האוניברסיטה העברית בירושלים
The Hebrew University of Jerusalem

המרכז למחקר בכלכלה חקלאית
The Center for Agricultural Economic Research

המחלקה לכלכלה חקלאית ומנהל
The Department of Agricultural Economics and Management

Discussion Paper No. 6.09

Identifying Determinants of Income Inequality in the Presence of Multiple Income Sources: The Case of Korean Farm Households

by

Yuko Arayama, Jong Moo Kim and Ayal Kimhi

P.O. Box 12, Rehovot 76100

ת.ד. 12, רחובות 76100

Identifying Determinants of Income Inequality among Korean Farm Households

by

Yuko Arayama
Nagoya University, Japan

Jong Moo Kim
Sung Kyun Kwan University, Korea

Ayal Kimhi*
The Hebrew University, Israel

This draft: May 2013

Abstract

We extend the existing regression-based inequality decomposition methods to account for different income sources and different income regimes, and adequately correct for selectivity into the different income regimes. We apply these extensions to data on Korean farm households, and find that they lead to different and more informative conclusions. We also find that the correction for selectivity is essential. In particular, our results show that much of the inequality in farm household income comes through variations in family size and composition and in land ownership. However, family size and land ownership contribute to income inequality mostly through farm income, while family composition contributes mostly through non-farm labor income. We also found that education contributes to income inequality mainly through its effect on non-farm labor income, but this result was obtained only after differentiating the decomposition results by income regimes. Overall, we found that non-farm labor income is an equalizing source of income while farm income is disequalizing. Our results imply that a continued increase in the variability of landholding distribution could worsen income inequality among farm households in Korea.

---

* Corresponding author (ayal.kimhi@mail.huji.ac.il). The order of authors is alphabetic and does not imply seniority. This research was initiated while the third author was visiting at the School of Economics at Sung Kyun Kwan University, and was completed while he served as a Visiting Professor at the Economic Research Center, School of Economics, Nagoya University. The Financial support of these institutions and the Center for Agricultural Economic Research at the Hebrew University is gratefully appreciated.

Introduction

The purpose of this paper is to suggest an empirical framework for identifying determinants of income inequality in societies in which significant fractions of households have multiple income sources. This is relevant for many low- and middle-income countries, as well as for rural areas in developed countries, and we illustrate the usefulness of this framework using data on farm households in Korea.

The issue of inequality received much attention in the economic literature in the last three decades, motivated by the recognition that inequality is not only an outcome of growth but also a determinant of growth. Recently, much concern has been expressed with regard to increased inequality in fast-growing economies such as China, India and Vietnam. The increased availability of suitable data has led to numerous empirical studies of inequality based on cross-country data, labor force surveys, household surveys, and population censuses (Kimhi 2004). Much of this effort has been devoted to low and middle income countries. However, the methodologies used were in many cases adopted from more developed countries. For example, analysis of demand and supply factors in the evolution of wage inequality is perhaps suitable for an advanced economy in which the vast majority of the population is engaged in full-time wage employment, but not for a developing country with considerable self employment, informal employment and multiple jobholding.

In this paper we adopt and extend regression-based inequality decomposition methods for the case of multiple income sources, and demonstrate their usefulness using data on Korean farm households. As can be seen in figure 1, income inequality is more pronounced among Korean farm households than in the Korean economy as a whole. Using a micro data set collected in 2003, we examine the contribution of various income sources, and their determinants, to overall income inequality of farm households. Heshmati (2004) reports that inequality can be "decomposed by sub-groups, income sources, causal factors and by other socio-demographic characteristics" (page 1). Decomposition by population groups is perhaps the most popular of these, and will not be dealt with in this paper. Regarding decomposition by income sources, Shorrocks (1983) has shown that the "natural" decomposition rule of the Gini index of inequality is $G(\mathbf{y}) = \Sigma_k\{2\Sigma_i[i-(n+1)/2]y_i^k/n^2/\mu\}$, where $\mathbf{y}_k$ is income derived from source k, $\mathbf{y}$ is total income, G is the Gini index, μ is mean income, n is the number of households, and i is the rank of the household in the total income distribution. Therefore, the term inside the curled brackets, denoted S^k, is the contribution of $\mathbf{y}_k$ to G($\mathbf{y}$), and the proportional contribution of $\mathbf{y}_k$, or the share of income from source k in total inequality, is $s^k = S^k$ /G($\mathbf{y}$). Further, Lerman and Yitzhaki (1985) have shown that the change in G($\mathbf{y}$) resulting from a percentage change in $\mathbf{y}_k$ is ($s^k - \mu_k/\mu$)G($\mathbf{y}$), where μ_k is the mean of $\mathbf{y}_k$. It should be noted that several authors (e.g., Davis et al., 2010) misinterpret the decomposition results in that they treat the proportional contributions to inequality as if they were the marginal effects. Kimhi (2011) offers a more thorough discussion of this misinterpretation.

Table 1 shows the income shares and the proportional and marginal contributions to the Gini index of income inequality of several income sources of Korean farm households. One can see that farm business income, the main single source of income of these households, contributes more than half of the total income inequality, proportionately more than its income share. Moreover, a uniform one-percent increase in farm business income would increase total income inequality by six percentage points. On the other hand, non-farm labor income contributes to inequality less than its income share, and a uniform one-percent increase in non-farm labor income would decrease total income inequality by three percentage points. This implies that non-farm labor is an equalizing source of income. Non-farm business income and capital income contribute to inequality more or less proportionally to their income shares, and their marginal effects on inequality are quantitatively negligible. Transfer income and irregular income also reduce inequality, but not as much as non-farm labor income. A similar conclusion is obtained by looking at Gini coefficients for different groups of households defined by income regime. As can be seen in table 2, farm households that derive income from non-farm labor (regimes 2 and 3) have lower per-capita income Gini coefficients than other farm households.

Off-farm income was found as an equalizing income source in other countries as well, including the U.S. (See El-Osta et al., 1995, and references therein), China (Zhu and Luo, 2006), the Republic of Georgia (Kimhi, 2007), Egypt (Adams, 2001), Taiwan (Chinn, 1979), and the Philippines (Leones and Feldman, 1998). Gallup (2002), on the other hand, found that income other than farming contributed positively to inequality in Vietnam, and similar results were obtained by Elbers and Lanjouw (2001) for Ecuador. de Janvri and Sadoulet (2001) found that in Mexico, non-farm income as a whole reduced household income inequality, but not-agricultural wages in particular increased inequality. On the contrary, Canagarajah et al. (2001) found that in Ghana and Uganda, non-farm self-employment income was much more disequalizing than non-farm wages. Estudillo et al. (2001) found that nonfarm income changed from an equalizing to a disequalizing source as it became a major income source in Philippine rice villages.

Morduch and Sicular (2002) proposed a general approach to regression-based inequality decomposition. This approach brings together inequality decomposition by income source (Shorrocks 1982) and decomposition by population sub-groups (Shorrocks 1984). Adams (2001) extended the regression-based decomposition method of Morduch and Sicular (2002) to the case in which the composition of income by the different sources (e.g., labor, capital, transfers) is observed. As explanatory variables may have different effects on the different sources of income, he computed the income-source-specific contribution to inequality of each explanatory variable. The income from each source was estimated by a Tobit model, since not all households in his sample had positive income from all sources. Bardham and Boucher (1998) treated the selectivity problem differently. In particular, they were interested in the earnings equation of non-migrants in order to derive the counter-factual earnings of migrants. They estimated a Bivariate Probit selection model for non-migration and for labor force participation, and then corrected the earnings equation for selectivity using the method introduced by Tunali (1986).

In this paper, we carry the regression-based inequality decomposition method a step forward in two directions. First, we propose a decomposition method that allows the source-specific contributions to inequality of Adams (2001) to be aggregated and comparable to the Morduch and Sicular (2002) aggregate contributions. Second, we refine the selectivity-correction scheme, and correct the source-specific income-generating equations using the semiparametric method suggested by Dahl (2002), which offers a more general specification of the selectivity mechanism than Bardham and Boucher (1998). In particular, the method uses a series expansion of nonparametric choice probabilities to correct the income-generating equations for selectivity. As a result, it is straightforward to apply it to more than two choices.

We demonstrate our method with data for farm households in Korea which were collected in 2003. This choice of data is particularly suitable for our purpose since, as in many other countries, non-farm income is an important source of income for Korean farm households (Suh, 2004). Thus, many farm households derive income from the farm as well as from non-farm businesses and/or non-farm labor activities, and each of these income sources is likely to have a unique income-generating equation. We proceed by describing the methodology in the next section. After that we present the data. Next, we move to the empirical application. We present the estimated income-generating equations and the regression-based inequality decomposition results, first by income source, then by income regime, and finally by both source and regime. We further demonstrate the importance of differentiating the decomposition by income sources and income regimes by examining the marginal effect of landholdings on inequality. The last section summarizes the paper, proposes several policy implications and portrays avenues for future research.

Methodology

We start with the regression-based decomposition method suggested by Morduch and Sicular (2002), which is relevant for inequality indices that can be written as a weighted sum of household incomes:

(1) $I(\mathbf{y})=\Sigma_i a_i(\mathbf{y})y_i,$

where a_i are the weights. Income is expressed as a linear regression:

(2) $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon},$

where $\mathbf{X}$ is a matrix of explanatory variables, $\boldsymbol{\beta}$ is a vector of coefficients, and $\boldsymbol{\varepsilon}$ is a vector of residuals. Given a vector of consistent estimated coefficients $\mathbf{b}$, income can be expressed as a sum of predicted income and a prediction error according to:

(3) $\mathbf{y} = \mathbf{X}\mathbf{b}+\mathbf{e}.$

Substituting (3) into (1) and dividing through by $I(\mathbf{y})$, we obtain that the share of inequality attributed to explanatory variable m is:

(4) $s^m = b_m\Sigma_i a_i(\mathbf{y})x_i^m/I(\mathbf{y}).$[1]

[1] Wan (2004) extended this method to account for the contribution of the intercept of the income regression to inequality. Wan and Zhou (2005) presented an alternative method. It should be noted that Morduch and Sicular (2002) suggested a simple procedure to compute standard errors of s^m, but the

The partial derivatives of the Gini index of inequality with respect to an overall change in each explanatory variable can be derived by adapting the Lerman and Yitzhaki (1985) result described above to the formulation of (3) and (4). In particular, the partial derivative corresponding to x^m is $(s^m - \mu_m/\mu)\mathrm{G}(\mathbf{y})$, where μ_m is the sample mean of $b_m x_i^m$.

Moving to inequality decomposition differentiated by income sources, we specify the k^{th} source-specific income-generating function as:

(5) $\mathbf{y}_k = \mathbf{X}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k$,

where $\boldsymbol{\beta}_k$ could include zero elements corresponding to explanatory variables that do not affect the k'th source of income. Since $\mathbf{y} = \Sigma_k\mathbf{y}_k = \mathbf{X}\Sigma_k\boldsymbol{\beta}_k + \Sigma_k\boldsymbol{\varepsilon}_k$, using consistent estimates $\mathbf{b}_k$ of $\boldsymbol{\beta}_k$ and substituting into (1), the fraction of the inequality contribution of explanatory variable m in overall inequality is:

(6) $s^m = (\Sigma_k b_{km})\Sigma_i a_i(\mathbf{y})x_i^m/I(\mathbf{y})$.

This can be broken down to source-specific contributions of each explanatory variable to overall inequality, denoted s^{mk}, which is implicitly defined by:

(7) $s^m = \Sigma_k[b_{km}\Sigma_i a_i(\mathbf{y})x_i^m/I(\mathbf{y})] = \Sigma_k s^{mk}$.

It is easy to see that (4), the decomposition proposed by Morduch and Sicular (2002), is a special case of (6), in the case of identical income-generating equations for all income sources. However, this only holds when all households derive income from all sources. Otherwise, (5) has to be estimated using selectivity-correction methods, and therefore b_{km} measures the effect of x_i^m on y_{ik}^*, which is the *latent* income of household i from source k. In this case, the equalities in (6) and (7) do not hold, if x_i^m affects not only income from source k but also the tendency of household i to have income from source k. The intuitive reason is that the contribution of x_i^m to overall

procedure turns out to be incorrect. At least for the Gini index of inequality, it is not straightforward to compute standard error of the index itself (See Modarres and Gastwirth 2006 and references therein), so it is reasonable to expect that computing standard errors of *components* of that index would not be straightforward either.

income inequality is also affected by the effects of x_i^m on getting in and out of the different corner solutions. Deriving these effects is beyond the scope of this paper. However, the source-specific inequality shares s^{mk} are still informative for the channels through which x_i^m contributes to overall income inequality, hence we derive and present them in the empirical analysis below.

As mentioned above, it is reasonable that not all households derive income from all potential income sources, and this brings up the issue of selectivity correction. While each source-specific income-generating equation can be estimated by Tobit maximum likelihood, for example, the potential multiplicity of corner solutions may require a specification of regime-specific income-generating equations, where each regime is defined by a specific combination of corner and internal solutions. This essentially leads to a switching regression model. Specifically, we define each regime-specific income-generating function as:

(8) $$\mathbf{y}_r = \mathbf{X}^r \boldsymbol{\beta}_r + \boldsymbol{\varepsilon}_r,$$

where $\mathbf{y}_r = \mathbf{D}^r\mathbf{y}$, $\mathbf{X}^r = \mathbf{D}^r\mathbf{X}$, $\boldsymbol{\varepsilon}_r = \mathbf{D}^r\boldsymbol{\varepsilon}$, and $\mathbf{D}^r$ is a selection matrix with ones on the principal diagonal corresponding to observations that belong to regime *r* and zeros elsewhere. By construction, $\Sigma_r\mathbf{D}^r = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix, and hence $\Sigma_r\boldsymbol{\varepsilon}_r = \boldsymbol{\varepsilon}$ and $\Sigma_r\mathbf{y}_r = \mathbf{y}$. Note that $\mathbf{y}_r$, $\mathbf{X}^r$ and $\boldsymbol{\varepsilon}_r$ have the same number of rows as $\mathbf{y}$, $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, but have zeros corresponding to observations that are not part of income regime *r*.

Applying the same reasoning as in the case of decomposition by income source, we note that $\mathbf{y} = \Sigma_r\mathbf{y}_r = \Sigma_r\mathbf{X}^r\boldsymbol{\beta}_r + \Sigma_r\boldsymbol{\varepsilon}_r$, and hence the fraction of explanatory variable *m* in overall income inequality is:

(9) $$s^m = \Sigma_r b_{rm} \Sigma_i a_i(\mathbf{y}) x_i^{rm} / I(\mathbf{y}) = \Sigma_r b_{rm} \Sigma_i a_i(\mathbf{y}) d_i^r x_i^m / I(\mathbf{y}),$$

where d_i^r is the i^{th} element on the diagonal of $\mathbf{D}^r$. The difference between (9) and (6) is that here we have d_i^r that is indexed by both *i* and *r*, and hence we cannot separate the summation over *i* and over *r*. Still, deriving the regime-specific share of inequality of each explanatory variable as:

(10) $$s^{mr} = b_{rm} \Sigma_i a_i(\mathbf{y}) d_i^r x_i^m / I(\mathbf{y})$$

is useful for assessing the differential importance of each explanatory variable in each income regime. Of course, as in the earlier case, the s^{mr} will not in general sum up to s^m, if x_i^m affects the choice of income regime.

Finally, if the contribution of each explanatory variable to overall income inequality varies by income source and by income regime, it makes sense to estimate each source-specific income generating equation separately in each income regime, i.e. as a switching regression. This leads to regression coefficients $\mathbf{b}_{rk}$ that are indexed by both source and regime, and allows a decomposition of overall inequality in both dimensions. Specifically, write $\mathbf{y}_k = \Sigma_r \mathbf{y}_{rk}$ where $\mathbf{y}_{rk}$ is a vector of income from source k in regime r. Total income can therefore be written as:

$$(11) \qquad \mathbf{y} = \Sigma_k \mathbf{y}_k = \Sigma_k \Sigma_r \mathbf{y}_{rk} = \Sigma_k \Sigma_r \mathbf{X}^r \boldsymbol{\beta}_{rk} + \Sigma_k \Sigma_r \boldsymbol{\varepsilon}_{rk},$$

which leads to the following inequality decomposition:

$$(12) \qquad s^m = \Sigma_k \Sigma_r b_{rkm} \Sigma_i a_i(\mathbf{y}) x_i^{rm} / I(\mathbf{y}) = \Sigma_r (\Sigma_k b_{rkm}) \Sigma_i a_i(\mathbf{y}) d_i^r x_i^m / I(\mathbf{y}).$$

The source- and regime-specific share of explanatory variable m in total income inequality can be evaluated as:

$$(13) \qquad s^{mkr} = b_{rkm} \Sigma_i a_i(\mathbf{y}) d_i^r x_i^m / I(\mathbf{y})$$

Naturally, the same qualifications apply here, namely that this decomposition ignores the effect of explanatory variables on the choice of income regime and hence the summation in (12) will not hold in general.

Note that for each case in which inequality shares are defined, marginal effects of explanatory variables on inequality can be derived using an appropriate modification of the Lerman and Yitzhaki (1985) formula described above.

Data

We use data from the 2003 nationally-representative farm book-keeping survey that included 3,200 farm households. A farm household is defined as a household engaged in farming for the purpose of making a living, in which the farm operator manages at least 300 pyeong (about 0.1 ha) of cultivated land and generates annual sales of at least 500,000 Won (roughly $420). Excluded are single-person households, foreigners, and those employing more than five full-time employees. The survey provides information about household income from various farm and non-farm sources, as well as assets, expenditures, and demographics.

The variables we use to explain per-capita income are listed in table 3. We include age of the head of household and its squared value, to account for life-cycle effects. We also include a dummy indicator for the household head being a female. Next, we use a set of binary indicators of the educational level of the head of household. Household size and composition are represented by three variables: family size, the fraction of working-age males in the family, and the fraction of working-age females in the family. The working age was determined to be from 19 to 64. The economic resources of the household are represented by per-capita land owned and by a dummy indicator for landless households. We have experimented with a set of regional dummies, and eventually decided to include dummy indicators for center-east regions and for south-west regions.

Inequality decomposition by income source

The first column of table 4 shows the coefficients of the per-capita income generating function (2) for our sample. Most coefficients are statistically significant and have the expected sign. Age has a nonlinear effect, first positive and subsequently negative, education has a positive effect, and female-headed households have lower income per-capita than male-headed households, but the difference is not significantly different from zero. Per-capita income decreases with family size, but increases with the fraction of working-age adults (males or females). Income per-capita is lower for landless households, and increases with the amount of land owned per-capita. Households located in the southern and western regions have lower per-capita income than in the rest of the country.

Next, we estimated separate income-generating functions (5) for each source of income, except for irregular income which we consider as a residual source of income. Except for the case of farm income, which was reported for all households, each function was estimated by the Tobit maximum likelihood model in order to account for censoring from below. We have used the same set of explanatory variables in all the equations, because these equations are essentially reduced-form equations (encompassing elements of labor allocation, asset ownership, and returns to labor and assets), and hence exclusion restrictions do not follow naturally. The results are shown in the remaining columns of table 4, and it is quite clear that the effects of explanatory variables on the different sources of income are substantially different. Examining the three major sources of income, namely farm income, non-farm business income and non-farm labor income, we find statistically significant effects in opposite directions of female headship, family size and landholdings. Education does not have a statistically significant effect on any of these income sources. Regional differences are statistically significant for non-farm income but not for farm income. In summary, the importance of several determinants of income varies considerably across income sources.

We now turn to the decomposition of inequality by determinants of income. The first column in table 5 shows the decomposition of the Gini index of total income inequality using (4).[2] We find that only 18% of income inequality is explained by the set of explanatory variables as a whole. This is not too bad, given that only 13% of the variance in income is explained by these explanatory variables (table 4). We find that major contributions to inequality are assigned to family size and composition and to land ownership. The remaining columns in table 5 show the source-specific contributions of income determinants, computed according to (7). We find that land ownership and family size contribute the most to income inequality through farm income. In fact, land ownership has a dominant role in explaining income inequality through farm income, but not through any other source of income. This is of course not surprising. Family size also has a positive contribution to income inequality through non-farm labor income. Again, this is not surprising. Family composition, and specifically the fraction of working-age family members, has a relatively large contribution to income inequality through non-farm labor income, but a much smaller contribution through other sources of income. Contributions to income inequality through other sources of income, including non-farm business income, capital income, and transfer income, are relatively minor.

The decomposition results in table 5 have shown that different explanatory variables have different contributions to income inequality from different sources of income. For example, education has a positive contribution to overall income

[2] Computing standard errors of the decomposition results is not an easy task (see previous footnote), and is left for future research.

inequality but its contribution through transfer income is negative. This can provide useful information about potential inequality implications of various policy measures.

Inequality decomposition by income regime

The estimation results in table 4 ignore possible cross-equation dependencies that are implied by the fact that most farm households derive income from more than one source. In fact, the literature on labor allocations in farm households (e.g., Fafchamps and Quisumbing 1999) implies that the farm income-generating function could be different for farm households who also derive income from non-farm sources. In order to examine whether the way in which the income generating equations are estimated is important for inequality decompositions, we estimate a separate per-capita income generating equation (8) for each of the income regimes described in table 2. Note that this does not include all possible income regimes. This would be outside the scope of this paper. We focused on regimes determined mostly by labor allocation decisions.

The results are in table 6. As in the case of the sources of income, we find qualitatively different income-generating equations in the different income regimes. The effect of age, for example, is statistically significant only among farm households who have both non-farm business income and non-farm labor income. The fraction of adult males is more important among farm households who have non-farm labor income, while the fraction of adult females is more important among farm households who do not have non-farm labor income. Land owned (but not landlessness) is not important for farm households who do not have non-farm business or labor income. Regional location is only important for households who have non-farm labor income. This implies that the contributions of the different income determinants to income inequality are likely to vary across the income regimes.

Table 7 shows the regime-specific contributions of explanatory variables to the Gini index of inequality. The first column shows the aggregate decomposition by (4) from table 5, while the other columns present the contributions implied by (10), using the coefficients in table 6. We find that the contribution of age to income inequality is predominantly among households who have only farm income, while the contribution of education is predominantly among households who have non-farm business income but no non-farm labor income. The contributions of the household size and composition variables vary in a similar way to what was observed in table 6.

The combined contribution of landlessness and amount of land owned is lowest for households who have only farm income.

Selectivity correction

The shortcoming of the preceding analysis is that the choice of income regime is implicitly assumed to be exogenous. Since farm households make their labor allocation decisions based upon expected income, and these labor allocations determine the income regime of the household, this analysis suffers from potential selectivity bias. To correct this bias, we need to correct for selectivity when we estimate the regime-specific income generating equations. Dahl (2002) offers an attractive method for our purpose. The method is semiparametric in that it uses a series expansion of choice probabilities to approximate the selectivity-correction terms in the income equations, and hence does not assume a certain joint distribution of the selection rule and the outcome equations. It also uses nonparametric estimates of the choice probabilities in order to avoid making distributional and functional form assumptions in the choice model.

The choice probabilities are derived as cell fractions of the different choices, after the sample is divided into cells defined by explanatory variables. We define cells in our sample by age (up to 50, 51-60, 61-66, and 67+), education (up to elementary, middle school and above), existence and number of children (none, only children up to 18 years of age, at least one child older than 18), land owned (up to ½ ha, between ½ and 1 ha, more than 1 ha), and transfer income (up to 1,000 Won, more than 1,000 Won). Adjacent cells with relatively few observations were merged, so that no cell includes less than 11 observations. A total of 73 cells were generated in this way. The distributions of the derived choice probabilities are shown in figure 2. The spread of each choice probability in the sample illustrates the usefulness of the cell-generation method. A degenerate distribution would imply that there are no significant differences between the observations in the different cells. We observe that the spread of the probabilities of the more frequent choices (see table 2) is larger.

We estimated the models in table 6 with the correction to selectivity, using both first- and second-degree polynomials of the choice probabilities. F-tests revealed that the second-degree polynomial did not add explanatory power, and hence we used the model with the first-degree polynomial. An F-test of the corrected model versus the uncorrected model showed that only in regimes 1 and 3, those with positive non-farm business income, the correction is statistically significant. We conducted the regression-based inequality decomposition for these two regimes using the coefficients of the corrected model, and the results were not much different from those in table 5. These regression results and the related decomposition results are available from the authors upon request.

We now want to combine the two levels of differentiation of the income-generating equations, by income source and by income regime. In particular, we estimate source-specific and regime-specific income generating equations as implied by (11), including a correction for selectivity as described above. We tested these models in three levels. First, a test of the aggregated results in table 4 versus disaggregated results. Second, a test of the disaggregated results versus the same model corrected for selectivity. Finally, we tested a first-degree polynomial of choice probabilities in selectivity correction versus a second-degree polynomial. For all three major income sources (farm income, non-farm business income and non-farm labor income) the disaggregated results were significantly different from the aggregated results (p values close to zero). For disaggregated farm income, the results with selectivity correction were significantly different from the results without selectivity correction, with the exception of households that do not have non-farm business or labor income. In general, the results corrected using a second-degree polynomial of choice probabilities were significantly different from those corrected with a first-degree polynomial. For disaggregated non-farm business income and non-farm labor income, the corrected results were significantly different from the uncorrected results, but those corrected with a second-degree polynomial were not significantly different from those corrected with a first-degree polynomial. Therefore, we present in tables 8 and 9 the corrected disaggregated results, with farm income corrected with a second-degree polynomial (table 8) and non-farm income corrected with a first-degree polynomial (table 9).[3]

The regression results in table 8 show several notable differences in the farm income generating equation across income regimes. In particular, the negative effect of female headship exists only among households who have non-farm business income (regimes 1 and 3). Also, the quadratic age effect, which was statistically

[3] Note that using the Dahl (2002) selectivity correction requires a correction of the standard errors of the estimated coefficients, which we did not apply here. Hence, the t-values reported in tables 8 and 9 may be biased upwards. We plan to apply this correction in subsequent versions of this paper.

significant in the aggregate results, is not significant in any of the income regimes. The negative effect of family size also becomes insignificant. The positive effect of the fraction of adults in the household is statistically significant only among households who have non-farm labor income. The negative effect of landlessness is statistically significant only among households who do not have non-farm business or labor income. The positive effect of land is not significant among households who have both non-farm business income and non-farm labor income. Note that the coefficient of middle school is peculiarly large in regime 1. This is likely to be due to multicollinearity in this relatively small sub-sample, hence we will not assign too much importance to this effect.

Considerable differences are also observed in table 9. For non-farm business income there is one notable difference, namely that the positive effect of land exists only for households who have non-farm labor income as well. For non-farm labor income the differences are more pronounced. As in the case of farm income, the quadratic age effect, which was statistically significant in the aggregate results, is not significant in any of the income regimes. The positive effect of education, which was not significant in the aggregate results, becomes statistically significant when the results are allowed to differ by income regime (although it is quite similar across the regimes). The positive effects of female headship and of family size become insignificant in the regime-specific results. The positive effect of the fraction of adult females and the negative effect of land are significant only among households who have non-farm business income as well.

Some of the differences in the coefficients of the income-generating equations across income regimes are also reflected in the contributions to inequality of the corresponding explanatory variables. These are reported in tables 10 and 11, for farm income and non-farm income, respectively. In the following discussion we will focus on explanatory variables that were mentioned earlier as having notably different coefficients in the source- and regime-specific income generating equations (tables 8 and 9). Beginning with the contributions to inequality through farm income (table 10), we find that the positive contribution of age (including age squared) to income inequality is mostly among households that do not have non-farm labor income. The positive contribution of family size to inequality is mostly among households that have both non-farm business and non-farm labor income. However, this contribution is based on a non-significant coefficient in the relevant income-generating equation.

The positive contribution of the family composition variables to inequality also varies across the income regimes. The positive contribution of landlessness to inequality is most pronounced among household who do not have non-farm business or labor income, while the positive contribution of land owned is least pronounced among these same households.

Turning to the contributions to inequality through non-farm business income (table 11), the only notable difference is that of land owned, which has a relatively large contribution among households who also have non-farm labor income and a slight negative contribution among households who do not. The case of non-farm labor income is more interesting in this respect. The contribution of female headship to income inequality, which was positive in the aggregate sample, becomes negative in the regime-specific case. The contribution of education to inequality, which was very small or even negative in the aggregate sample, becomes positive, regardless of having non-farm business income or not. Finally, the contribution of land owned, which was positive in the aggregate sample, becomes negative in the regime-specific case.

Marginal effects

In order to further illustrate the usefulness and importance of differentiating the inequality decomposition by income sources and regimes, we compare the partial derivatives of the Gini index of inequality with respect to an overall change in land owned per-capita in each of the decomposition routines. The results are in table 12. When a single income-generating equation for total household income is estimated, we find that a one-percent increase in landholdings would reduce income inequality by two percentage points. However, when income is differentiated by source, the marginal effect of landholdings on inequality is only half of that. Moreover, we found that an increase in landholdings reduces inequality mostly through farm income, while it increases inequality through non-farm labor income.

When we differentiate by income regime, we find a positive effect of landholdings on inequality, mostly contributed by farms that have non-farm business income in addition to farm income. When we differentiate farm income by regime, we find that the marginal effect of landholdings operating through farm income is in fact positive, and, in this case as well, contributed mostly by farms that have non-farm business income in addition to farm income. We conclude that within income regimes landholdings increases inequality, so that the overall negative marginal effect is probably caused by regime switching behavior cause by the increase in landholdings. This supports our earlier conclusion that adequately controlling for income regimes is important in the analysis of inequality determinants.

Summary and conclusions

In this paper, we have proposed an extension of the regression-based inequality decomposition method suggested by Morduch and Sicular (2002), which differentiates between contributions to inequality of determinants of income operating through different sources of income, and in different income regimes. We found that this differentiation does lead to different conclusions. We also found that adequate correction for selectivity when estimating regime-specific income-generating equations is essential.

In the case of Korean farm households, we found that non-farm labor income is an inequality-decreasing source of income, relative to farm business income. Decomposing aggregate income inequality into components attributable to the different determinants of income, we found that not a large fraction of total income inequality could be explained by the explanatory variables. This is related to the fact that the explanatory power of these explanatory variables in the income-generating equations is not large. As a fraction of the explained inequality, family size (23%), family composition (33%) and land ownership (23%) are the major contributors. However, when looking at specific sources of income, we find that family size and land ownership are mostly contributing to income inequality through farm income, while family composition is mostly contributing to income inequality through non-farm labor income.

By breaking the contributions to inequality further by income regime, and appropriately correcting for selectivity, we find that the contribution of family size to income inequality through farm income is most pronounced among households that have both non-farm business and non-farm labor income. This makes sense because deriving income from multiple sources is facilitated by using the labor contributions of relatively many household members. However, the family size coefficient in the relevant income-generating equation is not statistically significant, and therefore this result is subject to doubt. In the case of land ownership, we found that its contribution to income is more widespread across the income regimes, but is more pronounced among households who have non-farm business income. Perhaps land serves as an asset that can be relied upon in non-farm business activities, i.e. as collateral for credit. As for family composition, we found that its contribution to inequality does not change much across the income regimes.

Other determinants of sources of income become significant, both quantitatively and statistically, when differentiating by income source and income regime. For example, through non-farm labor income, we find a positive contribution of high school and higher education to income inequality, a negative contribution of land owned among households who also have non-farm business income, and a larger contribution of the fraction of adult males than the contribution of the fraction of adult females, to income inequality.

Despite the fact that we are not able to explain a large fraction of inequality, our results can be used for policy analysis, because the parts that we are able to explain are related to important policy variables such as education and landholdings. In other words, we explain the part of inequality that is related to inequality in resources and opportunities, and this is most relevant for policy makers. The unexplained part of inequality could be due to unobserved preference variability that is less interesting. In this case, our results have several policy implications. First, we found that land ownership is one of the major contributors to income inequality, mostly through farm income, and in particular for households who also have non-farm business income. It seems like farming and non-farm family businesses are related activities. As the size distribution of rice farms in Korea has changed very little in the last few decades, income inequality could have increased through the expansion of non-farm business activity. However, income inequality could further increase if inequality of landholdings starts increasing in the future, as it did in many other countries.

Second, family composition contributes to income inequality mostly through non-farm labor income inequality. Over the years, the extent of off-farm work on Korean farm households has increased remarkably. While there doesn't seem to be much impact of policy on family composition, it should be noted that to the extent that farm households are multi-generational, the tendency of farmers' offspring to join their parents on the family farm depends largely on their income opportunities. We can expect to find more adult offspring, and as a result higher fractions of adult family members, on more profitable farms, and this process could lead to increased income inequality in the long run. To counteract this effect, authorities should design policies to make farming more attractive to younger generations, especially when the objective prospects of farming are less favorable.

Finally, the role of education in contributing to inequality through non-farm labor income should not be overlooked. Given that our results imply that non-farm labor income is an equalizing source of income, the increased tendency by farm household members to work off the farm could reduce income inequality. One of the key policy tools for achieving this is rural education. However, if rural education is not expanded in an equitable way, this could lead to an increase rather than a decrease in farm household income inequality.

Methodologically, this research can be completed and expanded in three dimensions. First, there is a need to compute standard errors of the regression-based decomposition results (see footnotes 2 and 3). Second, the regression-based decomposition by income regime will be more informative if it also computes the effects of explanatory variables on regime choices. Third, our results call for an extension of this analysis in the time dimension. In particular, it would be very useful to examine the trends of income inequality and its determinants over time, along the lines of Bourguignon, Fournier, and Gurgand (2001). In this way it might be possible to endogenize the trends in some of the income determinants. Empirically, more detailed information on non-farm labor supply could enable one to differentiate between the effect of labor supply, which could be endogenous, and the effect of the returns to labor, which are largely exogenous but may be affected by public policy. Finally, the framework used in this research and its extensions could be applied to other countries in Eastern Asia, and other parts of the world.

References

Adams, Richard H, Jr. (2001). “Nonfarm Income, Inequality, and Land in Rural Egypt.” *Economic Development and Cultural Change* 50: 339-363.

Adams, Richard H, Jr., and Jane J. He (1995). *Sources of Income Inequality and Poverty in Rural Pakistan*. International Food Policy Research Institute Research Report 102, Washington, DC.

Barham, Bradford, and Stephen Boucher (1998). “Migration, Remittances, and Inequality: Estimating the Net effects of Migration on Income Distribution.” *Journal of Development Economics* 55: 307-331.

Bourguignon, Francois, Martin Fournier, and Marc Gurgand (2001). “Fast Development with a Stable Income Distribution: Taiwan, 1979-94.” *Review of Income and Wealth* 47: 139-163.

Canagarajah, Sudharshan, Constance Newman, and Ruchira Bhattamishra (2001). “Non-Farm Income, Gender, and Inequality: Evidence from Rural Ghana and Uganda,” *Food Policy* 26: 405-420.

Chinn, Dennis L. (1979). “Rural Poverty and the Structure of Farm Household Income in Developing Countries: Evidence from Taiwan.” *Economic Development and Cultural Change* 27: 283-301.

Dahl, Gordon B. (2002). “Mobility and the Return to Education: testing a Roy Model with Multiple Markets.” *Econometrica* 70: 2367-2420.

Davis, Benjamin, Paul Winters, Gero Carletto, Katia Covarrubias, Esteban J. Quiñones, Alberto Zezza, Kostas Stamoulis, Carlo Azzarri and Stefania DiGiuseppe (2010). "A Cross-Country Comparison of Rural Income Generating Activities." *World Development* 38: 48-63.

El-Osta, Hisham, G. Andrew Bernat Jr., and Mary C. Ahearn (1995). “Regional Differences in the Contribution of Off-Farm Work to Income Inequality.” *Agricultural and Resource Economics Review* 24: 1-14.

Elbers, Chris, and Peter Lanjouw (2001). “Intersectoral Transfer, Growth, and Inequality in Rural Ecuador.” *World Development* 29: 481-496.

Estudillo, Jonna P., Agnes R. Quisumbing, and Keijiro Otsuka (2001). “Income Distribution in Rice-growing Villages During the Post-Green Revolution Periods: The Philippine Case, 1985 and 1998.” *Agricultural Economics* 25: 71-84.

Fafchamps, Marcel, and Agnes R. Quisumbing (1999). “Human Capital, Productivity, and Labor Allocation in Rural Pakistan.” *The Journal of Human Resources* 34: 369-406.

Gallup, John Luke (2002). *The Wage Labor Market and Inequality in Vietnam in the 1990s*. World Bank Policy Research Paper 2896, Washington, DC.

Heshmati, Almas (2004). *A Review of Decomposition of Income Inequality*. IZA Discussion Paper No. 1221.

de Janvri, Alain, and Elisabeth Sadoulet (2001). ”Income Strategies Among Rural Households in Mexico: the Role of Off-farm Activities.” *World Development* 29: 467-480.

Kimhi, Ayal (2004). *Growth, Inequality and Labor Markets in LDCs: A Survey*. CESifo Working Paper No. 1281.

Kimhi, Ayal (2011). “Comment: On the Interpretation (and Misinterpretation) of Inequality Decompositions by Income Sources.” *World Development* 39: 1888-1890.

Leones, Julie P., and Shelley Feldman (1998). “Nonfarm Activity and Rural Household Income: Evidence from Philippine Microdata.” *Economic Development and Cultural Change* 46: 789-806.

Lerman, Robert I., and Shlomo Yitzhaki (1985). “Income Inequality Effects by Income Source: A New Approach and Applications to the United States.” *Review of Economics and Statistics* 67: 151-156.

Modarres, Reza, and Joseph L. Gastwirth (2006). “A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality.” *Oxford Bulletin of Economics and Statistics* 68: 385-390.

Morduch, Jonathan, and Terry Sicular (2002). “Rethinking Inequality Decomposition, with Evidence from Rural China.” *The Economic Journal* 112: 93-106.

Shorrocks, Anthony F. (1982). “Inequality Decomposition by Factor Components.” *Econometrica* 50: 193-211.

Shorrocks, Anthony F. (1983). “The Impact of Income Components on the Distribution of Family Incomes.” *Quarterly Journal of Economics* 98: 311-326.

Shorrocks, Anthony F. (1984). “Inequality Decomposition by Population Subgroups.” *Econometrica* 52: 1369-1385.

Suh, Jin Kyo (2004). “Republic of Korea.” In Tongroj Onchan (ed.), *Non-Farm Employment Opportunities in Rural Areas in Asia*. Asian Productivity Organization, Tokyo, Japan.

Tunali, Insan (1986). “General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Remigration.” In R.G. Ehrenberg (ed.), *Research in Labor Economics* 8B: 235-284. JAI Press, Greenwhich, CT.

Wan, Guanghua (2004). “Accounting for Income Inequality in Rural China: A Regression-Based Approach.” *Journal of Comparative Economics* 32: 348-363.

Wan, Guanghua, and Zhangyue Zhou (2005). “Income Inequality in Rural China: Regression-Based Decomposition Using Household Data.” *Review of Development Economics* 9: 107-120.

Zhu, Nong, and Xubei Luo (2006). *Nonfarm Activity and Rural Income Inequality: A Case Study of Two Provinces in China*. World Bank Policy Research Paper 3811, Washington, DC.

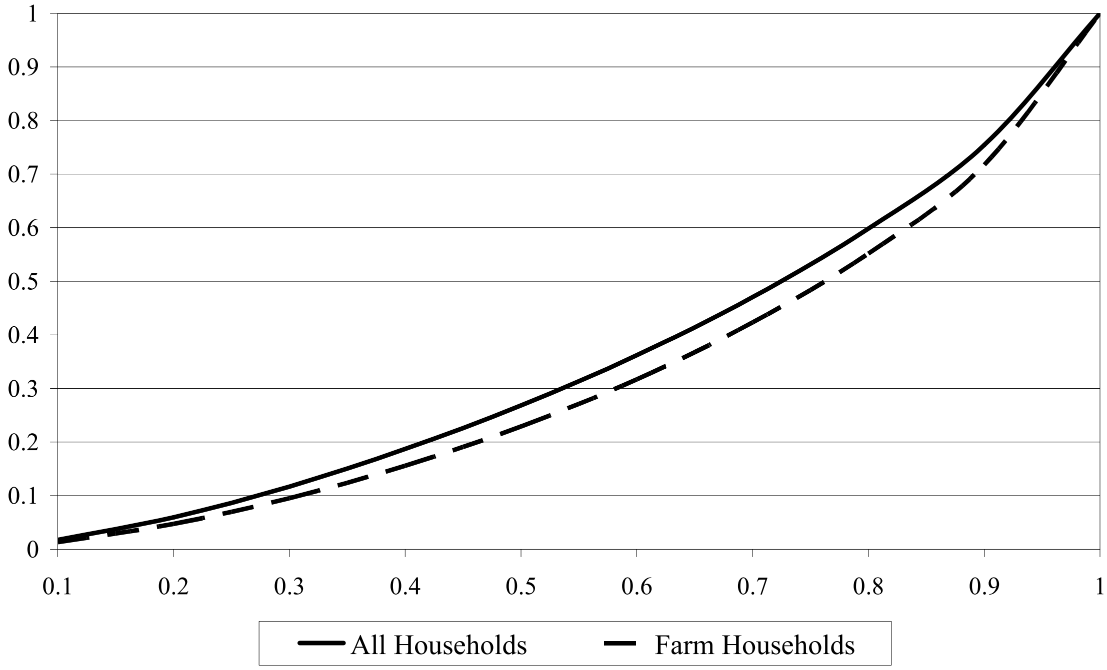

Figure 1. Lorenz curves

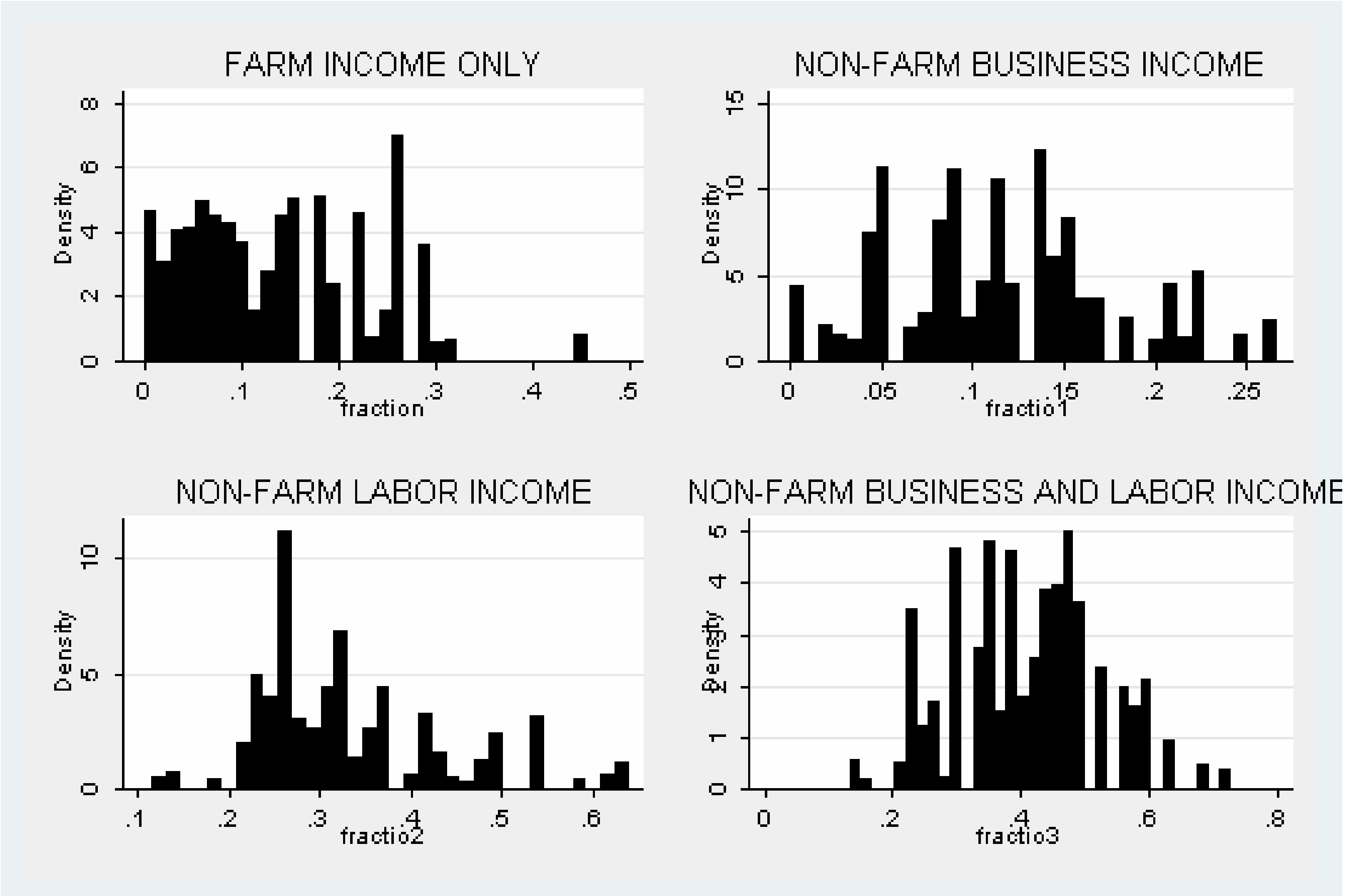

Figure 2. Distributions of nonparametric choice probabilities

Table 1. Sources of Farm household Income and their Contribution to Inequality

Income Component	Income Share	Proportional Contribution to Gini	Marginal Contribution to Gini
Farm business income	0.4247	0.581	0.1560**
			(0.0148)
Nonfarm business income	0.0778	0.083	0.0057
			(0.0105)
Nonfarm labor income	0.1987	0.118	-0.0809**
			(0.0075)
Capital income	0.0300	0.023	-0.0069**
			(0.0023)
Transfer income	0.0846	0.045	-0.0396**
			(0.0035)
Irregular income	0.1843	0.150	-0.0343**
			(0.0078)

Notes:
bootstrapped standard errors in parentheses.
* significant at 5%; ** significant at 1%.

Table 2. Per-Capita Income and Inequality by Income Regime

Group of households	Mean per-capita income	Gini coefficient	Number of observations
All cases	9.06	0.4147	3,042
Regime=0	8.96	0.5050	423
Regime=1	9.76	0.4564	350
Regime=2	8.29	0.3911	1,016
Regime=3	9.52	0.3863	1,253

Notes:
Income is measured in millions of Won. All farm households have farm income by definition.
Regime=0: household with no income from non-farm business or labor
Regime=1: household with income from non-farm business only
Regime=2: household with income from non-farm labor only
Regime=3: household with income from both non-farm business and non-farm labor

Table 3. Explanatory Variables[a]

Variable	Mean	Std. Dev.	Min	Max
Female head	0.036	0.188	0	1
Age	58.8	10.7	27	89
Elementary school	0.410	0.492	0	1
Middle school	0.195	0.396	0	1
High school	0.233	0.423	0	1
Higher education	0.039	0.194	0	1
Family size	3.21	1.401	2	10
Fraction adult males	0.288	0.226	0	1
Fraction adult females	0.303	0.201	0	1
Landless	0.043	0.204	0	1
Land owned per capita[b]	0.476	0.742	0	21.5
Center-east	0.337	0.473	0	1
South-west	0.441	0.497	0	1

a. 3,042 households
b. Land is measured in hectares.
.

Table 4. Source-Specific Per-Capita Income Generating Equations

Variable	Total income	Farm income	Non-farm business income	Non-farm labor income	Capital income	Transfer income
Female head	-1.146	-2.532**	-0.201	0.941**	-0.002	0.182
	(-1.50)	(-4.05)	(-0.42)	(2.83)	(-0.02)	(1.24)
Age	0.504**	0.250*	-0.023	0.259**	0.026	0.029
	(3.27)	(1.99)	(-0.24)	(3.68)	(0.98)	(0.98)
Age squared/100	-0.449**	-0.275*	-0.023	-0.224**	-0.004	0.005
	(-3.28)	(-2.46)	(-0.27)	(-3.55)	(-0.18)	(0.18)
Elementary school	0.628	0.333	0.070	-0.367	0.178*	0.313**
	(1.31)	(0.85)	(0.23)	(-1.71)	(2.21)	(3.45)
Middle school	1.183*	0.267	0.356	-0.252	0.235*	0.529**
	(2.11)	(0.58)	(1.02)	(-1.01)	(2.50)	(4.96)
High school	1.488**	-0.244	0.099	0.102	0.472**	0.644**
	(2.61)	(-0.52)	(0.28)	(0.40)	(4.95)	(5.95)
Higher education	4.188**	0.498	-0.178	0.747	0.959**	1.288**
	(4.80)	(0.70)	(-0.33)	(1.92)	(6.67)	(7.69)
Family size	-0.988**	-0.660**	0.043	0.306**	-0.040*	-0.132**
	(-8.10)	(-6.62)	(0.58)	(5.73)	(-1.96)	(-5.68)
Fraction adult males	4.219**	1.467*	1.098*	3.024**	0.339*	-0.477**
	(5.30)	(2.26)	(2.23)	(8.61)	(2.55)	(-3.15)
Fraction adult females	4.151**	1.965**	0.799	1.737**	0.036	-0.135
	(4.89)	(2.84)	(1.52)	(4.60)	(0.25)	(-0.84)
Landless	-3.156**	-1.583**	-0.753	-0.308	-0.727**	-0.369**
	(-4.43)	(-2.72)	(-1.68)	(-0.98)	(-5.35)	(-2.64)
Land owned per capita	1.532**	1.598**	0.551**	-0.546**	0.018	0.005
	(7.72)	(9.85)	(4.72)	(-6.04)	(0.56)	(0.12)
Center-east	-0.317	0.625	-0.573*	-0.637**	0.163*	0.107
	(-0.81)	(1.96)	(-2.40)	(-3.69)	(2.47)	(1.43)
South-west	-1.767**	-0.296	-0.765**	-0.773**	0.023	0.029
	(-4.71)	(-0.97)	(-3.34)	(-4.67)	(0.35)	(0.41)
Intercept	-4.487	-0.752	0.539	-7.518**	-1.716*	-1.096
	(-1.09)	(-0.22)	(0.21)	(-4.00)	(-2.45)	(-1.37)
R^2	12.87%	9.50%	1.02%	2.95%	1.81%	3.05%
No. (%) of positive cases	3,042	3,042	1,603	2,269	1,995	2,724
	(100%)	(100%)	(52.7%)	(74.6%)	(65.6%)	(89.5%)

Notes: OLS estimates for total income and farm income, Tobit estimates for other income sources. All farm households have farm income by definition; R^2 in Tobit results is Pseudo R^2; t-statistics in parentheses; * coefficient significant at 5%. ** coefficient significant at 1%.

Table 5. Regression-Based Source-Specific Contributions to Inequality

Variable	Total income	Farm income	Non-farm business income	Non-farm labor income	Capital income	Transfer income
Female head	0.094	0.819	0.013	0.166	0.000	-0.004
Age	-2.834	0.204	0.584	-8.220	0.480	2.061
Age squared/100	5.007	0.627	0.712	8.982	-0.087	0.397
Elementary school	0.023	0.143	-0.005	0.144	-0.015	0.277
Middle school	0.111	0.037	0.077	-0.069	-0.003	-0.116
High school	0.371	0.056	0.023	0.036	0.069	-0.616
Higher education	0.613	-0.042	-0.010	-0.006	0.167	-0.142
Family size	4.214	2.628	0.052	1.341	0.098	0.899
Fraction adult males	3.187	0.416	0.528	2.812	0.028	0.410
Fraction adult females	2.926	0.738	0.229	1.015	0.002	0.056
Landless	1.245	0.595	0.139	-0.027	0.357	0.092
Land owned per capita	2.933	4.315	0.277	0.848	0.018	0.007
Center-east	-0.055	0.206	-0.016	0.112	0.086	0.046
South-west	0.445	-0.036	0.183	0.325	-0.001	0.008
Intercept	0.000	0.000	0.000	0.000	0.000	0.000
Residual	81.721	[a]	[a]	[a]	[a]	[a]

a. Source-specific contributions are computed as percentages of overall inequality, hence they do not sum up to total source-specific inequality.

Table 6. Regime-Specific Per-Capita Income Generating Equations

Variable	All cases	Regime=0	Regime=1	Regime=2	Regime=3
Female head	-1.146	-2.045	-2.611	-1.146	0.180
	(-1.50)	(-0.57)	(-0.78)	(-1.18)	(0.16)
Age	0.504**	0.296	0.162	0.384	0.711**
	(3.27)	(0.61)	(0.32)	(1.51)	(3.01)
Age squared/100	-0.449**	-0.359	-0.193	-0.305	-0.612**
	(-3.28)	(-0.87)	(-0.43)	(-1.35)	(-2.87)
Elementary school	0.628	-0.121	1.595	0.903	0.787
	(1.31)	(-0.08)	(0.86)	(1.31)	(1.09)
Middle school	1.183*	0.567	6.175**	0.665	1.053
	(2.11)	(0.28)	(2.91)	(0.83)	(1.27)
High school	1.488**	0.563	3.776	0.918	2.336**
	(2.61)	(0.30)	(1.72)	(1.11)	(2.73)
Higher education	4.188**	3.906	3.280	3.791**	5.454**
	(4.80)	(1.54)	(1.10)	(2.83)	(4.03)
Family size	-0.988**	-1.231*	-1.133*	-0.734**	-1.134**
	(-8.10)	(-2.12)	(-2.44)	(-4.11)	(-6.69)
Fraction adult males	4.219**	1.931	2.294	3.679**	5.025**
	(5.30)	(0.63)	(0.68)	(3.28)	(4.38)
Fraction adult females	4.151**	6.341*	8.839**	3.374**	2.778*
	(4.89)	(2.09)	(2.82)	(2.79)	(2.20)
Landless	-3.156**	-6.858**	-1.910	-3.022**	-2.269*
	(-4.43)	(-2.74)	(-0.56)	(-3.17)	(-2.18)
Land owned per capita	1.532**	0.697	1.605*	0.905**	2.582**
	(7.72)	(0.89)	(2.57)	(3.59)	(7.45)
Center-east	-0.317	1.984	1.013	-0.196	-1.432*
	(-0.81)	(1.36)	(0.73)	(-0.35)	(-2.54)
South-west	-1.767**	0.525	-1.130	-1.441**	-2.841**
	(-4.71)	(0.38)	(-0.81)	(-2.66)	(-5.25)
Intercept	-4.487	4.973	3.859	-3.524	-9.707
	(-1.09)	(0.35)	(0.28)	(-0.51)	(-1.57)
R^2	12.87%	11.86%	14.40%	10.99%	19.64%
Number of cases	3,042	423	350	1,016	1,253

Notes: Dependent variable: total household income per capita. All farm households have farm income by definition; t-statistics in parentheses; * coefficient significant at 5%. ** coefficient significant at 1%.
Regime=0: household with no income from non-farm business or labor.
Regime=1: household with income from non-farm business only.
Regime=2: household with income from non-farm labor only.
Regime=3: household with income from both non-farm business and non-farm labor.

Table 7. Regression-Based Regime-Specific Contributions to Inequality[a]

Variable	All cases	Regime=0	Regime=1	Regime=2	Regime=3
Female head	0.094	0.080	0.217	0.036	-0.023
Age	-2.834	-9.429	-2.432	2.297	3.254
Age squared/100	5.007	15.315	4.131	-0.837	-0.353
Elementary school	0.023	0.017	-0.582	0.389	-0.004
Middle school	0.111	0.067	3.807	-0.032	-0.049
High school	0.371	0.270	1.489	0.101	0.477
Higher education	0.613	1.713	-0.315	0.557	0.801
Family size	4.214	1.282	1.859	4.074	7.347
Fraction adult males	3.187	1.485	1.009	2.689	4.255
Fraction adult females	2.926	6.100	6.531	2.107	1.816
Landless	1.245	2.382	0.436	1.888	0.801
Land owned per capita	2.933	0.689	3.457	1.860	5.858
Center-east	-0.055	0.962	-0.156	-0.046	-0.229
South-west	0.445	0.038	-0.468	0.349	1.767
Intercept	0.000	0.000	0.000	0.000	0.000
Residual	81.721	[b]	[b]	[b]	[b]

a. Dependent variable: total household income per capita. All farm households have farm income by definition.
b. Regime-specific contributions are computed as percentages of overall inequality, hence they do not sum up to total Regime -specific inequality.
Regime=0: household with no income from non-farm business or labor.
Regime=1: household with income from non-farm business only.
Regime=2: household with income from non-farm labor only.
Regime=3: household with income from both non-farm business and non-farm labor.

Table 8. Regime- and Source-Specific Per-Capita Farm Income Generating Equations[a]

Variable	All cases[b]	Regime=0	Regime=1	Regime=2	Regime=3
Female head	-2.532**	-0.189	-6.097*	-1.585*	-1.711
	(-4.05)	(-0.06)	(-2.27)	(-2.10)	(-1.90)
Age	0.250*	0.052	0.197	0.198	0.382*
	(1.99)	(0.11)	(0.50)	(1.00)	(2.06)
Age squared/100	-0.275*	-0.209	-0.267	-0.186	-0.321
	(-2.46)	(-0.54)	(-0.75)	(-1.05)	(-1.91)
Elementary school	0.333	-0.782	1.417	0.380	0.521
	(0.85)	(-0.55)	(0.97)	(0.72)	(0.93)
Middle school	0.267	-0.750	5.057**	-0.402	0.803
	(0.58)	(-0.40)	(2.96)	(-0.64)	(1.23)
High school	-0.244	-0.270	1.090	-0.891	0.708
	(-0.52)	(-0.15)	(0.62)	(-1.39)	(1.05)
Higher education	0.498	0.298	1.359	-0.371	1.097
	(0.70)	(0.13)	(0.57)	(-0.36)	(1.04)
Family size	-0.660**	-0.911	-0.487	-0.238	-0.611**
	(-6.62)	(-1.62)	(-1.19)	(-1.49)	(-3.98)
Fraction adult males	1.467*	3.407	-0.212	2.183*	3.103**
	(2.26)	(1.13)	(-0.07)	(2.28)	(3.04)
Fraction adult females	1.965**	4.568	4.052	2.692**	0.877
	(2.84)	(1.64)	(1.64)	(2.84)	(0.88)
Landless	-1.583**	-5.542*	-0.523	-1.219	-0.475
	(-2.72)	(-2.39)	(-0.19)	(-1.65)	(-0.59)
Land owned per capita	1.598**	1.620*	2.116**	1.046**	1.387**
	(9.85)	(2.06)	(4.13)	(5.11)	(4.95)
Center-east	0.625	0.981	1.775	0.302	-0.164
	(1.96)	(0.74)	(1.59)	(0.70)	(-0.37)
South-west	-0.296	-0.505	-0.100	-0.197	-0.967*
	(-0.97)	(-0.40)	(-0.09)	(-0.48)	(-2.32)
Intercept	-0.752	c	c	c	c
	(-0.22)				
R^2	9.50%	15.86%	20.23%	11.92%	15.94%
Number of cases	3,042	423	350	1,016	1,253

a. t-statistics in parentheses; * coefficient significant at 5%; ** coefficient significant at 1%.
b. from table 4.
c. the intercept is not separable from the set of selectivity-correction coefficients, hence it is not comparable to the first column and is not reported. Selectivity corrected using Dahl (2002) semiparametric procedure with second-degree polynomial of nonparametric choice probabilities. Regimes: (0) no income from non-farm business or labor; (1) income from non-farm business; (2) income from non-farm labor only; (3) income from both non-farm business and non-farm labor.

Table 9. Regime- and Source-Specific Per-Capita Non-Farm Income Generating Equations[a]

	Non-Farm Business Income			Non-Farm Labor Income		
Variable	All[b]	Reg. 1	Reg. 3	All[b]	Reg. 2	Reg. 3
Female head	-0.201	0.838	0.468	0.941**	0.289	0.416
	(-0.42)	(0.51)	(0.85)	(2.83)	(0.64)	(0.94)
Age	-0.023	-0.024	-1.111	0.259**	-0.031	0.150
	(-0.24)	(-0.10)	(-0.98)	(3.68)	(-0.26)	(1.65)
Age squared/100	-0.023	-0.021	0.072	-0.224**	0.030	-0.120
	(-0.27)	(-0.10)	(0.70)	(-3.55)	(0.29)	(-1.46)
Elementary school	0.070	-0.128	-0.087	-0.367	-0.076	0.085
	(0.23)	(-0.14)	(-0.25)	(-1.71)	(-0.24)	(0.31)
Middle school	0.356	0.559	0.489	-0.252	0.190	0.204
	(1.02)	(0.54)	1.23)	(-1.01)	(0.51)	(0.64)
High school	0.099	1.119	0.412	0.102	0.981**	0.892**
	(0.28)	(1.05)	(1.00)	(0.40)	(2.59)	(2.71)
Higher education	-0.178	1.427	-0.263	0.747	2.707**	2.179**
	(-0.33)	(0.99)	(-0.41)	(1.92)	(4.41)	(4.22)
Family size	0.043	-0.135	-0.037	0.306**	0.055	-0.083
	(0.58)	(-0.55)	(-0.40)	(5.73)	(0.59)	(-1.13)
Fraction adult males	1.098*	2.553	0.901	3.024**	2.447**	1.759**
	(2.23)	(1.42)	(1.43)	(8.61)	(4.28)	(3.50)
Fraction adult females	0.799	2.918	0.988	1.737**	0.930	1.350**
	(1.52)	(1.92)	(1.60)	(4.60)	(1.63)	(2.73)
Landless	-0.753	-1.946	-0.260	-0.308	-0.762	-0.654
	(-1.68)	(-1.17)	(-0.52)	(-0.98)	(-1.73)	(-1.64)
Land owned per capita	0.551**	-0.274	1.286**	-0.546**	-0.214	-0.522**
	(4.72)	(-0.88)	(7.53)	(-6.04)	(-1.79)	(-3.82)
Center-east	-0.573*	-1.602*	-0.598*	-0.637**	-0.710**	-0.417
	(-2.40)	(-2.37)	(-2.22)	(-3.69)	(-2.73)	(-1.93)
South-west	-0.765**	-1.414*	-0.642*	-0.773**	-1.105**	-0.608**
	(-3.34)	(-2.10)	(-2.48)	(-4.67)	(-4.44)	(-2.94)
Intercept	0.539	c	c	-7.518**	c	c
	(0.21)			(-4.00)		
R^2	1.02%	14.38%	8.35%	2.95%	16.52%	12.76%
Number of cases	1,603	350	1,253	2,269	1,016	1,253

Notes: see table 8.

Table 10. Regime- and Source-Specific Contributions to Inequality via Farm Income[a]

Variable	All cases[b]	Regime=0	Regime=1	Regime=2	Regime=3
Female head	0.819	0.007	0.506	0.050	0.220
Age	0.204	-1.649	-2.965	1.185	1.750
Age squared/100	0.627	8.899	5.700	-0.511	-0.186
Elementary school	0.143	0.109	-0.517	0.164	-0.003
Middle school	0.037	-0.089	3.118	0.020	-0.037
High school	0.056	-0.130	0.430	-0.098	0.144
Higher education	-0.042	0.131	-0.130	-0.054	0.161
Family size	2.628	0.948	0.799	1.325	3.961
Fraction adult males	0.416	2.621	-0.093	1.596	2.628
Fraction adult females	0.738	4.394	2.994	1.682	0.573
Landless	0.595	1.925	0.119	0.762	0.168
Land owned per capita	4.315	1.603	4.557	2.150	3.148
Center-east	0.206	0.475	-0.274	0.071	-0.026
South-west	-0.036	-0.037	-0.041	0.048	0.602
Intercept	0.000	0.000	0.000	0.000	0.000

a. Regimes: (0) no income from non-farm business or labor; (1) income from non-farm business; (2) income from non-farm labor only; (3) income from both non-farm business and non-farm labor.
b. from table 5.

Table 11. Regime- and Source-Specific Contributions to Inequality via Non-Farm Income[a]

	Non-Farm Business Income			Non-Farm Labor Income		
Variable	All[b]	Reg. 1	Reg. 3	All[a]	Reg. 2	Reg. 3
Female head	0.013	-0.032	-0.063	0.166	-0.008	-0.049
Age	0.584	0.649	-0.526	-8.220	-0.026	0.735
Age squared/100	0.712	0.391	0.042	8.982	0.008	-0.083
Elementary school	-0.005	0.078	0.001	0.144	-0.026	0.000
Middle school	0.077	0.222	-0.020	-0.069	-0.011	-0.006
High school	0.023	0.386	0.073	0.036	0.113	0.163
Higher education	-0.010	-0.124	-0.046	-0.006	0.407	0.314
Family size	0.052	0.154	0.126	1.341	-0.192	0.644
Fraction adult males	0.528	1.105	0.760	2.812	1.767	1.470
Fraction adult females	0.229	2.018	0.624	1.015	0.614	0.921
Landless	0.139	0.354	0.094	-0.027	0.451	0.227
Land owned per capita	0.277	-0.497	2.979	0.848	-0.481	-1.091
Center-east	-0.016	0.248	-0.096	0.112	-0.161	-0.070
South-west	0.183	-0.529	0.399	0.325	0.267	0.382
Intercept	0.000	0.000	0.000	0.000	0.000	0.000

a. Regimes: (1) income from non-farm business; (2) income from non-farm labor only; (3) income from both non-farm business and non-farm labor.
b. from table 5.

Table 12. Marginal Effect of Land Owned Per Capita on Total Gini

Variable	Total Income	Income by Source	Income by Regime[a]	Income by Source and Regime
Aggregate effect	-0.021	-0.010	+0.013	
Farm business		**-0.017**		**+0.016**
Regime 0				+0.001
Regime 1				+0.012
Regime 2				+0.002
Regime 3				+0.002
Non-farm business		**-0.005**		**-0.001**
Regime 1				-0.002
Regime 3				+0.001
Non-farm labor		**+0.012**		**+0.000**
Regime 2				+0.000
Regime 3				+0.000
Capital		**+0.000**		
Transfers		**+0.000**		
Other		**+0.000**		
Regime 0			+0.000	
Regime 1			+0.009	
Regime 2			+0.002	
Regime 3			+0.002	

a. Regimes: (1) income from non-farm business; (2) income from non-farm labor only; (3) income from both non-farm business and non-farm labor.