AMERICAN UNIVERSITY
WASHINGTON, DC

Department of Economics
Working Paper Series

The Effect of Non-Farm Income on Investment in Bulgarian Family Farming

By:

Tom Hertz
Department of Economics, American University

No. 2007-15

First Draft, July 2006
Revised for Presentation at FAO Conference on Migration and Agriculture, Rome, January, 2007
This Revision, June 2007

http://www.american.edu/academic.depts/cas/econ/workingpapers/workpap.htm

https://doi.org/10.17606/m8bp-de82

The Effect of Non-Farm Income on
Investment in Bulgarian Family Farming

Abstract

This paper documents a relationship between non-farm income (primarily earnings and pensions) and agricultural outlays in Bulgaria, using the 2003 Multitopic Household Survey. The outcomes analyzed are expenditures on working capital (variable inputs such as feed, seed, and herbicides) and investment in livestock. I find that while non-farm income has no significant effect on the probability of purchasing variable inputs, it does have an effect on the amount spent if positive, with an estimated elasticity of 0.14. Non-farm income also has an effect on the number of households that purchase farm animals, with an estimated elasticity of 0.35. The use of non-farm income for farm investment is consistent with the presence of credit constraints, as is the fact that less than one per cent of farmers report outstanding debts for agricultural purposes. Yet it is also noted that many farm households take out large unsecured loans for other purposes, suggesting that a lack of demand for agricultural borrowing may also be part of the problem.

Many thanks to Paul Winters and the Food and Agriculture Organization for financial support of this work, to Angel Bogushev for translation and data preparation, and to Mieke Meurs for her insight into the Bulgarian farm economy. Thanks also to Gero Carletto for helpful comments on the first draft and to John Maluccio for comments on the January draft.

Please address correspondence to:
Prof. Tom Hertz
Dept. of Economics, American University,
4400 Mass. Ave. NW, Washington, DC 20016, USA.
Email: hertz@american.edu
Tel. (202) 885-2756
Fax (202) 885-3790
Word count: 7573 (text and references); 1655 (tables and figures).

1. Introduction and Summary of Findings

One of the most enduring challenges in development economics has been the question of how to encourage productivity-enhancing investments in small-scale farming, which provides the main source of income for the majority of the world's poor, and additional food security for many low- and middle-income households as well. A prominent obstacle is that family farmers are typically credit-constrained, i.e. unable to borrow to finance productive agricultural investments (Ghosh, Mookherjee, and Ray 2000). Owning land should help relax the credit constraint, but where markets for land are thin or missing, as they are in many countries emerging from socialism, land is of limited value as collateral. Moreover, even saleable land does not guarantee access to credit: additional factors such as total debt service burden in relation to income are taken into account by informal as well as formal lenders (Zeller 1994).

Bulgaria is a country in which 58 per cent of households reported owning farm land in 2003, and yet, of these, just 0.6 per cent reported having agricultural loans outstanding, whether from a bank, an individual or any other kind of lender. Roussenova and Nenkov (2001) enumerate a host of reasons, summarized below, why banks are reluctant to lend to farmers in general, and to small farmers in particular, many of whom are elderly pensioners, and the majority of whom sell none of their output.

Given this supply-side credit constraint, we should expect to observe a relationship at the household level between non-farm income and investments in agriculture, since non-farm cash relaxes the budget constraint. In this paper I attempt to quantify this relationship, using data from Bulgaria's 2003 Multitopic Household Survey, and looking at two investment-related outcomes. First, I analyze the connection between non-farm income and the sum of reported annual expenditures on 13 different variable inputs, such as feed, seed, herbicides, and so forth, which represent the working capital needs of the farm. I find that non-farm income has no significant effect on the probability of spending at least some money on variable inputs, but does have an effect on the amount spent, conditional on its being positive, with an estimated elasticity of 0.14. Next I look at the relation between non-farm income and the purchase of farm animals during the last year, which is a true measure of investment, i.e. of an increment to the (live) capital stock. Here I find that non-farm income has a positive effect on the probability of purchasing farm animals during the year, but no statistically significant effect on the amount spent. The coefficients imply that a ten per cent increase in non-farm income raises the mean household's probability of purchasing livestock by about 0.0075. Given that the observed probability of purchase in the sample is 0.21, this represents about a 3.5 per cent increase in the number of households buying livestock, or an elasticity of that figure with respect to non-farm income of 0.35. I conclude that non-farm income does provide a non-trivial source of finance for agricultural investments and expenditures on working capital.

These finding are broadly consistent with about half of the previous empirical work on the question of linkages between farm and non-farm activities. For example, they square with the findings of de Janvry, Sadoulet, and Zhu (2005), who document a connection in China between participation in off-farm activities and rising agricultural productivity (a different, but related, outcome measure). On the other hand, a study from Ethiopia finds that access to rural non-farm employment leads to a lack of attention to the ecological management of farm resources, and subsequent soil erosion (Holden, Shiferaw, and Pender 2004). This highlights the fact that off-farm employment clearly draws labor away from farming, a point which has implications for policy, as well as for the econometrics of estimating the effects of non-farm income, as will be discussed below. A larger literature has studied the links between migration, remittances, and agricultural investment, with equally mixed results. Some find that remittances have little effect on productive activities (Durand, *et al.* 1996; Taylor, *et al.* 1996) while others find positive effects (Taylor, Rozelle, and De Brauw 2003; Black, King, and Tiemoko 2003). A plausible interpretation is that the effects of non-farm cash depend on the agricultural context, which is specific to time and place.

In present-day Bulgaria, there is reason to believe that credit constraints are not the only impediment to the growth of the family farm sector: demand-side obstacles may also be relevant. In particular, I note that while fewer than one per cent of farm households have agricultural debt, about 17 per cent report outstanding loans for other purposes, primarily to cover current bills. The fact that the bulk of these are *unsecured* consumer loans suggests that a lack of interest in borrowing for agricultural purposes may be as much of a reason for the absence of farm lending as the lack of access to credit *per se*.

The next section gives a brief overview of the past and present state of Bulgarian farming. Section 3 describes the survey dataset, and presents some initial descriptive findings. Section 4 discusses the econometric issues involved; results follow in Sections 5 and 6, and conclusions are drawn in Section 7.

2. Historical Background and Current Conditions

Meurs (2001), from which much of what follows is drawn, notes that prior to the Second World War, virtually all Bulgarian households owned land, but farms were extremely small (69 per cent of all landholdings were under 5 hectares) and methods of farming were crude. Markets for the private sale and rental of land existed, as did the practice of hiring labor, but these were limited in scope. Agricultural cooperative organizations, however, were widespread, incorporating about one-third of the economically active population. These provided credit, primarily, but also supplies, distribution and processing services, and to a lesser extent, engaged in cooperative production.

Under socialism, about 95% of arable land (not counting the ubiquitous small household plots, which averaged 0.3 hectares) came under state control in the form of larger-scale, mechanized agricultural collectives (as opposed to cooperatives); the private ownership of agricultural machinery was largely eliminated. The small household plots, however, continued to serve as important sources of food for household consumption, and still do. Despite reforms in the 1970s and 1980s, production by family farmers for sale in the market remained extremely limited, while production for own consumption remained the norm.

In 1991, the Ownership and Use of Farm Land Act required that state-held land be returned to whoever owned it in 1944, or their heirs, a remarkable social fact in itself, and one that lends a degree of exogeneity to the current distribution of land holdings. All those who worked for the collectives were eligible to receive a share of the proceeds of their liquidation. Agricultural equipment and fixed capital was sold, often to members of the former collective who reorganized themselves as private cooperatives, in a return to the pre-war arrangement. However, despite these cooperative efforts to maintain the viability of mechanized farming, and although attempts were made to avoid recreating the scattered plots that characterized the pre-war era, the immediate result of restructuring (combined with the elimination of subsidies to agriculture, the loss of export markets, and a dramatic deterioration in the ratio of output prices to input costs) has been the de-capitalization of agriculture, and a return to subsistence farming. From 1991 to 2000, although agriculture's share of GDP fell from 15 to 13 per cent, its share of employment *rose* from 19 to 26 per cent, indicating a dramatic drop in labor productivity in agriculture in relation to the rest of the economy. Over this period yields of cereals and vegetables declined by about 60%, while output of meat, dairy and eggs fell by even more (Roussenova and Nenkov 2001).

Although some consolidation of land holdings has since taken place, the median farm household in the sample soon to be described owns just 0.5 hectares of land, and 99 per cent own less than 10 hectares.[2] Virtually all commentators have noted that a middle class of more prosperous private commercial farmers has yet to emerge, and continuing market failures in both land and credit are often cited as causes. Roussenova and Nenkov explain that banks insist that farmers borrow against their residential real estate, rather than their agricultural holdings, since the markets for residential land are better developed, yet farmers are reluctant to do so. They also argue that privatization led to increased risk-aversion among banks, as they sought to avoid deterioration in their portfolios prior to their sale, and to avoid investing in activities in which they had no prior experience, e.g. small private farms, whose finances neither they nor the farmers could assess accurately. Legal reforms have also resulted in uncertainties regarding creditors' rights, and tighter prudential regulations have discouraged risky lending. Although some of these problems have since been somewhat ameliorated, including by an expansion in the government's provision of credit to farmers, it is clear that most if not all of the limitations just described were significant during 2003, when the survey data used here were collected.

[2] These figures, however, should understate the size of farms as operating units, since many households rent their lands to cooperatives, which presumably consolidate them.

3. Data, Definitional Matters, and Descriptive Statistics

The 2003 Multitopic Household Survey is a nationally representative, stratified and clustered sample of some 3000 households.[3,4] The modules that I use measure economic participation and earnings, describe agricultural land use and farm production, estimate income, assets and liabilities, and contain some subjective indicators of interest.

Total agricultural income in the survey can be broken down into (a) the estimated value of crops, livestock, and animal products sold; (b) the estimated value of crops (but not animal products) that were paid-in-kind for rented land or hired labor, given to friends or relatives, fed to animals, stored, or consumed by the household; and (c) the value of payments received for land rented to, or otherwise farmed by, other parties, including agricultural co-ops. Unfortunately, the measure of consumption of the household's own production is limited to the two-dozen crops that were itemized in the crops section of the agricultural module, and does not correspond to the detailed item-by-item accounting of self-produced food consumption that appears in the consumption module.[5] Thus, for example, consumption of farm-raised eggs and dairy products is not counted in agricultural production. As a result, some 215 households that owned land and produced non-crop food solely for own consumption are not counted as having any agricultural income, and are not included in my sample. Some 662 households that reported positive own-consumption but reported having *no* farm land, or did not consider their plots as such (i.e. are probably more gardeners than farmers) are likewise omitted. These exclusions raise issues of sample selection bias, which are discussed in the final section.

Tables 1 through 3 describe the sample of what I will call agricultural households, defined as those who reported that they owned or rented farm land, at least some of which was

---

[3] The survey was designed to be self-weighting and was stratified by Bulgaria's 28 districts: the district proportions thus agree very closely with those of the 2001 census. Clusters consist of up to five households, but 64% of clusters have fewer than five households represented. If this is due to non-response, then reweighting may be called for. Still, the basic race and gender proportions of the sample agree fairly closely with those of the 2001 census. Standard errors in all regression are heteroskedasticity-robust, and take account of clustering (i.e. non-independence of observations within the same cluster). However, the slight efficiency gains due to stratification are not exploited: reported standard errors are thus marginally higher than necessary, which errs on the side of conservatism.

[4] The raw number of households is 3023; I exclude those with no non-absent members (5) and those with zero reported income (10) for a final household sample size of 3008. The analytic samples are subsets of this number.

[5] The data from the consumption module are flawed as well. The problem is that not all weights and measures relating to self-produced food have been converted to monetary values. Instead, kilograms and liters appear to have been treated as money, and simply added to agricultural income. I netted these out from total household income, causing it to fall by about eight percent.

devoted to crops (i.e. planted), and also reported that they had positive agricultural *production*, defined as the sum of items (a) and (b), above. Agricultural rental income is excluded, and 191 households whose only farm income was from the rental of their land are thereby dropped.[6] I require that the households had non-farm income of at least 100 leva, and I drop one household that reported expenditures on variable inputs of greater than 100,000 leva. These two exclusions affect a total of 16 outliers, and improve the robustness of my results to the choice of functional form. Lastly, five households that lacked information of the education level of the head of household were excluded. The final number of agricultural households then stands at 1206, which is 40 per cent of the total number of households in the survey.

Table 1 describes the two outcomes to be analyzed. We see that 75 per cent of households spent at least some money on variable inputs, with the average amount spent (for those who purchased) being $623. The most common categories, each populated by at least 20 per cent of households, were for feed, seed, contracted seeding and plowing, veterinary services, fertilizers, and herbicides. Interestingly, just three per cent of farm households reported hiring in labor, indicating that farm labor markets are not of great relevance to this analysis. A smaller share, 21 per cent, purchased livestock during the year, and the average amount spent was $172.

Table 2 summarizes the right-hand-side variables that will be used to predict these expenditures, starting with non-farm income, the main covariate of interest, which ranged between about $200 and $41,000, with a mean of $6167, and clear positive skewness. This variable will enter in logarithms in all regression analyses in order to prevent the outliers in the long right tail from exerting undue leverage over the results. Next are listed the shares of non-farm income from each of four broad sources: pensions, earnings, social transfers, and non-agricultural rental income. Of these, the first two are clearly of greatest importance.[7] Note that the calculation of total household income does not appear to have made use of data collected on self-employment incomes (other than from farming) and remittances; these two elements are thus not counted as non-farm income in this analysis.

---

[6] Although rental income includes payments from agricultural co-ops, it does not appear that family members also *work* on these co-ops: 96 percent of adults in this category reported not having worked on the "land owned or rented by the household" or having bred livestock in the last year.

[7] Some people with jobs that were described as occurring on a "farm owned or rented by the household" also reported labor earnings for these jobs, which may represent a conflation of family farm and non-family-farm employment. I adjusted the data on non-farm earnings to omit these amounts, so as to avoid creating a spurious correlation between farm and non-farm activities.

First among the basic control variables are measures of land ownership and land area under cultivation; these serve both as measures of wealth (which influences the ability to self-finance investment, and to obtain credit) and of scale of operations (which is a determinant of the need for inputs). We see that the mean farm size is somewhat less than one hectare, and that, as usual, this distribution is also highly skewed, which we will again handle using logarithms. Next are three other wealth measures: an indicator for households who reported having money in the bank; one for those with other financial investments; and an estimate of the value of all consumer durables owned. This latter is highly skewed, but often zero; taking the square root preserves the zeroes while reducing the skewness.

The survey also asked for an itemization of major agricultural implements owned, drawn from the following list: tractor up to 15 Hp, tractor more than 15 Hp, autocombine, planting machine, thresher, truck, trailer, mill, milking machine, mechanical plow, incubator, and other machinery. Only eleven per cent of households reported owning *any* of the above (results not shown in table), with the most common asset being a mill (at five per cent). Fewer than three per cent of households owned a tractor of any description, about the same proportion as owned a milking machine, or a trailer. Rather than attempt to value these assets, I use a simple count of their number, whose mean stands at 0.19, and which works quite well as a regressor, with each extra item corresponding to about a 37 per cent increase in expenditures on variable inputs. This may reflect a direct causal connection – such as buying fuel for a tractor – but also may simply be one more way of capturing the overall scale and intensity of farm operations.

In the next block of the table, the labor inputs to farming are estimated. The survey asked this question in various ways, yielding about a dozen overlapping yet still incomplete measures of labor usage. I chose those that consistently performed well as predictors of variable inputs. The first is an estimate of the number of full time equivalents devoted to farming, based on two questions from the agriculture module: one asked how many household members were "engaged in farming all the time" and the other asked how many were "engaged in farming from time to time." I combined these into a single measure by assuming that the part-time farmers counted as 0.3 full time equivalents.[8] The next variable is the reported number of hours worked per week,

[8] Some 128 of the 1206 households (11 per cent) did not report anyone in either category. Fourteen of these zeros were replaced with estimates based on the next variable to be discussed, namely, reported hours spent in farming as a main or secondary occupation. The remaining 114 (9.5 per cent of the 1206 households) were assumed to have

taken from the labor module, for all those who reported that they worked on "a farm owned or rented by [your] household." Fully half of our farm households reported zero in this category, doubtless reflecting the fact that many do not count their family farm work as employment, despite having the option to do so on the questionnaire.[9] For those who report positive values, the responses are once again highly skewed, and I take the square root of reported hours for the reasons mentioned above.

The third and fourth labor measures come from the module that defines labor force participation, in which adults (15 or older) were asked, among other things, whether they had bred livestock in the last seven days, or the last 12 months, and these results are reported in the table. For the regression analyses of variable inputs, I created a variable that counted the number reporting livestock activities in the last 12 months but not the last seven days, after observing that this construct performed better as a predictor than the simple 12-month variable, and that the seven-day measure had little predictive power of its own, once the other measures described above were included. In the livestock equations I also enter the seven-day measure, which performs better in this context. A separate indicator variable is included that flags the 70 per cent of households who reported any livestock breeding activity (from the agriculture module), who consistently had higher expenditures on variable inputs. This must be omitted from the livestock equations, however, as only those who said yes to this question could report spending money on animals.

Next is a constructed indicator for those clusters where one or more households reported having either no public sewerage or water supply, or no electricity. This flags the most rural households who might lack ready access to markets for inputs, and who are in general poorer. Not shown are the shares in each of the 28 regions of the country, indicators for which are included in the regression analyses in order to capture the major differences in geography and crop mix that one observes across the country.

Last are a set of demographic indicators, which serve both as measures of the nature of human resources available to the household (age, health, gender, education), and perhaps also as measures of cultural traditions and attitudes towards farming, as captured by indicators for each

---

one family member engaged part-time in farming. For the analysis of those with positive expenditures on variable inputs, the share of imputed values for this variable was just two per cent (20 out of 907).

[9] The question was asked separately for "main" and "secondary" jobs, and hours from both categories were added, provided they were specified as being spent on the family farm.

of the four main language groups. It is noteworthy that the average age of all members of agricultural households is quite high, at 53 years, which emphasizes a key point: participation in small-holder agriculture in Bulgaria is biased towards the older generations. Among the adult members of our agricultural households, the estimated probability of having participated in any farm activity in the last seven days, controlling only for gender and education, peaks at age 58 for men, and age 55 for women. These figures compare to age 40, for both men and women, for participation in off-farm paid employment (results not shown in table).

Not shown in the table are the proportions of households receiving each of the four sources of income (as opposed to the shares of income from each source). Fully 85 per cent of agricultural households also received earnings; two-thirds of agricultural households received retirement pension incomes[10]; and smaller shares received pensions of other kinds as well as social transfers. This represents a considerable overlap between farm activity and non-farm income, which should facilitate identification.

Subject to the measurement problems described above, mean annual household income (at a 2003 PPP exchange rate of 0.60 leva per $US) is estimated, in Table 3, at $7,347 for agricultural households, of which $6,167 (84 per cent) is non-farm income. Per capita income stood at $2,749.   Following World Bank practice in the region, poverty thresholds are set at $4.30 per person per day, as well as $2.15/day; I also report the international standard $1.00/day estimates. These rates work out to 18, three, and one per cent respectively, which are somewhat lower than the corresponding national estimates from this survey (21, six, and two per cent.)[11]

---

[10] With some exceptions, retirement pension eligibility was from age 60 for men and 55 for women, until pension reforms were enacted in 2000, which have gradually raised the age of eligibility. By 2009 the ages of retirement will stand at 63 and 60 (International Social Security Association 2002).

[11] Note that the World Bank estimated consumption-based poverty rates for Bulgaria in 2003, at PPP, to be 33 per cent at $4.30/day and four per cent at $2.15/day, based on the 2003 Household Budget Survey (Alam, *et al.* 2005).

4. Econometric Specifications and Identification Strategies

I use two-part models to predict expenditures on variable inputs and livestock, where the first part consists of a probit specification of the probability of expenditures being positive, and the second part uses ordinary least squares (OLS) on the natural logarithms of the positive values. This choice is based on arguments that are developed most clearly in the health economics literature concerning the relative merits of the different ways of dealing with outcomes, like these, whose distributions are characterized by a large mass at zero and a long right tail. (See, for example, Duan, *et al.* 1984, and the references cited therein.)[12] The model may be written as:

[1] $$\Pr(y > 0 \mid X) = \Phi(X\gamma),$$

[2] $$\text{and, for } y > 0\text{: } \ln y = X\beta + u \text{ with } E(u \mid X) = 0.$$

Here y is expenditures on either variable inputs or livestock, Φ is the cumulative standard normal distribution, and X contains the variables listed in Table 2, including $\ln x_{nf}$, the log of non-farm income. The marginal effect of a change in log non-farm income on the probability of observing positive expenditures, evaluated at the sample mean, is then:

[3] $$\left.\frac{\partial \Phi(X\gamma)}{\partial \ln x_{nf}}\right|_{X=\overline{X}} = \gamma_{nf}\,\phi(\overline{X}\gamma)$$, where ϕ is the standard normal density, and

γ_{nf} is the probit coefficient for $\ln x_{nf}$.

---

[12] Alternative models include the Tobit approach (Tobin 1958) and models based on the work of Heckman (1979). I avoid the former because of their hypersensitivity to departures from homoskedastic normality in the error term, and because they often give misleading estimates of the effect of a given covariate on the probability of observing a positive outcome (as they do in this case). Further, following Deb, Manning, and Norton (2005) I argue that Heckman models are not needed if one is interested in modeling *observed* outcomes (treating the zeros as valid data, which represent an optimizing agent's corner solution) as opposed to *potential* outcomes (treating the zeros as censored observations of a latent outcome). Heckman models are also suspect when one lacks instruments that are predictive of observing a positive outcome but not predictive of the level of the outcome conditional on its being positive (Deaton 1997), as is often the case, the present application included. Two simpler alternatives are to use OLS and include the zero values (in levels, or k^{th} roots), or to add an arbitrary positive constant to the outcome and take logarithms: $ln(y+c)$. In addition to their being an improbable process by which the large spike at zero could have been generated (Tobin's original point) the problem with linear models in levels of y is that its skewness renders the estimates highly sensitive to a few influential observations, and may lead to poor finite-sample behavior of the estimator. On the other hand, while it handles skewness well, the problem with the $ln(y+c)$ approach is that the results are often highly sensitive to the choice of c, which is arbitrary. Lastly, the alternative of taking square (or higher order) roots of the outcome variable to reduce its skewness, while feasible, poses a retransformation problem when calculating marginal effects. We are not generally interested in the effect of a covariate on the expected value of the square root of y, but calculations of the corresponding effect on the expected value of y itself are sensitive to the presence of heteroskedasticity, a point which is often ignored in applied research (Deb, *et al.* 2005).

For the logged positive values, the relevant marginal effects are straightforward, and invariant across households. In particular, because non-farm income enters in logs, its coefficient is an elasticity:[13]

[4] $$\frac{\partial E[\ln(y \mid X, y>0)]}{\partial \ln x_{nf}} = \beta_{nf}$$

Potential sources of inconsistency in estimating the effect of non-farm income

There are a number of reasons why we should be concerned that reduced-form equations such as [1] and [2] might not yield consistent estimates of the effect of non-farm income on the purchase of variable inputs or livestock. First, because non-farm income is not randomly assigned, we must be alive to the possibility that it may be correlated with unobserved factors that are also predictive of these expenditures, and whose omission is thus a source of bias. In the present context, once we have controlled for various measures of the scale and nature of agricultural operations, and any other likely determinants of agricultural expenditures or farmers' preferences that we can capture (i.e. all of the variables listed in Table 2), we are left with many different candidates for omitted confounders. Perhaps the most common assumption is that "general ability" has a positive influence on most economic activities, including both non-farm earnings and the intensity of expenditures on farming, all else equal, leading to an upward bias in our estimate of the non-farm income effect. However, other factors might cut the other way: for example, risk aversion could lead to a reluctance to make large cash outlays on agriculture, given the risk of crop failure, combined with a desire to maximize off-farm income, as a means of income diversification, creating a negative bias. On the other hand, risk averse households might pursue non-farm income in order to spend *more* on agricultural inputs, perhaps buying more

[13] As in all such log-log models, however, this is an elasticity of the conditional mean of the log of y (the conditional geometric mean of y), not of the arithmetic mean of y, with respect to x. The expression for the elasticity at the arithmetic mean takes expectations and logs in the other order: $\frac{\partial \ln E[y \mid X, y>0]}{\partial \ln x_{nf}}$. Fortunately, however, this is equivalent to [4] if u and x_{nf} are independent, as well as under the weaker assumption that $\frac{\partial E[\exp(u) \mid X]}{\partial x_{nf}} = 0$.

Following a suggestion by Mullahy (1998) this conditional expectation can be modeled as an exponential function of the elements of X, and this zero restriction can then be tested using non-linear least squares, with the outcome variable being the antilogged residuals from [2] (Hertz 2007).

expensive drought-resistant seeds, or veterinary services, in order to reduce the chance of losing their crops or herds.[14] In short, there are many plausible stories one might tell about the role of unobserved factors, and no single dominant mechanism comes readily to mind, especially given that non-farm income is composed of so many different sources, some depending on ability, others age, and others *dis*ability. As a result, I have no initial sense of the net direction or magnitude of omitted variables bias that may plague our estimates of the non-farm income effect, and while I do not assume that this bias is necessarily large enough to invalidate the results of the two-part model outlined above, it is nonetheless clear that the question merits further investigation.

With this in mind, I present an augmented specification, inspired by the work of Cox Edwards and Ureta (2003), who faced similar issues of omitted variables in their efforts to estimate the effect on children's schooling of migrants' remittances in El Salvador. They postulate that households that receive remittances (in our case, households that receive a particular form of non-farm income) might differ in systematic ways from those that do not. They then propose including in their equation not only the level of remittances (which may be zero, and which is the variable whose effect they wish to estimate), but also an indicator variable for whether remittances were indeed positive. They argue that:

> More generally, the indicator variable will capture any additional effect of remittances on children's schooling that acts through channels other than the budget constraint, and any systematic differences in attitudes toward the schooling of children across families that do and those that do not receive remittances [p. 443].

Note, however, that while the remittance indicator does capture any linear, additive, conditional effects of differences between remitters and non-remitters, it does not eliminate all potential for bias. In particular, it might be that migrants from families who place a high value on education (this attitude being an omitted variable) would send *more* remittances home, for the express purpose of educating their siblings, *and* that children from these families would get more education, conditional on remittance income, than would children from families that place less value on education, yet who nonetheless receive *some* remittances. Thus differences *among*

[14] The complexity of the interrelations between farm and non-farm income generating activities has been emphasized in the theoretical literature (Singh, Squire, and Strauss 1986) and may account for the conflicting results of the studies cited above.

remitters are not controlled for. Still, if there are indeed systematic and relevant differences between remitters and non-remitters, it is better that these should be accounted for than ignored.

I implement a variant of this approach by including a set of variables that describe the composition of non-farm income, quantifying the share that was derived from each of the 16 different sources that were itemized in the Bulgarian survey (with one category omitted, since the shares sum to one).[15] These variables should capture the difference between households who do and do not receive (say) pension incomes, and do so in a way that differentiates between households that receive varying degrees of pension income, as a proportion of their total non-farm income. Given that many forms of non-farm income are determined by categorical eligibility (such as retirement pensions, disability payments, or income supports) it is plausible that the income shares will convey important information about differences in the human resources available to households.

A second and equally important concern is that decisions relating to expenditures on variable inputs and livestock are presumably made simultaneously with the decision of how much labor to devote to farming (versus to non-farm employment, or other pursuits), which could lead to simultaneity biases. In particular, suppose households that spend an especially large amount on, say, herbicides, creating a large positive error term in equation [2], are able to reduce the amount of labor time devoted to farming, by spending less time weeding. If this newly freed-up labor were then allocated to non-farm employment, this could create a positive correlation between the error term and the non-farm-income variable, leading to an upward bias in our estimates of its effect. On the other hand, it is also possible that the additional inputs could be complements rather than substitutes for agricultural labor, in which case their purchase might draw labor back away from non-farm employment, reducing non-farm income and creating a downward bias in its estimated effect. Note that these arguments apply only to non-farm *earnings*, since the other forms of non-farm income are transfers that require no labor input; yet earnings are an important component of the total, as shown in Table 2.

To address this problem I use an instrumental variables (IV) approach that attempts to isolate exogenous variations in non-farm income. This idea was motivated by the work of Case

[15] There are two measures of earnings (main and secondary jobs), four kinds of pensions (retirement, survivors, disability, and other), nine social transfers (unemployment benefits, family allowances, child allowances, educational assistance, heating assistance, rental assistance, medical assistance, income support payments, and other) and, finally, non-agricultural rental income.

(2001) on the relation between pension incomes and children's health in South Africa. In the 1990s, as apartheid was dismantled, the race-specific schedule of retirement pension benefit levels was gradually phased out, resulting in black households receiving unexpectedly large pensions. Case argues convincingly that this offers an ideal context in which to study the effects of exogenous differences in income on the health of those who do or do not co-reside with pensioners. She uses the presence of an age-eligible family member as an instrument for the receipt of a pension's worth of income, in order to avoid the possibility that the 80 per cent of eligibles who took up the pension differ in unobserved ways from the 20 per cent that did not.

I adapt this approach by using the number of people in each of seven ten-year age categories, starting at 21-30 and continuing to 81-90, as a set of instruments to predict total non-farm income from all sources. The relevance of the instruments stems from the fact that each additional adult of prime working age increases the expected value of household earnings, while each additional elder increases the expected value of retirement income, these being the two main sources of non-farm income.[16] Thus a count of household members by age category should predict total non-farm income rather well, and indeed it does: by themselves, these seven counts are able to explain 31 per cent of the variance in the log of non-farm income in the full sample of 1206 households, and 27 per cent of that variance among households with positive spending on variable inputs. Moreover, after conditioning on a long list of other covariates, F-tests of their joint significance reveal that these are indeed strong instruments, as detailed below. Thus the relevance of the instruments is established; the harder question, as always, is whether they are exogenous, i.e. whether, after conditioning on all other covariates, their effect on the two outcomes operates only through their effect on non-farm income. As is well known, instrumental exogeneity cannot be proven empirically, and must be defended by other means.

The advantage of using counts by age category as instruments is that each serves, in a sense, as an eligibility variable – being of working age makes you eligible to work – but these simple counts should not be correlated with those unobservable idiosyncratic attitudes that determine the amount of energy or ability devoted to paid employment, or that determine who takes advantage of pensions and other transfers for which they are eligible. This alone, however,

[16] One might imagine that each additional child or teenager would increase the expected value of family allowances and education assistance, but counts of the number of people in the 0-10 and 11-20 age categories were not significant predictors of non-farm income, and were thus redundant as instruments.

does not establish exogeneity. In fact, the number of people in the household clearly *does* influence agricultural production in ways other than by generating more non-farm income, since it increases the potential agricultural labor pool. And because agricultural production, in turn, is highly predictive of our outcome variables, it would appear that these instruments are flawed. But once we condition on the amount of labor actually supplied to agriculture we may then argue that the only remaining channel through which an extra body influences expenditures on variable inputs or livestock is through their bringing in more non-farm income.

A second threat to the validity of these instruments is that they convey information about age, and age is predictive of the degree of involvement in farming. Once again, however, after controlling for both the age of the head of the household, and the average age of all household members, as well as the measured amount of labor devoted to farming, we can make a plausible case for the conditional exogeneity of the instruments. I also control for the self-reported health status of the head, and the gender composition of the household, both of which may have a bearing on labor supply and labor productivity.

A third possible objection is that although an individual's age is exogenous insofar as it is not subject to manipulation, the number of people of each age in a household is clearly not exogenous. For instance, it is possible that the adult children of families that have a greater propensity to invest in their farms would be more likely to stay at home, on the ancestral land. Then larger households, with higher predicted levels of non-farm income labor, would be associated with this unobservable propensity, invalidating the instruments. I argue, however, that these endogenous family formation effects do not pose a problem. If they are relevant at all, they should relate most strongly to young adults, who may choose to leave the household for other pursuits. If the number of young adults in a household is endogenous (for our particular outcomes, and given all included covariates) but the number of older people is not, then we would expect to get different results when using the counts of young adults as our instruments than when using the older age groups only. This can be tested empirically, since the system is over-identified, and the results reported below do not present any evidence that different subsets of the instruments yield different results. These tests are based on heteroskedasticity-robust and clustered standard errors, and the regressions are estimated using the two-step feasible general method of moments, which offers efficiency gains over traditional two-stage least squares in over-identified systems.

5.1 Results: Expenditures on Variable Inputs

Table 4 reports the results of the four OLS and one IV specification of equation [2], modeling the log of expenditures on variable inputs for the sample of 907 agricultural households who spent some money on these items. In the first row is the elasticity of variable inputs to non-farm income, and its associated standard error, which is robust to heteroskedasticity and takes account of the clustered sample design. In the first column, log non-farm income is the only variable in the equation; the elasticity is 0.15 and it is statistically significant at the five per cent level.

The second column adds the covariates that describe the physical and financial resources of the household, the indicator for those who reported raising livestock, as well as the public services and district indicators. The log of the amount of land owned and planted, the number of agricultural implements owned, whether one bred livestock, and the (square root of the) value of consumer durables owned were all significant positive predictors of expenditures, but the financial wealth variables were not, perhaps because they are crude binary measures. The district fixed effects were jointly significant in all specifications, and the addition of this set of covariates raised the elasticity of interest to 0.207, and reduced its standard error. The share of variance explained rose to 0.44.

In the third column I add the full set of labor and demographic controls, as well as squared terms for the land, tools, and wealth (durables) variables. The estimated elasticity of variable inputs to non-farm income barely budges, at 0.198. The quadratic terms are jointly significant (p=0.022) and their addition was calculated to satisfy Ramsey's specification test, which had failed in the previous two columns, indicating the importance of omitted non-linearities which might potentially distort both this equation and the IV specifications to follow (see row labeled "No Nonlinearities?").[17] All three measures of labor utilization are significant predictors of expenditures on variable inputs, and all have positive effects, implying that positive scale effects outweigh negative substitution effects. The demographic variables were jointly insignificant (p=0.393) and the only individually significant finding was a one per cent reduction

[17] The Ramsey RESET test adds the 2^{nd}, 3^{rd}, and 4^{th} powers of the fitted values back into the original equation and tests their joint significance. Under the null hypothesis that the linear specification is correct, these higher order terms, which include all possible cross-products of the right-hand-side variables, should have no effect on the outcome. Rejection (statistical significance) indicates that non-linear terms are needed.

in predicted expenditures for each extra year of age of the head of household (p=0.046, not shown in table). This finding is surprising: I had expected both education and gender to play a role in determining the nature of farm operations, and had also expected the ethnicity indicators to capture unobserved wealth differences, as well as possible cultural differences in crop choices and farming methods.

In the next column I add the set of 15 variables measuring the composition of non-farm income, which are jointly significant (p=0.047 in row labeled "Income Shares"). These source-of-income variables reduce the estimated elasticity by six points, to, 0.138, a figure, however, which remains statistically distinguishable from zero at the ten per cent level. It thus appears that the Cox Edwards and Ureta-inspired approach does capture important differences between recipients of greater and lesser shares of the various forms of non-farm income, and in so doing reduces its estimated effect on agricultural expenditures.

In the final column I instrument for non-farm income using the seven counts of household members by age category, which causes the elasticity in question to more than double, to 0.315. The F-statistic testing the joint significance of the seven instruments in the first stage regression stands at 27.6 (see row labeled "Weak Instruments?"). Using the critical values developed by Stock and Yogo (2002), this value is large enough to reject the hypothesis that weak instrument bias is greater than five per cent of the magnitude of the endogeneity bias under OLS, at the five per cent level of significance.[18] Thus the instruments appear strong enough to merit consideration.

Below this I report Hansen's J statistic for over-identification, whose null hypothesis is that the instruments are jointly valid, conditional on at least one of them being valid. The equation "passes" this test with flying colors (p=0.935), indicating either that our instruments are all valid, or that none of them are, or that some are and some are not, but we lack the power to detect this inconsistency. I also tested each subset of six instruments and in no case was the seventh instrument shown to be inconsistent with the other six, nor were any of them shown to be redundant (in the sense of adding nothing to the asymptotic efficiency of the estimator). As I

---

[18] See their Table 1, but note an apparent typo: the column headed K_2, referring to the number of instruments, begins at 3, whereas all similar tables begin at 1, which is consistent with the ability to handle one endogenous variable but not more, as indicated by the blank cells. Thus the labels in that column would appear to be off by 2; for seven instruments, I look to row nine. The critical value needed to reject the hypothesis (at the five per cent level) of a weak instrument bias of five per cent or more in relation to the OLS (endogeneity) bias is 20.5.

argued above, I take the mutual consistency of the different age-category instruments as evidence that the problem of endogenous family formation is not serious in this application.

In the final row, I report a heteroskedasticity and cluster-robust form of the Durban-Wu-Hausman test for the exogeneity of the non-farm income variable (conditional on instrumental validity), which is not rejected (p=0.263). This means that while we have produced no test statistics that challenge the validity of the IV results, neither have we any evidence that non-farm income is endogenous in the OLS equation in the first place. Hence we have no compelling reason to prefer the IV estimate (whose standard error is on the order of 0.17) to the OLS results (whose standard errors are less than half as large). My preferred estimate of the elasticity of variable inputs with respect to non-farm income, for those with positive expenditures, is thus the figure from column OLS-4, or 0.138, whose 90% confidence interval stretches from 0.014 to 0.263.

Table 5 reports the analogous results for probit models that predict which of the 1206 farm households spent money on variable inputs. In the initial columns we see a positive effect of non-farm income, but once the full set of controls are added, in column (3), the effect vanishes. The addition of the income shares, in column (4), makes no difference. In the next column, instrumenting for non-farm income (using the maximum likelihood approach to a probit model with endogenous regressors) raises the estimated effect, as it did in the previous table, but also greatly amplifies its standard error, so that it remains statistically insignificant. Also as before, the hypothesis that non-farm income is in fact exogenous in the conventional probit equation is not rejected (p=0.335, from the Wald test for the existence of a correlation between the error terms of the main equation and the "first stage" equation that includes the extra instruments).

The results in columns (3) and (4) are not only statistically insignificant, they are trivial in magnitude: they imply that a ten per cent increase in non-farm income would raise the probability of positive expenditure by 0.001. Given that 75 per cent of households had positive expenditures, this ten per cent change in income corresponds to about a 0.14 per cent change in the number of purchasing households, for an elasticity of 0.014. It thus appears that non-farm income plays no significant role in the decision as to whether to purchase variable inputs, but does influence the amount spent for those who purchase.

5.2 Results: Investment in Livestock

Table 6 looks at expenditures on livestock, again starting with OLS models for those households who had positive expenditures; this subset is a fairly small, numbering just 259. No significant effects of non-farm income are found, but this is due as much to large standard errors (which are not helped by the small sample size) as it is to smaller point estimates. In column (4) the estimated elasticity is 0.16, which is similar to that found for variable inputs, although not significant at the ten per cent level. This equation, however, did not pass Ramsey's specification test, due to non-linear effects of the income share variables. This was rectified by adding quadratic terms for each income share, in column (4B), generating an elasticity estimate of 0.12, which is again close to the result for variable inputs, yet insignificant.

The IV estimator in column (5) roughly doubles this estimated elasticity (as before), but it remains statistically indistinguishable from zero. In this equation several of the seven instruments were redundant and were thus omitted; the combination that performed best was to start at age 30 and group people into 15 year age brackets: 30-44, 45-59, 60-74, and 75 or older. The F-statistic of the joint significance of these four counts (15.0) was large enough to reject the hypothesis of weak instrument bias of ten per cent or more.[19] Hansen's J statistic again provides no evidence of instrumental invalidity, nor was any single instrument inconsistent with the other three, or redundant. However, as with variable inputs, the IV specification is not sufficiently different from the OLS to warrant the conclusion that non-farm income is endogenous in the first place. Hence the OLS results remain the preferred estimates.

Table 7 reports the probit estimates of the probability of making livestock purchases of any size. In order to avoid selection bias, these are run for the full sample of 1206, not just for the subset of 855 households who reported breeding livestock (even though the non-breeders had no purchases by definition). As a result they capture the effect of non-farm income both on the probability of breeding livestock at all, and on the probability of purchasing livestock, in a given year.

The results for the effect of non-farm income are positive and significant, and non-trivial in magnitude. The estimate from column (4) (whose covariates include the squared income shares that were seen to be important in the previous table) implies that a ten per cent change in

[19] The Stock-Yogo critical values for four instruments and one endogenous regressor are 11.1 for ten per cent maximal bias (in relation to the endogeneity bias under OLS) and 19.3 for five per cent.

non-farm income leads to an increase in the probability of purchasing livestock of 0.0075. Given that the observed probability of purchase is 0.215, this represents a 3.5 per cent increase in the number of households who purchase. Thus the elasticity of this figure with respect to non-farm income is on the order of 0.35.

In the final column, the IV estimator (which uses the full set of seven instruments, given the larger sample size) produces exactly the same point estimate, but with a standard error that is more than ten times as large as in the conventional probit. This washes away the statistical significance of non-farm income, but, again, the endogenous-regressor probit model does not actually detect any endogeneity in non-farm income (p=0.318), and thus provides no reason to doubt the conventional probit results. Although no formal test of instrumental relevance is reported, recall that these seven instruments explain nearly one-third of the variance of log non-farm income in this sample.

Taken together, Tables 6 and 7 suggest that non-farm income aids in the purchase of livestock. The effect appears strongest for the decision of whether or not to buy, and is harder to detect in relation to the amount spent, in part due to the small sample size of this subset. These conclusions hold after controlling for a detailed itemization of the sources of non-farm income (the 15 income shares and their squares) which should help capture some of the differences among households that are not picked up by the other covariates.

6. Demand-Side Factors

Table 8 summarizes the number of loans outstanding, and their amounts, according to the purpose and source of the loan. For households in general, and for the subset of agriculturally productive households, the dominant reason for borrowing is to cover current bills. Among agricultural households, these unsecured consumer loans outnumber agricultural loans by twelve to one. Thus unsecured credit is available, and in amounts that are non-trivial: the median consumer loan for farm households ($1667) was large enough to cover about 3.5 times the average amount spent on variable inputs ($469, Table 1). It is hard to see how households who have access to fairly large unsecured consumer loans would not have access to credit for farming. The lack of borrowing for agriculture would thus seem to reflect not merely problems with collateral and balky bankers, but a general lack of interest in borrowing to finance agricultural activities.[20] This squares with the views of Roussenova and Nenkov (2001) who include "Insufficient motivation for investment projects" among their explanations for the low level of agricultural lending. The high average age of Bulgaria's family farmers may well be a factor in explaining this lack of motivation, whether due to the difficulty of "unlearning" past practices (Mishev and Kostov 2003) or to the canonical economic logic that dictates that investments be made at younger ages.

Table 9 provides another piece of evidence to suggest that farming is not widely viewed as a viable route out of poverty. Families were much more likely to report an interest in finding more and, in particular, better-paid, employment, or in relying on the assistance of family and friends, than in taking up farming as a means of raising their standards of living.[21] Convincing younger Bulgarians that private farming is the road to riches may not be an easy task. Still, we should not discount the importance of Bulgaria's small farm economy to providing, if not a road to riches, at least a modicum of food security.

---

[20] One possible counter-argument is that some of the borrowing that is described as covering current bills is actually covering farm expenses, since most households probably do not keep separate books for their farm activities. Yet "Agricultural activities" was actually the *first* option on the questionnaire, and "Current bills" was the last…

[21] Unfortunately, expanding the scale of *existing* farm operations was not a listed option, but one gets the impression it would not score much higher.

7. Comments

A few words are in order about the statistical robustness of my findings. The OLS and probit results were generally robust to alternative specifications, and their standard errors were calculated conservatively – in particular, the variance-inflating effects of clustering are recognized throughout, as is the possibility of heteroskedasticity, whereas the slight efficiency gains from incorporating stratification are not exploited. The IV results, as is usually the case, were far more sensitive to specification and to the choice of instruments. Although they never serve as my preferred estimates, the conclusion that they cast no doubt on the OLS or standard probit results is open to challenge if one doubts the conditional exogeneity of the instruments. The most vulnerable aspect of that claim is that it relies strongly on the ability to control for labor inputs to agriculture, yet these are imperfectly measured. The size and direction of bias that this imparts cannot be determined without a better understanding of the structure of measurement error in labor inputs.

Another covariate that is likely to be poorly measured is non-farm income itself, especially its earnings component. Even in carefully executed studies such as the Panel Study of Income Dynamics, the reliability of log annual earnings falls in the range of 0.70 to 0.85 (Bound, *et al.* 1994; Duncan and Hill 1985). No validation studies are available for the present survey to determine the reliability of the non-farm income data, but various factors suggest that it is more likely to fall near the lower end of this range than the upper. First, as noted above, self-employment income and remittances are not counted. Second, the estimates are based in part on monthly figures, whose true values are inherently more volatile than are their annual sums. This is important because the annual figures (or perhaps even multi-year averages) are likely more relevant to the determination of annual agricultural expenditures than are the monthly amounts. Thus even if the monthly data were perfectly measured they should still be interpreted as noisy measures of the variable of interest. Third, there is anecdotal evidence of poor performance by some enumerators, who may have struggled with a challenging income module.[22] In principle measurement error should lead to an understatement of the impact of non-farm income, but this

[22] Incomes were recorded in a mixture of annual and monthly terms, and the layout of the questionnaire makes it very easy to confuse the two. For the monthly data, a second question records the number of months out of the year that this level of income is earned (a hard question to answer if the monthly amount varies), which is later multiplied by the monthly amount. This will yield the correct annual total only if the monthly figure is at its annual mean.

conclusion is not certain, since other variables (such as labor inputs, farm assets, financial wealth) are likely to be poorly measured as well, and attenuation bias in any given coefficient cannot be guaranteed when multiple variables are measured with error (Garber and Klepper 1980).[23] If the OLS estimates are indeed downwardly biased by measurement error, that could explain why the IV results, which should be immune to classical measurement error, are so much larger.

Last comes the issue of selection bias. I have argued, following Duan *et al* (1984), that this is not a problem *among* farm households – the two-part model that treats observed zero outcomes as meaningful results does not need to be replaced with a selection model that treats the observed zeros as missing data. However, in my initial definition of farm households I made a number of exclusions based on incomplete data on farm production, and even if these data were perfect, one might wonder if the unobserved factors that determine *whether* one is a farmer might also determine one's expenditures on livestock, feed, seed, herbicides, and so forth – i.e. whether the error terms in the selection equation and the substantive equation are correlated. If they are, then the possibility for selection bias exists, and its sign will be the opposite of the sign of the product of the correlation in the error terms and the partial correlation of non-farm income with the probability of being a farmer, net of all other covariates in the main equation of interest (Stolzenberg and Relles 1997). The first of these two terms, the correlation in the error terms, is likely to be positive: for a given land holding, endowment of human and physical capital, and level of non-farm earnings, we would expect that the more one was willing to spend on livestock and working capital, the more likely one would be to engage in farming in the first place, and to do so in a way that generated either saleable output, or crop production whether sold or not, which are the criteria for inclusion in my sample.

The sign of the second term is harder to intuit: without an explicit model of the selection process, it is difficult to know whether non-farm income is positively or negatively partially correlated with the probability of being a farmer, after controlling for all the other variables in the model. To investigate this, I constructed a rough-and-ready selection model, based on the same variables as appear in the main equations, but allowing the land variables to enter in levels rather than logs so as not to eliminate the landless, and excluding the agricultural labor variables,

[23] The lower the reliability of measurement of a given variable, and the lower its (true) correlation with other poorly measured ones, the more likely it is that the standard assumption of attenuation bias will hold.

which are hardly exogenous to the decision to farm. The sign of the partial correlation we require can then be read from the coefficient on log non-farm income in a regression whose dependent variable is the predicted probability of sample inclusion, and whose other covariates are the same as appear in our substantive estimates. This sign was positive, which, together with the assumption of a positive correlation among error terms, implies that selection bias should be negative. In other words, my estimated positive effects of non-farm income on agricultural outlays are likely to be lower bounds.

This positive relationship may seem at odds with my argument in the preceding section that farmers are not in fact credit constrained. One way to reconcile the two is to note that financing farm expenditures out of non-farm income will generally be cheaper than borrowing: the credit constraint need not be absolute in order to see a link between cash-on-hand and agricultural expenditures, especially if there are not alternative vehicles for saving, or if the spread between the interest rates on savings versus loans is large. Another possible answer is that farmers who spent most of their careers on collective farms are not used to, or comfortable with, borrowing on an individual basis, from private sources, for agricultural purposes. Institutional inertia on both the supply and demand side of the agricultural lending market may partly explain why households borrow so little money for farming. Although a more detailed analysis of the reforms to land, credit and insurance markets that might increase agricultural investment of borrowed funds lies beyond the scope of this paper, the results of the last section suggest that attitudinal considerations must also be taken into account, and thus that institutional reforms may be slow to produce changes in behavior.

Whatever the explanation, the finding that there exists a significant elasticity between non-farm income and expenditures on farming provides support for the proposition that raising non-farm earnings (itself not an easy task) would also stimulate agriculture. Yet, as others have noted, increased off-farm employment might also draw prime-age labor away from farming. Because our estimates hold constant the amount of labor devoted to agriculture, they do not capture this potential effect of an expanding rural non-farm economy.

Given that the two main sources of non-farm income are pensions and earnings, it is tempting to ask which of these is more important for farm finance, a question which cannot easily be separated from the question of which types of households, older or younger, invest more heavily in their farms. Among those who spent money on variable inputs, and using the specification from column (4) of Table 4, I find a statistically significant elasticity with respect to earnings (of 0.09, with a p-value of 0.01), for those with positive earnings (results not shown in tables). For those receiving pension incomes, the estimated elasticity with respect to this source of income is similar (0.08) but is not statistically significant (p=0.42). The difficulty in identifying the pension effect is due to the fact that the variance of pension income is quite small, as compared to that of earnings, even though its mean is higher.[24] These results thus do not shed much light on the debate over whether public transfers *per se* serve as a source of finance for agricultural investment.[25]

However, just as the stimulation of agriculture seems a roundabout justification for the need for more non-farm employment, so too does it seem an unnecessary defense of the pension system. Bulgaria's pensions provide about 28 per cent of income received by poor households, and one-quarter of households receive 73 per cent of more of their income from pensions. They clearly play a substantial role in reducing poverty, and it is worth noting that old-age-based transfers are an inherently well-targeted intervention, since one's potential alternative sources of income dwindle rapidly with age. Moreover, because they benefit those for whom the long term and the short term are not all that different in duration, the question as to whether they generate longer run growth in agriculture would seem moot, at least to current beneficiaries.

---

[24] The variance of log pension incomes in the sample from Table 4 is 0.32; for log earnings it is 2.23.

[25] Prior evidence to this effect has been found in relation to cash transfers to the poor and elderly, as well as public works employment, in Africa (Devereux 2002), and elsewhere, implying that increased spending on social safety net programs, which are clearly effective in reducing poverty in the short term, may lead to productive investments that reduce poverty in the longer term as well. A related argument is made by Case (2001) who notes that state pensions in South Africa appear to promote poverty-alleviating investment, not in farming, but in human capital, in the form of children's health. Yet there is also evidence that safety-net spending in Hungary has little long-term impact on poverty (Ravallion, Walle, and Gautam 1995). Poverty-trap theorists argue that reducing transient, short-term poverty itself can prevent at least some families from falling into chronic, persistent poverty, not by promoting investment, but by forestalling the sell-off of productive assets in response to short-term income shocks. Carter and Barrett (2006) note that when these fire sales push the family's stock of assets below a critical threshold (which, following Lipton, they term the "Micawber Threshold") the dynamic re-accumulation of assets becomes essentially impossible.

Table 1: Annual Expenditures on Agricultural Inputs and Livestock (in $US PPP 2003)

Working Capital Expenditures	Mean	SD	Min	Max	Number of HH's	Share of HHs
Feed	452	693	8	8333	543	0.45
Seed	94	227	1	2000	470	0.39
Seeding, ploughing, digging	158	290	8	3333	413	0.34
Veterinary services	71	99	2	833	389	0.32
Fertilizers	115	255	3	3500	383	0.32
Herbicides	59	99	4	833	265	0.22
Transportation	103	100	8	500	123	0.10
Other expenditures	176	230	3	1667	104	0.09
Fuel for agriculture use	360	660	5	3333	90	0.07
Rental of land	356	971	8	7500	62	0.05
Manure	48	78	5	500	46	0.04
Hiring labour	633	1051	33	6000	40	0.03
Rental of equipment	342	439	25	1667	30	0.02
Any of above, if positive	623	1201	5	18333	907	0.75
Any of above, all households	469	1075	0	18333	1206	1.00
Investment in Livestock						
Expenditures if positive	172	460	10	6500	259	0.21
Expenditures, all households	37	223	0	6500	1206	1.00

Table 2: Summary Statistics for 1206 Agricultural Households (Costs expressed in $US PPP 2003)

Independent Variables	Mean	SD	Min	Max
Non-Farm Income & Average share from...	6167	4566	197	40500
Pensions	0.50		0	1
Employment earnings	0.42		0	1
Social transfers	0.07		0	1
Non-agricultural rental income	0.01		0	1
Area of land owned (hectares)	1.5	4.9	0.003	120
Area of land planted (hectares)	0.8	4.9	0.002	80
Household has bank deposits	0.21		0	1
Household has other financial investments	0.05		0	1
Value of consumer durables	1683	3590	0	57533
Number of major agricultural implements	0.19	0.74	0	7
Estimated FTE's of farm labor	1.1	0.7	0.3	4.0
Hours per week worked on farm	22.5	33.3	0	256
People breeding livestock, last 7 days	0.65	0.82	0	5
People breeding livestock, last 12 months	0.73	0.88	0	5
Household breeds livestock	0.71		0	1
Cluster lacks public services	0.69		0	1
Health status of head (1=excellent, 5=poor)	3.6	0.9	1	5
Age of head	60	14	20	91
Average age of all household members	53	17	12	91
Female headed household	0.20		0	1
Share of adults who are women	0.48		0	1
Mother tongue (Omitted: Bulgarian)				
Turkish	0.11		0	1
Roma	0.03		0	1
Russian/Other	0.01		0	1
Head's education (Omitted: None)				
Initial	0.12		0	1
Primary	0.40		0	1
Secondary	0.17		0	1
Vocational	0.19		0	1
Tertiary	0.10		0	1

Table 3: Income and Poverty Statistics for 1206 Agricultural Households (Costs expressed in $US PPP 2003)

	Mean	SD	Min	Max
Annual household income ($US, PPP)	7,347	5,312	238	61,578
Non-farm income	6,167	4,566	197	40,500
Farm income	1,179	2,698	1	55,278
Pension income	2,272	1,971	0	12,138
Household size	2.8	1.5	1	9
Income per capita	2,749	1,657	161	20,526
Non-farm income per capita	2,317	1,375	95	14,436
Farm income per capita	432	1,001	0.2	18,426
Pension income per capita	1,067	980	0	6,069
Share Poor (< $4.30/day)	0.18			
Share Very Poor (< $2.15/day)	0.03			
Share Ultrapoor (< $1/day)	0.01			

PPP exchange rate in 2003 was 0.60 leva per $US. Official exchange rate was 1.73 leva/$

Table 4: OLS and IV Models of Log Expenditures on Variable Inputs

	(OLS-1)	(OLS-2)	(OLS-3)	(OLS-4)	(IV-5)
Log Non-Farm Income	0.150**	0.207***	0.198***	0.138*	0.315*
(Standard Error)	(0.066)	(0.059)	(0.067)	(0.076)	(0.166)
[p-value]	[0.024]	[0.000]	[0.003]	[0.068]	[0.058]
Log Land Owned		0.056*	0.025	0.022	0.017
Log Land Owned Squared			0.006	0.007	0.007
Log Land Planted		0.379***	0.312***	0.310***	0.307***
Log Land Planted Squared			-0.010	-0.010	-0.009
Breeds Livestock		0.993***	0.785***	0.795***	0.829***
Number of Agric. Tools		0.209***	0.405***	0.391***	0.440***
Number of Agric. Tools Sq’d.			-0.048**	-0.044**	-0.052**
Has Deposits		0.002	0.000	0.002	0.003
Has Investments		0.149	0.149	0.131	0.103
Value of Durables (Sqrt.)		0.004*	0.009**	0.008*	0.006
Value of Durables			-0.0001**	-0.0001**	-0.0001**
Cluster Lacks Services		0.113	0.140	0.167	0.169
District Fixed Effects	No	Yes	Yes	Yes	Yes
Farm Labor FTEs			0.177**	0.166**	0.171**
Hours/week on farm (Sqrt.)			0.066**	0.064***	0.064***
No. Bred Livestock (Year)			0.203***	0.211***	0.194***
Quadratic Terms [p-value]	None	None	[0.022]	[0.042]	[0.017]
Demographic Terms [p-value]	None	None	[0.393]	[0.423]	[0.181]
Income Shares [p-value]	None	None	None	[0.047]	[0.078]
No Nonlinearities? [p-value]	[0.000]	[0.011]	[0.839]	[0.873]	n/a
Weak Instruments? [F-stat.]					27.6
Hansen’s J [p-value]					0.935
NonFarm Income Exog.? [p-value]					[0.263]
Observations	907	907	907	907	907
Clusters	405	405	405	405	405
R-squared	0.006	0.435	0.492	0.503	0.499

Notes: *Significant at 10%; **Significant at 5%; ***Significant at 1%; robust, clustered standard errors.

Table 5: Estimates of the Probability of Purchasing Variable Inputs

(Marginal effects: ∂Pr(Y>0)/∂x at sample mean)

	Probit-1	Probit-2	Probit-3	Probit-4	IV Probit-5
Log Non-Farm Income	0.032*	0.037**	0.011	0.011	0.057
(Standard Error)	(0.018)	(0.019)	(0.017)	(0.019)	(0.055)
[p-value]	[0.076]	[0.043]	[0.524]	[0.560]	[0.295]
Log Land Owned		-0.005	-0.008	-0.008	-0.008
Log Land Owned Squared			0.001	0.001	0.001
Log Land Planted		0.060***	0.044***	0.044***	0.044***
Log Land Planted Squared			-0.009***	-0.009***	-0.008***
Breeds Livestock		0.294***	0.179***	0.173***	0.178***
Number of Agric. Tools		0.200***	0.173**	0.192***	0.200***
Number of Agric. Tools Sq'd.			-0.017	-0.019	-0.020
Has Deposits		-0.008	0.008	0.009	0.007
Has Investments		0.054	0.044	0.043	0.043
Value of Durables (Sqrt.)		0.001	0.000	0.000	0.000
Value of Durables			0.000	0.000	0.000
Cluster Lacks Services		0.005	-0.007	-0.009	-0.004
District Fixed Effects	No	Yes	Yes	Yes	Yes
Farm Labor FTEs			0.095***	0.092***	0.089
Hours/week on farm (Sqrt.)			0.007*	0.006	0.007
No. Bred Livestock (Year only)			0.251***	0.253***	0.252
Quadratic Terms [p-value]	None	None	[0.002]	[0.003]	[0.002]
Demographic Terms [p-value]	None	None	[0.137]	[0.153]	[0.145]
Income Shares [p-value]	None	None	None	[0.971]	[0.957]
Non-Farm Income Exog.? [p-value]					[0.335]
Observations	1206	1206	1206	1206	1206
Clusters	494	494	494	494	494
Pseudo R-squared	0.003	0.287	0.354	0.358	n/a

Notes: *Significant at 10%; **Significant at 5%; ***Significant at 1%; robust clustered standard errors. See text for description of test statistics.

Table 6: OLS and IV Models of Log Expenditures on Livestock

	(OLS-1)	(OLS-2)	(OLS-3)	(OLS-4)	(OLS-4B†)	(IV-5†)
Log Non-Farm Income	0.115	0.090	0.087	0.161	0.120	0.233
(Standard Error)	(0.075)	(0.088)	(0.101)	(0.120)	(0.142)	(0.247)
[p-value]	[0.126]	[0.307]	[0.388]	[0.181]	[0.399]	[0.344]
Log Land Owned		0.045	-0.019	-0.029	-0.041	-0.037
Log Land Owned Squared			0.021	0.022	0.022	0.018
Log Land Planted		0.104*	0.051	0.066	0.061	0.081
Log Land Planted Squared			-0.018	-0.023	-0.027	-0.028**
Number of Agric. Tools		0.019	-0.086	-0.236	-0.383	-0.409**
Number of Agric. Tools Sq'd.			0.014	0.040	0.062	0.066**
Has Deposits		-0.216	-0.135	-0.178	-0.138	-0.088
Has Investments		0.014	0.017	0.153	0.254	0.278
Value of Durables (Sqrt.)		0.001	0.003	0.009	0.007	0.006
Value of Durables			0.000	0.000	0.000	0.000
Cluster Lacks Services		0.110	-0.047	-0.042	-0.055	0.008
District Fixed Effects	No	Yes	Yes	Yes	Yes	Yes
Farm Labor FTEs			0.093	0.088	0.150	0.128
Hours/week on farm (Sqrt.)			0.034	0.013	0.001	-0.007
No. Bred Livestock (Week)			0.173	0.228**	0.240**	0.255***
No. Bred Livestock (Year only)			0.059	0.144	0.125	0.142
Quadratic Terms [p-value]	None	None	[0.643]	[0.260]	[0.208]	[0.077]
Demographic Terms [p-value]	None	None	[0.072]	[0.344]	[0.292]	[0.040]
Income Shares [p-value]	None	None	None	[0.002]	[0.000]†	[0.000]†
No Nonlinearities? [p-value]	[0.139]	[0.235]	[0.304]	[0.055]	[0.186]	n/a
Weak Instruments? [F-stat.]						15.0
Hansen's J [p-value]						0.624
NonFarm Income Exog.? [p-value]						0.709
Observations	259	259	259	259	259	259
Clusters	144	144	144	144	144	Not Used‡
R-squared	0.007	0.202	0.342	0.418	0.455	0.449

Notes: *Significant at 10%; **Significant at 5%; ***Significant at 1%; robust, clustered standard errors.
†Includes squared income share terms. ‡ Too few observations to support clustered analysis.
See text for description of test statistics.

Table 7: Estimates of the Probability of Purchasing Livestock
(Marginal effects: ∂Pr(Y>0)/∂x at sample mean)

	Probit-1	Probit-2	Probit-3	Probit-4†	IV Probit-5†
Log Non-Farm Income	0.040**	0.064***	0.059***	0.075***	0.075
(Standard Error)	(0.017)	(0.019)	(0.019)	(0.023)	(0.291)
[p-value]	[0.017]	[0.001]	[0.002]	[0.001]	[0.798]
Log Land Owned		0.030***	0.006	0.006	0.028
Log Land Owned Squared			0.005	0.005	0.024
Log Land Planted		0.021**	0.026**	0.024**	0.113**
Log Land Planted Squared			-0.007**	-0.007**	0.032**
Number of Agric. Tools		-0.003	-0.028	-0.026	-0.157
Number of Agric. Tools Sq'd.			0.003	0.002	0.016
Has Deposits		-0.012	-0.007	-0.005	-0.009
Has Investments		-0.065	-0.055	-0.054	-0.291
Value of Durables (Sqrt.)		0.000	0.001	0.001	0.006
Value of Durables			0.000	0.000	0.000
Cluster Lacks Services		0.074**	0.048	0.045	0.181
District Fixed Effects	No	Yes	Yes	Yes	Yes
Farm Labor FTEs			0.019	0.025	0.120
Hours/week on farm (Sqrt.)			-0.003	-0.003	-0.021
No. Bred Livestock (Week)			0.089***	0.085***	0.401***
No. Bred Livestock (Year only)			0.051*	0.052*	0.241*
Quadratic Terms [p-value]	None	None	[0.253]	[0.228]	[0.232]
Demographic Terms [p-value]	None	None	[0.000]	[0.000]	[0.000]
Income Shares [p-value]†	None	None	None	[0.405]†	[0.538]†
Non-Farm Income Exog.? [p-value]					0.318
Observations	1206	1206	1206	1206	1206
Clusters	494	494	494	494	494
Pseudo R-squared	0.005	0.184	0.256	0.275	n/a

Notes: *Significant at 10%; **Significant at 5%; ***Significant at 1%; robust clustered standard errors.
†Includes squared income share terms.
See text for description of test statistics.

Table 8: Loans Outstanding

	All Households (N=3008)				Agricultural Households (N=1206)			
Purpose of loan	**Number of HHs**	**Pcnt**			**Number of HHs**	**Pcnt**		
Current bills	386	12.8%			133	11.0%		
Other purposes	49	1.6%			21	1.7%		
Purchase of a house	33	1.1%			9	0.7%		
Education	29	1.0%			10	0.8%		
Starting own business	25	0.8%			6	0.5%		
Building a house, villa, etc.	17	0.6%			8	0.7%		
Agricultural activities	12	0.4%			11	0.9%		
All households with any loan	532	17.7%			189	15.7%		
Lender								
Bank	409	13.6%			133	11.0%		
Personal	78	2.6%			35	2.9%		
Other	80	2.7%			31	2.6%		
Amounts ($US PPP 2003)	**N**	**Median**	**Min**	**Max**	**N**	**Median**	**Min**	**Max**
All Loans	532	3333	25	1333333	188	2000	33	45833
Agricultural Loans	12	2583	750	11667	11	2500	750	11667
Current Bills	386	2500	6	45833	132	1667	6	45833

Note: Numbers of loans by purpose or by lender add to more than the total number of households with loans, because some households had loans for/from more than one purpose/lender.

Table 9: What are your and other members of your household's plans for overcoming the financial problems your household has?

To find a better paid job	0.43
Hope for some financial help from relatives, friends	0.24
To find a job	0.19
Will claim social benefits	0.15
To start cultivating own land, breeding livestock	0.14
To start own business	0.07
Will spend savings	0.04
Will let out property, land	0.03
Will sell property, land	0.02

Note: questions were asked separately, so totals add to more than 100%.
Based on full sample of households (n=3008)

References

Alam, Asad, *et al.* 2005. *Growth, Poverty, and Inequality: Eastern Europe and the Former Soviet Union.* Washington: World Bank.

Black, Richard, Russell King, and Richmond Tiemoko. 2003. "Migration, Return and Small Enterprise Development in Ghana: A Route of Poverty?" Sussex Centre for Migration Research, Brighton UK, Working Paper No. 9.

Bound, John, *et al.* 1994. "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, 12(3):345-368.

Carter, Michael R. and Christopher B. Barrett. 2006. "The Economics of Poverty Traps and Persistent Poverty: An Asset-Based Approach." *Journal of Development Studies*, 42(2):178-199.

Case, Anne. 2001. "Does Money Protect Health Status? Evidence from South African Pensions." NBER, Working paper w8495, October.

Cox Edwards, Alejandra and Manuelita Ureta. 2003. "International Migration, Remittances, and Schooling: Evidence from El Salvador." *Journal of Development Economics*, 72(2):429-461.

de Janvry, Alain, Elisabeth Sadoulet, and Nong Zhu. 2005. "The Role of Non-Farm Incomes in Reducing Rural Poverty and Inequality in China." Department of Agricultural & Resource Economics, University of California at Berkeley, CUDARE Working Paper 1001.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore and London: Johns Hopkins University Press & World Bank.

Deb, Partha, Willard Manning, and Edward Norton. 2005. "Modeling Health Care Costs and Counts." Paper presented at the International Health Economics Association 2005 World Congress, Barcelona.

Devereux, Stephen. 2002. "Can Social Safety Nets Reduce Chronic Poverty." *Development Policy Review*, 20(5):657-675.

Duan, Naihua, Jr. Willard G. Manning, Carl M. Morris, and Joseph P. Newhouse. 1984. "Choosing between the Sample Selection Model and the Multi-Part Model." *Journal of Business & Economic Statistics*, 2(3):283-289.

Duncan, Greg J. and Donald H. Hill. 1985. "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data." *Journal of Labor Economics*, 3(4):508-532.

Durand, Jorge, William Kandel, Emilio A. Parrado, and Douglas S. Massey. 1996. "International Migration and Development in Mexican Communities." *Demography*, 33(2):249-264.

Garber, Steven and Steven Klepper. 1980. "Extending the Classical Normal Errors-in-Variables Model." *Econometrica*, 48(6):1541-1546.

Ghosh, Parikshit, Dilip Mookherjee, and Debraj Ray. 2000. "Credit Rationing in Developing Countries: An Overview of the Theory," in *A Reader in Development Economics*, edited by Dilip Mookherjee and Debraj Ray. London: Blackwell.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47:153-161.

Hertz, Tom. 2007. "Estimating Elasticities in a Two-Part Log-Log Model." Unpublished paper.

Holden, Stein, Bekele Shiferaw, and John Pender. 2004. "Non-Farm Income, Household Welfare, and Sustainable Land Management in a Less-Favoured Area in the Ethiopian Highlands." *Food Policy*, 29:369-392.

International Social Security Association. 2002. Social Security Programs Throughout the World: Europe. Online at http://www.ssa.gov/policy/docs/progdesc/ssptw/2002-2003/europe/bulgaria.pdf, as of August 2006.

Meurs, Mieke. 2001. *The Evolution of Agrarian Institutions: A Comparative Study of Post-Socialist Hungary and Bulgaria*. Ann Arbor: The University of MIchigan Press.

Mishev, Plamen and Philip Kostov. 2003. "Decision Making Pattern of Subsistence Farmers in Bulgaria," in *Subsistence Agriculture in Central and Eastern Europe: How to Break the Vicious Circle?*, edited by Steffen Abele and Klaus Frohberg, pp. 71-85: Institut für Agrarentwicklung in Mittel- und Osteuropa (IAMO).

Mullahy, John. 1998. "Much Ado About Two: Reconsidering Retransformation and the Two Part Model in Health Econometrics " NBER, Technical Working Papers 228.

Ravallion, Martin, Dominique van de Walle, and Madhur Gautam. 1995. "Testing a Social Safety Net." *Journal of Public Economics*, 57:175-199.

Roussenova, Lena and Dimiter Nenkov. 2001. "Agricultural Finance and Institutional Reforms in Bulgaria." European Institute.

Singh, Inderjit, Lyn Squire, and John Strauss, eds. 1986. *Agricultural Household Models : Extensions, Applications, and Policy*. Baltimore: Johns Hopkins University Press.

Stock, James H. and Motohiro Yogo. 2002. "Testing for Weak Instruments in Linear I.V. Regression." NBER, Technical Working Paper 284.

Stolzenberg, Ross M. and Daniel A. Relles. 1997. "Tools for Intuition About Sample Selection Bias and Its Correction." *American Sociological Review*,, 62(3):494-507.

Taylor, J. Edward, *et al.* 1996. "International Migration and Community Development." *Population Index*, 62(3):397-418.

Taylor, J. Edward, Scott Rozelle, and Alan De Brauw. 2003. "Migration and Incomes in Source Communities: A New Economics of Migration Perspective from China." *Economic Development and Cultural Change*, 52:75-101.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica*, 26(1):24-36.

Zeller, Manfred. 1994. "Determinants of Credit Rationing: A Study of Informal Lenders and Formal Credit Groups in Madagascar." Food Consumption and Nutrition Division, International Food Policy Research Institute, Discussion Paper no. 2. Washington.