background). Given this distinction, he defines “equality of opportunity” essentially as a situation in which important outcomes—which he calls “advantages”—are distributed independently of circumstances.<sup>2</sup>

Such a distinction between inequality of opportunity and the more standard concept of inequality of outcomes is of interest to economists for at least three sets of reasons. First, there is an increasingly widespread normative view that it is inequality of opportunity, and not that of outcomes, which should inform the design of public policy. Inequality of opportunity is, in this view, the appropriate “currency of egalitarian justice” (Cohen, 1989). Public action need not necessarily aim to eliminate all outcome inequalities, but may be justified in seeking to reduce those that arise from unequal opportunities: “economic inequalities due to factors beyond the individual responsibility are inequitable, and should be compensated by society” (Peragine, 2004, p. 11). To the extent that this view, which is already popular among social choice theorists and political philosophers, gains traction among policymakers, will behoove economists to provide tractable empirical measures of the concept.

Second, if the degree of inequality of opportunity affects popular attitudes to outcome inequality, then it may affect beliefs about social fairness and attitudes to redistribution. These beliefs and attitudes may in turn affect the extent of redistribution actually implemented in society, and thus the level of investment and output generated: Alesina and Angeletos (2005) and Bénabou and Tirole (2006) provide examples of models where such beliefs and attitudes themselves play a key

<sup>1</sup>Space constraints prevent us from exploring these differences in nuance here, but they have been reviewed extensively elsewhere (see, e.g. Roemer, 1993a).

<sup>2</sup>Roemer phrases his definition somewhat differently. He argues that, if it were possible to partition the population into circumstance-homogeneous groups (which he calls “types”), and if the only variable that differed across individuals *within* each type was their effort level, then equality of opportunity would attain only if the distributions of advantage across all such groups were identical. Under Roemer’s assumptions, the requirement of identical distributions of advantage regardless of type is equivalent to stochastic independence between advantage and circumstances. Although this should be intuitively clear, we return to this argument more formally in Section 2.

role in generating multiple equilibria, with very different objective economic characteristics.

Third, it has also been suggested that inequality of opportunity might be a more relevant concept (than income inequality) for understanding whether aggregate economic performance is worse in more unequal societies—and if so, why. In addition to the role of beliefs and attitudes to redistribution, it is possible that the kinds of inequality that are detrimental to growth (such as inequality in access to good schools, or to financial markets) are more closely associated with the concept of opportunities, while other components of outcome inequality—such as those arising from returns to different levels of effort—may actually have a positive effect on growth (World Bank, 2006; Bourguignon *et al.*, 2007b). Perhaps one of the reasons why the cross-country empirical literature on inequality and growth is so inconclusive is that it conflates the two kinds of inequality.<sup>3</sup> In fact, a recent study by Marrero and Rodríguez (2009) finds that if one decomposes total income inequality into an “opportunity” component and an “effort” component, both terms have statistically significant coefficients in a growth regression estimated for 23 states of the United States in the last two decades. But while the coefficient on inequality of opportunity has a negative sign, the opposite is true for inequality of efforts.

But in order to make empirical use of the concept of inequality of opportunity—whether for the design of taxation and public expenditures or in the study of the determinants of cross-country growth differences—it is first necessary to measure it appropriately. The recent literature contains at least three different approaches to the measurement of inequality of opportunity.<sup>4</sup> Bourguignon, Ferreira, and Menéndez (2007a)—henceforth BFM—estimate a linear model of advantage (earnings) as a function of circumstances and efforts, and use it to simulate counterfactual distributions where the effect of circumstances is suppressed. By comparing the actual earnings distribution with different counterfactuals, they decompose overall earnings inequality in Brazil into a component due to five observed circumstance variables, and a residual. The circumstance (or inequality of opportunity) component is further decomposed into a direct effect and an (indirect) effect that operates through the influence of circumstances on the choice of efforts. Crucially, BFM seek to estimate the contribution of the five *specific* circumstances observed in their dataset: race, mother’s schooling, father’s schooling, region of birth, and father’s occupation. By imposing certain restrictions on coefficient signs and on their variance–covariance matrix, they estimate bounds on the possible biases arising from the omission of other, unobserved circumstance variables. The procedure is therefore interpreted as estimating the contribution of those *specific observed circumstances* to overall earnings inequality.

A second approach to decomposing overall inequality into an opportunity component and an “ethically acceptable” component is to rely on standard between-group inequality decompositions. Checchi and Peragine (2010)—

<sup>3</sup>See Banerjee and Duflo (2003) on the inconclusiveness of that literature.

<sup>4</sup>See also van de Gaer *et al.*, (2001) for an earlier treatment of inequality of opportunity in the context of intergenerational mobility.

henceforth CP—show that if groups are defined by circumstance characteristics—so that they correspond to Roemer’s types—then the between-group component can be interpreted as an “ex-ante” measure of inequality of opportunity. Conversely, if groups are defined in terms of their relative position in the effort distributions across types, then inequality within groups (called “tranches” by CP) corresponds to an “ex-post” measure of inequality of opportunity.<sup>5</sup> The authors present both estimates for the distribution of earnings in Italy.

Finally, a third approach, associated with Lefranc *et al.* (2008), relies on stochastic dominance comparisons of distributions conditional on types for assessing whether inequality of opportunity is present in a society. These authors also propose a Gini of Opportunities index for the scalar measurement of inequality of opportunity.

This paper combines elements from the first two approaches, and shows that a variant of the parametric approach in BFM and the non-parametric ex-ante approach of CP are effectively alternative procedures for estimating the same quantity. For this to hold, however, it is necessary to interpret both estimates as yielding a lower-bound on the set of possible true measures of (ex-ante) inequality of opportunity. It is necessary, in other words, to treat the share of inequality associated with the circumstances one observes as a lower-bound on the share accounted for by *all circumstances*, observed and unobserved, rather than as the share corresponding to those *specific* observed circumstances. Under this interpretation, a variant of the index proposed by BFM and that due to CP are simply parametric and non-parametric alternatives for the measurement of (lower-bound) inequality of opportunity.<sup>6</sup>

We derive this index from the pioneering definitions of equality of opportunity due to Roemer (1993b, 1998) and van de Gaer (1993), and define two versions for it: an absolute measure of the *level* of inequality of opportunity (IOL), and one measure that is *relative* to overall outcome inequality (IOR). Both versions are computed for six countries in Latin America—Brazil, Colombia, Ecuador, Guatemala, Panama, and Peru—using both the non-parametric and the parametric estimation procedures. Although the two methods tend to generate robustly similar results for large samples, the parametric approach yields more conservative estimates of the lower-bound for inequality of opportunity in smaller samples. The lower-bound estimates of the inequality of opportunity ratio (IOR) range from 25 percent of total inequality in household consumption in Colombia, to 51 percent in Guatemala.

Finally, we also define and compute opportunity profiles and opportunity-deprivation profiles for all six countries: these are rankings of types within society which may be of direct practical relevance for the implementation of the concepts of “equal opportunity policy” found in the literature, and to Roemer’s proposed

<sup>5</sup>See the discussion in Section 2, as well as Fleurbaey and Peragine (2009), on the conceptual distinction between ex-ante and ex-post approaches to inequality of opportunity.

<sup>6</sup>An added advantage of this approach is that it eliminates the need for the strong assumptions made by BFM to estimate confidence intervals around their coefficient estimates, so that those could be interpreted causally.

criterion for assessing economic development.<sup>7</sup> The profiles are shown to vary substantially across countries (with ethnicity being fundamental in Brazil but much less important in Colombia, for instance).

The remainder of the paper is structured as follows. Section 2 relates our index of inequality of opportunity explicitly to both Roemer's and van de Gaer's conceptual definitions of equality of opportunity. It also defines the relative and absolute versions of the index, and describes some of their properties. Section 3 then discusses the two alternative procedures for estimating the index in practice: the non-parametric approach of CP, and (a version of) the BFM parametric approach. It also establishes that the index is a lower-bound estimator of the true measure of inequality of opportunity. Section 4 provides some information on the six household survey datasets used in our empirical application, and discusses issues of cross-country comparability. Section 5 presents the main empirical results, while Section 6 discusses the opportunity and opportunity-deprivation profiles. Section 7 concludes.

## 2. A CONCEPTUAL FRAMEWORK

Consider a finite population of discrete agents indexed by  $i \in \{1, \dots, N\}$ , where  $N$  is large. Each individual  $i$  is characterized by a set of attributes  $\{y_i, C_i, e_i\}$ , where  $y$  denotes an advantage,  $C$  denotes a vector of circumstance characteristics, and  $e$  denotes an effort level. We follow Roemer (1998) in considering a single advantage variable (which we will later associate with household per capita income or consumption), and in representing effort as a scalar.<sup>8</sup> In this section, and for ease of exposition, we will also follow Roemer (1998) in treating effort as a continuous variable, while the vector  $C_i$  consists of  $J$  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,  $x_j, \forall i$ .<sup>9</sup>

This permits us to partition the population into Roemerian *types*, i.e. population subgroups that are homogeneous in terms of circumstances. This partition is given by  $\Pi \in \{T_1, T_2, \dots, T_K\}$ , such that  $T_1 \cup T_2 \cup \dots \cup T_K = \{1, \dots, N\}$ ,  $T_l \cap T_k = \emptyset, \forall l, k$ , and  $C_i = C_j, \forall i, j | i \in T_k, j \in T_k, \forall k$ . Naturally, the maximum possible number of types is given by  $\bar{K} = \prod_{j=1}^J x_j$ .<sup>10</sup> It will prove useful to denote the joint distribution of advantages and circumstances over the population by  $\{y, C\}$ , and the space of such distributions by  $\Omega$ . The marginal distribution of advantages, of course, is given simply by the vector  $y = (y_1, \dots, y_N)$ . Similarly, denote the space of possible population partitions  $\Pi$  by  $\Lambda$ .

<sup>7</sup>Roemer (2006) suggested that "the rate of economic development should be taken to be the rate at which the mean advantage level of the worst-off types grows over time" (p. 243).

<sup>8</sup>We later show that the proposed index does not hinge on effort being a scalar, and is perfectly consistent with alternative representations, such as a vector of efforts,  $E$ .

<sup>9</sup>For clarity, subscripts applied to  $C$  denote individuals, while superscripts denote elements in the  $C_i$  vector. While our treatment of circumstances as discrete variables is common to most of the literature, see O'Neill *et al.* (2000) for an alternative approach that relies on a single continuous circumstance variable.

<sup>10</sup> $K < \bar{K}$  if some cells in the partition are empty in the population.

In his original formal definition of equality of opportunity, Roemer (1998) defines the distribution of effort within each type  $T_k$ ,  $G_\rho^k(e)$  as “a probability measure on the set of effort levels, which are non-negative real numbers” (p. 10). Since he is primarily concerned with defining an equal-opportunity *policy*, or set of allocation rules, he adds a subscript to indicate that the distribution of efforts is conditional on some policy  $\rho$ . He denotes the advantage level enjoyed by a person in quantile  $\pi = G_\rho^k(e)$  of the effort distribution in type  $k$ , given policy  $\rho$ , as  $y^k(\pi, \rho)$ . His analysis is then couched in terms of seeking an equal-opportunity policy  $\rho^*$ , which he ultimately proposes should maximize the average (over quantiles of the effort distribution) of minimum levels of advantage across all types, at each given quantile:

$$(1) \quad \rho^* = \arg \max_{\rho} \int_0^1 \min_k y^k(\pi, \rho) d\pi.$$

Although Roemer (1998) does not actually write down a formal definition of equality of opportunity itself—only that of the equal opportunity *policy*—his equation (1) has been widely interpreted to imply that equal opportunities would attain if the levels of advantage were the same across all types, at each and every quantile of the effort distribution:

$$(2) \quad y^k(\pi, \rho) = y^l(\pi, \rho), \forall \pi \in [0, 1]; \forall T_k, T_l \in \Pi.$$

Equation (2) clearly accords with Roemer’s informal statement 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” (Roemer, 1998, p. 12).

Now denote the cumulative distribution function of *advantage* in type  $k$ , under policy  $\rho$ , by  $F_\rho^k(y)$ . Note that the effort rank ( $\pi$ ) and the advantage rank must be the same within each type because, given circumstances, advantage is fully and monotonically determined by effort. Dropping the policy subscript  $\rho$ , which is not the focus of our analysis, and noting that  $y^k(\pi)$  is simply the inverse function of  $\pi = F^k(y)$ , (2) then implies:

$$(3) \quad F^k(y) = F^l(y), \forall l, k | T_k \in \Pi, T_l \in \Pi.$$

This is presented as Roemer’s “strong criterion” definition of equal opportunities in Bourguignon *et al.* (2007b), and by Lefranc *et al.* (2008).<sup>11</sup>

If equality of opportunity corresponds to a hypothetical situation in which advantage distributions are identical across types, then the measurement of *inequality* of opportunity must, in some sense, seek to capture the extent to which

<sup>11</sup>Lefranc *et al.* (2008) refer to equation (3)—albeit obviously in slightly different notation—as “a compelling case of equality of opportunity [that] corresponds to the definition of equality of opportunity adopted by Roemer (1998)” (p. 517).

$F^k(y) \neq F^l(y)$ , for  $k \neq l$ . An obvious first step would be to test for the *existence* of inequality of opportunity, by examining whether the conditional distributions of advantage differ across types. This is precisely what Lefranc *et al.* (2008) do, using stochastic dominance techniques and the associated statistical tests to compare conditional income distributions across types in a number of OECD countries, where the types are defined by the level of education (or, in a couple of cases, the occupation) of a person's father.

Theirs is a very interesting approach to ascertaining whether or not individual countries, or other populations, could be described as having equality of opportunity. (In their sample, the null hypothesis of equal opportunities can be rejected for every country, except Sweden.) It also allows for a (partial) ranking of *types* within each country, relying on the dominance comparisons. This partial ranking is complemented by a scalar index for inequality of opportunity, which is based on a variant of the Gini coefficient, defined over mean advantage levels for each type and adjusted for "within-type inequality" (see Lefranc *et al.*, 2008).

However, while this reliance on stochastic dominance comparisons across type-specific advantage distributions is desirable in terms of robustness, it does come at a practical cost, given usual sample sizes. Because the estimation of distribution functions (or generalized Lorenz curves) requires a reasonable number of observations within each type, the partition  $\Pi$  of the population must perforce be quite coarse. Lefranc *et al.* (2008) work with  $K = 3$  in all countries. This implies a rather limited treatment of inequality of opportunity, since any inequality within those three types is then associated with differences in efforts. These would include, for example, any income differences associated with gender, race, or birthplace that might exist within types defined solely on the basis of father's education.

An alternative approach is to adopt a weaker criterion for the empirical identification of equality of opportunity, namely that mean advantage levels should be identical across types. If we define  $\mu^k(y) = \int_0^y dF^k(y)$ , then this weaker criterion for equality of opportunity is written:

$$(4) \quad \mu^k(y) = \mu^l(y), \forall l, k | T_k \in \Pi, T_l \in \Pi.$$

This criterion is consistent with Roemer's original definition, given by equation (3), in the sense that it is always implied by that equation. Whenever (4) does not hold, so that the hypothesis of equality of opportunity is rejected empirically, we can be confident that the theoretical definition is not satisfied either (subject to the usual confidence margins associated with statistical inference): Since  $F^k(y) = F^l(y), \forall k, l \Rightarrow \mu^k(y) = \mu^l(y), \forall k, l$ , it follows that  $\mu^k(y) \neq \mu^l(y), \exists k, l \Rightarrow F^k(y) \neq F^l(y)$ .

However, (4) is clearly weaker than (3), since two different conditional distributions may happen to have the same mean. It is possible therefore, that the empirical test in (4) will fail to reject the hypothesis of equality of opportunity even though it is false according to the original definition in (3). This kind of "type 2 error" in the empirical identification of equality of opportunity is not exclusive to this method, or to approaches that rely on the mean rather than the entire distribution. Precisely the same issue arises with empirical identification criteria based

on first- or second-order stochastic dominance.<sup>12</sup> The use of an empirical criterion for assessing whether or not equality of opportunity holds that is somewhat weaker than (3) seems to be the price to be paid for applying these concepts to datasets with realistic sample sizes.<sup>13</sup>

The transition from equation (3) to equation (4) can therefore be justified on the basis of practical considerations: in practice, sample sizes are generally too small to allow for the estimation of type-specific distribution functions when the number of types becomes realistically large. There is also an alternative, and rather different, justification for moving from (3) to (4), which has to do with the conceptual distinction between the *ex-post* and *ex-ante* approaches to equality of opportunity. In the *ex-post* approach, inequality of opportunity is viewed as inequality among people who have exerted the same degree of effort, regardless of circumstances. Measuring this inequality would require aggregating outcome differences among people at the same effort quantile across types, *for each quantile*. Full equality of opportunity would imply equality at each quantile, and hence for the whole distribution (as in (3)).

The *ex-ante* approach, on the other hand, sees inequality of opportunity as inequality between groups of people who share the same circumstances (i.e. between types). Conceptually, the *ex-ante* approach does not require observing effort, or comparing individuals from different types at each percentile of their effort distributions. But it does require agreement on some valuation of the opportunity set faced by people in each type. Van de Gaer (1993) proposed that the opportunity set of each type could be valued by its mean level of advantage.<sup>14</sup> Full equality of opportunity, in this case, would imply equality at the mean (as in (4)).

In this paper, we are agnostic about whether one adopts the weaker criterion for equal opportunities (in equation (4)) on the basis of a conceptual preference for the *ex-ante* approach (and van de Gaer's use of mean outcomes to value a type's opportunity set), or for practical reasons to do with the difficulties associated with estimating full type-specific distributions for many types in most datasets.<sup>15</sup> Once one does accept (4) as identifying *equality* of opportunity, however, the measurement of *inequality* of opportunity must now seek to capture the extent to which  $\mu^k(y) \neq \mu^l(y)$ , for  $k \neq l$ . This is an easier task, since it appears to call

<sup>12</sup>See the discussion in Lefranc *et al.* (2008, pp. 517–18).

<sup>13</sup>The cases in which the proposed empirical identification criterion and Roemer's definition would clash ( $\mu^k(y) = \mu^l(y)$  but  $F^k(y) \neq F^l(y)$ ) appear, in any case, to be rare in practice, at least in Latin America. Conditional means were not found to be equal across types in *any* of the cases investigated in Section 5. In addition, the weaker nature of the empirical criterion is consistent with the lower-bound interpretation of the scalar indices of inequality of opportunity that build on it, as discussed in Section 3. We are grateful to Marc Fleurbaey for a helpful discussion on this point.

<sup>14</sup>This is why van de Gaer's equal opportunity policy is defined somewhat differently than Roemer's: instead of taking the minimum (across types) at each centile of the conditional distribution of advantages, and *then* averaging across centiles (equation (1)), in the so-called "mean of mins" approach, van de Gaer (1993) proposed first averaging across centiles, and then taking the minimum across types (a "min of means"):  $\rho_{VDG}^* = \arg \max_{\rho} \min_k \int_0^1 y^k(\pi, \rho) d\pi = \arg \max_{\rho} \min_k \mu^k(y^k)$ . For further discussion of the *ex-ante* and *ex-post* approaches to measuring inequality of opportunity, see Checchi and Peragine (2010), Ooghe *et al.* (2007), and Ferreira *et al.* (forthcoming).

<sup>15</sup>In the latter case, the criterion would still be consistent with Roemer's *ex-post* approach, subject to the "type 2 error" caveat discussed above.

for an inequality index defined not on the marginal distribution of advantages,  $y = (y_1, \dots, y_N)$ , but on the corresponding *smoothed* distribution.

A smoothed distribution, which we denote  $\{\mu_i^k\}$ , was originally defined by Foster and Shneyerov (2000), drawing on the earlier inequality decomposition literature associated with Bourguignon (1979), Cowell (1980), and Shorrocks (1980). It was introduced to the measurement of inequality of opportunity by Checchi and Peragine (2010). The smoothed distribution  $\{\mu_i^k\}$  is obtained from a distribution of advantages  $y$  and a partition  $\Pi$  by replacing each individual advantage  $y_i^k$  with the group-specific mean,  $\mu^k(y)$ . So, with  $N$  individuals and  $K$  types,  $\{\mu_i^k\} = (\mu_1^1, \dots, \mu_{n_1}^1; \dots; \mu_i^K, \dots, \mu_N^K)$ , with  $\mu_g^k = \dots = \mu_i^k = \dots = \mu_{n_g}^k \forall k, g = 1 + \sum_{l=1}^{k-1} n_l$ , and  $h = \sum_{l=1}^k n_l$ .

The weak identification criterion for equality of opportunity in (4) and the definition of a smoothed distribution immediately give rise to a candidate scalar measure of inequality of opportunity, which maps from a joint distribution of advantage and circumstances  $\{y, C\}$  and from the associated partition  $\Pi$ , to the non-negative real line. This index is given by  $\theta_a : \Omega \times \Lambda \rightarrow \mathfrak{R}^+$ :

$$(5) \quad \theta_a = I(\{\mu_i^k\}).$$

Associated with the absolute index  $\theta_a$ , is a relative version of the index:  $\theta_r : \Omega \times \Lambda \rightarrow [0,1]$ :

$$(6) \quad \theta_r = \frac{I(\{\mu_i^k\})}{I(y)}.$$

$\theta_a$  is a measure of the absolute level of inequality of opportunity (IOL), while  $\theta_r$  measures that level in relation to total inequality, and is thus an inequality of opportunity ratio (IOR). The latter is, of course, CP's measure of inequality of opportunities in the types, or "ex ante" approach. In (5) and (6),  $I()$  is any inequality index that satisfies the axiomatic properties which are now standard in the literature on the measurement of relative inequality (see, e.g. Cowell, 1995). These properties include symmetry (or anonymity); the transfer principle; scale invariance; population replication; and, crucially, additive decomposability. This last property requires that  $I(y) = I(\{\mu_i^k\}) + \sum_k w_k I(y_k)$ , where  $y_k$  denotes the income vector within each type  $T_k$ , and  $w_k$  denotes type-specific weights, subject to  $\sum_k w_k = 1$ .<sup>16</sup>

For any inequality index  $I()$  that satisfies these properties, it is easy to check that both  $\theta_a$  and  $\theta_r$  satisfy:

<sup>16</sup>The treatment of effort as a continuous variable, and the ensuing notation with continuous within-type distributions  $G^k(e)$  and  $F^k(y)$ , were useful primarily to relate our conceptual framework to the existing theory of equality of opportunity (in particular to Roemer, 1998). From this point onwards, with effort only in the background of the analysis, we revert to a fully discrete notation, and use  $y_k$  as the within-type income vector. There is no other change in notation, and the marginal and joint distributions of advantage and circumstances defined earlier are unchanged.

- (i) *Principle of population*: the index is invariant to a replication of the population  $\{1, \dots, N\}$ .
- (ii) *Scale invariance*: the index is invariant to the multiplication of all advantages by a positive scalar.
- (iii) *Normalization*: if the smoothed distribution  $\{\mu_i^k\}$  is degenerate, so that equation (4) holds, then the index takes a value of zero.
- (iv) *Within-type symmetry*: the index is invariant to any permutation of two individuals within a type.

Furthermore, the IOL measure  $\theta_a$  satisfies:

- (v) *Within-type transfer insensitivity*: the index is invariant to any mean-preserving spread in advantages within a type.
- (vi) *Between-type transfer principle*: the index weakly rises with any transfer from any individual  $i$  to  $j$ , if  $i \in T_k, j \in T_l$ , with  $\mu^k < \mu^l$ .

The class of indices  $I()$  that satisfy symmetry (or anonymity), the Pigou–Dalton transfer principle, scale invariance, population replication, and additive decomposability, reduces to a well-known class of inequality measures. Shorrocks (1980) and Foster (1985) show that (under a regularity condition) an inequality measure satisfies the four basic properties and additive decomposability if and only if it is a positive multiple of a member of the Generalized Entropy ( $E_\alpha$ ) class.

Nevertheless, that is still a large class of measures. As is well known, an inequality decomposition by population subgroup, for a given distribution of advantages and for a given partition, will in general differ for different indices  $I()$  in the Generalized Entropy family, implying that  $\theta_a$  and  $\theta_r$  are not uniquely defined. So, for a given smoothed distribution—that is, for a given joint distribution  $\{y, C\}$  and partition  $\Pi$ —one could obtain different values for both the absolute and relative versions of our inequality of opportunity index, by selecting different inequality measures  $I()$  from the set of indices that satisfy the previously imposed axioms. Since these measures are sensitive to different parts of the distribution, different choices of  $I()$  could in principle lead to different rankings across two smoothed distributions.

Fortunately, there is an eminently plausible further requirement which allows us to refine the set of eligible indices to a singleton, namely Foster and Shneyerov's (2000) path-independent decomposability axiom. Just as we previously defined a smoothed distribution, we now define a *standardized distribution*, denoted  $\{v_i^k\}$ , as the distribution which is obtained from a distribution of advantages  $y$  and a partition  $\Pi$ , by replacing  $y_i^k$  with  $y_i^k \frac{\mu}{\mu^k}$  (where  $\mu$  is the grand mean). Just as a smoothed distribution eliminates all within-group inequality by construction, a standardized distribution eliminates all between-group inequality, by appropriately rescaling all subgroup means. One might therefore wish to impose the requirement that  $I(\{\mu_i^k\}) = I(y) - I(\{v_i^k\})$ . This requirement is the axiom of path-independent decomposability.

Foster and Shneyerov (2000) fully characterize the “*path-independent decomposable*” class of inequality measures. They show that when the set of inequality indices  $I()$  under consideration is restricted to those that use the arithmetic mean as the reference income, and that satisfy the Pigou–Dalton transfer axiom, this class

reduces to a single inequality measure, the mean logarithmic deviation, which we denote  $E_0$  since it is a member of the generalized entropy class, when its parameter is set to zero.<sup>17</sup> By adding path-independent decomposability to the list of axioms that the inequality indices  $I()$  must satisfy, we are able to restrict the two versions of our scalar measure of inequality of opportunity (IOL and IOR) to two unique indices:

$$(5') \quad \theta_a = E_0(\{\mu_i^k\})$$

and

$$(6') \quad \theta_r = \frac{E_0(\{\mu_i^k\})}{E_0(y)}$$

These two scalar measures of inequality of opportunity have a number of appealing features. First, they follow directly from van de Gaer's (1993) ex-ante approach to inequality of opportunity, but can also be seen as consistent with an identification criterion for equality of opportunity which is weaker than, but implied by, Roemer's (1998) ex-post definition. Second, the indices satisfy a range of desirable properties, listed above as axioms (i) through (vi) for  $\theta_a$  (and (i) through (iv) for  $\theta_r$ ), as well as path-independence. Third, they are extremely simple to calculate, and are identical to the between-group component ( $\theta_a$ ) or share ( $\theta_r$ ) of the standard Theil-L decomposition by population subgroups, *provided that the population is partitioned by circumstance variables only, as in our earlier definition of*  $\Pi = \{T_1, T_2, \dots, T_K\}$ .

Property (v), namely within-type transfer insensitivity, also sets  $\theta_r$  apart from other measures in current use, such as Lefranc *et al.*'s Gini of Opportunities, which is sensitive to "risk"—or inequality—within types. The approach here is to take seriously the notion that the only kind of ethically objectionable inequality is that associated with opportunities, i.e. that which occurs between types. The index is therefore deliberately insensitive to within-group inequality. Within-type transfer insensitivity may be seen as a kind of "focus axiom" for (ex-ante) inequality of opportunity measurement: if types are well-defined, so that individuals are homogeneous in circumstances within each type, then within-group inequality should be ignored, much as incomes above the poverty line are ignored by virtue of the focus axiom in poverty measurement.

<sup>17</sup>It is easy to see why the two decomposition paths yield different results for other generalized entropy measures. The decomposition of total inequality for these measures can be written as follows:  $E_\alpha(y) = E_\alpha(\{\mu_i^k\}) + \sum_{k=1}^K \frac{n^k}{N} \left(\frac{\mu^k}{\mu}\right)^\alpha E_\alpha(y^k)$ , where  $n^k$  and  $y^k$  denote, respectively, the population and the advantage distribution in type  $k$ , and  $\alpha$  is the generalized entropy parameter. The first term in the right-hand side of this equation—the between-group component—is inequality in the smoothed distribution. The second term is the within-group component. Clearly, for  $\alpha \neq 0$ , the rescaling of subgroup means implied by standardization  $\left(y_i^k \frac{\mu}{\mu^k}\right)$  not only drives the first term (the between-group component) to zero, but also affects the weights in the within-group term. So, for  $\alpha \neq 0$ ,  $E_\alpha(\{\mu_i^k\}) \neq E_\alpha(y) - E_\alpha(\{y_i^k\})$ .

A similar argument applies to property (vi), the between-type transfer principle, which requires the inequality of opportunity index to rise if a transfer is made from someone in a “poorer” type to someone else in a “richer” type—regardless of whether the first person is individually richer or poorer than the second.

Although we find these two axioms conceptually appealing for a measure that seeks to isolate and quantify inequality of opportunity, they do not apply to  $\theta_r$ , which is decreasing in within-type inequality by construction. While IOL ( $\theta_a$ ) is our “preferred” version, we nevertheless follow CP and use IOR ( $\theta_r$ ) in our empirical application below, as a complementary measure. Obviously, if one insists on axioms (v) and (vi), then only  $\theta_a$  should be used, with no reference to  $\theta_r$ .

### 3. ESTIMATING IOL AND IOR IN PRACTICE

Given a sample with information on the advantage and circumstance variables in the joint distribution  $\{y, C\}$ , and agreement on a partition  $\Pi$ ,  $\theta_a$  and  $\theta_r$  can be calculated immediately by any algorithm that computes  $E_0(\{\mu_i^k\}) = \frac{1}{N} \sum_{i=1}^N \log \frac{\mu}{\mu_i^k}$ , the between-group component in the standard decomposition of the mean logarithmic deviation by population subgroups.

This standard non-parametric approach is certainly optimal for most common sample sizes, provided that there are relatively few types in the partition  $\Pi$ . It was, for instance, the method used by CP, who had  $K = 5$  types. As noted earlier, however, a small  $K$  requires assuming a very limited role for circumstances. In both CP and Lefranc *et al.* (2008), inequality of opportunity is associated only with differences between 3 or 5 groups, defined by a coarse categorization of parental background. In both cases the circumstance vector  $C_i$  is actually a scalar ( $J = 1$ ), and  $x_j$  is either 3 or 5.

Such a restrictive approach to partitioning the population into types is likely to lead to an underestimate of inequality of opportunity. Any inequality associated with race, gender, birthplace, or family wealth, which may remain within those three to five types, would be attributed to effort. As we will see in the next section, many surveys do contain information on a number of other variables which can be unambiguously classified as circumstances. In addition to mother’s and father’s education, surveys often contain information on parental occupation, race or ethnicity, gender, and place of birth. As  $J$  and  $x_j$  rise,  $K$  increases geometrically.

As the number of types increases, the frequency of sample observations per type (or cell) tends to diminish quite rapidly. In the empirical applications that follow, with five circumstance variables ( $J = 5$ ), and two or three possible values per circumstance ( $x_j = 2$  or  $3$ ), we end up with  $\bar{K} = 108$ . In two of our six countries, this led to there being over a quarter of all types for which there were fewer than five observations in the sample, causing the precision of the estimates of mean advantage per type to become unacceptably low.

As is often the case when sample sizes are insufficient for fully flexible, non-parametric estimation, a parametric alternative is available that permits efficient estimation, at the cost of some functional form assumptions. This was the route followed by BFM, who noted that Roemer’s view of advantages as determined by circumstances and efforts (plus possibly luck, or unobserved random

terms) would be consistent with a stylized model of advantage of the general form  $y = f(C, E, u)$ . Since circumstances are economically exogenous by definition—in the sense that they cannot be affected by individual decisions—and given that efforts may be, and generally are, influenced by circumstances, one would rewrite this more fully as:<sup>18</sup>

$$(7) \quad y = f[C, E(C, v), u].$$

For the purpose of measuring inequality of opportunity—rather than of estimating any causal relationship between circumstances, efforts, and advantages—one can simply write the reduced form of (7) as  $y = \phi(C, \varepsilon)$ .<sup>19</sup> A log-linearized version of this equation,  $\ln y = C\psi + \varepsilon$ , can be estimated by OLS. As in BFM, such an equation must be interpreted as a *reduced form* of model (7), so that the parameters  $\psi$  encompass both the direct effect of circumstances on the advantage  $y$ , and the indirect effect of circumstances through efforts. Once estimates for the reduced-form coefficients  $\psi$  have been obtained, one can construct a parametric estimate of the smoothed distribution as:

$$(8) \quad \tilde{\mu}_i = \exp[C_i \hat{\psi}].$$

Here, a hat indicates the parameter estimate from an OLS regression, and the tilde indicates a counterfactual advantage level. The vector  $\tilde{\mu}$  (whose elements are given by (8) for each  $i$ ) is a parametric analogue to the smoothed distribution  $\{\mu_i^k\}$  because, by eliminating the residuals, (8) replaces individual advantage levels with their predictions (i.e. their averages conditional on certain values for  $C$ ). Predicted advantage is, of course, the same for all individuals with identical circumstances.

Similarly, the parametric estimate of the standardized distribution would be given by:

$$(9) \quad \tilde{v}_i = \exp[\bar{C}_i \hat{\psi} + \hat{\varepsilon}_i].$$

Here, the overbar indicates an average of circumstances across all individuals. By assigning the vector of average circumstances to all individuals, but retaining within-type variation (through  $\hat{\varepsilon}_i$ ), the vector  $\tilde{v}$  becomes a parametric analogue to the standardized distribution  $\{v_i^k\}$ .

We can thus define parametric (smoothed) estimates for our inequality of opportunity indices as follows:

$$(10) \quad \theta_a^p = E_0(\tilde{\mu})$$

and

<sup>18</sup>The stochastic terms  $u$  and  $v$  can be thought to account for luck and other random factors. For an excellent recent treatment of the role of luck in the theory of equality of opportunity, see Lefranc *et al.* (2009). In empirical applications, these terms will also capture variation in unobserved determinants.

<sup>19</sup>This is why, as noted earlier, our approach to the measurement of inequality of opportunity is perfectly consistent with a view of efforts as a vector,  $E$ , rather than a scalar,  $e$ .

$$(11) \quad \theta_r^p = \frac{E_0(\tilde{\mu})}{E_0(y)}$$

Parametrically standardized estimates are obtained as:

$$(10') \quad \theta_a^{PS} = E_0(y) - E_0(\tilde{v})$$

$$(11') \quad \theta_r^{PS} = 1 - E_0(\tilde{v})/E_0(y)$$

Although (10) and (10'), and (11) and (11'), are estimates for the same path-independent measures, the fact that they are estimated parametrically, involving linear functional form assumptions, means they are not exactly identical. However, they are generally very similar, and the parametric estimates for IOL and IOR that we report in Section 5 are obtained from the parametrically standardized distributions, through (10') and (11'), respectively.

Two important methodological considerations remain, before we can turn to the empirical application. First is the issue of omitted circumstance variables. Realistically, the vector  $C_i$  observed in any particular dataset is likely to be a sub-vector of the theoretical vector  $C_i^*$  of all possible circumstances (observed and unobserved) that help determine a person's advantage. "True" measures of inequality of opportunity (call them  $\theta_a^*$  and  $\theta_r^*$ ) would require that all relevant circumstance variables, and all relevant values for those circumstances, be used to define the partition  $\Pi$ . This is unlikely ever to be the case in practice for almost any conceivable dataset. It is certainly not the case for the six countries in our application below, even though we work with a much finer partition of circumstances than any other study we are aware of.

The implication is that the empirical estimates defined in this section—whether parametric or non-parametric—should be interpreted as *lower-bound* estimates of inequality of opportunity. Whenever the dimension of the observed vector  $C_i$  is less than the dimension of the "true" vector  $C_i^*$  ( $J \leq J^*$ ), then  $\theta_a$  and  $\theta_r$  are lower-bound estimators of true inequality of opportunity—the inequality that would be captured by the same indices if the full vector  $C_i^*$  were observed. This result is formalized for the non-parametric case in the proposition and corollary below:

*Proposition:* The IOL measure  $\theta_a(\{y, C\})$  is a lower-bound estimator of the true inequality of opportunity level,  $\theta_a^*(\{y, C^*\})$ .

*Proof:* Recall that  $\theta_a = I(\{\mu_i^k\})$  is defined for an observed joint distribution  $\{y, C\}$  and partition  $\Pi$ , with the dimension of  $C_i$  given by  $J$ , and the number of types given by  $K \leq \bar{K} = \prod_{j=1}^J x_j$ . Note that the dimension of the vector of observed circumstances can be no greater than that of the true vector of circumstances,  $C_i^* : J \leq J^*$ .

Write the smoothed distribution for  $\{y, C\}$ :  $\{\mu_i^k\} = (\mu_1^1, \dots, \mu_{n_1}^1; \dots; \mu_1^k, \dots, \mu_{n_k}^k)$ , where  $\mu_g^k = \dots = \mu_h^k = \dots = \mu_h^k | \forall k, g = 1 + \sum_{l=1}^{k-1} n_l$ , and  $h = \sum_{l=1}^k n_l$ .

Consider a single unobserved circumstance  $C^{J+1}$ , so that  $C_i^* = (C_i, C_i^{J+1})$ . Then  $\theta_a^*({y, C^*}) = I(\{\mu_i^{*k}\}) = I(\mu_1^{*1}, \dots, \mu_{n_1}^{*1}, \dots; \mu_i^{*K^*}, \dots, \mu_N^{*K^*})$ , with  $K \leq K^* \leq x_{J+1}\bar{K}$ .  $\{\mu_i^{*k}\}$  is obtained from  $\{\mu_i^k\}$  by replacing each subvector  $(\mu_g^k, \dots, \mu_h^k)$  with  $(\mu_g^{*k1}, \dots, \mu_{g-1+n_{k1}}^{*k1}; \dots; \mu_i^{*kx_{J+1}}, \dots, \mu_h^{*kx_{J+1}})$ . Since  $(\mu_g^k, \dots, \mu_h^k)$  and  $(\mu_g^{*k1}, \dots, \mu_{g-1+n_{k1}}^{*k1}; \dots; \mu_i^{*kx_{J+1}}, \dots, \mu_h^{*kx_{J+1}})$  have the same mean,  $\mu^k$ , but  $I(\mu_g^k, \dots, \mu_h^k) = 0$  and  $I(\mu_g^{*k1}, \dots, \mu_{g-1+n_{k1}}^{*k1}; \dots; \mu_i^{*kx_{J+1}}, \dots, \mu_h^{*kx_{J+1}}) \geq 0$ , it must be possible to obtain  $(\mu_g^{*k1}, \dots, \mu_{g-1+n_{k1}}^{*k1}; \dots; \mu_i^{*kx_{J+1}}, \dots, \mu_h^{*kx_{J+1}})$  from  $(\mu_g^k, \dots, \mu_h^k)$  by a sequence of mean-preserving spreads. Since this is true for all  $k \in [1, K]$ , and since  $I(\cdot)$  satisfies the principle of transfers, it follows that

$$I(\mu_1^{*1}, \dots, \mu_{n_1}^{*1}, \dots; \mu_i^{*K^*}, \dots, \mu_N^{*K^*}) \geq I(\mu_1^1, \dots, \mu_{n_1}^1; \dots; \mu_i^K, \dots, \mu_N^K).$$

The same argument holds *a fortiori* for  $J^* = J + p, p \in N, p \geq 1$ . QED.

*Corollary:* The IOR measure  $\theta({y, C})$  is a lower bound estimator of true inequality of opportunity ratio,  $\theta_r^*({y, C^*})$ .

*Proof:* The denominator of  $\theta({y, C})$ ,  $I(y)$ , is invariant in changes to the vector  $C$ . QED.

The intuition for the proof of the above proposition is very simple. Imagine that an additional circumstance, previously unobserved, now becomes observed, raising the dimension of  $C$  from  $J$  to  $J + 1$ . This causes every cell in the partition  $\Pi$  to be further subdivided (into  $x_{J+1}$  cells), increasing the maximum number of types,  $\bar{K}$ , by a factor of exactly  $x_{J+1}$ . The effect of this on  $E_0(\{\mu_i^k\}) = \frac{1}{N} \sum_{i=1}^N \log \frac{\mu}{\mu_i^k}$  cannot be negative. Observing a previously omitted circumstance variable cannot lower the between-group inequality share and, unless the additional element is orthogonal to the measure of advantage, will raise it.<sup>20</sup>

The parametric estimates are, as noted above, merely alternative estimates for the same quantities, which rely on linear regressions to economize on data. They are also, necessarily, lower-bound estimates. Although we do not provide a formal proof for the parametric case, the intuition is analogous to the one underlying the proposition above. Consider including an additional element of  $C$  in the regression  $\ln y = C\psi + \varepsilon$ . This cannot reduce—and will in general increase—the share of the variation in  $y$  which is accounted for by  $\tilde{\mu}_i = \exp[C_i\psi]$ . Subvectors of  $\tilde{\mu}$  which were previously constant now contain variation, given a new element in  $C_i$ . Including previously unobserved circumstances will in general raise  $\theta_a^p, \theta_r^p$ , and their

<sup>20</sup>A similar effect would arise from refining the partition of the population into more categories *within* each circumstance variable in  $C$ —i.e. increasing  $x_j$  for a given  $J$ . An example from our empirical analysis below is the classification of parental occupations into only two cells: “agricultural worker” or “other.” For most circumstance variables, international comparability required aiming for “common denominator,” relatively aggregated classifications. Like adding other circumstance variables, further subdivision of these categories within each circumstance might also increase (but could not reduce) the share of inequality attributed to opportunities.

standardized analogues. This makes all empirical estimates given by (5'), (6'), (10'), or (11') lower-bound estimates.<sup>21</sup>

The second methodological consideration worthy of note is that the parametric approach *might* permit the estimation of the partial effects of one (or a subset) of the circumstance variables, controlling for the others, by constructing alternative counterfactual distributions, such as:

$$(12) \quad \tilde{v}_i^J = \exp[\bar{C}_i^J \hat{\psi}^J + C_i^{j \neq J} \psi^{j \neq J} + \hat{\varepsilon}_i]$$

in the case of a parametrically standardized decomposition. In equation (12), instead of holding *all* circumstance variables to a constant value, as in (9), only one circumstance ( $J$ ) is equalized across individuals, while all others are allowed to take their actual values. The resulting counterfactual distribution allows us to compute circumstance  $J$ -specific inequality shares, or “partial IORs”:

$$(13) \quad \theta_r^J = 1 - E_0(\tilde{v}^J) / E_0(y).$$

However, such partial shares *do* rely on the validity and unbiasedness of specific reduced-form coefficients  $\psi$ . These are not, therefore, lower-bound estimates of anything. They are meaningful only as estimates of the (total) contribution of a particular circumstance to inequality of opportunities under the much stronger assumption that any circumstance variables omitted from the reduced-form regression  $\ln y = C\psi + \varepsilon$  are orthogonal to  $C$ . While we report some of the partial shares given by (13) in Section 5, we do not place much weight on them, given their strong assumption requirements.

We now apply this approach to measuring inequality of opportunity for household welfare in six Latin American countries. For each country, we report—and compare—both non-parametric (equations (5') and (6')) and parametric estimates (equations (10') and (11')) for IOL and IOR. We also report some partial shares for individual circumstances, subject to the caveat discussed immediately above. Before presenting the results in Sections 5, the next section briefly describes the datasets.

#### 4. THE DATA

We use data from six nationally representative household surveys in Latin America, namely the Brazilian *Pesquisa Nacional por Amostra de Domicílios* (PNAD) 1996; the Colombian *Encuesta de Calidad de Vida* (ECV) 2003; the Ecuadorian *Encuesta de Condiciones de Vida* (ECV) 2006; the Guatemalan *Encuesta Nacional sobre Condiciones de Vida* (ENCOVI) 2000; the Panamanian

<sup>21</sup>It is of course possible that the share of inequality attributed to a *specific set* of (observed) circumstances is overestimated—say, because some unobserved circumstance variable is positively correlated with all observed ones. But the share of inequality attributed to *all* circumstances (rather than to the observed subset) cannot fall by enlarging the circumstance set. This emphasis on the lower-bound measure of the effect of all circumstances is a major departure from Bourguignon *et al.* (2007a), who sought to estimate the effect of a specific, observed set of circumstances, on opportunities. That objective required them to use Monte-Carlo simulations to estimate bounds around the possible biases in specific coefficients. If one is interested in a lower-bound for the *overall* effect of all circumstances, that procedure is unnecessary.

TABLE 1  
HOUSEHOLD SURVEY NAMES, DATES, AND SAMPLE SIZES

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Survey	PNAD 1996	ECV 2003	ECV 2006	ENCOVI 2000	ENV 2003	ENAHO 2001
Sample of 30 to 49 year-olds	85,692	22,517	12,650	6,956	6,339	17,030
Sample of heads and spouses, aged 30 to 49 years	73,847	18,069	10,719	6,067	5,105	13,947
Of those, observations with income/consumption and circumstances	70,521	17,979	10,719	5,988	4,556	13,621
(share of original sample)	0.823	0.798	0.847	0.861	0.719	0.800

*Encuesta de Niveles de Vida* (ENV) 2003; and the Peruvian *Encuesta Nacional de Hogares* (ENAHO) 2001. This particular group of surveys was selected for containing information on family background for adult individuals, such as their parents' education levels, father's occupation, or both.

In all countries, we restrict the sample to individuals aged 30 to 49, who report being household heads or their spouses.<sup>22</sup> Sample sizes for each survey, both before and after excluding observations with missing data, are reported in Table 1. Sample sizes with complete information for our analysis range from about 4500 (for Panama) to 70,000 observations (for Brazil).

Household wellbeing, proxied by household per capita income (and consumption expenditure, where available) is used as our measure of advantage ( $y$ ). Household incomes are computed as the sum of all household members' individual incomes, and include earnings from all jobs, plus all other reported incomes, such as those from assets, pensions, and transfers. Consumption expenditure data is not available for Brazil. Table 2 reports means and standard deviations (in domestic currencies) for the two advantage variables in our six countries.

As always, international comparisons of household wellbeing are problematic. Different surveys define concepts slightly differently, and questions are never asked in exactly the same manner. In our set of six countries, income and consumption are adjusted for differences in the local cost of living in most Living Standards Measurement Study (LSMS) datasets (Ecuador, Guatemala, and Panama) and in the Brazilian PNAD and Peruvian ENAHO datasets. They are not so adjusted in Colombia. LSMS surveys (Colombia, Ecuador, Guatemala, and Panama) and the ENAHO survey also include imputed rents for owner-occupied housing in both consumption and income aggregates, whereas the PNAD does not.

<sup>22</sup>The age restriction enables us to focus on the cohorts with the highest proportion of employed persons. The household head or spouse restriction seeks to improve comparability across countries, since in both Brazil and Peru the family background information was only collected for these individuals.

TABLE 2  
DESCRIPTIVE STATISTICS: ECONOMIC ADVANTAGE

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Currency unit	Reais 1996	Pesos 2003 (thousands)	USD 2006	Quetzal 2000	Balboas (USD) 2003	Sols 2001
Per capita total household income	302 (538)	324 (562)	167 (265)	678 (1260)	254 (370)	376 (683)
Per capita consumption		347 (512)	125 (134)	602 (715)	187 (195)	307 (353)

*Note:* Means (and standard deviations) for the advantage variables in the population, in domestic currencies.

*Source:* All six surveys, samples for analysis of per capita income and consumption.

There are also some differences in reporting periods. For consumption expenditures, the reference period is the year everywhere, but some expenditures are captured on a weekly or monthly basis. For incomes, the reference period is the month for wage earnings in all surveys. But the reference period for earnings from self-employment varies: it is the month in Brazil, Colombia, and Peru; depends on the frequency of payments in Panama; and is the year in Ecuador and Guatemala.

Despite these methodological differences, the surveys used in this study are likely more comparable with one another than is usually the case in a developing country context. They are all from Latin America, a region where national statistical institutes have made some effort toward methodological consistency.<sup>23</sup> Our samples are also closely comparable to the harmonized samples contained in the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), which is jointly maintained by The World Bank and the Centro de Estudios Distributivos, Laborales y Sociales (CEDLAS) of the University of La Plata (Argentina). The SEDLAC database currently represents the most systematic attempt at generating a harmonized, comparable collection of surveys across Latin America. Table A1 (in the Appendix) reports mean per capita income, the Gini coefficient, and the mean log deviation from the CEDLAS data, and from the full samples in our own data. It also reports the same statistics for the more restricted samples used in our analysis (household heads and spouses aged 30–49, with complete information on advantages and circumstances). Of course, means and inequality indices are not identical between our dataset and the standardized database—in part because we have deliberately retained certain income sources that SEDLAC excludes from their income aggregates—but they are generally quite close, providing welcome external validation of the data we use.<sup>24</sup>

<sup>23</sup>Some such efforts were undertaken during the 1990s and 2000s under the aegis of the MECOVI program of statistical training, co-sponsored by the World Bank and the Inter-American Development Bank.

<sup>24</sup>One exception is Colombia, where we use a different survey altogether from the one contained in SEDLAC. Mean incomes are much higher in the ECV (which we use) than in the Encuesta de Condiciones de los Hogares (ECH), contained in SEDLAC. We retain the ECV, because it has better information on circumstance variables, and is a well-reputed survey in Colombia. Nevertheless, the discrepancy between means reported by ECH and ECV for the same year is a source of concern.

Turning to the circumstance vector ( $C$ ), the surveys contain information on the following common set of circumstances: (a) three variables related to family background: father's and mother's education and father's occupation during the person's childhood; (b) ethnicity (or race); (c) region of birth (or type of area of birth); and (d) gender. The only exception is that the father's occupation variable is not available for Colombia or Peru, and results must be interpreted with this caveat in mind.

The gender variable requires special treatment. Since our advantage indicators are defined at the level of the household, and the gender of the household head is endogenous—both because in some countries the reported headship is an interviewee choice, and because household formation (e.g. whether or not one marries) is, at least in part, a matter of choice—it is not a circumstance variable. Because there are personal choices involved in household headship, including gender as a circumstance variable in the analysis of opportunity for household income or consumption would invalidate our claim that our scalar indices of inequality of opportunity are lower-bound estimators. However, given the importance of gender as a circumstance at the individual level, Section 5 does contain some results from an analysis of inequality of opportunity when labor earnings is the advantage variable.<sup>25</sup>

Parental education variables are coded into three categories: no education (or unknown); primary (incomplete or complete, depending on the country); and complete primary or secondary and more.<sup>26</sup> Father's occupation is recoded into two categories: agricultural workers and others. Ethnicity (coded in two categories) is captured either by self-reported ethnicity or by the ability to speak an indigenous language. Region of birth is coded into three broad regions (one being generally the capital area) but is captured by the type of area (urban/rural) for Panama. Table 3 describes the specific definitions of the circumstance variables in each survey in greater detail. Table 4 presents the corresponding descriptive statistics.

The number of categories ( $x_j$ ) for each circumstance variable was restricted to three or fewer, so as to reduce the number of types with zero or very few observations in the sample. As discussed in Section 3, this is important for the non-parametric analysis, which crucially relies on the precision of the estimates of conditional means for each type (or "cell" of the partition). As sampling variance is high for cells containing few observations, estimated between-type inequality may become inflated, thereby inducing an overestimation of inequality of opportunity. Table 5 shows the maximum number of types in each country, the number of types actually observed in each sample (i.e. the complement of the number of empty cells), the mean cell size, and the proportion of cells with fewer than five observations. Despite observing only six circumstance variables and

<sup>25</sup>A fuller discussion of inequality of opportunity for individual labor earnings can be found in the working paper version of this article (Ferreira and Gignoux, 2008).

<sup>26</sup>Whether complete primary attainment was included as part of the middle or upper grouping for parental education depends on relative group sizes. An effort was made to prevent the top grouping from becoming too small relative to other countries, to enhance comparability. None of the results are particularly sensitive to these decisions.

TABLE 3  
DEFINITION OF CIRCUMSTANCE VARIABLES, BY COUNTRY

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Ethnicity						
Category 1	Self reported white ethnicity	Other	Self-reported ethnicity: white, mixed blood ("mestizo") or other	European maternal language	Other	European maternal language
Category 2	Self reported black ("negro") and mixed blood ("pardo") ethnicity	Self-reported minority ethnicity: "indígena, gitano, archipiélago o palenquero"	Self-reported ethnicity: indigenous, black ("negro" or "mulato")	Indigenous maternal language	Speaks indigenous language	Indigenous maternal language
Father's occupation						
Category 1	Agricultural worker	Missing	Agricultural worker or domestic worker	Agricultural worker	Agricultural worker	Missing
Category 2	Other		Other	Other	Other	
Mother's and father's education						
Category 1	None or unknown	None or unknown	None or unknown	None or unknown	None or unknown	None or unknown
Category 2	Completed grade 1 to 4	Primary incomplete	Primary	Primary incomplete	Primary	Primary incomplete
Category 3	Completed grade 5 or more	Primary complete or more	Secondary or more	Primary complete or more	Secondary or more	Primary complete or more
Birth region						
Category 1	Sao Paulo and Federal district	Departments at the periphery	Sierra and Amazonia provinces	Guatemala City, Northeast departments and El Petén	Cities and intermediate urban centers	Inland non-southern departments
Category 2	South East, Center-West, and South	Central departments(a)	Costa and Insular provinces	North and Northwest departments	Other urban centers	Southern and other coastal departments
Category 3	North-East, North or missing	Bogota, San Andres, and Providencia islands and foreign country	Pichincha province (with Quito) and Azuay province	Southeast, Southwest, and Center departments	Rural areas	Arequipa, Callao, and Lima

Note: Central departments are Boyaca, Caldas, Cauqueta, Cundinamarca, Huila, Meta, Norte de Santander, Quindio, Risaralda, Santander, Tolima, and Valle del Cauca.

TABLE 4  
DESCRIPTIVE STATISTICS: CIRCUMSTANCES

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Gender						
Male	47.4	44.8	48.1	47.3	46.8	47.6
Female	52.6	55.2	51.9	52.7	53.2	52.4
Ethnicity						
Majority	59.8	90.6	87.6	68.0	92.6	72.2
Minority	40.2	9.5	12.4	32.0	7.4	27.8
Father's occupation						
Agricultural worker	35.0	missing	54.1	52.2	38.8	missing
Other	65.0		45.9	47.8	61.2	
Father's education						
None or unknown	50.2	36.9	29.6	69.1	22.3	30.9
Primary	40.3	49.8	55.4	16.1	54.7	32.2
Primary complete/secondary	9.6	13.3	15.0	14.8	23.0	36.9
Mother's education						
None or unknown	53.1	32.2	30.9	78.2	25.6	48.7
Primary	38.0	55.0	55.5	11.7	54.6	24.9
Primary complete/secondary	9.0	12.8	13.6	10.1	19.8	26.4
Birth region						
Region 1	17.6	44.2	32.6	26.4	28.1	45.5
Region 2	47.3	47.1	49.5	21.2	22.2	35.7
Region 3	35.2	8.7	17.8	52.4	49.7	18.8

*Notes:* All entries are population shares. Sample: household heads and spouses, aged 30–49, with positive income and information on a set of circumstances; father's occupation missing for Colombia and Peru.

*Source:* All six surveys.

TABLE 5  
DESCRIPTION OF THE SAMPLE PARTITION

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Maximum number of types	108	54	108	108	108	54
Number of types observed	108	53	101	96	82	53
Mean number of observations per type	653.0	339.2	106.1	62.4	55.6	257.0
Proportion of types with fewer than 5 observations	0.065	0.038	0.188	0.260	0.305	0.075

*Note:* Types are defined by crossing the circumstance variables in Tables 3 and 4.

*Source:* All six surveys.

exercising considerable parsimony in the partitioning of the population, we still have two surveys—from Guatemala and Panama—for which over 25 percent of cells have fewer than five observations. By contrast, Brazil's PNAD survey, with a sample size one order of magnitude larger, has 6.5 percent of cells with fewer than five observations. Colombia and Peru also have relatively few sparsely populated cells. These differences underscore the importance of the parametric estimates in validating (or refuting) the non-parametric results presented below.

## 5. RESULTS

Total household income (or consumption expenditure) per capita is by no means the only—or perhaps even the most important—advantage that people have cause to value. Even in terms of measuring purely economic welfare these aggregates are incomplete, since they generally do not include a valuation for access to public or publicly provided goods (such as public safety, or free public education and health care, respectively). They do, however, provide a reasonable measure of a household's command over private goods, which is an important dimension of well-being. They are often the best available indicators of well-being available in regular household surveys, and are the two measures of economic advantage used in this study.

How unequal is the distribution of opportunity for economic advantage in the set of countries for which we have data? Table 6 presents both the IOL ( $\theta_a$ ) and IOR ( $\theta_r$ ) measures of inequality of opportunity for household income (Panel A) and consumption expenditures per capita (Panel B). The first row in both panels reports total inequality (measured by  $E_0$ ) in the sample for each country. As expected, these measures are always higher for income than for consumption, for the usual reasons associated with (likely) greater measurement error and a larger

TABLE 6  
SCALAR INDICES OF INEQUALITY OF OPPORTUNITY

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
<b>Panel A: Household income (per capita)</b>						
Total inequality ( $E_0$ )	0.692 (0.013)	0.572 (0.033)	0.580 (0.028)	0.593 (0.036)	0.630 (0.029)	0.557 (0.022)
Non-parametric estimates						
IOL	0.227 (0.008)	0.144 (0.023)	0.164 (0.022)	0.213 (0.031)	0.213 (0.024)	0.163 (0.015)
IOR	0.329 (0.008)	0.252 (0.026)	0.283 (0.023)	0.359 (0.030)	0.338 (0.026)	0.293 (0.018)
Parametric estimates						
IOL	0.223 (0.008)	0.133 (0.019)	0.150 (0.020)	0.199 (0.028)	0.190 (0.023)	0.156 (0.014)
IOR	0.322 (0.008)	0.232 (0.023)	0.259 (0.023)	0.335 (0.030)	0.301 (0.028)	0.279 (0.018)
<b>Panel B: Household consumption expenditures (per capita)</b>						
Total inequality ( $E_0$ )		0.462 (0.018)	0.359 (0.015)	0.415 (0.025)	0.381 (0.018)	0.351 (0.013)
Non-parametric estimates						
IOL		0.123 (0.015)	0.124 (0.013)	0.221 (0.024)	0.156 (0.016)	0.123 (0.010)
IOR		0.265 (0.021)	0.346 (0.021)	0.532 (0.023)	0.409 (0.025)	0.351 (0.018)
Parametric estimates						
IOL		0.114 (0.014)	0.117 (0.012)	0.213 (0.022)	0.144 (0.015)	0.119 (0.009)
IOR		0.247 (0.021)	0.326 (0.022)	0.514 (0.022)	0.377 (0.026)	0.339 (0.017)

*Notes:* Sample: household heads and spouses, aged 30–49, with positive income and information on a set of circumstances; bootstrap standard errors (taking into account stratification and clustering) in parentheses; father's occupation missing for Colombia and Peru.

*Source:* All six surveys.

variance for transitory components in the distribution of incomes (see, e.g. Deaton, 1997). The next two rows report the non-parametric estimates of IOL ( $\theta_a$ ) and IOR ( $\theta_r$ ), followed by two rows with the parametric (standardized) estimates for the same indices:  $\theta_a^{PS}$  and  $\theta_r^{PS}$ . In all cases, bootstrapped standard errors are reported in parentheses, taking into account sampling weights, stratification, and clustering.

Table 7 reports the OLS coefficients of the reduced-form equation that was used to generate the parametrically standardized distributions (given by equation (9)) of household per capita income. An analogous table for the reduced-form regression of consumption expenditures on circumstances is available from the authors on request. All coefficients in these regressions have the expected signs, and most are significant at the 1 percent level. Coefficient sizes are consistent with a reduced-form specification.

The non-parametric estimates of IOL for household incomes range from 0.14 (in Colombia) to 0.23 (in Brazil). IOR ranges from 25 percent (in Colombia) to 36 percent (in Guatemala). The parametric estimates are only slightly lower in both cases: from 0.13 in Colombia to 0.22 in Brazil for IOL, and from 23 percent in Colombia to 34 percent in Guatemala for IOR. The differences between the indices generated parametrically and non-parametrically are never statistically significant, which provides a welcome sense of robustness. In addition, country rankings are also quite robust, both to the choice of index (IOL vs. IOR) and to the estimation method (parametric or not). The rank-correlation between parametric and non-parametric estimates (for IOL) is 0.89, and between parametric estimates of IOL and IOR it is 0.94.

For consumption, our non-parametric (parametric) estimates of IOL are: 0.12 (0.11) in Colombia, 0.12 (0.12) in both Ecuador and Peru, 0.16 (0.14) in Panama, and 0.22 (0.21) in Guatemala.<sup>27</sup> The corresponding estimates for the shares of inequality of opportunity are: 27 percent (25 percent) in Colombia, 35 percent (33 percent) in Ecuador, 35 percent (34 percent) in Peru, 41 percent (38 percent) in Panama, and 53 percent (51 percent) in Guatemala.<sup>28</sup> The rank-correlation between parametric and non-parametric estimates for IOL is 0.90, and between parametric estimates of IOL and IOR it is 1.00.

We also investigate the extent to which these results are sensitive to different assumptions about equivalence scales, by adopting two alternative scale economy parameters in the Buhmann *et al.* (1988) scale ( $\alpha = 0.5$ ;  $\alpha = 0.75$ ).<sup>29</sup> All of our estimates of total inequality and inequality of opportunity (IOL and IOR) for both income and consumption were recomputed for the distributions of equivalized income corresponding to those two scales; results are presented in Table 8.

<sup>27</sup>The differences between the levels of inequality of consumption opportunity observed in Colombia, Ecuador, and Peru are not statistically significant.

<sup>28</sup>With the exception of the difference between Ecuador and Peru, all cross-country differences are significant at the 5 percent level, on the basis of the bootstrapped standard errors.

<sup>29</sup>The Buhmann *et al.* equivalence scale is a parametric class of scales given by the simple transformation  $y_{eq} = \frac{y}{n^\alpha}$ , where  $y$  is total household income,  $n$  is household size, and  $\alpha \in [0, 1]$ . Given the lack of robustness of specific econometric scales, different values of the parameter  $\alpha$  are often used to test the sensitivity of inequality measures to different assumptions about equivalence scales. See Coulter *et al.* (1992a, 1992b).

TABLE 7  
REDUCED-FORM OLS REGRESSION OF HOUSEHOLD INCOME ON OBSERVED CIRCUMSTANCES

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Member of an ethnic minority	-0.416*** (0.009)	0.004 (0.035)	-0.210*** (0.033)	-0.335*** (0.037)	-0.871*** (0.063)	-0.229*** (0.023)
Father not an agricultural worker	0.296*** (0.009)		0.442*** (0.025)	0.160*** (0.035)	0.224*** (0.039)	
Father primary education	0.249*** (0.010)	0.200*** (0.025)	0.217*** (0.029)	0.211*** (0.044)	0.229*** (0.051)	0.190*** (0.026)
Father secondary education	0.699*** (0.018)	0.661*** (0.046)	0.527*** (0.046)	0.353*** (0.087)	0.505*** (0.066)	0.441*** (0.031)
Mother primary education	0.297*** (0.010)	0.238*** (0.029)	0.337*** (0.029)	0.336*** (0.065)	0.247*** (0.049)	0.216*** (0.026)
Mother secondary education	0.789*** (0.018)	0.713*** (0.045)	0.736*** (0.046)	0.877*** (0.093)	0.726*** (0.069)	0.573*** (0.034)
Birth region 2	-0.377*** (0.012)	0.176*** (0.024)	-0.192*** (0.024)	-0.266*** (0.053)	-0.019 (0.051)	0.058*** (0.021)
Birth region 3	-0.737*** (0.013)	0.454*** (0.032)	0.214*** (0.032)	-0.237*** (0.049)	-0.328*** (0.051)	0.375*** (0.030)
Constant	5.079*** (0.013)	11.583*** (0.024)	3.948*** (0.028)	5.923*** (0.050)	4.480*** (0.063)	4.916*** (0.021)
Observations	70,521	17,979	10,719	5,988	4,556	13,621
R-squared	0.324	0.175	0.249	0.258	0.320	0.248

Notes: Standard errors in parentheses. \*Significant at 10%; \*\*significant at 5%; \*\*\*significant at 1%. Omitted categories are: male, ethnic majority, father agricultural worker, father and mother with no or unknown education, and birth region 1 (see Table 3 for the country-specific definitions).  
Source: All six surveys.

TABLE 8  
ROBUSTNESS TO THE APPLICATION OF AN EQUIVALENCE SCALE

	Brazil		Colombia		Ecuador		Guatemala		Panama		Peru	
	$\alpha = 0.50$	$\alpha = 0.75$										
Equivalence scale												
<b>Panel A: Household income (per capita)</b>												
Total inequality ( $E_0$ )	0.609 (0.009)	0.644 (0.009)	0.490 (0.032)	0.524 (0.034)	0.503 (0.021)	0.534 (0.024)	0.527 (0.032)	0.554 (0.036)	0.530 (0.027)	0.572 (0.028)	0.474 (0.019)	0.508 (0.020)
Non-parametric estimates												
IOL	0.197 (0.007)	0.212 (0.007)	0.121 (0.021)	0.133 (0.024)	0.139 (0.015)	0.152 (0.018)	0.170 (0.026)	0.191 (0.031)	0.176 (0.022)	0.194 (0.022)	0.129 (0.012)	0.145 (0.014)
IOR	0.324 (0.007)	0.329 (0.007)	0.248 (0.026)	0.253 (0.029)	0.277 (0.019)	0.284 (0.021)	0.322 (0.030)	0.345 (0.033)	0.332 (0.027)	0.340 (0.026)	0.272 (0.017)	0.286 (0.020)
Parametric estimates												
IOL	0.194 (0.007)	0.208 (0.007)	0.112 (0.018)	0.122 (0.020)	0.129 (0.014)	0.140 (0.017)	0.157 (0.024)	0.178 (0.029)	0.158 (0.021)	0.174 (0.021)	0.123 (0.012)	0.139 (0.013)
IOR	0.318 (0.008)	0.323 (0.008)	0.228 (0.025)	0.234 (0.025)	0.258 (0.018)	0.262 (0.021)	0.297 (0.032)	0.320 (0.034)	0.298 (0.029)	0.304 (0.025)	0.259 (0.017)	0.273 (0.019)
<b>Panel B: Household consumption expenditures (per capita)</b>												
Total inequality ( $E_0$ )			0.378 (0.020)	0.413 (0.018)	0.291 (0.014)	0.318 (0.015)	0.332 (0.021)	0.368 (0.024)	0.295 (0.017)	0.331 (0.017)	0.289 (0.011)	0.313 (0.012)
Non-parametric estimates												
IOL	0.103 (0.015)	0.113 (0.014)	0.103 (0.014)	0.113 (0.014)	0.107 (0.011)	0.116 (0.012)	0.180 (0.019)	0.200 (0.022)	0.120 (0.014)	0.138 (0.015)	0.096 (0.008)	0.109 (0.008)
IOR	0.273 (0.025)	0.273 (0.025)	0.273 (0.025)	0.273 (0.025)	0.367 (0.022)	0.363 (0.020)	0.542 (0.023)	0.545 (0.023)	0.407 (0.025)	0.416 (0.025)	0.333 (0.018)	0.349 (0.016)
Parametric estimates												
IOL	0.097 (0.014)	0.106 (0.013)	0.097 (0.014)	0.106 (0.013)	0.102 (0.011)	0.110 (0.011)	0.173 (0.017)	0.193 (0.020)	0.113 (0.013)	0.128 (0.014)	0.093 (0.008)	0.106 (0.008)
IOR	0.257 (0.024)	0.256 (0.022)	0.257 (0.024)	0.256 (0.022)	0.350 (0.022)	0.345 (0.020)	0.521 (0.021)	0.525 (0.021)	0.382 (0.025)	0.387 (0.025)	0.322 (0.017)	0.337 (0.016)

Notes: Sample: household heads and spouses, aged 30–49, with positive income and information on a set of circumstances; bootstrap standard errors (taking into account stratification and clustering) in brackets; father's occupation missing for Colombia and Peru.  $\alpha$  denotes the economy of scale parameter in the Buhmann *et al.* (1988) equivalence scale.

Source: All six surveys.

As expected, estimates of total inequality decline as the degree of scale economies allowed for within the household rises (i.e. as  $\alpha$  declines from 1) (see Coulter *et al.*, 1992a, 1992b). The same is true, albeit to a lesser extent, of the levels of inequality of opportunity estimated by IOL. With both the numerator and denominator declining, however, the ratios of inequality of opportunity (IOR) are generally stable. For incomes, they are marginally higher with  $\alpha = 0.75$  than with  $\alpha = 1.0$  (Table 6), but a little lower with  $\alpha = 0.5$ . For consumption, interestingly, while IOLs are still lower for  $\alpha < 1.0$  than in the per capita case, IORs are actually generally somewhat higher. In Guatemala, for example, which already had the highest IOR in our sample, the (non-parametric) inequality of opportunity ratio reaches 0.542 for  $\alpha = 0.5$  (0.545 for  $\alpha = 0.75$ ). In Ecuador, the parametric estimate of IOR rises from 0.326 ( $\alpha = 1.0$ ) to 0.350 ( $\alpha = 0.5$ ).

Overall, it seems that inequality of opportunity levels, much like levels of overall income inequality, are somewhat lower if one allows for a greater degree of scale economy in household consumption. But the share of outcome inequality accounted for by inequality of opportunities is either stable (in the case of incomes) or even higher (in the case of consumption) at these more generous scales.

Two features of our benchmark results in Table 6 warrant further remarks. First, these are relatively large estimates of inequality of opportunity, particularly since they are lower bounds: between one fifth and one third of all income inequality, and between one quarter and one half of all consumption inequality is associated with opportunities in these six countries. This compares, for instance, with just under 20 percent for incomes in Italy, in similar estimates by Checchi and Peragine (2010).

Second, the differences in IOL and IOR between income and consumption are interesting. Inequality of opportunity *ratios* are higher for consumption than for income in all five countries, and the difference is often substantial (e.g. in the order of 20 percent for Ecuador, Panama, and Peru, and even higher for Guatemala). However, this is driven entirely by much larger within-type components of inequality in the income decomposition. Inequality of opportunity *levels* are actually generally lower for consumption than for income (with the exception of Guatemala). IORs are lower for income because the residual inequality in the income distribution is considerably higher, which is consistent with the view that there is greater measurement error, and transitory income variance, in that variable. This suggests the possibility that income-based IORs may underestimate lifetime (or permanent income) inequality of opportunity, since transitory income variance (and likely higher measurement error) is effectively counted as inequality due to “efforts and luck.”<sup>30</sup>

Table 9 presents our estimates of the circumstance-specific opportunity shares of inequality (partial IORs) as defined by equation (13), for both household income (Panel A) and consumption expenditure (Panel B) per capita. Panel C presents analogous results for *individual* labor earnings. As discussed above, the interpretation of these partial shares requires the much stronger additional

<sup>30</sup>See Bourguignon *et al.* (2007b) for a discussion. The finding is analogous to the well-known fact that inter-generational mobility estimates are much higher when based on single-period wages for parents and children, than when based on longer earnings histories. See, inter alia, Solon (1999) and Mazumder (2005).

TABLE 9  
CIRCUMSTANCE-SPECIFIC OPPORTUNITY SHARES (PARTIAL IORS)

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
<b>Panel A: Household income (per capita)</b>						
Race/ethnicity	0.086 (0.004)	0.000 (0.002)	0.016 (0.005)	0.079 (0.014)	0.065 (0.010)	0.044 (0.008)
Father's occupation	0.047 (0.003)		0.099 (0.012)	0.043 (0.012)	0.052 (0.014)	
Father's education	0.133 (0.006)	0.142 (0.019)	0.113 (0.016)	0.112 (0.044)	0.111 (0.024)	0.120 (0.014)
Mother's education	0.144 (0.007)	0.147 (0.021)	0.146 (0.017)	0.212 (0.034)	0.160 (0.026)	0.172 (0.015)
Birth region	0.079 (0.005)	0.035 (0.013)	0.029 (0.013)	0.057 (0.018)	0.085 (0.019)	0.075 (0.012)
<b>Panel B: Household consumption expenditures (per capita)</b>						
Race/ethnicity		0.001 (0.002)	0.035 (0.007)	0.148 (0.015)	0.111 (0.013)	0.054 (0.007)
Father's occupation			0.106 (0.012)	0.082 (0.015)	0.061 (0.015)	
Father's education		0.147 (0.017)	0.136 (0.017)	0.179 (0.027)	0.116 (0.025)	0.142 (0.012)
Mother's education		0.161 (0.017)	0.180 (0.018)	0.290 (0.030)	0.174 (0.027)	0.204 (0.017)
Birth region		0.043 (0.011)	0.040 (0.012)	0.109 (0.017)	0.103 (0.019)	0.108 (0.013)
<b>Panel C: Individual labor earnings</b>						
Gender	0.036 (0.005)	0.002 (0.010)	0.037 (0.019)	0.058 (0.026)	0.009 (0.012)	0.019 (0.013)
Race/ethnicity	0.076 (0.004)	0.001 (0.002)	0.007 (0.004)	0.032 (0.012)	0.035 (0.007)	0.023 (0.007)
Father's occupation	0.070 (0.003)		0.065 (0.012)	0.007 (0.011)	0.047 (0.014)	
Father's education	0.107 (0.005)	0.110 (0.018)	0.069 (0.015)	0.072 (0.030)	0.079 (0.020)	0.073 (0.014)
Mother's education	0.119 (0.006)	0.108 (0.018)	0.102 (0.014)	0.095 (0.030)	0.094 (0.022)	0.099 (0.013)
Birth region	0.065 (0.005)	0.023 (0.012)	0.016 (0.010)	0.022 (0.016)	0.069 (0.019)	0.044 (0.010)

*Notes:* The sample includes heads and spouses aged 30–49 with positive income and consumption (Panels A and B) or labor earnings (Panel C) and information on a set of circumstances. Bootstrapped standard errors (taking into account stratification and clustering) are in parentheses. Father's occupation is missing for Colombia and Peru. The levels of overall inequality in labor earnings are measured by mean log deviations (resp. Gini indexes) of 0.616 (0.574) in Brazil, 0.616 (0.548) in Colombia, 0.665 (0.554) in Ecuador, 0.786 (0.601) in Guatemala, 0.586 (0.510) in Panama, and 0.675 (0.571) in Peru. The regressions of earnings and consumption on observed circumstances are available from the authors upon request.

*Source:* All six surveys.

assumption that any omitted circumstance variable is orthogonal to those variables included in the regressions reported in Table 7 (and the analogous regressions for consumption). Subject to that caveat, the results suggest that family background characteristics are associated with the largest share of inequality of opportunity. The share of consumption inequality accounted for by mother's education alone is 16 percent or higher in all countries, and as high as 29 percent in Guatemala. In general, father's education is also more important than ethnicity or region of birth in most countries. The higher levels of inequality of opportunity

observed in Guatemala and Panama, however, are associated with larger partial shares for ethnicity and region of birth (which is also important in Peru).

Because co-residence and household headship involve individual choices, the gender of the head of a household cannot be treated as a circumstance variable, and this has prevented us from accounting for the role of gender in our measures of inequality of opportunity so far. This omission can be remedied in part by considering a different kind of economic advantage, namely individual earnings from labor. While labor earnings are a less satisfactory measure of economic wellbeing than those which account for sharing within households, earnings are important in their own right, both as a key income source and, arguably, as a source of self-esteem. They would likely qualify as an alternative measure of “economic advantage,” in Roemer’s terminology, and a fuller analysis of inequality of opportunity for earnings (which is omitted here due to space constraints) can be found in Ferreira and Gignoux (2008).

Panel C of Table 8 presents circumstance-specific opportunity shares of inequality in labor earnings, among the population of occupied individuals, analogously to Panels A and B for income and consumption. After controlling for the other observed circumstances (race, birthplace, family background), the share of overall inequality accounted for by gender differences is actually fairly small, ranging from 0.2 percent in Colombia to 5.8 percent in Guatemala. This is a much smaller share than that associated with family background variables, such as mother’s and father’s education. In one case (Guatemala), the gender share is higher than the share associated with father’s occupation. In two cases (Ecuador and Guatemala), it is higher than the birth region share, and in three cases it is higher than the race/ethnicity share. Overall, it seems that the inclusion of gender as a circumstance variable does not drastically alter our conclusions about the measurement of inequality of opportunity for economic outcomes, nor does it challenge the apparent importance of family background variables in driving that inequality.<sup>31</sup>

## 6. OPPORTUNITY PROFILES: IDENTIFYING THE LEAST ADVANTAGED GROUPS

The scalar indices of inequality of opportunity proposed and computed above allow us to assess the degree of this kind of inequality in a particular society, and possibly to compare different countries (or regions) for which there are comparable joint distributions of advantage and circumstance variables. But the partition  $\Pi = \{T_1, T_2, \dots, T_K\}$  that is used to calculate these indices can also be used directly to identify the least-advantaged social groups in a given population. In a sense, this is closer to the original objective that motivated Roemer (1993b, 1998) and van de Gaer (1993) to formalize these concepts. They were interested in identifying an equal-opportunity *policy*, and tended to think of it as the set of allocation rules that maximized advantage for the worst-off type(s).

But which is the worst-off type? Given a partition  $\Pi$ , various criteria can be used to rank types. As previously discussed, one obvious such ranking would be

<sup>31</sup>These results do not preclude, of course, that gender differences may be important in other dimensions, including educational and employment opportunities. In fact, Ferreira and Gignoux (2010) find that gender is the single most important circumstance driving inequality of opportunity for educational attainment in Turkey.

given by (first- or second-order) stochastic dominance relationships between types. However, rankings based on stochastic dominance suffer from two problems. First, any such ranking is perforce partial and incomplete, since distribution functions (or generalized Lorenz curves) may cross (see Atkinson, 1970). The second issue, more practical in nature, is that the distribution of cell sizes partly summarized in Table 5 makes it impossible to estimate the conditional distributions for the full set of 53–108 types in our partitions. This will be a common problem with usual household survey sample sizes, whenever the analysis requires (or permits) the use of more than two or three circumstance variables. Such practical considerations are of the essence for empirical applications of the theory of equality of opportunity, and should not be overlooked.

An alternative ranking algorithm is to use a particular moment of  $F_k(y)$ , such as the mean—or a particular percentile, such as the median, the first quartile, etc—to rank across types. The type’s mean advantage,  $\mu^k(y)$ , is a natural candidate for such a ranking criterion. It is the metric proposed by van de Gaer (1993) for evaluating the opportunity set faced by individuals in type  $k$ . It also corresponds to the metric Roemer (2006) suggested as a suitable measure of economic development (see footnote 7). It is also central in defining smoothed and standardized distributions, and thus for the construction of our scalar IOL and IOR indices.

We therefore define an *opportunity profile* as the ordered partition  $\Pi^* = \{T_1, T_2, \dots, T_K\} | \mu^1 \leq \mu^2 \leq \dots \leq \mu^K$ , corresponding to any original partition  $\Pi$ . 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  $\Pi^*$  that includes only a certain fraction  $\pi$  of the population that belongs to the lowest-ranked types. Formally:

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

If, for example  $\pi = 0.1$ , 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 percent.

An opportunity-deprivation profile is therefore simply a list of types: those with the lowest mean advantage levels, up to some arbitrary population share threshold. Table 10 presents the opportunity-deprivation profile (with  $\pi = 0.1$ ) for our Brazilian sample, by specifying the full set of circumstances that define each type (ethnicity, mother’s and father’s education levels, father’s occupation and birthplace). It also contains an estimate of the population and mean income in each type (using sample expansion weights), both in absolute terms and as a share of the total.<sup>32</sup>

<sup>32</sup>Similar tables for the other five countries are presented in Ferreira and Gignoux (2008), the working paper version of this article. The number of types in each opportunity-deprivation profile varies substantially across countries: there are five types in Guatemala and Peru, six in Brazil, ten in Colombia, 16 in Ecuador, and 25 in Panama. Ferreira and Gignoux (2008) also discuss a comparison of opportunity-deprivation and standard poverty profiles for this set of countries.

TABLE 10  
THE OPPORTUNITY-DEPRIVATION PROFILE FOR BRAZIL ( $\pi = 0.1$ )

Type	Ethnicity	Father's Occupation	Father's Education	Mother's Education	Place of Birth	Estimated Population	Share of National Population	Mean Advantage (HPCY)	Share of Overall Mean
1	Black and mix-race	Agricultural worker	None or unknown	None or unknown	Nordeste or North	2,237,946	0.06751	79.2	0.262
2	Black and mix-race	Agricultural worker	Upper primary (5) or more	None or unknown	Sao Paulo or Federal District	1,417	0.00004	86.5	0.286
3	Black and mix-race	Agricultural worker	None or unknown	Lower primary	Nordeste or North	310,540	0.00937	101.7	0.337
4	Black and mix-race	Agricultural worker	Lower primary	None or unknown	Nordeste or North	349,183	0.01053	102.0	0.338
5	Black and mix-race	Agricultural worker	Upper primary (5) or more	None or unknown	Nordeste or North	7,564	0.00023	107.0	0.354
6	Black and mix-race	Other	None or unknown	None or unknown	Nordeste or North	2,028,884	0.06120	107.8	0.357

Source: Pnad 1996 survey.

Such a profile permits identifying the social groups at which equal-opportunity policies should presumably be aimed (at least according to van de Gaer's criterion; see footnote 14), and whose welfare should be monitored to assess the pace of economic development (according to Roemer, 2006). They are generally quite informative of which combinations of predetermined, morally irrelevant circumstances lead to the greatest opportunity deprivation in a given society.

The "full" form of the profile, while possibly the most useful to policy-makers in a particular country, does not lend itself as easily to inter-country comparisons. This is attempted by means of Table 11, which summarizes the composition of the opportunity deprivation profiles in all six countries (again for  $\pi = 0.1$ ), in terms of the frequency with which various circumstances are observed in each profile. These profiles are constructed using household per capita consumption as the advantage variable, except for Brazil, where income is used, as in Table 10.

Three common traits are salient. First, members of ethnic minorities form the vast majority of the population in these disadvantaged groups. In three of the six countries, these groups are composed exclusively of members of racial or ethnic minorities: black and mixed-race in Brazil; and native speakers of indigenous languages in Guatemala and Peru. In two other countries, ethnic minorities are still a majority of the opportunity-deprived: 72 percent of the opportunity-deprivation profile in Panama consists of native speakers of indigenous languages; and 62 percent of self-reported indigenous, black, or mixed-race ethnicity in Ecuador. Colombia is the only country in our sample where ethnic minorities are not the majority among the opportunity-deprived but, even there, the proportion of minorities, 43 percent, is much higher than in the population as a whole.<sup>33</sup>

Second, family background is also strongly associated with opportunity-deprivation. In the four countries where this information is available, never fewer than 82 percent of the opportunity-deprived are daughters and sons of agricultural workers; this proportion reaches very nearly 100 percent in Guatemala. Almost the same holds for mother's education: in all countries, more than 90 percent of the opportunity-deprived are daughters and sons of women who did not go to school—99 percent in Guatemala and Peru, 98 percent in Ecuador, 93 percent in Panama, 92 percent in Colombia, and 91 percent in Brazil.

Third, opportunity deprivation is remarkably spatially concentrated. A majority of the opportunity-deprived are often natives of the same specific regions. In Brazil, all persons in our profile were born in the Northeast or North regions; in Colombia, 99 percent hail from peripheral departments; in Guatemala, 99 percent come from either the North or the Northwestern departments; and in Panama, 96 percent were born in a rural area. The similarity in profiles across countries reflects a high degree of correlation among certain circumstance

<sup>33</sup>The composition of the opportunity-deprivation profile depends both on distribution of income across types *and* on the marginal distribution of types. In other words, both the population and the income shares of the indigenous group affect their representation in the deprivation profile. The fact that ethnic minorities are a lower proportion of the Colombian profile than in other countries shown in Table 11 is clearly related to their smaller population share (see Table 4). We are grateful to an anonymous referee for this observation.

TABLE 11  
SUMMARY COMPOSITION OF OPPORTUNITY-DEPRIVATION PROFILES ( $\pi = 0.1$ )

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
Member of ethnic minority	100.0	43.0	61.7	100.0	71.6	100.0
Father's agricultural occupation	87.7		97.9	99.9	82.8	
Other father's occupation	12.3		2.1	0.1	17.2	
Father without education	89.2	65.2	86.2	99.4	82.4	99.8
Father's primary education	10.5	34.8	11.9	0.3	12.2	0.2
Father's secondary education (or complete primary)	0.3	0.0	1.9	0.3	5.4	0.0
Mother without education	90.6	92.2	97.5	99.2	93.4	99.4
Mother's primary education	9.4	7.8	1.0	0.3	4.8	0.0
Mother's secondary education (or complete primary)	0.0	0.0	1.5	0.5	1.8	0.6
Birth regions	Northeast and North (100%)	Periphery (99%)	Coast and insular (51%) Sierra and Amazonia (48%)	North and Northwest (99%)	Rural areas (96%)	South and Coast (58%) Inland (42%)

Source: All six surveys.

variables: the combination of characteristics that are markers of deprivation is relatively similar across these six countries.

## 7. CONCLUSIONS

Drawing on two (previously unrelated) approaches to the measurement of inequality of opportunity, this paper shows that the ex-ante measure of Checchi and Peragine (2010) and a variant of the index proposed by Bourguignon *et al.* (2007) are conceptually equivalent. Both measure inequality of opportunity as the between-type share of overall outcome inequality: CP estimate it non-parametrically, while our version of BFM estimates it parametrically. We formally derive the measure from Roemer's (1998) definition of the equal opportunity policy and show that the measure follows naturally from it, but only provided that the empirical identification criterion for equality of opportunity is substantially weakened, from equality of conditional distributions to equality of conditional means. Such a weakening is consistent with van de Gaer's (1993) ex-ante approach to inequality of opportunity, and with his "min of means" definition of the equal-opportunity policy.

Although the previous literature refers only to the share (or relative) version of the index, an absolute version can also be defined and, in fact, satisfies two additional (and appealing) properties: within-type transfer insensitivity and the between-type transfer principle. We refer to the absolute version of the index as inequality of opportunity level (IOL), and to the relative version as inequality of opportunity ratio (IOR). Both are shown to be lower-bound estimators of true inequality of opportunity. This is because any unobserved circumstance would lead to a finer partition of the population and thus to greater inequality in the smoothed distribution.

We estimate both versions of this lower-bound index of inequality of opportunity for six countries in Latin America, whose household surveys contain information on a number of predetermined, morally irrelevant circumstances, namely: gender, race or ethnicity, birthplace, mother's and father's education, and father's occupation. IOL and IOR indices were calculated for both household per capita income and consumption expenditure. In all cases, both parametric and non-parametric estimates were computed and, reassuringly, they were generally quite similar. Differences across estimation methods were never statistically significant, suggesting a reasonable degree of robustness of the results. With small sample sizes, however, the parametric approach provides the preferred lower-bound estimates for inequality of opportunity.

When household per capita income was used as the advantage variable, IOL ranged from 0.13 in Colombia to 0.22 in Brazil. IOR ranged from 0.23 in Colombia to 0.34 in Guatemala. For consumption expenditures, IOL (IOR) ranged from 0.11 (0.25) in Colombia to 0.21 (0.51) in Guatemala. The IOR results indicate that, in this sample, between one quarter and one half of observed consumption inequality is due to differences in opportunities, as a lower bound. Although inequality of opportunity *levels* in each country were either similar for the income and consumption distributions, or slightly lower for consumption, inequality of opportunity *ratios* were always larger for

consumption, reflecting lower within-type inequality in that distribution. Inequality of opportunity ratios were also remarkably robust to changes in the economies of scale parameter in the Buhmann *et al.* (1988) equivalence scale.

We also defined opportunity and opportunity-deprivation profiles, which are essentially ordered partitions and partition subsets, where types are ranked by their mean levels of advantage. By appropriate choice of the threshold quantile in the smoothed distribution ( $\pi$ ), the opportunity-deprivation profile can be used to empirically identify the “worst-off types” that the previous conceptual literature has suggested should both be the focus of equitable policy-making and provide the metric for assessing the rate of economic development. In an illustration for our Latin American sample we found that, for  $\pi = 0.1$ , the entire set of the opportunity-deprived consisted of ethnic or racial minorities in three out of six countries. They were also fairly homogeneous in terms of mother’s education, father’s occupation and region of birth, suggesting that targeting a number of policies on the basis of such a social ranking might in fact be feasible, at least from the purely practical point of view of the identification of intended recipients.

## APPENDIX

TABLE A1

A COMPARISON OF SUMMARY STATISTICS FOR PER CAPITA HOUSEHOLD INCOME, USING OUR DATA AND THE SEDLAC DATABASE

	Brazil	Colombia	Ecuador	Guatemala	Panama	Peru
<b>CEDLAS database: full samples</b>						
Survey	PNAD 1996	ECH 2003	ECV 2006	ENCOVI 2006	ECH 2003	ENAHO 2001
Mean per capita total household income	255	181	157	591	172	309
Gini index	0.593	0.545	0.535	0.542	0.561	0.524
MLD index	0.666	0.563	0.556	0.536	0.607	0.508
<b>Our datasets: full samples</b>						
Survey	PNAD 1996	ECV 2003	ECV 2006	ENCOVI 2006	ENV 2003	ENAHO 2001
Mean per capita total household income	246	289	141	585	209	347
Gini index	0.593	0.558	0.545	0.545	0.576	0.523
MLD index	0.671	0.563	0.588	0.545	0.663	0.503
<b>Our datasets: sample of heads and spouses aged 30–49</b>						
Mean per capita total household income	302	324	167	678	254	376
Gini index	0.597	0.561	0.546	0.567	0.568	0.550
MLD index	0.692	0.572	0.580	0.593	0.630	0.557

*Notes:* The discrepancy between the estimates of mean per capita household income between the ECH and ECV 2003 has been confirmed by the CEDLAS team. All public SEDLAC information can be obtained from <http://sedlac.econo.unlp.edu.ar/eng/>.

## REFERENCES

- Alesina, Alberto and George-Marios Angeletos, "Fairness and Redistribution," *American Economic Review*, 95(4), 960–80, 2005.
- Arneson, Richard, "Equality of Opportunity for Welfare," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 56(1), 77–93, 1989.
- Atkinson, Anthony B., "On the Measurement of Inequality," *Journal of Economic Theory*, 2(3), 244–63, 1970.
- Banerjee, Abhijit and Esther Duflo, "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth*, 8(3), 267–99, 2003.
- Bénabou, Roland and Jean Tirole, "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 121(2), 699–746, 2006.
- Bourguignon, François, "Decomposable Income Inequality Measures," *Econometrica*, 47(4), 901–20, 1979.
- Bourguignon, François, Francisco H. G. Ferreira, and Marta Menéndez, "Inequality of Opportunity in Brazil," *Review of Income and Wealth*, 53(4), 585–618, 2007a.
- Bourguignon, François, Francisco H. G. Ferreira, and Michael Walton, "Equity, Efficiency and Inequality Traps: A Research Agenda," *Journal of Economic Inequality*, 5(2), 235–56, 2007b.
- Buhmann, Brigitte, Lee Rainwater, Guenther Schmaus, and Timothy M. Smeeding, "Equivalence Scales, Well-Being, Inequality and Poverty: Sensitivity Estimates Across Ten Countries using the Luxembourg Income Study database," *Review of Income and Wealth*, 34, 115–42, 1988.
- Checchi, Daniele and Vito Peragine, "Inequality of Opportunity in Italy," *Journal of Economic Inequality*, 8(4), 429–50, 2010.
- Cohen, Gerry A., "On the Currency of Egalitarian Justice," *Ethics*, 99(4), 906–44, 1989.
- Coulter, Fiona A. E., Frank A. Cowell, and Stephen P. Jenkins, "Equivalence Scale Relativities and the Extent of Inequality and Poverty," *Economic Journal*, 102(414), 1067–82, 1992a.
- , "Differences in Needs and Assessment of Income Distribution," *Bulletin of Economic Research*, 44, 77–124, April 1992b.
- Cowell, Frank A., "On the Structure of Additive Inequality Measures," *Review of Economic Studies*, 47(3), 521–31, 1980.
- , *Measuring Inequality*, 2nd edition, Prentice Hall/Harvester Wheatsheaf, Hemel Hempstead, 1995.
- Deaton, Angus, *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, John Hopkins University Press, Baltimore, 1997.
- Dworkin, Ronald, "What is Equality? Part 2: Equality of Resources," *Philosophy and Public Affairs*, 10(4), 283–345, 1981.
- Ferreira, Francisco H. G. and Jérémie Gignoux, "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America," World Bank Policy Research Working Paper, # 4659, 2008.
- , "Inequality of Opportunity for Education: Turkey," in Ravi Kanbur and Michael Spence (eds), *Equity in a Globalizing World*, Commission on Growth and Development, Washington, 131–56, 2010.
- Ferreira, Francisco H. G., Jérémie Gignoux, and Meltem Aran, "Measuring Inequality of Opportunity with Imperfect Data: The Case of Turkey," *Journal of Economic Inequality*, forthcoming.
- Fleurbaey, Marc and Vito Peragine, "Ex Ante versus Ex Post Equality of Opportunity," ECINEQ Working Paper, # 141, 2009.
- Foster, James E., "Inequality Measurement," in H. Peyton Young (ed.), *Fair Allocation*, American Mathematical Society, Providence, 31–68, 1985.
- Foster, James and Artyom Shneyerov, "Path Independent Inequality Measures," *Journal of Economic Theory*, 91(2), 199–222, 2000.
- Lefranc, Arnaud, Nicolas Pistoletti, and Alain Trannoy, "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?" *Review of Income and Wealth*, 54(4), 513–46, 2008.
- , "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France," *Journal of Public Economics*, 93(11–12), 1189–207, 2009.
- Marrero, Gustavo and Juan G. Rodríguez, "Inequality of Opportunity and Growth," Instituto de Estudio Fiscales and Departamento de Economía Aplicada, Universidad Rey Juan Carlos, Unpublished Manuscript, 2009.
- Mazumder, Bhakshar, "The Apple Falls Even Closer to the Tree than We Thought: New and Revised Estimates of the Intergenerational Inheritance of Earnings," in Samuel Bowles, Herbert Gintis, and Melissa O. Groves (eds), *Unequal Chances: Family Background and Economic Success*, Princeton University Press, Princeton, 2005.

- O'Neill, D., O. Sweetman, and Dirk van de Gaer, "Equality of Opportunity and Kernel Density Estimation: An Application to Intergenerational Mobility," in Thomas B. Fomby and R. Carter Hill (eds), *Advances in Econometrics*, JAI Press, Stanford, Series 14, 259–74, 2000.
- Ooghe, Erwin, Erik Schokkaert, and Dirk van de Gaer, "Equality of Opportunity versus Equality of Opportunity Sets," *Social Choice and Welfare*, 28, 209–30, 2007.
- Peragine, Vito, "Ranking Income Distributions According to Equality of Opportunity," *Journal of Economic Inequality*, 2(1), 11–30, 2004.
- Roemer, John E., *Theories of Distributive Justice*, Harvard University Press, Cambridge, 1993a.
- , "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy and Public Affairs*, 22(2), 146–66, 1993b.
- , *Equality of Opportunity*, Harvard University Press, Cambridge, 1998.
- , "Review Essay: The 2006 World Development Report: Equity and Development," *Journal of Economic Inequality*, 4(2), 233–44, 2006.
- Sen, Amartya, *Commodities and Capabilities*, North Holland, Amsterdam, 1985.
- Shorrocks, Anthony F., "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48(3), 613–25, 1980.
- Solon, Gary, "Intergenerational Mobility in the Labor Market," in Ashenfelter Orley and David Card (eds), *Handbook of Labor Economics*, North-Holland, Amsterdam, 1999.
- van de Gaer, Dirk, *Equality of Opportunity and Investment in Human Capital*, PhD Dissertation, Catholic University of Louvain, 1993.
- van de Gaer, Dirk, Erik Schokkaert, and Michel Martinez, "Three Meanings of Intergenerational Mobility," *Economica*, 68(272), 519–37, 2001.
- World Bank, *World Development Report 2006: Equity and Development*, The World Bank and Oxford University Press, Washington, 2006.