Are Intrahousehold Allocations Efficient? Evidence from Monte Carlo Simulations

Jorge M. Agüero *
University of California, Riverside

January 22, 2008

Abstract

Models explaining household decisions assume that the member's bargaining process, although unknown, leads to efficient outcomes. The empirical literature has not been able to reject this hypothesis when tested in several datasets, including those from poor, rich and emerging countries. This paper presents Monte Carlo simulations to show that the methods used for testing are inadequate for two reasons. First, the performance of the test statistic, and in particular the Type II error, depends on the algebraic formulation of the restriction to be tested. Second, none of the formulations dominate the other alternatives. These two factors prevent us from concluding whether there is enough evidence supporting the assumption of efficient intrahousehold allocations. Alternative approaches to validate the test are discussed.

Keywords: Collective Models, Nonlinear Tests, Power Function, Multiple Equations.
JEL codes: D13, C15, C12, C30.

---

*Department of Economics, University of California, Riverside, 4108 Sproul Hall, Riverside CA 92521; email:jorgea@ucr.edu. I would like to thank Michael Carter and Kenneth West for their comments and suggestions on an earlier version of this paper. I also benefited from conversations and comments from Maurizio Mazzocco, James Walker, Bruce Hansen and from seminar participants at the University of Wisconsin-Madison, the University of KwaZulu-Natal, the meetings of the American Agricultural Economics Association and LACEA. Michele Back helped me edit this manuscript, however, all remaining errors are my own.

"[The simplicity of the data needed] suggests, more generally, that testing 'collective' models of household behavior may not be as difficult a task as it was sometimes suggested." -Bourguignon, Browning, Chiappori, and Lechene (1993, p. 152).

1 Introduction

Models of household behavior have departed from the unitary framework developed by Becker (1991) towards models where bargaining among members takes place in order to allocate resources such as labor supply (Chiappori, Fortin, and Lacroix 2002), consumption (Attanasio and Lechene 2002) or human capital investments (Thomas 1990). While the exact nature of the bargaining remains unknown and a matter of debate[1], collective models of household behavior are built under one critical assumption: the decision is in the Pareto frontier (Browning, Bourguignon, Chiappori, and Lechene 1994). The empirical evidence can not reject the null hypothesis of efficiency, despite having been evaluated in different regions and in poor and rich countries[2]. Does this mean that intrahousehold allocations are indeed efficient? To answer this question, this chapter analyzes the validity of the methods used to test for efficiency in intrahousehold allocations.

I use Monte Carlo simulations to explore the performance of the statistical tests used in the literature. Because the null hypothesis of efficiency has not been rejected, I explore the power function of the test to measure the Type

[1] See Manser and Brown (1980), McElroy and Horney (1981), and Chiappori (1988, 1992) for alternative approaches.

[2] See Table 1 below.

II error: the probability of failing to reject a false hypothesis.

The efficiency assumption derived from collective models of household behavior imposes a nonlinear restriction on the parameters of the household member's labor supply or consumption demand. Thus, the marginal rate of substitution between spouses' earnings (or between two distributional factors) is the same across goods (or spouses' labor supply functions). As mentioned in the opening quote, the relative simplicity of the data required to implement the test allows researchers to use widely available standard household surveys.

Also, the test can be easily done by using Wald statistics. This statistic does not require a second estimation of the model under the null hypothesis, as opposed to the Likelihood Ratio and Generalized Method of Moments. However, it is this simplicity of the Wald method that complicates deriving robust conclusions about the assumption of efficiency.

There is a vast literature suggesting that this method, while simple to implement, is not adequate for nonlinear restrictions. The nonlinearity of the restriction allows for different, although equivalent, algebraic formulations. For example, testing for $\lambda_1 = 1/\lambda_2$ is equivalent to $\lambda_1\lambda_2 = 1$.[3] Unfortunately, the numerical value of the Wald statistic is not invariant to the choice of the formulation.[4]

I extend the current literature on the performance of the Wald statistic

[3] Under the null, λ_2 can be close, but never equal to zero.

[4] See for example Phillips and Park (1988) and more recently in Dufour (1997) and Hansen (2006).

for nonlinear restrictions by analyzing a model of multiple equations instead of a single equation used by Gregory and Veall (1985), Lafontaine and White (1986) and Hansen (2006). By implementing Monte Carlo simulations for the particular form of the nonlinearity of the restriction imposed by efficient collective models of household bargaining I am able to test the validity of the current findings in the literature.

The results suggest that failing to reject the efficiency of intrahousehold allocations could be generated by the used of an inadequate test. Different algebraic formulations of the Wald statistic have different power functions. Most importantly, I show that none of the explored formulations dominates the other for all specifications and sample sizes.

The rest of the paper is organized as follows. In section 2, I present the restrictions imposed by efficient models of collective bargaining. Section 3 reviews the empirical implementation for testing the model, which implies evaluating a nonlinear restriction. The econometrics of nonlinear restrictions is analyzed in section 4. The Monte Carlo simulations are presented in section 5. Given the low performance of the Wald statistic I also discuss possible alternatives to overcome the low power of those tests that do not require the estimation of the model under the null hypothesis. Section 6 summarizes the findings of this paper.

2 The test for efficient allocations

Here I consider the case of a household formed by only two individuals, indexed by $i = 1, 2$ but the analysis could be extended for more household members.[5] Each member has a twice-continuously differentiable utility function $u_i(\boldsymbol{c}_i, \ell_i)$ where $\boldsymbol{c}_i = (c_i^1, \ldots, c_i^J)$ represents a vector of privately consumed goods indexed by j and ℓ_i is leisure. In the collective model the household problem is to assign consumption and leisure for each member given a vector of prices ($\boldsymbol{p} = p_1, \ldots, p_J$), wages ($w_i$), non-labor income ($y_i$) and their unit of labor endowment. Formally the problem is presented in equation (1)

$$\max_{\boldsymbol{c}_1, \boldsymbol{c}_2, \ell_1, \ell_2} u_1(\boldsymbol{c}_1, \ell_2) \tag{1}$$

$$\text{s.t.} \qquad \boldsymbol{p} \cdot (\boldsymbol{c}_1 + \boldsymbol{c}_2) = y_1 + y_2 + w_1 h_1 + w_2 h_2 \qquad [a]$$

$$u_2(\boldsymbol{c}_2, \ell_2) \geq \overline{u}_2(\boldsymbol{\phi}) \qquad [b]$$

In this problem, restriction $[a]$ is the typical budget constraint.[6] Restriction $[b]$ implies that member 2's utility will not be less than his reservation utility $\bar{u}_2(\boldsymbol{\phi})$ (i.e., the value of his outside option). This reservation utility is allowed to change with the exogenous parameters of the model $(\boldsymbol{p}, w_1, w_2, y_1, y_2)$ but it can also change with other *distribution factors* $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_K)$ such as sex ratios and divorce laws (Chiappori, Fortin, and

[5] The model considered here is static. For dynamic versions of collective models see Mazzocco (2004) and Dubois and Ligon (2005). See also Udry (1996) for examples on the production side.

[6] Throughout this paper I assume the existence of interior solutions.

Lacroix 2002). Hence $\boldsymbol{\phi} = \phi(p, w_1, w_2, y_1, y_2, \zeta)$.

Under the assumption of efficiency, the problem in (1) can be seen as a two-stage problem (Chiappori 1992). In the first stage the household decides how to distribute its non-labor income. The solution is a set of transfers $\boldsymbol{\theta} = (\theta_1, \theta_2)$. The transfers depend on the parameters of the model so $\theta_i = \theta_i(\phi)$ and they add up to the household non-labor income so $\theta_1 + \theta_2 = Y = y_1 + y_2$. In the second stage, each member maximizes their own utility given the prices (including wages) and the transfers. The solution of this problem is a set of household demands and labor supply equations expressed also as a function of the transfers θ_i for $i = 1, 2$ as follows:

$$c^j = c_1^j + c_2^j = g_1^j(\boldsymbol{p}, w_1, w_2, Y, \theta_1) + g_2^j(\boldsymbol{p}, w_1, w_2, Y, \theta_2) \qquad j = 1, \ldots, J \quad (2)$$

$$1 - \ell_i = h_i(\boldsymbol{p}, w_1, w_2, \theta_i) \qquad i = 1, 2 \quad (3)$$

where the transfers θ_i are not constant but functions as defined above.

The test of efficient intra-household allocations is derived by evaluating the impact of non-labor income (or any other distribution factor) on the demand functions or individual labor supply function in equations (2) and (3), respectively.

For example, the test can be constructed by looking at the ratio of (non-labor) income effects $(\partial c^j / \partial y_i)$ for $i = 1, 2$ holding Y constant. Note that $\frac{\partial c^j}{\partial y_1} = \frac{\partial \theta_1}{\partial y_1} \left[\frac{\partial g_1^j}{\partial \theta_1} - \frac{\partial g_2^j}{\partial \theta_1} \right]$ and the expression in brackets cancels out when divided by $\partial c^j/\partial y_2$. Thus, efficient allocations imply

$$\eta_j = \frac{\partial c^j/\partial y_1}{\partial c^j/\partial y_2} = \frac{\partial \theta_1/\partial y_1}{\partial \theta_1/\partial y_2}. \tag{4}$$

The test relies on the fact that the right hand side of (4) does not depend on commodity j, hence should be the same across consumption demands as follows,

$$\eta_1 = \frac{\partial c^1/\partial y_1}{\partial c^1/\partial y_2} = \frac{\partial c^j/\partial y_1}{\partial c^j/\partial y_2} = \eta_j \qquad \forall j = 2, \dots, J \tag{5}$$

As will be clarified below, some studies, suspecting the endogeneity of non-labor income, used distributional factors to test for efficiency. In that case condition (5) is given by

$$\eta_j = \frac{\partial c^j/\partial \zeta_1}{\partial c^j/\partial \zeta_2} = \frac{\partial \theta_1/\partial \zeta_1}{\partial \theta_1/\partial \zeta_2} \qquad \forall j \tag{6}$$

Thus, as in equation (4), η_j represents the marginal rate of substitution between both member's earnings or between two distributional factors in the sharing rule (i.e., the right hand side of equations (4) and (6)).

This fact has been exploited by Chiappori, Fortin, and Lacroix (2002), Browning, Bourguignon, Chiappori, and Lechene (1994) and Bourguignon, Browning, Chiappori, and Lechene (1993) to recover the sharing rule (which is not observable) up to a scale transformation.

A rejection of (5) or (6) is considered evidence against the hypothesis that allocations are efficient. Otherwise, we cannot reject the efficiency of intra-household allocations. The intuition for this result is that member's earnings (holding total earning constants) or distributional factors affect the labor supply or consumption decisions through only the location chosen on the Pareto frontier.[7]

The requirements on the type for data to test for efficient intrahousehold allocations is not extreme. As discussed by Bourguignon, Browning, Chiappori, and Lechene (1993) leading to the opening quote of the chapter, all that is needed is a cross sectional survey with information about consumption. The data does not need to have a complete labor supply model and the test can be performed with or without assignable goods. This simplicity, together with the relevance of opening the black box about how household decisions are made, has motivated researchers to test this hypothesis in countries around the world.[8] In the next section I briefly summarize evidence from papers testing for the efficiency of intrahousehold allocations.

3 Evidence of efficiency tests

Most studies implement the test for efficiency introduced above using a set of equations on labor supply or Engel curves. Without loss of generality let

[7]The model in (1) can be expressed also as problem where a weighted average of each member's utility function is maximized subject to the household budget constraint in [a], where the weights are given by the sharing rule θ_i. In this case, an equivalent interpretation of (5) is that member's earnings or the distributional factors affect the household decision "through the implicit weighting of each spouse's utility" (Chiappori, Fortin, and Lacroix 2002, p. 44).

[8]However, it is clear that when using non-labor income to test for efficiency instrumental variable methods are preferred due to endogeneity problems.

us assume that the linear system has two equations as follows, for a sample of n households $i = (1, \ldots, n)$

$$
\begin{aligned}
x_i^1 &= \alpha_1 + \beta_1 s_i^1 + \beta_2 s_i^2 + \delta_1 z_i + \varepsilon_i^1 \\
x_i^2 &= \alpha_2 + \beta_3 s_i^1 + \beta_4 s_i^2 + \delta_2 z_i + \varepsilon_i^2
\end{aligned}
\tag{7}
$$

where x_i^1 and x_i^2 could represent individual labor supply or the household consumption of two goods, z_i refers to vector of variables affecting the decision such as prices, individual wages and other preference parameters (such as age, education, household composition, etc.). The parameters of interest are $\beta_1, \beta_2, \beta_3$ and β_4 as they represent the marginal impact of s_i^1 and s_i^2. These two variables can be seen as distribution factors or non-labor income as discussed above.

Testing the hypothesis of efficiency implies $\eta_1 = \eta_2$ from equation (4). Using the model in (7) this is represented by evaluating the null hypothesis

$$
\begin{aligned}
\mathrm{H}_0: &\quad \frac{\beta_1}{\beta_2} = \frac{\beta_3}{\beta_4} \\
\mathrm{H}_1: &\quad \frac{\beta_1}{\beta_2} \neq \frac{\beta_3}{\beta_4}
\end{aligned}
\tag{8}
$$

Table 1 below presents a summary of the literature testing for the null hypothesis of efficiency in the household using the collective model.

As shown there, the test for efficiency has been implemented in rich and poor countries, from Asia to Europe, from Africa to North America. It has also been done using consumption goods, labor supply and health-related outcomes. Notably, in all cases the null hypothesis of efficient intrahousehold allocation has not been rejected once.

It is important to note that all but one study uses the Wald statistic to evaluate the null hypothesis of efficiency. As explained below this choice might be driven by the simplicity of the test: only one (unrestricted) estimation of the model is needed. This contrasts with the requirement of other statistics where estimations of restricted and unrestricted models are needed.

The nonlinearity of the restriction makes it possible to be written using two equivalent algebraic expressions. Let $g_R(\beta) = \beta_1/\beta_2 - \beta_3/\beta_4$ the ratio-type expression and $g_M(\beta) = \beta_1\beta_4 - \beta_3\beta_2$ the multiplicative version. Most papers in Table 1 used $g_R(\beta)$. Only Thomas, Contreras, and Frankenberg (2002) and Rangel and Thomas (2005) used $g_M(\beta)$. They do so because they are concerned about the performance of nonlinear restrictions of the Wald statistics. As I will show later, neither of the two algebraic expressions evaluated for the hypothesis of efficiency is preferred over the other when considering the power of the test.

Table 1: Empirical evidence for the hypothesis of efficient intrahousehold allocation

Author	Data	Commodity	Variable	Method	Formula	Result
Bourguignon et al (1993)	France	Consumption	NL income§	Wald	Ratio	Accepted
Thomas and Chen (1994)	Taiwan	Consumption	NL Income§	Wald	Ratio	Accepted
Chiappori, Fortin, and Lacroix (2002)	USA	Labor supply	Sex ratio†	GMM	n.a.	Accepted
Thomas, Contreras, and Frankenberg (2002)	Indonesia	Health	Assets‡	Wald	Multiplicative	Accepted
Quisumbing and Maluccio (2003)	Bangladesh, Indonesia Ethiopia and South Africa	Consumption	Assets‡	Wald	Ratio	Accepted
Rangel and Thomas (2005)	Ghana and Senegal	Consumption	Land	Wald	Multiplicative	Accepted

§NL Income=Non-labor income. †Sex ratio and divorce laws. ‡Assets at marriage

4 The econometrics of nonlinear tests

The Wald test is a popular choice among researchers because of its simplicity. It requires to estimate the model only once, unlike the Likelihood Ratio or Generalized Method of Moments (GMM) statistics where the parameters need to be estimated for the restricted and unrestricted models. However, its simplicity comes with a cost in the case of nonlinear restrictions. This is important because testing for the efficiency of intrahousehold allocations relies on a nonlinear function of the parameters as shown in equation (8).

Consider the following example for a nonlinear restriction for a single equation. Let $H_0 : g_R(\lambda) = \lambda_1 - 1/\lambda_2 = 0$ and an equivalent formulation using the multiplicative form $H_0 : g_M(\lambda) = \lambda_1\lambda_2 - 1 = 0$. Gregory and Veall (1985) present Monte Carlo evidence where the size of the test differs for different formulations of the null hypothesis. Using simulations for $g_M(\lambda)$ and $g_R(\lambda)$, the authors show that the size of the test is sensitive to the choice of the algebraic formulation of the nonlinear restriction. The first formulation $g_R(\beta)$ has a systematic higher Type I error compared to the second one ($g_M(\beta)$). The authors present limited information about the power of the test. They conclude that no formulation is preferred regarding Type II error.[9]

[9] The differences in the numerical value (and the distribution) of the test is related to the fact that as λ_2 gets closer to zero there is an approximate violation of the continuity assumption for the test. Note that under the null, λ_2 can approach but never be equal to zero, so the problem is rooted at the continuity problem more than the non-existence of the derivative at $\lambda_2 = 0$.

In a related paper, Lafontaine and White (1986), use Monte Carlo simulations to show that it is possible to obtain any Wald statistic to either reject or not the null hypothesis by merely altering the algebraic formulation of the same nonlinear restriction. Phillips and Park (1988) use Edgeworth expansions to formally show that the Wald statistic is not invariant to the formulation of the nonlinear restriction. This property has also been studied by Critchley, Marriott, and Salmon (1996). These papers have mainly focused on the implications for single equation models.[10]

Nelson and Savin (1990) point out an additional possible problem with nonlinear restrictions using Wald statistics. In particular, they show that the power of test can be nonmonotonic: it increases at small deviations from the true value but decreases for higher deviations.

The particular nature of the restriction derived from the efficient collective model of household behavior requires multiple (instead of single) equations and a nonlinear restriction not commonly analyzed in the literature. Below I describe the model used to produce a Monte Carlo experiment for the specific case of testing efficiency in the allocations within the household.

5 Monte Carlo simulations

To show the size and power performance of the Wald statistic with nonlinear restrictions I use the following model

[10]Dagenais and Dufour (1991) proposed a formal proof for more general frameworks.

$$x_i^1 = \alpha_1 + \beta_1 s_i^1 + \beta_2 s_i^2 + \varepsilon_i^1 \tag{9}$$

$$x_i^2 = \alpha_2 + \beta_3 s_i^1 + \beta_4 s_i^2 + \varepsilon_i^2 \tag{10}$$

with $E(\varepsilon_i^1|s_i) = E(\varepsilon_i^2|s_i) = 0$. In the simulations I generate $s_i^1, s_i^2, \varepsilon_i^1$ and ε_i^2 mutually independent, iid, and distributed $N(0,1)$. This specification is the multiple equation counterpart of Gregory and Veall (1985) and Hansen (2006). Without loss of generality, in the simulations I set $\alpha_1 = 0 = \alpha_2$.

I consider two formulations of the Wald statistic based on the following two hypotheses

$$H_0^R : g_R(\beta) = 0 \qquad \text{with } g_R(\beta) = \frac{\beta_1}{\beta_2} - \frac{\beta_3}{\beta_4}$$

and

$$H_0^M : g_M(\beta) = 0 \qquad \text{with } g_M(\beta) = \beta_1\beta_4 - \beta_3\beta_2.$$

W^M and W^R, respectively, denote the corresponding Wald statistics to the two formulations of the null hypothesis where

$$W^t = g_t(\hat{\beta})'(G_t' V_n \hat{G}_t)^{-1} g_t(\hat{\beta}) \qquad t = M, R \tag{11}$$

$$\hat{G}_t = \frac{\partial}{\partial\beta} g_t(\hat{\beta}).$$

The estimates were constructed using the Eicker-White covariance matrix following what is commonly done in the empirical application of collective models.[11] Similar to Hansen (2006) I calculate the finite sample size of asymptotic 5% tests, with sample sizes $n = \{20, 30, 50, 100, 500, 1000\}$, from 100,000 Monte Carlo replications, hence the standard error for the estimated rejection frequencies is close to 0.0007.

5.1 Size

I first evaluate the size of the test. The simulations are conducted using equations (9) and (10) imposing the nonlinear restriction that $\beta_1\beta_4 - \beta_3\beta_2 = 0$. The results are presented in Table 2.

First, in almost all specifications the size of the distortion is statistically different from 5%. With a .0007 standard error for the replications, frequencies outside the .049 and 0.51 interval can be considered as distorted. Also, unlike the single-equation case, the distortions are on both sides: some specifications reject more than 5% of the cases and other much less than 5%. Hence, the Type I error is of considerable magnitude.

Second, as expected, the Wald statistic is not invariant to the formulation of the test. This discrepancy holds also as we increase the sample size, with the only exception when all the parameters are equal to one.

Third, W^M performs better than W^R for all specifications as the sample

[11]Likewise in (11), $V_n = (S'S)^{-1}\hat{\Omega}_n(S'S)^{-1}$, where $S = I_2 \otimes s$ with $s = [s^1, s^2]$; $\hat{\Omega}_n = \sum_{i=1}^{n} s_i s_i' \hat{\varepsilon}_i^2$ with $\hat{\varepsilon}_i = x_i - s_i'\hat{\beta}$, where $\hat{\beta}$ are the OLS estimates equation by equation.

Table 2: Size: Percentage Rejection at the 5% Asymptotic Level

Case	$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$	Test	$n = 20$	$n = 30$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
I	$\beta_1 = 1, \beta_2 = 1$	W^M	.116	.093	.073	.062	.052	.052
	$\beta_3 = 1, \beta_4 = 1$	W^R	.061	.055	.052	.050	.050	.051
II	$\beta_1 = 2, \beta_2 = 0.1$	W^M	.125	.097	.078	.067	.053	.051
	$\beta_3 = 12, \beta_4 = 0.6$	W^R	.212	.205	.196	.164	.096	.081
III	$\beta_1 = 3, \beta_2 = 0.6$	W^M	.122	.097	.077	.063	.052	.051
	$\beta_3 = 2, \beta_4 = 0.4$	W^R	.018	.013	.014	.022	.039	.044
IV	$\beta_1 = 2.5, \beta_2 = 0.5$	W^M	.120	.095	.076	.063	.052	.052
	$\beta_3 = 0.5, \beta_4 = 0.1$	W^R	.153	.155	.151	.137	.091	.077
V	$\beta_1 = 1, \beta_2 = 0.05$	W^M	.113	.092	.075	.063	.053	.052
	$\beta_3 = 1.5, \beta_4 = .075$	W^R	.042	.032	.026	.019	.007	.002

Note: The frequencies were constructed from 100,000 replications.

size increases. Recall that Thomas, Contreras, and Frankenberg (2002) and Rangel and Thomas (2005) used W^M citing evidence from the single-equation literature to choose W^M over W^R. Note also that as β_4 decreases, W^M rejects the null hypothesis just above the 5% level, while W^R either overrejects or underrejects depending on the specification.

However, for the intrahousehold literature is the power, not the size of the test that matters because the testing for the efficiency in the allocations requires a no rejection of the null hypothesis. Simulations for the power function are presented next.

5.2 Power function

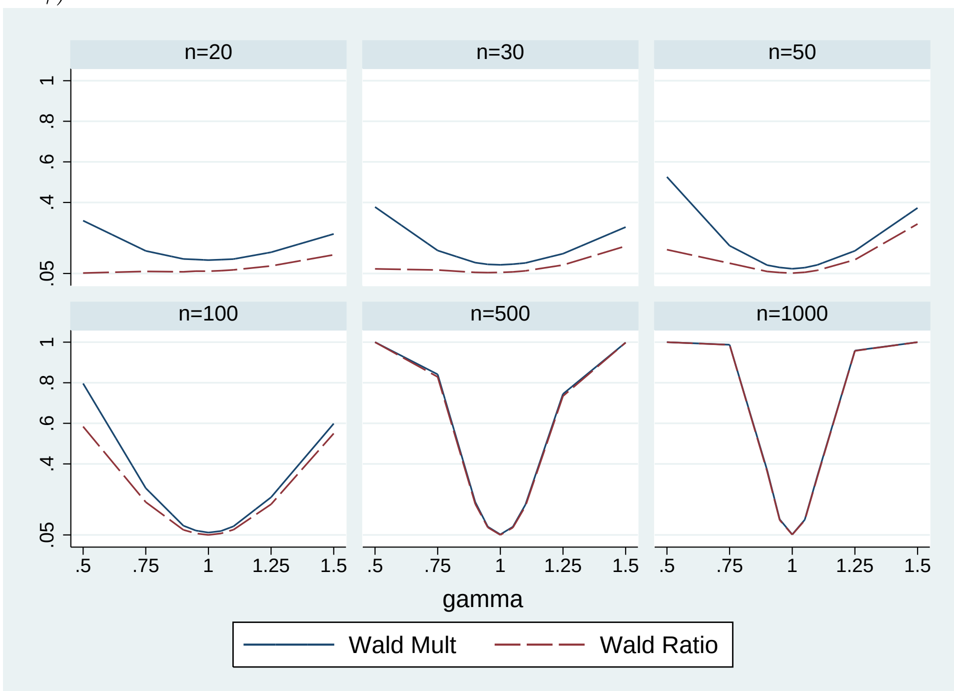
To evaluate the power (Type II error) of the nonlinear test I simulate the model described in (9) and (10) where $\beta_1\beta_4 - \beta_3\beta_2 \neq 0$. This is obtained by generating data where, by construction, the null hypothesis of efficiency does not hold. Deviations from the null hypothesis are generated by changing the parameter β_4 from the values on Table 2, keeping the other three parameters constant. In particular the parameter β_4 was scaled up (and down) by $\gamma \in (0.5, 0.75, 0.9, 0.95, 1.0, 1.05, 1.1, 1.25, 1.5)$. Clearly, when $\gamma = 1$ the simulations are the same as when evaluating the size of the test. For example, in Case IV the set of parameters is $\beta_4 = 0.1\gamma$ but with $\beta_1 = 2.5, \beta_2 = 0.5$ and $\beta_3 = 0.5$ (kept fixed) as in Table 2.

The results of these simulations are presented in Figures 1 to 3 where the frequency of rejections of the null hypothesis (vertical axis) is plotted against different values of γ (horizontal axis)[12]. For the test to have high power the rejection should increase rapidly as we deviate from the $\gamma = 1$ (where the null hypothesis of efficiency holds). Failure to do so will imply a failure to reject the efficiency hypothesis when we know, by construction, that it does not hold.

The main result is that for all specifications, the power of the test varies with the formulation of the null hypothesis. The exceptions are found in Figure 1 where $\beta_4 = \gamma$ but just for $n \geq 500$, for all other cases, the choice

[12]Due to space limitation, the power function for Cases II and V are not included here but reinforce the findings of this section.

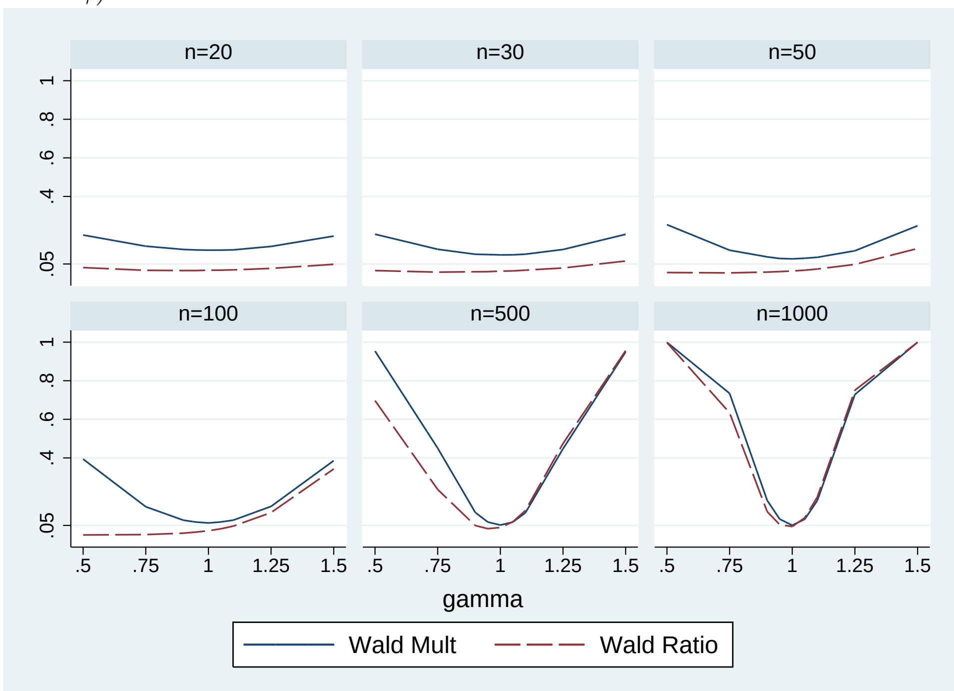
Figure 1: Power function: Rejections at the 5% Asymptotic Level (Case I, $\beta_4 = \gamma$)



of the algebraic expression for the efficiency test may alter the conclusion of the test.

A second result from these simulations, and unlike the discussion for the size of the test, is that W^M (the multiplicative form) is not always the preferred formulation compared to W^R (the ratio form). In particular, for small β_4 W^M dominates W^R but this advantage decreases as the sample size increases (figures 2 and 3). When $n \geq 500$ and for $\gamma > 1$ it is W^R that has a smaller Type II error. Only when all β's are equal to one, Figure 1 shows that W^M is preferred to W^R but just for small samples. For $n \geq 500$ that difference disappears.

Figure 2: Power function: Rejections at the 5% Asymptotic Level (Case III, $\beta_4 = 0.4\gamma$)

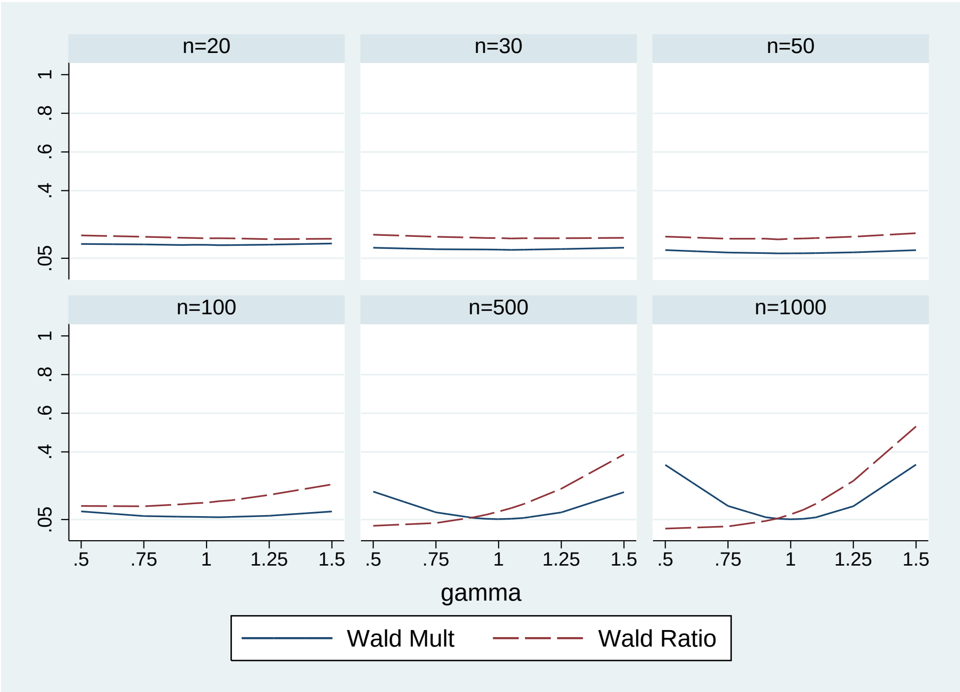


Also, it is important to note that the magnitude of the Type II error is not symmetric. W^R tends to show a lower Type II error when $\gamma > 1$ while the strength of W^M appears to take place for $\gamma < 1$.

Finally, for the range of deviations from the null hypothesis, $\gamma \in (0.5, 1.5)$, the percentage of rejections barely reaches the 95% level. That level is reached only when $n = 500$ and when $\beta = \gamma$ (Case I) or $\beta = 0.4\gamma$ (Case III), while for $\beta = 0.1\gamma$ (Case IV) the percentage of rejections reached was found in only 40% of the replications with $n = 500$ and 53% with $n = 1000$.

The implications of these results for testing the null hypothesis of efficiency in intrahousehold allocations are discussed next.

Figure 3: Power function: Rejections at the 5% Asymptotic Level (Case IV, $\beta_4 = 0.1\gamma$)



5.3 Discussion

The Monte Carlo simulations shown above suggest that we can not conclude that failing to reject the null hypothesis of efficient intrahousehold allocations actually implies that the allocations are indeed efficient. The Wald statistic, which is widely used to evaluate the null hypothesis, is not invariant to the formulation of the restriction.

Most importantly, from this experiment it is clear that there is no one formulation that can be considered better "behaved" when considering the Type II error of the test. This could explain why we do not see any difference in the outcome of the test from those authors using the W^R formulation compared to the few using W^M.

How then can we adequately test for the efficiency hypothesis? I consider here alternatives that do not require re-estimation of the model under the null hypothesis. The first and simplest alternative is to perform both formulations to understand how sensitive the results are. Second, Andrews (1989) proposes a methodology to check the power of the test performed introducing what he calls an *inverse power function*.

Suppose that one is interested in testing the null hypothesis $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$. The first step is to find a region with threshold $c > 0$ such that $\{\lambda : |\lambda| > c\}$. When the null hypothesis cannot be rejected it implies that with significant level, say 5%, $|\lambda|$ is less than c. If c is "close" to zero, then the test provides evidence that $|\lambda|$ is zero or "close enough," as desired.

Then Andrews proposes to find the region where the probability of Type II error is high, for example bigger than 0.5. That region is defined by $\{\lambda : 0 < |\lambda| \leq b\}$ for $b \in (0, c)$. Andrews's (1989) contribution is to tabulate the values for b and c for different tests, including the Wald statistic. The implementation is simple as one only needs an estimate of $g(\hat{\beta})$ and its standard error. With this information one can compute b and c and determine whether these values (b and c) are "close enough" to the desired value.

Another possibility is to calibrate the Monte Carlo experiment to match the moments of the data under study. In that sense the researcher can evaluate how and whether the performance of the different formulations of the null hypothesis affect the performance of the Wald statistics for their particular case under study. The difficulty to extend those results to other contexts precluded this paper to explore such an option, as the goal is to have a general understanding of the behavior of the Wald statistics.

Also, one could use bootstrap methods to derive the 1%, 5% and 10% critical values for either formulation of the null hypothesis. This will have the advantage of being tailored to the data under study and does not require the computation of the standard error of $g(\hat{\beta})$ as in the case using Andrews's (1989) inverse power function. This proposal is somehow close to the *"corrected" Monte Carlo distribution* discussed by Lafontaine and White (1986). Their method finds correct critical values via replications of the simulated data. Finally, one could use the Lagrange Multiplier test, which requires estimating the model only once: under the null hypothesis. Hence, the test is invariant to the formulation of the hypothesis as desired.

5.4 Approaches with instrumental variables

The simulations used here assumed that the regressors s_i^1 and s_i^2 are exogenous. However, in the empirical literature some few articles recognize the endogeneity of these or other variables. Then instrumental variables (IV) are used to estimated the parameters of the model. This paper does not evaluate the performance of the Wald statistics when IV methods are used.

Nonetheless, it is possible to infer that the performance will be close to the evidence presented here, if not worse. Dufour (1997) notes that IV estimates can be seen as nonlinear combinations of the underlying parameters of the model.[13] As the parameters of the first stage approaches zero (as in the case of weak instruments) the performance of the test on the parameters of interest could suffer from the same problems as discussed earlier in the case of the Wald statistic for single equations with nonlinear restrictions. Hence, one should expect that the results shown here will hold in the case of IV.

In the case when regressors are not exogenous, GMM methods can be considered (Newey and West 1987). In the literature of intrahousehold allocations, they have been implemented only by Chiappori, Fortin, and Lacroix (2002). GMM statistics have the advantage of being invariant to the formulation of the hypothesis (Hansen 2006), however it requires two estimations: one for the unrestricted model and one for the restricted one. While Dufour's (1997) paper did not address the case of GMM estimations, one could suspect that his conclusion may apply to this case as well. However, note that Hansen (2006) shows that theoretically GMM statistics are invariant to the formulation of the null hypothesis.

6 Conclusions

This paper critically reviews the existing literature on intrahousehold allocations. The challenge to the efficiency hypothesis from this paper has been based on methodological grounds. However there is a growing literature sug-

[13]I would like to thank Bruce Hansen for this reference.

gesting that the assumption of efficiency might not be correct due to the existence of information asymmetries between spouses (Ashraf 2005, Goldstein and Udry 1999).

The empirical work evaluating the null hypothesis of efficient collective household models cannot reject it, even when tested in poor, rich or emerging countries and regions.

I argue in this paper that such a consistent result can be explained by the choice of the statistic used to test the nonlinear hypothesis implied by the model. In particular, I show that the Type II error of the Wald statistic for nonlinear restrictions in multiple equations depends on the algebraic formulation of the null hypothesis. But unlike the results from the Type I error, none of the formulations considered behaves better in all settings. This finding reduces the certainty about the efficiency of household decisions observed in the literature.

The results presented here come from Monte Carlo simulations. The next step is to take existing data and compute the simulations using the moments from that data and then evaluate the performance of the test. This could be complemented by using the Lagrange Multiplier statistic in addition to the Wald statistic. The econometric approach of this paper requires a theoretical model of household decisions allowing for asymmetric in order to complement the findings of this paper.

The paper also presents possible solutions in order to evaluate, for the particular data under study, whether the power of the test used is valid. I also proposed bootstrap methods to derive the "correct" critical values for the test. All of these alternatives can be easily implemented and do not require additional estimation of the model under the alternative. I also proposed the use of the Lagrange multiplier test that unlike the Wald statistic is invariant to the formulation of the restriction but it is computed only once (under the null). Whether using these proposed alternatives or other GMM-based statistics used to test the hypothesis of efficiency, it seems clear that testing 'collective' models of household behavior may not be as *simple* a task as has been suggested.

References

Andrews, D. W. K. (1989): "Power in Econometric Applications," *Econometrica*, 57(5), 1059–90.

Ashraf, N. (2005): "Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines," Mimeo, Harvard Univeristy.

Attanasio, O., and V. Lechene (2002): "Tests of Income Pooling in Household Decisions," *Review of Economic Dynamics*, 5(4), 720–748.

Becker, G. S. (1991): *A Treatise on the Family*. Harvard University Press, Cambridge, 2nd edn.

Bourguignon, F., M. Browning, P.-A. Chiappori, and V. Lechene (1993): "Intra Household Allocation of Consumption: A Model and Some Evidence from French Data," *Annales d'Economie et Statistiques*, 29, 137–156.

Browning, M., F. Bourguignon, P.-A. Chiappori, and V. Lechene (1994): "Income and Outcomes: A Structural Model of Intrahousehold Allocation," *Journal of Political Economy*, 102(6), 1067–96.

Chiappori, P.-A. (1988): "Rational Household Labor Supply," *Econometrica*, 56(1), 63–90.

——— (1992): "Collective Labor Supply and Welfare," *Journal of Political Economy*, 100(3), 437–67.

Chiappori, P.-A., B. Fortin, and G. Lacroix (2002): "Marriage Market, Divorce Legislation, and Household Labor Supply," *Journal of Political Economy*, 110(1), 37–72.

Critchley, F., P. Marriott, and M. Salmon (1996): "On the Differential Geometry of the Wald Test with Nonlinear Restrictions," *Econometrica*, 64(5), 1213–22.

Dagenais, M. G., and J.-M. Dufour (1991): "Invariance, Nonlinear Models, and Asymptotic Tests," *Econometrica*, 59(6), 1601–15.

Dubois, P., and E. Ligon (2005): "Incentives and Nutrition for Rotten Kids: Intrahousehold Food Allocation in the Phillipines," Working paper, University of California at Berkeley.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1388.

Goldstein, M., and C. Udry (1999): "Gender and Land Resource Management in Southern Ghana," Mimeo, Yale University.

Gregory, A. W., and M. R. Veall (1985): "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 53(6), 1465–68.

Hansen, B. (2006): "Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, ed. by D. Corbae, S. N. Durlauf, and B. E. Hansen, pp. 9–35. Cambridge University Press.

Lafontaine, F., and K. White (1986): "Obtaining Any Wald Statistic You Want," *Economics Letters*, 21, 35–40.

Manser, M., and M. Brown (1980): "Marriage and Household Decision-Making: A Bargaining Analysis," *International Economic Review*, 21(1), 31–44.

Mazzocco, M. (2004): "Household Intertemporal Behavior: a Collective Characterization and a Test of Commitment," Mimeo, University of Wisconsin-Madison.

McElroy, M. B., and M. J. Horney (1981): "Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand," *International Economic Review*, 22(2), 333–349.

Nelson, F. D., and N. E. Savin (1990): "The Danger of Extrapolating Asymptotic Local Power," *Econometrica*, 58(4), 977–81.

Newey, W. K., and K. D. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28(3), 777–87.

Phillips, P. C. B., and J. Y. Park (1988): "On the Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica*, 56(5), 1065–83.

Quisumbing, A. R., and J. A. Maluccio (2003): "Resources at Marriage and Intrahousehold Allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa," *Oxford Bulletin of Economics and Statistics*, 65(3), 283–327.

Rangel, M., and D. Thomas (2005): "Out of West Africa: Evidence on the Efficient Allocation of Resources within Farm Households," Harris School Working Paper 05.15, University of Chicago.

Thomas, D. (1990): "Intra-Household Resource Allocation: An Inferential Approach," *Journal of Human Resources*, 25(4), 635–664.

Thomas, D., and C. Chen (1994): "Income Shares and Shares of Income: Empirical Tests of Models of Household Resource Allocations," Working Paper 94-08, RAND.

Thomas, D., D. Contreras, and E. Frankenberg (2002): "Distribution of Power Within the Household and Child Health," Working paper, UCLA.

Udry, C. (1996): "Gender, Agricultural Production, and the Theory of the Household," *Journal of Political Economy*, 104(5), 1010–46.