Separation Hypothesis Tests in the Agricultural Household Model

Kien T. Le

In an agricultural household model, farmers' production decisions can be either separated or nonseparated from preferences. Since previous studies on agricultural household behavior and policy effects have shown that the model with separation yields different results than the model with nonseparation, it is important to have a test to identify the correct model. This paper provides new tests that extend current tests in two directions. First, the new tests avoid issues that current tests have to address, such as simultaneity bias and the estimation of the production function. Second, the new tests make use of more information implied from the separation hypothesis than current tests.

Key words: agricultural household model, separation hypothesis test, shadow wage.

JEL codes: O12, Q12.

Rural households, especially in developing countries, are systematically exposed to market imperfections. These imperfections lead to what has been called nonseparability. A household behaves as in the *nonseparation model* (NM) if household's production decisions (e.g., choices of labor, production inputs and outputs) are affected by its preferences (e.g., consumer preferences, demographic composition). By contrast, in a *separation model* (SM), a household behaves as a pure profit-maximizing producer. The household's production decisions are not affected by its preferences.

Because of these differences, the analysis of household behavior and policy effects could yield different results between the SM and NM. Eswaran and Kotwal (1986) show that under the NM, households with different asset positions use different strategies of labor deployment, and then use these to establish social classes. Taylor and Adelman (2002) analyze the impact of Mexican trade and transfer policies. They find that these policy shocks under imperfections in the labor and food market have, contrary to expectations of policymakers, a remarkably small impact on production and rural incomes. Löfgren and Robinson (2002) show that large transaction costs in market participation by households in response to price and productivity shocks create discontinuities which differ markedly from smooth responses normally expected in the SM.

These differences in policy effects call for tests to identify the correct model. The first set of tests uses the relationship between production decisions and preferences. These tests yield mixed results depending on the country in the study. For example, Lopez (1984) rejects the SM with Canadian data, Benjamin (1992) cannot reject it with a sample of Javanese rural households, Bowlus and Sicular (2003) arrive at the same conclusion as Benjamin with a sample from China, but Grimard (2000) rejects it for Côte d'Ivoire. The second set of tests uses the relationship between shadow wages and market wages (Jacoby 1993; Skoufias 1994; Bhattacharyya and Kumbhakar 1997; Abdulai and Regmin 2000). All of these tests reject the equality between shadow wages and market wages, and hence reject the SM.

Although both sets of tests are based on a very sound theoretical model, the empirical application brings in some issues that the tests have to address. For example, Benjamin has to address the simultaneity bias when estimating the farm labor demand function, while Jacoby has to address the endogeneity problem when estimating the production function.

Kien T. Le is a research associate with the Social and Economic Survey Research Institute, Qatar University. This article is one chapter in the author's Ph.D. dissertation (2008) submitted to the University of Virginia. The author is grateful to John McLaren, John Pepper, and Sanjay Jain for their mentorship.The author is indebted to Jeffrey H. Dorfman and anonymous reviewers for insightful comments and helpful suggestions. The author is responsible for any remaining errors.

Amer. J. Agr. Econ. 92(5): 1420–1431; doi: 10.1093/ajae/aaq070
Received December 2008; accepted July 2010; published online September 10, 2010

Other tests have similar issues. This paper tries to extend current tests by avoiding these issues in the first place, so the empirical application will be simpler and less data intensive than current tests, and the final results can be more reliable. In addition, the tests proposed in this paper simultaneously use both relationships implied from the separation hypothesis, while current tests use only one relationship, either between production decisions and preferences or between shadow wages and market wages. Therefore, we can use more information implied from the separation hypothesis and so increase the test power.

A key observation to develop new tests comes from the fact that the shadow wage, an important variable in farmers' decisions, is the marginal product of labor at the optimal point on the production function. Accordingly under certain assumptions on the functional form of the production function, we can get the shadow wage without estimating the production function (Le 2009). These functional form assumptions are not restrictive, since they result in a semiparametric function that is more flexible than the Cobb-Douglas function, which is usually used in current tests.

In the empirical section, the new tests are applied to a sample of Vietnamese farmers. Vietnam is an interesting case, since there are several studies on the impact of trade liberalization on farmers' welfare.[1] These studies use Deaton's (1989) nonparametric method to evaluate effects of trade liberalization on household welfare. An important assumption in the method is the separation between farmers' production and consumption decisions. However, most market jobs in Vietnam are provided by state-owned enterprises beset by a lack of motivation, underdevelopment, and adherence to a very rigid wage structure (Van De Walle and Cratty 2003). So the validity of the separation assumption could be a concern. In the empirical section, I will use the tests proposed in this paper to investigate the validity of this assumption.

Theoretical Model

The theoretical model is based on the standard time allocation model. A farmer maximizes his utility function defined over leisure (l), consumption (c), and a vector of preference shifters A (e.g., number of children and number of adults): $U(c, l; A)$, subject to a budget constraint: $c = pQ + wm$, where p is the price of farm output.[2] In this budget constraint, the farmer receives income from two activities: working on his farm to receive the farm output (Q) and working in the labor market to receive the market wage (w). Thus, his labor supply (h) includes farm labor (L) and market labor (m): $h = L + m$. The sum of labor supply and leisure is the total stock of time: $T = h + l$, where T could be twenty-four hours a day.

The production function for the farm output is $Q(L, F)$, where F is a vector of the fixed inputs (e.g., land area, farm equipment). Variable inputs such as hired labor and fertilizer are suppressed in the production function for brevity. They will be included in the empirical section. Both functions U and Q are assumed to be increasing and concave in their arguments.

Imperfections are introduced into the model as the upper and lower constraints on the market labor: $0 \leq m \leq M$, where M is the maximum number of hours a farmer can work in the labor market. The farmer faces imperfections if either the lower constraint or the upper constraint is binding ($m = 0$ or $m = M$). If this is the case, the farmer's behavior is characterized by the NM. By contrast, the farmer faces no imperfection if neither the lower nor the upper constraint is binding ($0 < m < M$), and his behavior is consistent with the SM.

The above household maximization problem can be summarized as follows:

$$(1) \qquad \max_{L,m} U(pQ(L, F) + wm,\ T - L - m; A) \quad \text{subject to} : 0 \leq m \leq M.$$

The first-order condition for L states that

$$(2) \qquad w^* = pQ_L(L, F)$$

where Q_L is the derivative of output with respect to labor.

And the first-order conditions for m are:[3]

$$(3) \qquad w^* = w \quad if \quad 0 < m < M$$

$$(4) \qquad w^* \neq w \quad if \quad m = 0 \ \ or \ \ m = M$$

where

$$(5) \qquad w^* = \frac{U_l(pQ(L, F) + wm, T - L - m; A)}{U_c(pQ(L, F) + wm, T - L - m; A)}.$$

[1] See Minot (2000) and Benjamin (2004).

[2] Like previous studies on this topic, I assume that farmers maximize utility instead of expected utility.

[3] These first-order conditions are derived from the Largrangian function.

w^* is called the shadow wage or the opportunity cost of time, and it is the key variable in this time allocation model. The farmer has SM behavior in equation (3) if the constraints are not binding, and NM behavior in equation (4) if the constraints are binding. Our job is to identify which the correct model is.

- If the SM is correct, then $w^* = w$. We plug this into equation (2) to have: $pQ_L(L, F) = w$. Since A does not show up in this equation, the choice of L does not depend on A. In other words, the choice of L is separated from A, which is why this is called the SM.
- If the NM is correct, then $w^* \neq w$. We find L by substituting equation (2) into equation (5) to have $pQ_L(L, F) = \frac{U_l(pQ(L,F)+wm,\ T-L-m,A)}{U_c(pQ(L,F)+wm,\ T-L-m,A)}$ where $m = 0$ or $m = M$. Since A shows up in this equation, the choice of L depends on A. In other words, the choice of L is not separated from A, which is why this is called the NM.

In summary, we can identify the correct model from the relationship between w^* and w: w^* is equal or not equal to w; or the relationship between L and A: L depends or does not depend on A. The latter could be easily generalized to the relationship between any production decision (e.g., L, Q, variable inputs) and A.

Note that the imperfections are introduced into the model as the lower and upper constraints on the market labor ($0 \leq m \leq M$). These imperfections are chosen because they are cited in several studies about agricultural households (Benjamin 1992; Jacoby 1993; Skoufias 1994; Sonoda and Maruyama 1999; Abdulai and Regmin 2000; Le 2009). It is not difficult to show that the above relationships are basically unchanged for other market imperfections, such as credit constraints, transactions or search costs, or work location preferences. Interested readers can look at the study by Singh et al. (1986) for a survey of this agricultural household model and by de Janvry and Sadoulet (2006) for recent progress.

Current Tests in the Literature

Based on the above relationships, two sets of tests have been developed in the literature. The first set uses the relationship between production decisions and A (Lopez 1984; Benjamin 1992; Bowlus and Sicular 2003; Grimard 2000), and the second set uses the relationship between w^* and w (Jacoby 1993; Skoufias 1994; Bhattacharyya and Kumbhakar 1997; Abdulai and Regmin 2000). A prominent test for each of these approaches, those developed by Benjamin and Jacoby, is presented below to set the stage for comparison to the new tests.

Benjamin (1992) proposes a well-known test that uses the relationship between production decisions and A. He starts with a Cobb-Douglas production function: $Q = \lambda_0 L^{\lambda_L} z^{\lambda_z} F^{\lambda_F}$, where z is a vector of variable inputs (e.g., hired labor, fertilizer, seed). The household profit from the farm production is: $\pi = p\lambda_0 L^{\lambda_L} z^{\lambda_z} F^{\lambda_F} - w^* L - p_z z$, where the first term on the right side is the revenue from the farm output, the second term is the household farm labor opportunity cost, and the third term is the variable input cost. By maximizing farm profit, we find the farm labor demand: $L = \beta_0 p^{\beta_1} p_z^{\beta_2} (w^*)^{\beta_3} F^{\beta_4}$, where $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ are some functions of $(\lambda_0, \lambda_L, \lambda_z, \lambda_F)$.

Under the SM $w^* = w$ while under the NM $w^* \neq w$ and w^* depends on A, so Benjamin approximates the shadow wage with this linear function: $\log(w^*) = \log(w) + \alpha A$, where $\alpha = 0$ for the SM and $\alpha \neq 0$ for the NM. Plug this shadow wage into the farm labor demand to have:

$$(6) \quad \log(L) = \log(\beta_0) + \beta_1 \log(p) + \beta_2 \log(p_z) + \beta_3(\log(w) + \alpha A) + \beta_4 \log(F).$$

The test for the SM is whether $\alpha = 0$. The issue with this regression is that w is determined from both labor supply and labor demand, so there is a well-known simultaneity bias issue. Benjamin is able to find valid instruments to address this issue properly.[4]

While Benjamin uses the relationship between L and A to identify the correct model, Jacoby (1993) proposes a different test that uses the relationship between w^* and w. Since w^* cannot be observed from the data, it must be estimated. Jacoby notices from the first-order condition in equation (2) that w^* is identical to the marginal product of labor (MPL)

[4] Besides the labor demand in equation (6), an important contribution of Benjamin's paper is his careful analysis of several empirical issues that could arise during the estimation such as disaggregating farm work, aggregation of male and female labor, functional form and others. Since Benjamin concludes that these issues do not change the test results, I do not examine these issues in this paper, so do other papers using Benjamin approach.

regardless of market imperfections, so it is necessary to calculate this MPL from the farm production.[5] Therefore, he starts by estimating a Cobb-Douglas production function:[6]

$$(7)\qquad \log(Q) = \log(\lambda_0) + \lambda_L \log(L) + \lambda_z \log(z) + \lambda_F \log(F).$$

Because of the well-known endogeneity problem, instruments are necessary in this regression.[7] Having estimated the production function, the shadow wage is:

$$(8)\qquad w^* = p\partial Q/\partial L = p\lambda_L \hat{Q}/L$$

where $\hat{Q}$ is the fitted output from the production function regression. The fitted output is used instead of the real output Q, since farmers make decisions when they do not know the random weather shock ε and the real output Q. $\hat{Q}$ is a kind of farmers' prediction of Q.

Having calculated w^*, Jacoby identifies the correct model by running the following regression:

$$(9)\qquad w^* = \beta_0 + \beta_1 w.$$

The test for the SM is whether $\beta_0 = 0$ and $\beta_1 = 1$. Jacoby raises two issues in his discussion of the estimation of the production function. First, the estimation requires information on the quantities of all inputs and outputs on the farm. However, due to data limitation, the values (the product of quantity and price) instead of the quantities are in use, so any price variation across households can bias the estimation of the production function (Jacoby 1993, p. 910).[8] Second, finding valid instruments to address the well-known endogeneity issue is another source of concern.

The final shortcoming shared by both Benjamin and Jacoby tests is that each test makes use of only one relationship: either between production decisions and preferences or between shadow wages and market wages. That means these tests use only half of the information implied from the separation hypothesis. In addition, the results from Benjamin and Jacoby tests can contradict each other. For example, it is possible in an empirical study that for the same dataset, the Benjamin test rejects the separation hypothesis while the Jacoby test does not. Because of this shortcoming, it is necessary to have a test that simultaneously uses both relationships, so we can increase the power of the test and avoid the possibility of contradictory results. The next section will present tests to address this shortcoming.

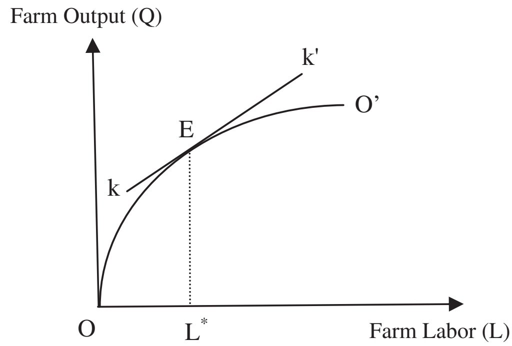

Figure 1. MPL at the optimal point on the production function

New Tests

As pointed out by Jacoby, the shadow wage is identical to the MPL. However, it is not any MPL, but the one at the optimal point on the production function. This observation is illustrated in figure 1.

In this figure the vertical axis is the farm output (Q) and the horizontal axis is the amount of farm labor (L). OO′ is the production function curve, which describes the amount of farm output for each choice of farm labor. The slope of this curve is the MPL.

Assuming that L* is the optimal choice of farm labor, then E is the optimal point on the production curve, and the farmer's shadow wage is the MPL at this point (the slope of kk′). If we estimate the whole production function to get the shadow wage, it is equivalent to estimating the MPL of all points on the production curve in order to find the MPL of only one point. By doing so, we would have to specify a parametric functional form for the production function (e.g., Cobb-Douglas function) and confront the issues mentioned in the

[5] Following Jacoby, most studies on farmers' behavior derive shadow wages from the MPL. The exception is Barrett, Sherlund, and Adesina (2008) who propose a different method to derive the shadow wage.

[6] Jacoby also uses a more flexible translog function.

[7] The endogeneity problem comes from the correlation between the unobserved factors in the regression error and variable inputs (z) and labor (L) in the production function.

[8] Many surveys do not provide information about the price and quantity of all inputs and outputs (though the information about values is usually available). In addition, the need to sum different kinds of farm outputs and inputs also requires the use of values instead of quantities.

Jacoby test. The following presents a simple method to get the shadow wage without estimating the production function. We start with this semiparametric production function:[9]

$$(10)\quad \overline{Q} = L^{\lambda_L} f(z, F, \sigma)$$

where $f()$ is a nonparametric function.

This production function is similar to the Cobb-Douglas function but is more flexible, since there is no assumption on the functional form for $f()$. The restriction on this production function is that the labor input has to enter in a Cobb-Douglas form (L^{λ_L}). We can further increase the flexibility of this production function by allowing λ_L to vary across households: $\lambda_L = e^{\lambda K+\xi}$ where K is a vector of observed variables accounting for the difference in λ_L across households and ξ is the unobserved variable which is assumed to be random. In the empirical section, we use a vector of dummy variables representing the difference in region (R) and types of farm products (O) for K, so λ_L in this case becomes:[10]

$$(11)\quad \lambda_L = e^{\lambda_0+\lambda_1 R+\lambda_2 O+\xi}.$$

The real output Q is different from the above production function $\bar{Q}$ due to the random weather shock ε: $Q = \bar{Q}e^{\varepsilon}$, where ε is normalized to have $E(e^{\varepsilon}) = 1$. Because of the weather shock, farmers do not know the real output Q, and their MPL is based on their expectation about Q:[11]

$$(12)\quad MPL = p\partial E(Q)/\partial L = p\partial \overline{Q}/\partial L = \lambda_L pQe^{-\varepsilon}/L.$$

Since pQ and L are observed and the observed values are the optimal values made by the farmers, we have arrived at the optimal MPL or the shadow wage without estimating the production function. Although this shadow wage is calculated up to an unknown parameter (λ_L) and a random term (ε), this information is enough to test the separation hypothesis.

Taking logs on both sides and replacing MPL by w^*, we have:

$$(13)\quad \log(pQ/L) = -\log(\lambda_L) + \log(w^*) + \varepsilon.$$

This equation can be used to develop several tests. The first test is based on Benjamin's approximation of the shadow wage: $\log(w^*) = \log(w) + \alpha A$. Substituting this into (13) yields the model:

$$(14)\quad \log(pQ/wL) = -\log(\lambda_L) + \alpha A + \varepsilon.$$

The test for the SM is whether $\alpha = 0$. We want to know if production decisions on the left side are affected by preference shifters on the right side. In the following, I will refer to this test as the simplified Benjamin test. The second test is based on the relationship between w^* and w as in the Jacoby test. By replacing $\log(w^*)$ by $\beta\log(w)$ in equation (13), we have the following regression:

$$(15)\quad \log(pQ/L) = -\log(\lambda_L) + \beta\log(w) + \varepsilon.$$

The test for the SM is whether $\beta = 1$. In the following, I will refer to this test as the simplified Jacoby test. The regressions in equations (14) and (15) can be modified to allow λ_L to vary across households by replacing λ_L with the expression in equation (11). The regressions in this case are slightly changed. Dummy variables for regions (R) and types of product (O) will be included in the explanatory variables, and the regression error term will include ε and ξ. Specifically, the regression in equation (14) becomes $\log(pQ/wL) = -\lambda_0 - \lambda_1 R - \lambda_2 O + \alpha A + \tau$ and the regression in equation (15) is $\log(pQ/L) = -\lambda_0 - \lambda_1 R - \lambda_2 O + \beta\log(w) + \tau$ where $\tau = \xi + \varepsilon$.

Compared with current tests in the literature, both simplified tests are less data intensive. For example, in the simplified Jacoby test (see equation (15)), we need information about only farm output (pQ), market wage (w), and farm labor (L), while the Jacoby test needs information on these variables plus production input variables (e.g., land areas and quality, farm equipment, fertilizer) and instrument variables to address the endogeneity issue in estimating the production function.

[9] Hired labor is included in z, so we assume that there is no substitution between hired and household labor.

[10] The specification of λ_L is similar to the random coefficient model in econometrics literature (Greene 2003, pp. 318–319). For households producing more than one type of product, the type that yields the largest income is selected.

[11] I assume a multiplicative error in the production function: $Q = L^{\lambda_L} f()e^{\varepsilon}$, so the expected output is $E(Q) = L^{\lambda_L} f()$ by assuming that $E(e^{\varepsilon}) = 1$. If we assume additive error, the production function would be $Q = L^{\lambda_L} f() + \varepsilon$ and the expected output would still be the same as $E(Q) = L^{\lambda_L} f()$ by assuming that $E(\varepsilon) = 0$. So we arrive at the same expected output in both cases (though the assumption on the mean of the error differs). There will be difference between the additive and multiplicative error if we calculate the variance of the output.

The estimation in the simplified tests is also simpler. In the simplified Benjamin test, w shows up in only the dependent variable, so we do not have to address the simultaneity bias that the Benjamin test has to address. In the simplified Jacoby test, the estimation of the production function is not required, so we can avoid issues that are mentioned in the Jacoby test, such as the endogeneity and the use of input values instead of quantities. In addition, both of the simplified tests are based on a semiparametric production function that includes the Cobb-Douglas function as the special case.[12] However, they have the same shortcoming as the current tests: each test makes use of only one relationship, while there are two relationships implied from the separation hypothesis. The following present two tests that make use of both relationships.

To develop a test that uses both relationships, it is natural to think about the combination of the two simplified tests. It should be noted that the two relationships are based on maximization of the same utility function, and hence both relationships might involve some of the same parameters. It might be difficult to combine the Benjamin and Jacoby tests, since they are developed in different ways, but the two simplified tests can be easily combined, since they are both developed from equation (13). We simply estimate a system of equations (14) and (15), and the test for the SM is whether $\beta = 1$ and $\alpha = 0$. Joint estimation of the two equations will involve imposing cross-equation constraints on some parameters, as well as recognizing that error terms are correlated. This issue will be discussed more in the empirical section. In the following, I will refer to this test as the combined test.

The second test that uses both relationships is based on a generalized specification of the shadow wage. Recall that in the Benjamin test and its simplified version, the shadow wage is $\log(w^*) = \log(w) + \alpha A$, while in the Jacoby test and its simplified version the shadow wage is $\log(w^*) = \beta \log(w)$. We can combine these two specifications of the shadow wage into one general specification as $\log(w^*) = \beta \log(w) + \alpha A$ and plug this into equation (13) to have:[13]

$$\text{(16)} \quad \log(pQ/L) = -\log(\lambda_L) + \beta \log(w) + \alpha A + \varepsilon.$$

The test for the SM is whether $\beta = 1$ and $\alpha = 0$. In the following, I will refer to this test as the generalized test.

Data

The paper makes use of the Vietnam Living Standard Survey in 1997–98, the extremely rich datasets for analyzing household economic behaviors and their link to policies. The survey was conducted by the Vietnam General Statistical Office with technical assistance from the World Bank. The survey includes 6,000 households, selected from a two-stage sampling process. In the first stage, 4,800 households were selected based on the 1989 census that 80% of the population was living in rural areas and 20% in urban areas; 120 communes in rural areas and 30 wards in urban areas were randomly selected with the probability proportional to the number of households in those villages or wards. In the second stage, 1,200 households were added to the sample, which was not proportional to the number of households or population. Therefore, the *weight* variable is needed in the calculation to ensure efficiency. In general, rural farmers are well represented in the sample.

The survey collects information about household characteristics and consumption and production activities. As in previous studies, households in urban areas, households with no land or farming activities, and households with no one working in the labor market are dropped out of the sample.[14] Households with missing information on some variables are also

[12] Note that the semiparametric production function still requires that the labor input enters the production function in the Cobb-Douglas form, so it cannot include the Translog function as the special case.

[13] I am very grateful to an anonymous referee who suggests this test. The test comes from the observation that in the shadow wage for the Benjamin test the coefficient on $\log(w)$ is restricted to unity while in the shadow wage for the Jacoby test the coefficients on preference shifters are restricted to zeros. The generalized specification for the shadow wage simply removes these restrictions to arrive at a more general formula for the shadow wage.

[14] The first-order condition in equation (4) indicates that households with no market labor behave as in the NM. Previous studies do not include these households in the test and so there might be a concern about the sample selection. Following previous studies, the test results in this paper are limited to the sample of households with market labor. However, I have also run tests for the whole sample by using commune wage for households with no market labor. The test results are unchanged and can be provided upon request.

dropped. The final sample for the tests is 2,552 households.[15]

All variables required for the tests are available in the data. First, the value of household farm output (pQ) is the aggregated sum of every farm product produced by the household in a year: $pQ = \sum p_k q_k$ (k is the index for products produced by the household). Unlike the information about the quantities of outputs and inputs, the information about the value ($p_k q_k$) is readily available in the data. Second, farm labor (L) is the annual total of household hours working on the farm. This variable is calculated from the information in the survey data about how many hours a farmer works on the farm a day, how many days a week, how many weeks a month, and how many months a year. Third, market wage (w) is the household annual income from market work divided by the corresponding working hours. The annual income from market work is available in the survey data, and the corresponding working hours are calculated in the same way as farm labor. Finally, preference shifter (A) includes the number of children fifteen years and younger and the number of males and females sixteen years and older. Appendix table A1 presents descriptive information about these variables.

Test Results

This section presents the results of the separation hypothesis tests that were developed above using the Vietnam Living Standard Survey in 1997–98.

The Simplified Benjamin Test

Table 1 presents the results of the simplified Benjamin test with different specifications for regression (14). The simplest specification is in column 1, where the explanatory variables include only preference shifters. The coefficients on the number of adult males and children are significant, while the coefficient on the number of adult females is not. Column 2 shows the result of another specification of equation (14) where regions and types of farm product are added to represent the difference in λ_L across households. The coefficients on these variables are significant, reflecting the difference in λ_L across regions and types of product in Vietnam. In this specification, the coefficients on the number of adult males and children are still significant.

The last column shows the results using least absolute deviation (LAD) regression, which is less sensitive than ordinary least squares (OLS) regression to the presence of outliers in the dependent variable, a common occurrence in household surveys (Deaton 1997). This is so because in LAD the residuals to be minimized are not squared as in OLS, therefore outliers receive less emphasis. If the error term of the regression is not distributed normally, LAD may be more efficient than OLS (Buchinsky 1998). Compared with the results in column (2), the coefficients on the number of adult males and females increase while the coefficient on the number of children decrease (in absolute

Table 1. The Simplified Benjamin Test
***Dependent variable:* $\log(pQ/wL)$**

Independent Variables	OLS (1)	OLS (2)	LAD (3)
Number of adult males	−0.0791	−0.0738	−0.0930
	(0.0229)	(0.0225)	(0.0283)
Number of adult females	−0.0273	−0.0463	−0.0806
	(0.0259)	(0.0256)	(0.0243)
Number of children	−0.0613	−0.0406	−0.0357
	(0.0159)	(0.0160)	(0.0173)
Types of farm product[a]			
Paddy		−0.0926	−0.1201
		(0.0553)	(0.0557)
Other crops		−0.0810	−0.1793
		(0.0618)	(0.0619)
Regions[b]			
Red River Delta		−0.6984	−0.7090
		(0.0661)	(0.0725)
North Central		−0.1699	−0.2561
		(0.0622)	(0.0643)
Central Coast		−0.3395	−0.3897
		(0.0637)	(0.0688)
Central Highlands		−0.2959	−0.3280
		(0.0740)	(0.0723)
Southeast		−0.5540	−0.5654
		(0.1056)	(0.0866)
Mekong Delta		−0.0964	−0.0502
		(0.0874)	(0.0756)
Constant	0.4581	0.7775	0.9062
	(0.0656)	(0.0897)	(0.0927)
F-test: Coefficients on preference shifters are simultaneously 0 (df = 3)	8.15	6.12	7.92

Note: [a]Livestock is the benchmark. [b]Northern Mountain is the benchmark. Standard errors appear in parentheses.

[15] The total number of dropped households is 3,448. This number includes 1,619 households in urban areas, 440 households without land or farming activities, and 1,389 households without labor market or with missing information on some variables.

Table 2. The Simplified Jacoby Test *Dependent variable:* log(w^*)

Independent Variables	OLS (1)	OLS (2)	LAD (3)	IV (4)
Log(wage)	0.1989 (0.0223)	0.1603 (0.0215)	0.1322 (0.0251)	0.5333 (0.1359)
Types of farm product[a]				
Paddy		−0.1263 (0.0439)	−0.1259 (0.0502)	−0.1311 (0.0518)
Other crops		−0.0796 (0.0490)	−0.1314 (0.0560)	−0.0566 (0.0560)
Regions[b]				
Red River Delta		−0.7577 (0.0522)	−0.7712 (0.0649)	−0.7361 (0.0706)
North Central		−0.3474 (0.0488)	−0.4177 (0.0576)	−0.2830 (0.0713)
Central Coast		−0.6172 (0.0505)	−0.6190 (0.0623)	−0.5761 (0.0787)
Central Highlands		−0.3933 (0.0588)	−0.4121 (0.0652)	−0.3908 (0.0686)
Southeast		−0.5557 (0.0829)	−0.5098 (0.0766)	−0.4889 (0.0667)
Mekong Delta		0.0014 (0.0693)	0.0236 (0.0684)	−0.0694 (0.0657)
Constant	0.8642 (0.0254)	1.3586 (0.0542)	1.4121 (0.0631)	1.1326 (0.1525)

Note: [a]Livestock is the benchmark. [b]Northern Mountain is the benchmark. Standard errors appear in parentheses.

value). In this specification, all coefficients on preference shifters are significant.

The F-tests in the last row of table 1 reject the null hypothesis that all coefficients on preference shifters are simultaneous zeros. Overall, the results from the simplified Benjamin test clearly reject the separation hypothesis regardless of the specifications and estimation methods. We also notice that all coefficients on the preference shifters are consistently negative. These negative signs can be interpreted by recalling that $w^* = \lambda_L pQe^{-\varepsilon}/L$, so the dependent variable in the regression is almost equivalent to the ratio between w^* and w. That means when households get larger in size (because of more adults or more children), w^* decreases relative to w. This implies the rejection of the separation hypothesis, since under this hypothesis w^* is tied to w regardless of the household size (see equation (3) in the theoretical model).

I also conducted the standard Benjamin test. The results of the regression (6) based on different estimation methods are presented in Appendix table A2. Column 1 shows the result using OLS estimation, while column 2 shows the result using LAD estimation. The last column is the result using instrumental variable (IV) estimation to address the simultaneity bias in w.[16] In all estimations, the coefficients on the number of adult males and females are highly significant, strongly rejecting the separation hypothesis. The F-tests further confirm this result. The positive signs on these coefficients indicate that when the household gets larger in size, farmers will work more on the farm to support a larger household. As mentioned in the theoretical model, this is the consequence of the labor market constraint, which prevents farmers from working more in the labor market and so forces them to work more on the farm, leading to the nonseparation between farm labor and preference shifters. If there were no constraint, farmers would have increased market labor and would keep farm labor unchanged, leading to the separation between farm labor and preference shifters.

The Simplified Jacoby Test

The results of the simplified Jacoby test are presented in table 2. Overall, the separation hypothesis is rejected regardless of the estimation methods. Columns 1 and 2 show the results of the regression model in equation (15) using OLS. In both columns, the estimated coefficient on *Log(wage)* is significantly different from 1. Column 3 presents the regression results from LAD estimation. The estimated coefficient on *Log(wage)* using this estimation is smaller than the OLS estimate and statistically significantly different from 1.

Since measurement error in *wage* is usually present in survey data, it is necessary to estimate the regression using IVs to ensure the consistency of the estimated coefficients. Column 4 shows the results using the IV estimation, where the instruments are the proportion of males who are wage earners in the household, household head education, and the commune wage.[17] In the first stage regression to determine the power of the instruments (see Appendix table A3), the coefficients on the instruments are highly significant and the

[16] The instrument to address the simultaneity bias (and possible measurement error) in w is the commune wage. Benjamin also uses population density in surrounding areas but this information is not available in my data.

[17] The commune wage and education are also used in Benjamin and Jacoby as instruments to address the measurement error in wage.

Table 3. The Combined Test Result

	GMM without Instrument		GMM with Instruments	
Independent Variable	Eq. (15)	Eq (14)	Eq. (15)	Eq. (14)
Log(wage)	0.1823(0.0225)		0.7493(0.0288)	
Number of adult males		−0.0919 (0.0234)		−0.0826 (0.0224)
Number of adult females		−0.0739 (0.0267)		−0.0416 (0.0209)
Number of children		−0.0391 (0.0160)		−0.0251 (0.0104)
Types of farm product[a]				
Paddy	−0.1363 (0.0494)		−0.1532 (0.0506)	
Other crops	−0.1001 (0.0562)		−0.1898 (0.0573)	
Regions[b]				
Red River Delta	−0.7558 (0.0569)		−0.6865 (0.0612)	
North Central	−0.3480 (0.0549)		−0.1864 (0.0603)	
Central Coast	−0.6248 (0.0549)		−0.3960 (0.0641)	
Central Highlands	−0.3778 (0.0623)		−0.3186 (0.0654)	
Southeast	−0.5892 (0.0837)		−0.5415 (0.0876)	
Mekong Delta	−0.0847 (0.0764)		−0.0512 (0.0789)	
Constant	1.3475 (0.0705)	0.9757 (0.0924)	0.7956 (0.0756)	0.8166 (0.0862)
F-test: The coefficient on Log(wage) is 1 and coefficients on preference shifters are 0 (df = 4)	1321.2		77.1	

Note: [a]Livestock is the benchmark. [b]Northern Mountain is the benchmark. Standard errors appear in parentheses.

F-statistics for these coefficients to be zero is 25.37, so we do not have the weak instrument problem. The coefficient on *Log(wage)* using the IV estimation is higher compared with OLS and LAD because of the well-known downward measurement error bias, but it is still significantly different from 1.

I also calculate the Jacoby test by first estimating the Cobb-Douglas production function in equation (7). The results are presented in Appendix table A4. Having estimated the production function, we can calculate the shadow wage from equation (8) and then run the regression in (9). The result of this regression is in Appendix table A5. The coefficients on wage and constant are significantly different from 1 and 0, respectively, regardless of the estimation methods. Therefore, this test also rejects the separation hypothesis.

The Combined Test

In this test, we simultaneously estimate the system of equations (14) and (15). In both equations, the coefficients on regions and types of product come from λ_L in the production function, so we have to impose cross-equation restriction on these coefficients in the estimation. The system is also estimated taking into account correlation between the two regression error terms. I use the standard generalized method of moments (GMM) technique to estimate this system of equation, and two sets of results are presented in table 3: GMM with no instruments and GMM with instruments to address the measurement error in wage (the same instruments used in the simplified Jacoby test).

We focus attention on the first four rows. The separation hypothesis requires that the coefficient on *Log(wage)* in equation (15) is 1 and the coefficients on the preference shifters in equation (14) are 0. The F-tests in the last row clearly reject this hypothesis either with or without instruments. There is a significant difference in the coefficient of *Log(wage)* between the GMM without instrument and the GMM with instruments because of the classical downward measurement error bias. In terms of the coefficients on the preference shifters, there are some differences between GMM without instruments and GMM with instruments, but they are still within 95% confidence interval of each other.

The Generalized Test

The results of the generalized test are presented in table 4. Column 1 shows the results of the regression in equation (16) using OLS. The estimated coefficient on *Log(wage)* is significantly different from 1 and the estimated coefficients on all preference shifters are significant. We have similar results in column 2 using

Table 4. The Generalized Test Result

Independent Variables	OLS (1)	LAD (2)	IV (3)
Log(wage)	0.1549	0.1368	0.7095
	(0.0214)	(0.0296)	(0.1469)
Number of adult males	−0.0585	−0.0579	−0.0725
	(0.0177)	(0.0259)	(0.0218)
Number of adult females	−0.0993	−0.0945	−0.0497
	(0.0203)	(0.0298)	(0.0285)
Number of children	−0.0237	−0.0040	−0.0418
	(0.0126)	(0.0184)	(0.0191)
Types of farm product[a]			
Paddy	−0.1151	−0.1367	−0.1235
	(0.0436)	(0.0590)	(0.0515)
Other crops	−0.0741	−0.1774	−0.0563
	(0.0487)	(0.0657)	(0.0558)
Regions[b]			
Red River Delta	−0.7833	−0.8419	−0.7295
	(0.0522)	(0.0770)	(0.0714)
North Central	−0.4070	−0.4804	−0.2822
	(0.0494)	(0.0686)	(0.0711)
Central Coast	−0.6477	−0.6700	−0.5362
	(0.0508)	(0.0739)	(0.0827)
Central Highlands	−0.3895	−0.4148	−0.3910
	(0.0584)	(0.0769)	(0.0681)
Southeast	−0.5616	−0.5504	−0.4796
	(0.0832)	(0.0921)	(0.0689)
Mekong Delta	−0.0062	0.0752	−0.0969
	(0.0689)	(0.0804)	(0.0646)
Constant	1.6637	1.7161	1.2520
	(0.0742)	(0.1027)	(0.1655)
F-test: The coefficient on Log(wage) is 1 and coefficients on preference shifters are 0 (df=4)	394.4	215.4	24.5

Note: [a]Livestock is the benchmark. [b]Northern Mountain is the benchmark. Standard errors appear in parentheses.

LAD estimation. The only difference is that estimated coefficient on the number of children is not significant in this column. The last column presents the regression results using IV estimation with the same instruments used in the simplified Jacoby test. The coefficient on *Log(wage)* is higher compared with the first two columns, but it is still significantly different from 1. Regardless of the estimation methods, the F-tests in the last row consistently reject the separation hypothesis.

Conclusion

This paper offers two simplified tests for the separation hypothesis in the agricultural household model. One is based on the Benjamin test using the relationship between production decisions and preferences, and the other is based on the Jacoby test using the relationship between shadow wages and market wages. Unlike the current tests, the two simplified tests are developed in a way that can avoid issues that current tests encounter, such as simultaneity bias and the estimation of the production function, so the empirical application of the tests is simpler and less data intensive, and results could be more reliable. However, like the current tests, the two simplified tests make use of only one relationship implied from the separation hypothesis. To make use of both relationships in one test, the paper proposes two new tests: the combined test and the generalized test. The former is a combination of the two simplified tests, while the latter is based on a generalized specification of the shadow wage. Since these new tests use both relationships implied from the separation hypothesis, we can improve the test power and avoid possible contradictory results.

When applied to a sample of Vietnamese farmers, all tests reject the SM regardless of the specifications and the estimation methods. This comes as expected due to the legacy of central planning in Vietnam. Since the tests reject the separation hypothesis, this raises a concern about results of previous studies on impact of trade liberalization in Vietnam, since these studies are based on the SM. It will be interesting to know if the impact of trade liberalization differs if NM is used instead of SM.

References

Abdulai, A., and P. Regmin. 2000. Estimating Labor Supply of Farm Households Under Non-Separability: Empirical Evidence from Nepal. *Agricultural Economics* 22: 309–320.

Barrett, C., S. Sherlund, and A. Adesina. 2008. Shadow Wages, Allocative Inefficiency, and Labor Supply in Smallholder Agriculture. *Agricultural Economics* 38(1): 21–34.

Benjamin, D. 1992. Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica* 60(2): 287–322.

Bhattacharyya, A., and S. Kumbhakar. 1997. Market Imperfections and Output Loss in the Presence of Expenditure Constraint: A Generalized Shadow Price Approach.

American Journal of Agricultural Economics 79(3): 860–71.

Bowlus, A., and T. Sicular. 2003. Moving Toward Markets? Labor Allocation in Rural China. *Journal of Development Economics* 71(2): 561–583.

Buchinsky, M. 1998. Recent Advances in Quantile Regression Models. *Journal of Human Resources* 33(1): 88–126.

de Janvry, A., and Sadoulet, E. 2006. Progress in the Modeling of Rural Households' Behavior Under Market Failures. In *Poverty, Inequality and Development*, ed. A. Janvry and R. Kanbur. New York: Springer.

Deaton, A. 1989. Rice Prices and Income Distribution in Thailand: A Non Parametric Analysis. *Economic Journal* 99: 1–37.

Deaton, A. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: Johns Hopkins University Press.

Eswaran, M., and A. Kotwal. 1986. Access to Capital and Agrarian Production Organization. *Economic Journal* 96: 482–498.

Greene, W. 2003. *Econometric Analysis.* Upper Saddle River, NJ: Prentice Hall.

Grimard, F. 2000. Rural Labor Markets, Household Composition, and Rainfall in Côte d'Ivoire. *Review of Development Economics* 4(1): 70–86.

Jacoby, H. 1993. Shadow Wages and Peasant Family Labor Supply: An Econometric Application to the Peruvian Sierra. *Review of Economic Studies* 60: 903–921.

Le, K. 2009. Shadow Wages and Shadow Income in Farmers' Labor Supply Function. *American Journal of Agricultural Economics* 91: 685–696.

Löfgren, H., and S. Robinson 2002. To Trade or Not to Trade: Non-Separable Household Models in Partial and General Equilibrium. Washington DC: International Food Policy Research Institute.

Lopez, R. 1984. Estimating Labour Supply and Production Decisions of Self-Employed Farm Producers. *European Economic Review* 24: 61–82.

Minot, N., and F. Goletti. 2000. Rice Market Liberalization and Poverty in Vietnam. Research report no. 114. Washington, DC: International Food Policy Research Institute.

Singh, I., L. Squire, and J. Strauss (eds.). 1986. *Agricultural Household Models*. Baltimore: Johns Hopkins University Press.

Skoufias, E. 1994. Using Shadow Wages to Estimate Labor Supply of Agricultural Households. *American Journal of Agricultural Economics* 76(2): 215–227.

Sonoda, T., and Y. Maruyama. 1999. Effect of the Internal Wage on Output Supply: A structural estimation for Japanese Rice Farmers. *American Journal of Agricultural Economics* 81: 131–143.

Taylor, J. E., and I. Adelman. 2003. Agricultural Household Models: Genesis, Evolution and Extensions. *Review of Economics of the Household* 1(1): 33–58.

Van De Walle, D., and D. Cratty. 2003. Is the Emerging Non-farm Market Economy the Route out of Poverty in Vietnam. Policy research paper, no. 2950. Washington, DC: World Bank.

Appendix

Table A1. Descriptive Information about Variables

Variables	Description	Unit	Mean	Std. Dev.
pQ	Farm output value	Dong	10909.1	10881.39
L	Farm labor	Hour	3329.68	2279.83
w	Market wage	Dong/hour	3.52	12.84
A includes:				
	Number of adult males	Person	1.39	0.87
	Number of adult females	Person	1.42	0.75
	Number of children	Person	1.50	1.27

Note: Dong is Vietnamese currency (in thousand).

Table A2. The Benjamin Test Result with Dependent Variable : log(*L*)

Independent Variables	OLS (1)	LAD (2)	IV[b] (3)
Number of adult males	0.1256	0.1386	0.1262
	(0.0149)	(0.0175)	(0.0214)
Number of adult females	0.1584	0.1744	0.2179
	(0.0166)	(0.0202)	(0.0288)
Number of children	0.0115	0.0177	−0.0029
	(0.0102)	(0.0122)	(0.0152)
Log(wage rate)	−0.0195	−0.0201	0.8792
	(0.0174)	(0.0196)	(0.2425)

Table A2. Continued

Independent Variables	OLS (1)	LAD (2)	IV[b] (3)
Log(fertilizer price)	0.0607	0.0657	0.1241
	(0.0739)	(0.0652)	(0.1075)
Log(pesticide price)	−0.0784	−0.0897	−0.0785
	(0.0257)	(0.0296)	(0.0369)
Log(land areas)	0.3515	0.2730	0.3102
	(0.0150)	(0.0168)	(0.0243)
Land quality index	0.0268	0.0225	0.0374
	(0.0105)	(0.0119)	(0.0154)
Log(farm equipment)	0.0451	0.0352	0.0394
	(0.0046)	(0.0050)	(0.0067)
Region[a]			
Red River Delta	0.7791	0.7124	0.8635
	(0.0458)	(0.0549)	(0.0696)
North Central	0.5590	0.4889	0.7964
	(0.0440)	(0.0500)	(0.0897)
Central Coast	0.5874	0.4917	0.8961
	(0.0454)	(0.0535)	(0.1054)
Central Highlands	0.4169	0.3877	0.4873
	(0.0489)	(0.0524)	(0.0728)
Southeast	0.5545	0.5622	0.5455
	(0.0674)	(0.0607)	(0.0969)
Mekong Delta	−0.1269	−0.0871	−0.2314
	(0.0567)	(0.0546)	(0.0862)
Constant	3.8610	4.9716	2.6134
	(0.6219)	(0.5999)	(0.9977)
F-test: Coefficients on preference shifters are simultaneously zeros (df = 3)	61.24	49.3	100.91

Note: [a]Northern Mountain is the benchmark. [b]The instrument for log(wage rate) is the log(commune wage rate).

Table A3. First-Stage Regression with Dependent Variable : log(w)

Variables	Coeff.	Std. error
Household head education	0.0569	0.0144
Proportion of males who are wage earners	0.2372	0.0405
Commune wage	0.1844	0.0335
Types of farm product[a]		
Paddy	−0.0298	0.0400
Other crops	−0.0025	0.0447
Regions[b]		
Red River Delta	−0.0853	0.0485
North Central	−0.2574	0.0459
Central Coast	−0.3024	0.0487
Central Highlands	−0.0588	0.0546
Southeast	0.0645	0.0757
Mekong Delta	0.1244	0.0630
Constant	0.4074	0.0799

Note: [a]Livestock is the benchmark. [b]Northern Mountain is the benchmark.

Table A4. Cobb-Douglas Production Function with Dependent Variable : log(pQ)

Independent Variables	OLS (1)	LAD (2)	IV (3)
Log household labor*	0.1236	0.0738	0.1274
	(0.0113)	(0.0098)	(0.0404)
Log hired labor*	0.0967	0.0940	0.0837
	(0.0088)	(0.0077)	(0.0663)
Log fertilizer cost*	0.3143	0.3042	0.2020
	(0.0129)	(0.0112)	(0.0782)
Log pesticide cost*	0.0783	0.0811	0.0276
	(0.0099)	(0.0086)	(0.0514)
Log other cost*	0.0782	0.0954	0.1309
	(0.0095)	(0.0082)	(0.0268)
Log equipment value	0.0262	0.0272	0.0485
	(0.0069)	(0.0060)	(0.0111)
Log land size	0.2719	0.2800	0.3059
	(0.0123)	(0.0107)	(0.0316)
Land quality index	0.0285	0.0141	0.0334
	(0.0067)	(0.0057)	(0.0094)
Household head age	0.0105	0.0148	0.0111
	(0.0046)	(0.0039)	(0.0054)
Household head age square	−0.0000	−0.0001	−0.0001
	(0.0000)	(0.0000)	(0.0000)
Household head gender	−0.0293	−0.0412	−0.0372
	(0.0219)	(0.0189)	(0.0244)
Household head education			
Primary education	−0.0518	−0.0451	0.0095
	(0.0323)	(0.0278)	(0.0400)
Secondary education	−0.0053	−0.0068	0.0572
	(0.0352)	(0.0304)	(0.0448)
High school and higher	0.0031	−0.0164	0.0519
	(0.0387)	(0.0334)	(0.0508)

Note: Asterisk (*) denotes endogenous variables. The sample used to estimate the production function includes households with no market labor. Dummy variables for year and month of interview, dummy variables for regions, and constant are included in the regression but not shown in the above table. A constant of one is added to all inputs before taking *log* because many kinds of inputs are not used by all households in the sample (avoiding taking *log* of zeros). The instruments for the regression are the number of adults in the household (age 16 and above), the number of children (less than 16), dummy variable for big commune (more than 700 households), and the prices of variable inputs (pesticide, hired labor, fertilizer).

Table A5. The Jacoby Test Result with Dependent Variable: Shadow Wage

Independent Variables	OLS (1)	LAD (2)	IV (3)
Wage	0.0016	0.0012	−0.1564
	(0.0019)	(0.0004)	(0.1678)
Constant	0.5654	0.3315	1.1346
	(0.0289)	(0.0064)	(0.6067)

Note: Standard errors appear in parentheses. The instruments for IV regression include household head education and age.