5-12-05

THE RELATIONSHIP BETWEEN FARM SIZE AND PRODUCTIVITY IN CHINESE AGRICULTURE

Zhuo Chen

Post-doctoral Research Associate
University of Tennessee, Knoxville

Wallace E. Huffman

C.F. Curtiss Distinguished Professor of Agriculture
and Professor of Economics, Iowa State University

Scott Rozelle

Professor of Agriculture and Resource Economics,
University of California at Davis

Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rohde Island, July 24-27, 2005

The Relationship between Farm Size and Productivity in Chinese Agriculture

By

Zhuo Chen, Wallace E. Huffman and Scott Rozelle

Abstract: This paper examines the relationship between farm size and productivity in China's agriculture. In developing agriculture where there is a broad range of farm sizes, farm size and productivity or output per unit of land are often found to be inversely related. In China, where average farm size is small and the distribution of farm sizes is relatively compact, farm size and productivity are weakly inversely related. However, when we utilize the egalitarian principle during land allocation in China and use imputed homogenous land area rather than actual land area in the regression, the inverse relationship between farm size and productivity disappears. Hence, the strong inverse relationship that some studies have found are undoubtly due to a number of methodological problems, including the failure to account properly for land quality differences and the method of land distribution. Applying the principal agency theory, we also discuss the possibility that market inefficiency may contribute to the inverse relationship. We examine the necessity and validity (Hahn & Hausman 2002) of the instrumental estimation applied in the paper. The corresponding variance estimates are adjusted as Murphy & Topel (1985) suggested.

In the development literature, farm size, e.g., land area, and productivity, e.g., output per unit of land, seem to be inversely related. The inverse relationship is formally defined as average grain yields fall when the size of farm increases.[1] Chayanov (1926) is credited with first noticing this relationship in Russian agriculture, but Sen (1962) is believed to be the earliest modern reference on this subject. Berry and Cline (1979) reviewed the early empirical evidences on farm size and productivity and econometric issues. In American agriculture, farm size and productivity are believed to be positively related or unrelated (Huffman and Evenson 2001; Hallam 1993). However, Ahearn, Yee and Huffman (2002) show that average farm size in the U.S. is negatively related to multifactor productivity over 1960-1996.

Sen (1962) explained the inverse relationship with labor dualism, where given the same technology, small-scale farmers have lower opportunity costs of their labor than operators of large farms. Deininger and Feder (2001) applied agency theory analysis on this subject. When a farm is small and labor markets are not functioning, small-scale farms use only family labor (Taylor and Adelman 2003). Hence, in the terminology of principal-agent theory, the principal and his family members supply all of the labor for the farm. These family members have a strong incentive to work because they share the farm output directly and in the long run can expect to inherit the farm. Here monitoring and incentive problems are minimal, and excess family labor would push the value of the marginal product below the off-farm wage thus may result the inverse relationship. Bhalla and Roy (1988) and Benjamin (1995) suggested that unobserved land quality is positively related to farm productivity but inversely related to farm size, which might explain the inverse relationship between farm size and productivity as well. Heltberg (1998) claimed that Bhalla and Roy's conclusions are undermined by their use of district aggregate data. However, using farm level data obtained in Haryana, India, Carter (1984)

---

[1] Benjamin (1995) regressed output on farm size and claimed that there exists inverse relationship if the coefficient of farm size is less than one.

found a significant within-village inverse relationship between farm size and productivity. Heltberg also noted that Benjamin (1995)'s first-stage regression has a very low R-square (0.12-0.14) and suggested that weak instruments undermined Benjamin's analysis (Bound et al. 1995).

For China's agriculture, few studies exist on the inverse relationship between farm size and productivity. Brandt (1985) briefly reviewed the relationships between farm size, productivity, and factor markets for the pre-war (1930s) northeastern China. Later, Benjamin and Brandt (2002) attributed the inverse relationship between size and productivity in China's agriculture[2] to local administrative land distribution policies and uneven off-farm work opportunities. Using a panel of farm households during the late 1990s, this paper examines the relationship between farm size and productivity in China's agriculture, where average farm size is small and technology is only slightly dynamic,. First, we utilize the fact that an egalitarian land distribution policy exists under the Household Responsibility System. We then develop some instruments for farm size and apply a Hahn-Hausman specification test for the instruments. Second, Murphy and Topel (1985) claimed that the traditional form of the covariance matrix estimator for a two-step estimator is biased. Using their methodology, we identify four different approaches for obtaining variance estimates of the instrumental variable estimate.

The rest of the paper is organized as follows: In the second section, we describe the data and examine the relationship between farm size and crop yields for Chinese grain farms. In the third section, we discuss the likely effects of unobserved land quality and argue that a two-step estimation procedure should be applied. In the fourth section, we perform the second-stage analysis and test the null hypothesis of no inverse relationship between farm size and productivity. We also derive the MT-type variance estimators. In Section 5, conclusions are presented.

---

[2] The areas they surveyed are in the geographic area same as or close to these of Brandt (1985).

The Data Set and Evidence on Size and Productivity

The Dataset

The data for our study are a sample from a large comprehensive survey of Chinese rural households conducted by the Research Center for Rural Economy (RCRE). This survey started in 1986, covered 29 provinces, and included about 20,000 households. Sample attrition has been small. The survey was temporarily discontinued in 1992 and 1994 for financial reasons. The data set for this study consists of 591 farm households randomly drawn from the large survey and observed in 1995 to 1999. They are from 29 villages and 9 provinces and randomly selected from the original dataset.

Sampling for the original data set was conducted by provincial offices under the Ministry of Agriculture (see Benjamin, Brandt, and Giles 2001). Each provincial research office first selected equal numbers of three types of counties: upper, middle and lower income; then they chose a representative village in each county. Forty to 120 households were randomly surveyed within each village. Village officers and accountants filled out a survey form on general village characteristics every year.

RCRE claimed that 80 percent of the households, which enrolled in 1986, remained in 1999. By comparing the characteristics of those households, Chen (2001) found several new households (accounted for less than one percent of the whole sample) used id of an old household. He assigned new ids to these households. Chen (2001) converted all monetary variables such as prices, income, and expenditures into real terms with 1986 as base year.

Summary statistics of our sample are presented in Table 1. Clearly, per capita or per household land area of rural China is very small compared to these of developed countries. The household size ranges from 1 to 11, with an average of about four members per family. Household labor ranges from 1 to 8 persons with an average of roughly 2.6 persons per household. Household agricultural productive asset ranges from RMB Yuan 0 to 65,000 with an average of RMB Yuan 1,567. The index illustrates the

extent of household specialization in grain production. About 5 percent of the households actually have at least one member holding a position as a village officer.

The Relationship between Farm Size and Productivity

Benjamin (1995) and Heltberg (1998) considered a simple regression of logarithm of grain output on logarithm of sown area:

$$\ln q_i = \alpha_0 + \gamma \ln l_i + \eta_i \tag{1}$$

where q is the real value of grain output, l is land area in grain production, η is the random disturbance. Note that if γ is equal to 1 in equation (1) then output per unit land is unrelated to the farm size. If γ is less than one, then grain yield per unit land declines as land area increases. Finally, if γ is larger than one, then grain yield per unit of land increases with land area. Benjamin (1995) argued that using the actual area harvested rather than total farm size reduces measurement error, which otherwise could introduce a spurious inverse relationship.

We fit this simple model to our dataset and obtain the estimate of γ as 0.890 with a standard error of 0.011 (see Table 2, column 1). Hence, we reject the null hypothesis that γ is one at the five percent significance level and conclude that output per unit of land declines as farm size increases. Excluded factors—climate, regional effects, population density, average land quality-- however, could bias the size of γ (e.g., downward as Bhalla and Roy (1988) showed). Village level dummy variables can be used to control for differences due to climate, soil quality, multiple cropping indexes and the regional irrigation systems. Benjamin (1995) used a dummy variable for use of HYV (high yield varieties) versus traditional varieties.

In our dataset, grain output is measured as the aggregation value of wheat, rice, corn, and soya harvested.[3] We observe that the wheat and corn have quite similar prices and

---

[3] Pooling grain productions together is better than estimating the crops separately since households may produce a small amount of the crops that is not best suited for the local soil and weather just to add some variety to their menu. When estimating the relationship between productivity and land size, this may result in a pseudo positive relationship.

yields per unit of land. Therefore we use dummy variables to indicate whether soya (rice) was planted to proxy the composition of grain output.

Cheng (1998) incorporated an indicator for a family member holding a "village official" position into the grain production function and found it is positive and statistically significant. He argued that the effects were most likely due to the local policy followed by collective ownership of large farm equipment and privileged access to state subsidized farm inputs.

Consider the modified models:

$$\ln q_i = \alpha_0 + \gamma \ln l_i + \beta_{D_{rice}} D_{rice,i} + \beta_{D_{soya}} D_{soya,i} + \beta_{D_{vo}} D_{vo,i} + \eta_i \qquad (2)$$

$$\ln q_i = \alpha_0 + \gamma \ln l_i + \beta_{D_{rice}} D_{rice,i} + \beta_{D_{soya}} D_{soya,i} + \beta_{D_{vo}} D_{vo,i} + \sum_v \beta_{D_v} D_{v,i} + \eta_i \quad (3)$$

$$\ln q_i = \alpha_0 + \gamma \ln l_i + \beta_{D_{rice}} D_{rice,i} + \beta_{D_{soya}} D_{soya,i} + \beta_{D_{vo}} D_{vo,i} + \sum_{hh} \beta_{D_{hh}} D_{hh,i} + \eta_i \qquad (4)$$

where D_{soya} takes a value of one if the household harvested soya (zero otherwise); D_{rice} is one if the household harvests rice (zero otherwise). D_{vo} is one if a member of the farm household held a position as a village officer (zero otherwise). D_v is one for region v (and zero otherwise, we use the first village as the reference group)[4], D_{hh} has a value of 1 for household h (and zero otherwise, note we use the first household as the reference group).

When equation (2) is fitted to our dataset, the estimate of γ is 0.93 with a standard error of 0.01 (see Table 2). Hence, the estimate of γ is significantly different from one, which supports the so-called inverse relationship. When we include village fixed effects as in equation (3), the estimate of γ is 0.92 and but still significantly different from one at the 5 percent level. Thus, by expanding the model from equation (1) to (2), and then (3), the size of the estimate of γ changes slightly--by only 0.03 and the R-square increases from 0.73 to 0.89. Since our dataset is a five-year panel, individual unobserved effects might exist. Equation (4), a household fixed effects model, can account for the unobserved effects. When equation (4) is fitted to the data, the estimate of γ is 0.84 with a

---

[4] Note that we use the first region (or anyone) as the reference group.

standard error of 0.05 (see Table 2). Hence, we reject the null hypothesis that γ is equal to one and conclude that output per unit of land declines as land area increases. The *R*-squared for this regression is 0.92. This relationship exists even after the introduction of village or household effects, which is consistent with Carter (1984). We, however, have not controlled for land quality differences.

A Role for Land Quality Differences

Land heterogeneity within Chinese villages may contribute to the inverse relationship. Benjamin (1995) presented the following scenario for rural Java: "If farms were subdivided through inheritance over time, egalitarian motives on the part of the benefactor would result in higher quality parcels being divided more often than low quality parcels", which would impart a negative correlation between farm size and land quality, particularly at the local or village level.

Under the Household Responsibility System, the majority of the arable land in rural China is owned by rural communities and managed by the village council (Agriculture Law 1993). Yao (2001) noted that the politics in rural China is characterized by a mixture of an authoritarian command system and grass-roots democracy, which establishes the legality of the egalitarian principle during land distribution and ensures its implementation. Furthermore, a recently passed law, Law Rural Land Contracting[5] (2002, LRLC henceforth), explicitly stated the egalitarian principle. It stresses that women should have the same right as men during land distribution (Article 6, LRLC) and forbids local governments to void the HRS contract by specifying that no reallocation is allowed during the term of the contract. Exceptions need to be ratified by at least two thirds of the village council (Article 48, LRLC).

Studies confirmed that the egalitarian principle has been adopted for land distribution in the majority of Chinese rural communities. Yao (2001) studied the effects of

---

[5] It was passed in August 2002 and has been effective since March 1, 2003.

egalitarian land distribution on migration of rural men in China. Using household level data for two distinct provinces, Jiangsu and Sichuan, Burgess (2001) failed to reject the hypothesis of universal and egalitarian access to land. With an egalitarian principle in place, it is likely that average land quality and farm size are negatively correlated.

Estimating Land Elasticity with Unobserved Land Quality

If distributing equal amount of "effective land units" or quality adjusted land area is the objective of the local public land distribution system, then we can write:

$$l_i = a_i h_i \tag{5}$$

where l is homogenous land or effective units of land, h is average land quality and a is nominal land areas. For simplicity, impose the following normalization on h, $\prod_{i=1}^{n} h_i = 1$. We can write equation (4) as:

$$\ln l_i = \ln a_i + \ln h_i. \tag{6}$$

Now if the "true relationship with productivity" is:

$$\ln y_i = \alpha_L + \gamma_L(\ln a_i + \ln q_i) + \eta_{Li} \tag{7}$$

then when one fits equation (1) the estimate of γ will obviously be different from γ_L. In particular, if note the n×1 vector ln$a$ as A and ln$h$ as H, we have:

$$\hat{\gamma} = \gamma_L + \rho \widehat{\gamma_L} \text{ while } \rho = (A'A)^{-1} A'H \tag{8}$$

It is possible that $\gamma_L > 1$ but $\hat{\gamma} < 1$ when a and h are negatively correlated, i.e. $\rho < 0$, which could occur under an egalitarian policy of local land distribution.

Benjamin (1995) used an instrumental variable approach to explain the inverse relationship with unobserved land quality. Under local egalitarian land distribution, each household is presumably allocated certain amount of land according to attributes of the household, e.g. household size, number of workers, and the extent of farm specialization in grain production. Denote the logarithm of those variables as X and assume:

$p\lim\frac{1}{n}\{X'X\} = Q_{XX}$, which is a finite, positive definite matrix,   (A1)

$p\lim\frac{1}{n}\{X'A\} = Q_{XA}$, is a finite matrix, and   (A2)

$p\lim\frac{1}{n}\{X'(\eta - H)\} = 0$.   (A3)

A1 and A2 are quite straightforward and A3 implies that: $p\lim\frac{1}{n}\{X'H\} = 0$, which means that X is uncorrelated with H, i.e., the household's characteristics are not correlated with unobserved land quality.

    Now under the egalitarian land distribution policy assume that:

$$\ln(l)_i = \mathbf{X}_i\boldsymbol{\theta} + \eta_i \tag{9}$$

but due to the unobservable nature of effective land units, we fit:

$$\ln(a)_i = \mathbf{X}_i\boldsymbol{\theta} + \eta_{Ai} \tag{10}$$

Hence, we have: $\widehat{\mathbf{A}} = \mathbf{X}_i\boldsymbol{\theta} + \mathbf{P}_X(\boldsymbol{\eta} - \mathbf{H})$

by A3, we obtain: $p\lim \mathbf{P}_X(\boldsymbol{\eta} - \mathbf{H}) = 0$

thus, $\widehat{\mathbf{A}}$ converges in probability to E(L), which is the linear best predictor of effective or constant quality land area.

Two-Step Estimation of the Land-Productivity Relationship

We call our estimation as two-step estimation instead of two-stage estimation or instrumental estimation since the former is more general, i.e., the first step in two-step estimation can be estimated with maximum likelihood estimation. However, if only linear estimation is considered and all exogenous variables in the second step are included in the first step as well, two-step estimation, two-stage estimation and instrumental estimation are in essence the same thing. In the terminology of instrumental estimation, there are four procedures we need to go through: searching for appropriate instruments, judging whether we need instrumental estimation instead of ordinary linear regression,

examining the validity of the instruments, making appropriate inference. We discuss the details of these issues in the following.

Searching for Appropriate Instruments

Earlier studies by Chen and Brown (2001), Lin (1988), as well as the Law of Rural Land Contracting (2002), show that rural land distribution in China is roughly as follows: First, "subsistence land" is distributed according to a assessment of a household's nutrient needs (Burgess 2001), which can be characterized by the number of people in the household. Second, the remaining land is classified as "contract land", which are auctioned off at meetings of the local village council.

A farmer that makes the highest bid wins the parcel of contract land and pays the village council in cash or grain output. The size of a household's labor force is an important factor in determining how much a farmer bids for contract land (e.g. Burgess 2001). Some rural Chinese households lease additional land (either from the community where land is relatively abundant or from these households that cannot fully operate the farmland being allocated) to specialize in cropping. Investment on agricultural instruments and machinery is a good indicator of the household's intention and ability to farm additional land. Burgess (2001) included a "cadre" dummy variable in his regression of land allocation. Chen and Rozelle (1999) showed that when a farm household had a member that serves as a local village officer might be benefited during the land distribution land.

Village dummy variable can be used to account for geographical differences in land quality.[6] They capture not only the difference of regional soil, weather, irrigation system, and multi-cropping index as Bhalla and Roy (1988) pointed out, but also the special

---

[6] Note by doing so, we assume that different villages have the same coefficient estimates. This is supported by the result of Burgess (2001) where his land allocation function estimates are quite similar for Sichuan and Jiangsu province. We refrain from regressing separately for each village for fear of the loss of large sample properties.

features of China's rural land distribution system[7]. Krusekopf (2002) demonstrated that tremendous heterogeneity exists in land tenure policies at the local level in rural China. Yao (2002) classified land tenure systems in rural China into six categories and suggested there is significant regional heterogeneity in the local land distribution systems. For example, some villages prefer to reallocate farmland using land rental markets while others prefer to reallocate through meetings of the local village council.

We also added dummy variables indicating whether the household had rice output, whether the household had soya output, and whether the household focused on agricultural production or not (household type).

Consider the first stage regression explaining nominal land area:

$$\ln a_i = \theta_0 + \theta_1 \ln p_i + \theta_2 \ln f_i + \theta_3 \ln k_i + \theta_4 D_{hht,i} + \theta_5 D_{rice,i} + \theta_6 D_{soya,i} + \theta_7 D_{vo,i} + \sum_v \theta_{D_v} D_{v,i} + \eta_i \text{- -(11)}$$

where a is the land area for grain production; p is the number of household members; f is the number of household labor at the current year, and k is the value of agricultural equipment and machinery. D_v and D_{vo} are the village and village officer dummy variables defined in the usual way. The θs are unknown parameters to be estimated.

The least-squares estimate of equation (11) is reported in Table 3. The most notable feature of this equation is that the R-square is 0.79, and the partial R-square is 0.38, which means that the instruments are not "weak" (Bound et al 1995; Heltberg 1998).

The equation we need to estimate in the second stage is:

$$\ln q_i = \beta_0 + \beta_1 \widehat{\ln a}_i + \beta_{D_{soya}} D_{soya,i} + \beta_{D_{rice}} D_{rice,i} + \beta_{D_{hht}} D_{hht,i} + \beta_{D_{vo}} D_{vo,i} + \sum \beta_{D_v} D_{v,i} + u_i \quad (12)$$

where q is grain output in real terms, $\widehat{\ln a}$ is the predicted logarithm of land area from equation (11), D_v and D_{vo} are defined as above, u is a random disturbance. The least-squares estimate of equation (12) is reported in Table 4.

---

[7] Burgess (2001): Village dummies are included to control for across village variation in unobservables which may affect the form of the land allocation rule. Village land quality and parameters of the contractual environment (e.g. grain quotas, land rent rates, land tax rates) can all be absorbed in this manner.

Endogeneity Test

Hausman (1978) provided a specification test to examine whether the OLS estimate of a parameter is a consistent estimator. The test statistics is:

$$H = (b_{IV} - b_{OLS})'\{Est.Asy.Var(b_{IV}) - Est.Asy.Var(b_{OLS})\}^{-1}(b_{IV} - b_{OLS})$$
$$= \frac{(b_{IV} - b_{OLS})'[(\widehat{X}'\widehat{X})^{-1} - (X'X)^{-1}]^{-1}(b_{IV} - b_{OLS})}{s^2} \tag{13}$$

where $\widehat{X}$ is the predicted value of X (i.e., least-squares predictions of explanatory variables) from the regression that might need an instrument, s^2 is a usual estimate of σ^2 under the null hypothesis. The test statistics H has an asymptotic chi-squared distribution with degrees of freedom being the number of regressors minus the number of instruments in the second stage, which is one in this case. For this study, the Hausman test statistics is 31.33 and the critical value of the test statistic is 3.84. Hence, the OLS estimator is not consistent and we need to use instrumental estimation.

Examining Instruments

Are the instruments appropriate? There are two aspects of this appropriateness. First, does the orthogonality condition between the instruments and the error term hold? An over-identification test can be used for this purpose (Ruud 2000, p573). The over-identification test statistics for this dataset is 3.17, which is distributed as chi-square with degree freedom 4. Hence, we cannot reject the null hypothesis that the orthogonality condition holds at 5% significance level.

   Second, are the instruments sufficiently correlated with the stochastic regressor *l*? Much discussions about "weak instruments" has emerged over the past two decade, e.g., Staiger and Stock (1997) and Nelson and Startz (1990). Weak instruments may make the second stage inference invalid. Bound, Jaeger, and Baker (1995) suggested that partial *R*-squared and the *F* statistics of the identifying instruments in the first stage estimation are useful indicators of the quality of instruments.

Consider the recursive model:

$$y_1 = \beta y_2 + \varepsilon_1 \tag{14}$$

$$y_2 = z\pi_2 + v_2 \tag{15}$$

It is straightforward to show that:

$$p \lim \widehat{\beta}_{OLS} = \beta + \frac{\sigma_{y_2,\varepsilon_1}}{\sigma_{y_2}^2} \quad \text{and} \quad p \lim \widehat{\beta}_{IV} = \beta + \frac{\sigma_{\hat{y}_2,\varepsilon_1}}{\sigma_{\hat{y}_2}^2} \tag{16}$$

which imply that:

$$\frac{p \lim \widehat{\beta}_{IV} - \beta}{p \lim \widehat{\beta}_{OLS} - \beta} = \frac{\sigma_{\hat{y}_2,\varepsilon_1} / \sigma_{y_2,\varepsilon_1}}{R^2_{y_2,Z}}, \tag{17}$$

where $R^2_{y2,Z}$ is the R-square (the partial R-square if there are included exogenous variables) from the regression of y_2 on z. Obviously, as R-square increases, given a particular data set, the bias becomes smaller. The partial R-square for this model is 0.38, which suggests the instruments have reasonable explanatory power for effective land area.[8]

Hahn and Hausman (2002) proposed a specification test to determine whether conventional IV asymptotics are reliable. The test compares the difference of the

---

[8] Hall, Rudebusch, and Wilcox (1996) indicated that any relevance measure probably has little practical merit, as its use may actually exacerbate the poor finite-sample properties of the IV estimator. Their result is possibly caused by the inclusion of instruments that is Granger caused by the stochastic regressor. For the case of unobserved land heterogeneity, the objective of the first stage is to predict a homogenous measure that is free of land quality problem. As Davidson and Mackinnon (1993) pointed out, when the goal is forecasting, forecasts of the variable Y_t may be conditional on the variables X_t if Y_{t-1} does not Granger cause X_t. If the number of instruments is much less than the number of observations, including more instruments will increase the precision of prediction thus reduces the bias in the second stage. However, in finite samples, to include more instruments has two types of danger. First, it is likely to include some instruments that are Granger caused by the stochastic regressor and thus introduce correlation between the predicted values with the unobserved latent variables/random disturbance. These instruments cannot eliminate the problem of the correlation between regressor and the error term though they may have a good fit at the first stage. Second, including more instruments is at risk of constructing a linear space that is not orthogonal to the space spanned by the unobserved latent variables. An extreme example is that when we have N instruments for a dataset of N observations. If the data matrix of the N instruments is not singular, the predicted value of the stochastic regressor will be exactly same as the observed values, which means that the R-square is one and first stage random disturbances are zero. Obviously, the predicted value is still correlated with the error term in the second stage. In this paper, however, we deem this danger less realistic since the first stage regression is based on our observation of China's agriculture.

conventional 2SLS estimate of the coefficient of the right-hand side endogenous variable with the reverse 2SLS estimate of the same unknown parameter. Under the null hypothesis that the conventional first-order asymptotics provide a reliable guide to inference, the two estimates should be very similar. The Hahn-Hausman specification test shows whether the resulting difference in the two estimates satisfies the results of second-order asymptotic theory. The test statistic[9] is:

$$m_1 = \frac{\hat{d}_1}{\hat{w}_1^{0.5}}, \text{ where } m_1 = \sqrt{n}(b_{2SLS} - \frac{1}{c_{2SLS}} - \hat{B}), \text{ and} \qquad (18)$$

$$\hat{w}_1 = 2\frac{K-1}{n-K}\frac{(\sum_{i=1}^{n}(y_{1i} - \beta_{LIML}y_{2i})^2)^2(y_2'P_zy_2 - \frac{K-1}{n-K}y_2'M_zy_2)^2}{(y_2'P_zy_2)^2(y_2'P_zy_1)^2}. \qquad (19)$$

The null hypothesis is:

$$H_0 : p\lim\sqrt{n}(b_{2SLS} - \frac{1}{c_{2SLS}} - \hat{B}) = 0$$

Hahn and Hausman (2002) proved the test statistics has an asymptotic t distribution. The Hahn-Hausman test statistics for this data set is 0.35. Hence we cannot reject the null hypothesis that the 2SLS provides reliable inference[10].

Estimates of the Variance of the Instrumental Variable Estimators

As Ruud (2000) pointed out, the variance estimate of the traditional IV estimator ignores the fact that parameters were estimated in the first step. These parameters are treated as constants, not random variables. This may produce a biased variance estimate in the second step.

---

[9] Please see Hahn and Hausman (2002) for details of the test.

[10] Hahn and Hausman (2002) suggested that when the null hypothesis is rejected, a similar test based on Nagar-type estimators should be performed. If the second specification test rejects or the two Nagar-type estimators differ substantially, neither 2SLS nor LIML may provide reliable results for inference. If the null hypothesis is not rejected in the second test while the first test rejects, LIML estimator is preferred over 2SLS estimator. We only performed the first specification test. Please refer to Hahn and Hausman (2002) for details on the second specification test.

The existing literature has two versions of the instrumental variable estimator. Staiger and Stock (1997) and others assumed that:

$$y_1 = \beta y_2 + \varepsilon_1$$

$$y_2 = z\pi_2 + v_2$$

where y_2 is the stochastic regressor and z is the instruments. We do not consider the case where included exogenous variables exist in the first equation since we can invoke Frisch-Waugh theorem to partial them out. This model uses all the exogenous variables as instruments. However, we may use more generalized two-step procedure where we do not need to include all exogenous variables in the first step and even the linear assumption in the first step can be relaxed.

Hahn and Hausman (2002) and Murphy and Topel (1985) specified that:

$$y_1 = \beta y_2 + \varepsilon_1 = \beta(z\pi_2 + v_2) + \varepsilon_1 = \gamma z\pi_2 + v_1$$

$$y_2 = z\pi_2 + v_2$$

Note that although numerically identical β and γ have different interpretations. β is the coefficient for the observed value y_2 while γ is the coefficient for the unobserved $z\pi_2$. An example is the case of quality adjusted land area. The instrumental estimates of the coefficient of the observed land measure and the coefficient of the imputed homogenous land have same point estimates when all second stage exogenous variables are included in the first stage regression. They differ if there are instruments excluded in the second step. In either case, their variance estimates are different as is shown in the follows.

Traditional variance estimates for the IV estimator are obtained as following:

$$\mathrm{var}(\hat{\beta}_{IV}) = \mathrm{var}\left(\left(\hat{y}_2' y_2\right)^{-1} \hat{y}_2' y_1\right) = \mathrm{var}\left(\left(\hat{y}_2' y_2\right)^{-1} \hat{y}_2'(\beta y_2 + \varepsilon_1)\right)$$

$$= \mathrm{var}\left(\left(\hat{y}_2' y_2\right)^{-1} \hat{y}_2' \varepsilon_1\right) = \left(\hat{y}_2' y_2\right)^{-1} \hat{y}_2' \mathrm{var}(\varepsilon_1)\hat{y}_2 \left(\hat{y}_2' y_2\right)^{-1}$$

when assuming that no heteroscedasticity occurs, we have:

$$\text{var}(\widehat{\beta}_{IV}) = \sigma_1^2 \left(\widehat{y}_2' y_2\right)^{-1} = \sigma_1^2 \left(\widehat{y}_2' \widehat{y}_2\right)^{-1} \tag{20}$$

The commonly used estimator for σ_1^2 is:

$$\widehat{\sigma_1^2} = \tfrac{1}{n-k}(y_1 - \widehat{\beta}_{IV} y_2)'(y_1 - \widehat{\beta}_{IV} y_2) \tag{21}$$

However, if our main interest is to obtain the variance estimate of γ, which is coefficient of the unobserved $z\pi_2$, we need to proceed in another direction.

$$
\begin{aligned}
\text{var}(\widehat{\gamma}_{2SE}) &= \text{var}\left(\left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2' y_1\right) = \text{var}\left(\left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2'(\gamma z\pi_2 + v_1)\right) \\
&= \text{var}\left(\left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2'(\widehat{y}_2\gamma - \gamma z(\widehat{\pi}_2 - \pi_2) + v_1)\right) \\
&= \text{var}\left(-\left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2'\gamma z(\widehat{\pi}_2 - \pi_2) + \left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2' v_1\right) \\
&= \gamma^2 \left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2' z \, \text{var}(\widehat{\pi}_2 - \pi_2) z' \widehat{y}_2 \left(\widehat{y}_2' y_2\right)^{-1} + \left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2' \, \text{var}(v_1)\widehat{y}_2 \left(\widehat{y}_2' y_2\right)^{-1} \\
&\quad -2\left(\widehat{y}_2' y_2\right)^{-1} \widehat{y}_2'\gamma z \, \text{cov}(\widehat{\pi}_2 - \pi_2, v_1)\widehat{y}_2 \left(\widehat{y}_2' y_2\right)^{-1}
\end{aligned}
\tag{22}
$$

Obviously, equations (20) and (22) are different as Murphy and Topel (1985) noted. Although they may be asymptotically equivalent, if the Newey condition holds, the difference in finite samples cannot be ignored when making inferences regarding the coefficient of the stochastic regressor (Ruud 2000).

The difference between the two variance estimates is partly due to the fact that the estimate of the variance of the IV estimator omits the correlation between y_2 and the error term. It ignores the fact that parameters were estimated in the first step, which may cause mis-specification of the sampling variance of the second stage estimator (Ruud 2000).

To extend the analysis to the estimation of the variance when there are included exogenous variables in the second stage, we use the following matrix form:

Stage 1: $L = X\theta + \varepsilon$ (23)

Stage 2: $Y = X_2\beta + \gamma X\theta + u$ (24)

Stage 2': $Y = X_2\beta + \gamma L + u^*$ (25)

where ε, u and u^* are random disturbances.

The two-step estimator[11] is:

$$(\hat{\beta}',\hat{\gamma})'_{2SE} = (Z'Z)^{-1}Z'(X_2\beta + \gamma X\hat{\theta} - \gamma X(\hat{\theta}-\theta)+u)$$
$$= (Z'Z)^{-1}Z'Z(\beta',\gamma)' - \gamma(Z'Z)^{-1}Z'X(\hat{\theta}-\theta)+(Z'Z)^{-1}Z'u \quad (26)$$
$$= (\beta',\gamma)' - \gamma(Z'Z)^{-1}Z'X(\hat{\theta}-\theta)+(Z'Z)^{-1}Z'u$$

where $Z = (X_2 \mid P_X L)$

The usual instrumental estimator (or two-stage estimator) usually includes all possible instruments thus gives the estimator for the parameters as:

$$(\hat{\beta}',\hat{\gamma})'_{IV} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}(X_2\beta + \gamma L + u^*)$$
$$= (\beta',\gamma)' + (\hat{Z}'\hat{Z})^{-1}\hat{Z}u^* \quad (27)$$

where $\hat{Z} = P_X(X_2 \mid L)$.

From the above expressions, including $L = X\theta + \eta$, we obtain $u^* = u - \gamma\varepsilon$. The difference between the estimate of the variance of the two-step estimator and instrumental variable estimator is:

$$(\hat{\beta}',\hat{\gamma})'_{2SE} - (\hat{\beta}',\hat{\gamma})'_{IV} = ((Z'Z)^{-1}Z' - (\hat{Z}'\hat{Z})^{-1}\hat{Z}')u - \gamma((Z'Z)^{-1}\hat{Z}' - (\hat{Z}'\hat{Z})^{-1}\hat{Z}')\varepsilon \quad (28)$$

Hence, the following proposition holds:

Proposition 1: $(\hat{\beta}',\hat{\gamma})'_{IV}$ equals to $(\hat{\beta}',\hat{\gamma})'_{2SE}$ if and only if X_2 is included in X.

The proof of this proposition is straightforward from the derivation. Two-step estimation is more general than instrumental estimation and two-stage estimation is a special case (most efficient without presence of heteroskedasticity) of instrumental estimation. However, when and only when the condition in Proposition 1 holds, the three methods produce same result.

---

[11] The example can be treated as a special case of Murphy and Topel (1985) Theorem 1. Here we specified a linear functional form in the first step.

Variance estimates for the two-step estimation and instrumental variable estimator proceeds in two different directions. The variance estimate of the instrumental estimator tries to construct a consistent estimator for u^* while the variance of the two-stage least squares estimator can be decomposed into two parts. Murphy and Topel (1985) argued that although the instrumental estimator (or two-stage least squares) yields a consistent estimator for second-stage parameters under fairly general conditions, the second-step standard errors and related test statistics based on this procedure are incorrect.

Previous econometric studies have regularly based inference on the traditional variance estimator of the instrumental variable estimator. For example, Benjamin (1995) used the robust standard error estimator to correct for arbitrary heterogeneity.

In our model, it is better to use equation (24) rather than equation (25) as the structural form. Recall that we are testing the null hypothesis that output per unit of land area is unrelated to size (land area) with the alternative hypothesis being that the output per unit land is inversely related to farm size. We are more concerned with the effect of effective land area rather than measured land area. The estimate of γ may be biased and inconsistent if measured land area rather than effective land area is used as the regressor.

We derive four estimators of the variance of the parameters in the second stage. Due to their tie to Murphy and Topel (1985), we label them as the MT variance estimates and present them in Table 4. Note that the adjusted standard error is much greater than the robust error given by 2SLS procedures. Therefore, using smaller variance estimates, we are inclined to falsely conclude that there exists an inverse relationship given the land elasticity estimate is less than one. We also find that the adjusted standard error with assumption of no correlation between the random components is identical (after rounding) to the adjusted estimates with correlation, which indicates that the random components are nearly independent, i.e., the random disturbance in the land distribution equation is independent of the random disturbance of grain output.

Certainly, "outliers" may spoil the sample. Hence, we limited our sample to

households that have land areas less than or equal to 25 Mu (95% percentile) in another set of regressions. The results are summarized in Table 6 and reveal a similar pattern. We also split the data into two samples and fit the basic model. One of the samples is the group with farmland under grain production greater than or equal to 15 Mu and the other less. We found that the inverse relationship in the sample with larger land holdings is less severe. In fact the coefficient estimates is greater than one though not statistically significant. Land elasticity estimates for the group with larger land endowment are either similar or greater than those of the group with less land, though not statistically significant.

Based on these results, we cannot reject the hypothesis that the coefficient of imputed land area is one. Hence, the empirical irregularity of the inverse relationship between crop yields and farm size (land area) diminishes when we use the instrumental estimator with adjusted standard error.

Discussion

Deininger and Feder (2001) summarized several studies that confirmed the inverse relationship between farm-size and productivity. They argue that supervision cost for hired labor that comes with a larger farm is particularly large in agricultural production due to spatial dispersion and thus contributed to the inverse relationship. This could be interpreted as one reason why China's agriculture was transformed from collective farming to household responsibility system in the 1980s. Microeconomics theory suggests there is an optimum size for most production processes. Empirical evidence as summarized by Deininger and Feder (2001) indicate that the optimum farm size in most developing countries, given the existence of imperfect input/output/credit markets, low real wage, static agricultural technologies, and land heterogeneities, is small relative to the optimal size of farm in high wage, technically dynamic developed countries.

Case of China

In China, we see a complicated picture. First, China has a very large rural population relative to the amount of arable land. The arable land per rural person is about 0.144 hectare for China, for India it is 0.221, the U.S. 2.729, and the world average 0.426. Second, China is different from most other developing countries in that land is collectively owned by the rural community instead of individuals. No large private-owned farms exist in China. One would expect large farms to be "specialized" and have subleased land from the community or other households. Both communities and households are more likely to sublease their less-productive plots. This contributes to the spurious inverse relationship between land productivity and farm size. Third, eastern and southern China has seen an economic boost in last few decades, and a large number of rural labors are now engaged in off-farm activities. The "land bank"'s in Zhejiang province functioned as a rental market and successfully transferred lands from those households that are less relied on farming to these that are more "specialized" on farming. Similar institutional arrangement emerged in Jiangsu, Anhui, and Hunan, where rural laborers move out of agriculture sector to take local off-farm business or migrate.

The rapid economic development in China in the 1990s may have improved the function of input and output markets and most likely contributes to the weakening of inverse relationship between farm size and productivity. As China's agriculture is mechanized and the input sector started to produce a steady stream of new technologies, larger farms will have a comparative advantage over small farms. Our results show that land heterogeneities contribute to the observed inverse relationship. This, as well as other studies (Benjamin 1995; Carter, 1984; Deininger and Feder 2001), points to an important conclusion: the inverse relationship between farm-size and output per unit of land is not inherent to developing countries, but rather a consequence of heterogeneous land, (labor) market imperfection, and unobserved factors. Therefore, a public policy of breaking up large farms is not justified (Deininger and Feder 2001). The hidden unemployment

problem can be improved by general economic development and investments in rural education (Huffman 2001; Huffman and Orazem 2002). A mechanism that consolidates land to exploit the benefits of more advanced technology and to share such benefits between landowners and farmers is needed. The "land bank" in Zhejiang province is a result of efforts seeking such mechanism or an instrument of such mechanism. Further investigation and research are needed to judge whether these efforts are successful.

Alternative Perspective: Principal Agent Theory Interpretation

Deininger and Feder (2001) used agency theory to explain the inverse relationship between farm size and productivity. When the farm has access to more land, family labor may not provide enough labor to equilibrate the value of the marginal product of farm labor to the off-farm wage. When labor markets are functioning, the farmer may hire outside workers. This, however, introduces agency theory problems—the farmer (principal) may hire workers (agents) to carry out farm work. The farmer then faces a decision as to the type of pay-scheme to use. The hired worker might be employed on a time-rate of pay or a piece-rate of pay. With a time-rate of pay scheme, the worker faces little risk of his work but has an incentive to shirk. Because of the spatial dispersion of crop farming, this makes monitoring costs significant. Hence, the principal must invest in costly monitoring of hired wageworkers and this drives up cost and reduces productivity. Alternatively, the farmer can pay a piece rate. This provides the worker with strong incentives to perform but he bears output risk, i.e., if the crop develops poorly it may take lots of labor per unit of output harvested. Also, a piece-rate pay scheme requires that the task be well defined and output can be easily measured, and it provides a strong incentive for workers to produce a large volume of low quality output (Gibbon 1999). Hence, with piece rate pay scheme, the principal must be able to easily monitor quality and measure output. Thus, agency theory provides another set of reasons why large farms might be less efficient than small family farms.

Some incentives, however, run in the opposite direction. When new agriculture technologies are being developed and sold to farmers, the adoption decision involves a fixed cost of learning about and experimenting with the new technology. The returns to adoption are positively related to size of the farming activity and the length of the farmer's planning horizon. Hence, other things being equal, large farms have a comparative advantage over smaller farms when agricultural technologies are dynamic (Huffman & Evenson 2001). In addition, some new technologies are embedded in large durable goods, e.g., self-propelled combines, and large farms can more fully utilize these machines than can small farms. Also, if credit markets are imperfect, large farms generally have an advantage over small farms in obtaining credit or providing self-financing.

An agent, however, can acquire a large machine and engage in custom harvesting for other (small) farms. This institution, however, raises new agency theory issues. For example, if the agent is hired on a time-rate of pay base, he faces a weak incentive to harvest output and may spend all this time "adjusting" the combine. If he is paid on a piece-rate, he has an incentive to harvest large quantities of "dirty" and low quality grain. Hence, a joint-pay scheme comprised of part piece-rate and part time-rate of pay schemes may be optimal (Holmstrom and Milgrom 1987; Lazear 1995). The custom work example provides another example of a negative relationship between farm size and productivity.

Conclusion

This paper has examined the empirical relationship between farm size (measured in land area) and farm productivity (measured as grain output per unit of land) in China. Given that the local community council holds the majority of farmland and makes local land allocation decisions, we choose to use a two-step estimation procedure to examine in detail the relationship between farm size and productivity in small-scale agriculture. The

data that we used are from a panel dataset for Chinese farm household in the late 1990s. We find an inverse relationship between farm size and land productivity, but as we adjust for land quality and the likely endogeneity of effective land units per farm, we find that the inverse relationship is partially or completely diminished. We also discuss the potential contribution to the inverse relationship by principle-agency problem.

The paper also advances the methodology of variance estimators for the instrumental estimation. We apply the Hahn-Hausman test (Hahn and Hausman 2002) to examine whether the two-stage least square estimator provides an appropriate estimates and proposed new variance estimators for the instrumental estimates. More work, however, remains to be done to examine the impact of the emerging new land institutions on agricultural productivity in rural China.

Table 1. Descriptive Statistics of the RCRE Data Set (2708 Observations)

	Mean	Std. Dev	Min	Max
Land Under Grain Production (Mu)	9.866	8.772	0.3	150.0
Land Under Agricultural Use (Mu)	11.257	9.371	0.3	150.0
Household Population (People)	4.168	1.399	1	11
Household Labors (People)	2.564	1.065	1	8
Grain Production (kilogram)	3159.2	3621.6	50	77000

Table 2. Evidence of Inverse Relationship (2708 Observations, OLS regression)

	Model 1	Model 2	Model 3	Model 4
Ln(Land)	0.89	0.93	0.92	0.84
	$(0.01)^{12}$	(0.01)	(0.02)	(0.05)
Village Officer Dummy		-0.03	-0.06	0.00
		(0.03)	(0.02)	(0.04)
Dummy (Soya)		-0.18	-0.05	-0.07
		(0.01)	(0.02)	(0.02)
Dummy (Rice)		0.24	0.08	0.03
		(0.01)	(0.03)	(0.04)
Village/Household and Time Dummies			Village	Household
			Effect	Effect
Constant	5.96	5.89	5.88	6.09.
	(0.02)	(0.04)	(0.04)	(0.09)
R-square	0.73	0.77	0.89	0.92
Adjusted R-square	0.73	0.77	0.89	0.90

---

[12] The numbers in bracket are robust (White) variance estimates.

Table 3. First Stage Regression: Land Allocation in Rural China

	Village Effect Model		Household Effect Model	
Ln(Labor)	0.13	(0.03)	0.12	(0.04)
Ln(Household size)	0.57	(0.02)	0.51	(0.05)
Ln(Agricultural Productive Asset)	0.03	(0.00)	0.03	(0.01)
Village Officer Dummy	-0.16	(0.03)	-0.04	(0.05)
Household Type (Agricultural)	0.20	(0.07)	0.01	(0.12)
Constant	0.31	(0.09)	0.48	(0.14)
Village/Household, Time, crop variety Dummies	Omitted		Omitted	
R-square	0.79		0.92	
Adjusted R-square	0.78		0.90	
Partial R-square (Excluded Instruments)	0.38			

Table 4. Inverse Relationship: Second Stage Regression, Village Effect Model

(Dependent Variable: ln$Y$)	Estimate	2SLS Std. Err.	Robust Std. Err.	MT1 Std. Err.	MT2 std. Err.	MT3 std. Err.	MT4 std. Err.
Constant	5.735						
Fitted value of Ln(Land)	0.998	0.018	0.022	0.031	0.031	0.047	0.047
Village Officer Dummy	-0.047	0.022	0.023	0.037	0.037	0.050	0.050
Dummy (Soya)	-0.068	0.017	0.018	0.028	0.028	0.036	0.036
Dummy (Rice)	0.046	0.029	0.033	0.048	0.048	0.087	0.087
Village/Time Dummies	Omitted						
Over-identification Test	Chi(4)		3.17				
Hausman Test	Chi(1)		31.33				
Hahn-Hausman Test	Asymptotic t		0.36				
R-square	0.89				Partial R-square (1st Step)		0.38

Table 5. Inverse Relationship: Second Stage Regression, Household Effect Model

(Dependent Variable: ln$Y$)	Estimate	2SLS Std. Err.	Robust Std. Err.	MT1 Std. Err.	MT2 std. Err.	MT3 std. Err.	MT4 std. Err.
Fitted value of Ln(Land)	0.990	0.037	0.049	0.063	0.063	0.084	0.084
Over-identification Test	Chi(4)		0.76				
Hausman Test	Chi(1)		14.53				
Hahn-Hausman Test	Asymptotic t		1.24				
R-square	0.90				Partial R-square (1st Step)		0.18

Table 6. Inverse Relationship: Reduced Samples (Dependent Variable: ln*Y*)

Coefficient of Land	OLS Estimate	IV Estimate	Partial R-square	Over-id Test	Hausman Test	Hahn-Hausman Test
Land < 25	N=2575					
Village Effect	0.91 (0.02)[13]	1.00 (0.05)[14]	0.38	4.55	33.00	-0.21
Household Effect	0.79 (0.03)	0.96 (0.11)	0.17	3.69	15.21	-0.27
Land < 15	N=2165					
Village Effect	0.90 (0.02)	0.99 (0.03)[15]				
Household Effect	0.80 (0.04)	0.94 (0.07)				
Land >= 15	N=543					
Village Effect	0.89 (0.05)	1.16 (0.11)				
Household Effect	0.83 (0.06)	0.97 (0.18)				

Appendix 1: MT Type Instrumental Variance Estimators

For equation (24), we consider the usual instrumental estimation when X_2 is the included exogenous variables. Using the methodology of Murphy and Topel (1985), we derive the estimators of the covariance matrix in the following.

Note, for simplicity, we use $(\hat{\beta}', \hat{\gamma})'$ rather than $(\hat{\beta}', \hat{\gamma})'_{2SE}$ in the following.

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \beta \\ \gamma \end{bmatrix} - (Z'Z)^{-1}Z'X(\hat{\theta} - \theta)\gamma + (Z'Z)^{-1}Z'u$$

In this case, since we have only one fitted value in the second stage, it is easy to derive the covariance estimator.

---

[13] Robust standard error.

[14] MT1 standard error.

[15] Here we just use the usual robust standard error.

$$\text{var}\begin{bmatrix}\hat{\beta}\\\hat{\gamma}\end{bmatrix} = \gamma^2(Z'Z)^{-1}Z'X\,\text{var}(\hat{\theta}-\theta)X'Z(Z'Z)^{-1} + (Z'Z)^{-1}Z'\,\text{var}(u)Z(Z'Z)^{-1}$$

$$-2\gamma(Z'Z)^{-1}Z'XCov(\hat{\theta}-\theta,Zu)(Z'Z)^{-1}$$

1. No Heteroscedasticity, Disturbances Not Correlated (MT1).

If we assume that there is no correlation between the first stage random disturbance and the second stage random error, we have $\text{cov}(\hat{\theta}-\theta,Zu)=0$. The adjusted covariance matrix would be:

$$\widehat{\text{var}}\begin{bmatrix}\hat{\beta}\\\hat{\gamma}\end{bmatrix} = \hat{\sigma}_\varepsilon^2\gamma^2(Z'Z)^{-1}Z'P_XZ(Z'Z)^{-1} + \hat{\sigma}_u^2(Z'Z)^{-1}$$

where $\sigma_\varepsilon^2 = \frac{1}{n}(L-X\hat{\theta})'(L-X\hat{\theta})$, and $\frac{1}{n}(Y-X_2\hat{\beta}-\hat{\gamma}\hat{L})'(Y-X_2\hat{\beta}-\hat{\gamma}\hat{L})$ is a consistently estimate of σ_u^2 (Murphy and Topel, 1985).

2. No Heteroscedasticity, Disturbances Correlated (MT2).

Murphy and Topel (1985) obtained an estimator of the correlation of $\text{cov}(\frac{1}{\sqrt{n}}g(\theta),\frac{1}{\sqrt{n}}Zu)$ by the law of large numbers as: $\frac{1}{n}\sum_{i=1}^{n}Z_i'\hat{u}_ig_i(\theta)$, similarly we have the estimator of $\text{cov}(\frac{1}{\sqrt{n}}(\hat{\theta}-\theta),\frac{1}{\sqrt{n}}Zu)$ as:

$$Est.\text{cov}(\tfrac{1}{\sqrt{n}}(\hat{\theta}-\theta),\tfrac{1}{\sqrt{n}}Zu) = Est.\text{cov}((X'X)^{-1}(\tfrac{1}{\sqrt{n}}X'\hat{\varepsilon}),\tfrac{1}{\sqrt{n}}Zu)$$
$$= (X'X)^{-1}\tfrac{1}{n}\sum_{i=1}^{n}X_i'\hat{\varepsilon}_iZ_i\hat{u}_i$$

Therefore we have:

$$\widehat{\text{var}}\begin{bmatrix}\hat{\beta}\\\hat{\gamma}\end{bmatrix} = \hat{\sigma}_\varepsilon^2\gamma^2(Z'Z)^{-1}Z'P_XZ(Z'Z)^{-1} + \hat{\sigma}_u^2(Z'Z)^{-1}$$

$$-2\gamma(Z'Z)^{-1}Z'X(X'X)^{-1}\sum_{i=1}^{n}\{X_i'\hat{\varepsilon}_iZ_i\hat{u}_i\}(Z'Z)^{-1}$$

3. Heteroscedasticity, Disturbances Not Correlated (MT3).

Considering heteroscedasticity, under some fairly general conditions, White (1980) showed that the matrix: $S_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i'$ where e_i is the ith least squares residual, is a consistent estimator of:

$$\Sigma = \frac{1}{n}\sigma^2 X'\Omega X = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 x_i x_i'.$$

Hence the White estimator for the first stage covariance matrix is:

$$Est.Var(\hat{\theta}) = n(X'X)^{-1}S_0(X'X)^{-1}.$$

Similarly we have the White estimator of $(Z'Z)^{-1}Z'\text{var}(u)Z(Z'Z)^{-1}$ as:

$$n(Z'Z)^{-1}S_0^Z(Z'Z)^{-1}, \text{ where } S_0^Z = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2 z_i z_i'$$

$$\widehat{\text{var}}\begin{bmatrix}\hat{\beta}\\\hat{\gamma}\end{bmatrix} = n\gamma^2(Z'Z)^{-1}Z'X(X'X)^{-1}S_0(X'X)^{-1}X'Z(Z'Z)^{-1} + n(Z'Z)^{-1}S_0^Z(Z'Z)^{-1}$$

The second term on the right hand side is the usual white estimator while the first term is the variability of including the fitted value as a regressor.

4. Heteroscedasticity, Disturbances Correlated (MT4).

If there is correlation between the first stage random disturbance and the second stage random error, we need to use the estimator of $\text{cov}(\frac{1}{\sqrt{n}}(\hat{\theta}-\theta), \frac{1}{\sqrt{n}}Zu)$.

Substitute this into the original estimator, we have:

$$\widehat{\text{var}}\begin{bmatrix}\hat{\beta}\\\hat{\gamma}\end{bmatrix} = n\gamma^2(Z'Z)^{-1}Z'X(X'X)^{-1}S_0(X'X)^{-1}X'Z(Z'Z)^{-1} + n(Z'Z)^{-1}S_0^Z(Z'Z)^{-1}$$

$$-2\gamma(Z'Z)^{-1}Z'X(X'X)^{-1}\sum_{i=1}^{n}\{X_i'\hat{\varepsilon}_i Z_i \hat{u}_i\}(Z'Z)^{-1}$$

Reference:

Ahearn, M., J. Yee, and W.E. Huffman. "R&D, Productivity, and Structural Change in U.S. Agriculture, 1960-1996," USDA, ERS, July 2002.

Benjamin,D. "Can Unobserved Land Quality Explain the Inverse Productivity Relationship." *Journal of Development Economics* 46(February 1995): 51-84.

Benjamin,D. and L.Brandt "Property rights, labour markets, and efficiency in a transition economy: the case of rural China." *Canadian Journal of Economics-Revue Canadienne D Economique* 35(November 2002): 689-716.

Benjamin,D., L.Brandt, and J.Giles "Income Persistence and the Evolution of Inequality in Rural China." *Paper prepared for the NEUDC Meetings*(2001).

Berry,R.A. and W.R.Cline *Agrarian Structure and Productivity and Developing Countries.*John Hopkins University Press, Baltimore, MD, 1979.

Bhalla,S.S. and P.Roy "Mis-Specification in Farm Productivity Analysis - the Role of Land Quality." *Oxford Economic Papers-New Series* 40(March 1988): 55-73.

Bound,J., D.A.Jaeger, and R.M.Baker "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90(June 1995): 443-50.

Brandt,L. "Chinese-Agriculture and the International Economy, 1870-1930S - A Reassessment." *Explorations in Economic History* 22(1985): 168-93.

Burgess,R. "Land and Welfare: Theory and Evidence from China." *Working Paper* Department of Economics and STICERD, London School of Economics, London(2001).

Carter,M.R. "Identification of the Inverse Relationship Between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production." *Oxford Economic Papers* 36(1984): 131-45.

Chayanov,A.V. "The Theory of Peasant Economy." D.Thorner, ed.Irwin, Homewood, IL, 1926.

Chen,H.Y. and S.Rozelle "Leaders, managers, and the organization of township and village enterprises in China." *Journal of Development Economics* 60(December 1999): 529-57.

Chen,J. "Readme file for RCRE data set." *Notes* University of California, Davis(2001).

Chen,K. and C.Brown "Addressing shortcomings in the Household Responsibility System - Empirical analysis of the Two-Farmland System in Shandong Province." *China Economic Review* 12(2001): 280-92.

Cheng,E. "Household heads, non-economic factors & Grain Production in China in the 1990s." *Working Paper #98/5, Center for Asian studies, Chinese Economy Research Unit, University of Adelaide*(1998).

China " Agirculture Law." *The People's Republic of China*(1993).

Davidson,R. and J.G.Mackinnon *Estimation and Inference in Econometrics.* Oxford University Press, 1993.

Deininger,K. and G.Feder "Land Institutions and land markets." *Handbook of Agricultural Economics Vol. 1.* B.Gardner and G.Rausser, eds., 287-331.Elsevier Science, 2001.

Gibbons,R. "Incentives in Organizations." *J. Econ. Perspective* 12(Fall 1998):115-32.

Hahn,J. and J.Hausman "A new specification test for the validity of instrumental variables." *Econometrica* 70(January 2002): 163-89.

Hall,A.R., G.D.Rudebusch, and D.W.Wilcox "Judging instrument relevance in instrumental variables estimation." *International Economic Review* 37(May 1996): 283-98.

Hallam, A., Size, Structure, and the Changing Face of American Agriculture. Westview Press, Boulder, CO. (1993)

Hausman,J. "Specification Tests in Econometrics." *Econometrica* 46(1978): 1251-71.

Heltberg,R. "Rural market imperfections and the farm size-productivity relationship: Evidence from Pakistan." *World Development* 26(October 1998): 1807-26.

Holmstrom, B. and P. Milgrom,"Aggregation and Lineaarity in the Provision of Intertemporal Incentives," *Econometrica* 55(March 1987):303-28

Huffman, Wallace.E.. "Human Capital: Education and Agriculture," in Handbook of Agricultural Economics, Vol. IA, Bruce L. Gardner and G.C. Rausser, eds., Amsterdam, The Netherlands: Elsevier Science/North HOlland, 2001, pp. 333-381.

Huffman, W.E. and R.E. Evenson "Structural and productivity change in US agriculture, 1950-1982." *Agricultural Economics* 24(January 2001): 127-47.

Huffman, Wallace E. and Peter F. Orazem. "The Role of Agriculture and Human Capital n Economic Growth: Farmers, Schooling, and Health," in Handbook of Agricultural Economic, Vol. III, Robert Evenson, ed., Amsterdam, The Netherlands: Elsevier Science/North Holland, forthcoming.

Krusekopf,C.C. "Diversity in land-tenure arrangements under the household responsibility system in China." *China Economic Review* 13(2002): 297-312.

Lazear, E. Personnel Economics, Cambridge, MA: MIT Press, 1995

Lin,J.Y. "The Household Responsibility System in China's Agricultural Reform: A Theoretical and Empirical Study." *Economic Development and Cultural Change* 36(1988).

Murphy,K.M. and R.H.Topel "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics* 3(1985).

Nelson,C.R. and R.Startz "The Distribution of the Instrumental Variables Estimator and Its T-Ratio When the Instrument Is A Poor One." *Journal of Business* 63(January 1990): S125-S140.

Ruud,P. *An Introduction to Classical Econometric Theory.*Oxford University Press, 2000.

Sen,A.K. "An Aspect of Indian Agriculture." *Economic Weekly* 14(1962).

Staiger,D. and J.H.Stock "Instrumental variables regression with weak instruments." *Econometrica* 65(May 1997): 557-86.

Taylor, J. Edward and I. Adelman, "Agricultural Household Models: Genesis, Evolution, and Extensions," *Review of Economics of the Household*, 1(Jan./April 2003): 33-59.

Yao,Y. "Egalitarian Land Distribution and Labor Migration in Rural China." *Working Paper*, China Center for Economic Research, Beijing University, Beijing (October 2001).

----- "The Rational Versus the Political Model in Collective Decision: The Case of Land Tenure Choice in Chinese Villages." *Working Paper No. E2002003* China Center for Economic Research, Beijing University, Beijing(2002).