971204

THE HAUSMAN-MACURDY CONTROVERSY

Why do results differ between studies?*

By
Matias Eklöf** and Hans Sacklén ***

Abstract

The two perhaps most influential empirical labor supply studies carried out in the U.S. in recent years, Hausman (1981) and MaCurdy, Green & Paarsch (1990), report sharply contradicting labor supply estimates. In this paper we seek to uncover the driving forces behind the seemingly irreconcilable results. Our findings suggest that differences with respect to the estimated income and wage effects can be attributed to the use of differing nonlabor income and wage measures, respectively, in the two studies. Monte Carlo experiments suggest that the wage measure adopted by MaCurdy et al might cause a severely downward biased wage effect such that data falsely refute the basic notion of utility maximization.

Keywords: Labor supply, Slutsky condition, maximum likelihood estimation
JEL Classification: J22

---

* Helpful comments on previous drafts from Sören Blomquist, Jerry Hausman, Arthur van Soest and seminar participants at Uppsala university are greatly appreciated.
** Uppsala University, Department of Economics, Box 513, S-751 20 Uppsala, phone: +46 (0)18 18 76 36, fax: +46 (0)18 18 14 74, email: matias.eklof@nek.uu.se
*** Trade Union Institute For Economic Research (FIEF), Wallingatan 38, S-111 24 Stockholm, phone: +46 (0)8 69 69 914, fax: +46 (0)8 20 73 13, email: h.sacklen@fief.se

1 Introduction

Empirical labor supply studies using either of the two dominating estimation techniques - the Hausman method[1] or some linearization technique with instrumental variables - have generated results that represent widely differing views on the work disincentive effects and dead weight losses associated with progressive income taxes. Of course, variations in the quality of data and the underlying theoretical framework may explain part of the divergence in results across studies. In general, estimators possess different small sample properties and may react differently to various types of measurement errors in variables. Thus, even if studies rely on a common data source and theoretical framework, a certain divergence in estimates obtained by various empirical methodologies is presumably just what might be expected.[2]

The two perhaps most influential empirical labor supply studies carried out in the U.S. in recent years, Hausman (1981) and MaCurdy, Green & Paarsch (1990), report quite irreconcilable results. This is rather more puzzling, since the two studies applied the same model specification (a linear supply function with random income effects), the same estimation technique (maximum likelihood) and also collected data from the same source (the 1975 PSID cross-section). Without having to constrain the estimated parameters, the results for prime aged married men presented by Hausman suggested that data could have been generated by utility maximization with globally convex preferences. Specifically, the results indicated a negligible wage effect and large income responses. In contrast, the results reported by MaCurdy et al. indicated trivial income responses

1 Originally developed by Burtless and Hausman (1978), the so called Hausman method takes account of the complete form of individuals' budget constraints and uses maximum likelihood techniques to estimate the parameters of the supply function.

2 It should be noted that there is an ongoing debate on how to reconcile differences in estimates obtained using linearization techniques or the Hausman method. The surveys by Hausman (1985) and Pencavel (1986) suggest that the Hausman method in general has produced larger uncompensated wage effects and more negative income effects. MaCurdy et al. (1990) argues that this is because the Hausman method implicitly imposes restrictions which essentially guarantee parameter estimates that are consistent with utility maximization with globally convex preferences. In contrast, Blomquist (1995) argues that there are no particular constraints inherent in the Hausman method that can explain differences in results obtained by the Hausman method and the linearization method. It is beyond the scope of our paper to pursue this issue any further.

and a wage effect that had to be constrained to a nonnegative value in order not to violate the Slutsky condition.

The divergent empirical results reported by Hausman and MaCurdy et al. have, for several reasons, attracted widespread attention in recent years. First, as described above, it has been noted that these studies represent sharply contradicting views on the size of the work disincentive effects induced by progressive taxation and the consistency of data with the economic theory of consumer choice. Second, MaCurdy et al. claimed that they had attempted to replicate Hausman's original study as closely as possible. The fact that Hausman's results could not be replicated by MaCurdy et al. has since then raised questions about the reliability of Hausman's original findings, and hence on much of the evidence offered to support proposals aimed at lowering marginal tax rates in the U.S. during the 1980's since Hausman (1981) constituted the perhaps most frequently cited source.

However, a careful reading of the two studies reveals that there are several striking dissimilarities between the data sets constructed by Hausman and MaCurdy et al., respectively. In fact, not only did they apply separate sample selection criteria, they also used widely differing measures for all the key economic variables. In this paper we try to pinpoint the crucial differences between these two influential studies and explain why they generated such irreconcilable results. Using the same estimation procedures as Hausman and MaCurdy et al., our findings highlight the sensitivity of parameter estimates to variations in variable definitions and sample selection criteria, and we demonstrate how different strategies may generate sharply dissimilar results and conclusions.

The paper is organized as follows. Section 2 contains a brief description of the 1975 U.S. income tax system. In section 3 we discuss sample selection procedures and variable definitions. Section 4.1 contains the econometric specification, sections 4.2-4.3 the empirical results, and in section 4.4 we perform a set of Monte Carlo simulations to study how various types of measurement errors in the pretax hourly wage rate affect parameter estimates. Finally, section 5 concludes the paper.

2 Taxes

Since there is an extensive presentation of the 1975 U.S. tax system in Hausman (1981) and MaCurdy et al. (1990) - henceforth referred to simply as Hausman and MaCurdy - we will just briefly recapitulate the main features of the tax system and the different simplifying assumptions adopted by Hausman and MaCurdy.

The federal income tax system is represented by thirteen income brackets with marginal tax rates increasing from 14 to 50%. Table 1 displays the income brackets and the corresponding federal tax rates for married couples filing jointly. In addition to the income tax there is a social security tax levied at a rate of 5.85% on earned income up to $14,100. The tax system includes two sets of tax credits, a nonrefundable $30 personal credit and a refundable earned income credit. If the earned income credit exceeds tax payments, the remainder is repaid as a subsidy to the household.

Table 1. Federal income brackets and marginal tax rates for married couples filing jointly

Taxable income (dollars)	Marginal tax rate
0 - 1,000	0.14
1,000 - 4,000	0.16
4,000 - 8,000	0.19
8,000 - 12,000	0.22
12,000 - 16,000	0.25
16,000 - 20,000	0.28
20,000 - 24,000	0.32
24,000 - 28,000	0.36
28,000 - 32,000	0.39
32,000 - 36,000	0.42
36,000 - 40,000	0.45
40,000 - 44,000	0.48
44,000 and over	0.50

The federal standard deduction equals $1,900 for incomes up to $11,875 and $2,600 for incomes above $16,250. Within this interval the deduction is proportional to the income at a rate of 16%. As an alternative to the standard deductions, individuals are allowed to claim an itemized deduction. However, in this paper we follow MaCurdy and ignore this feature.[3] In addition to the standard deduction there is a personal deduction of $750 that applies to each family member.

Since the majority of states also levy income taxes we need to take this into account when constructing individuals' budget sets. Some states apply a proportional tax on federal taxable income, while others apply a progressive tax schedule with more than 20 brackets. In order to simplify the tax system we harmonize the state income brackets to the federal brackets (this procedure is employed by both Hausman and MaCurdy). The contribution of the state tax rate to the total tax rate is calculated as a weighted average of the state tax rate within the federal tax brackets. Variations in deduction rules across states also complicate the construction of budget constraints. We apply the rules of federal deductions to state taxation as well, but following Hausman we also take into account that several states allow deduction of the federal tax liability.[4]

Finally, it should be noted that standard deductions and the social security tax generate a few nonconvexities in individuals' budget constraints. Following Hausman and MaCurdy, we convexify budget sets by taking the convex hull in order to simplify computations and estimation procedures in the following sections.

3 Data

The data used in estimation by Hausman and MaCurdy originate from the 1976 University of Michigan's Panel Study of Income Dynamics (PSID Wave IX). The information was collected during 1976, but it mainly considered the status of the respondents in 1975. In the following two

---

3 Hausman adopted an approximation of the itemized deduction for incomes above $20,000. He assumed that individuals claimed a deduction equal to the average of joint returns as reported in *Statistics of Income*. Preliminary Monte Carlo simulations indicate that the simplification carried out in this paper (and in MaCurdy et al.) is harmless.

4 MaCurdy appears to have ignored this feature of the income tax system. This implies that some workers might face a higher marginal tax rate with MaCurdy's tax system. However, we have performed Monte Carlo simulations which indicate that this simplification does not bias the labor supply estimates in any significant way.

subsections we attempt to replicate as closely as possible the sample selection criteria and variable definitions applied by Hausman and MaCurdy.

It should be emphasized that our primary objective is not to value Hausman's and MaCurdy's strategies, but to reconstruct the data sets they used in their empirical analyses. The extent to which we are able to accomplish this is, of course, dependent on how comprehensive these two studies are when it comes to describing procedures and the amount of sample statistics supplied. The MaCurdy study is impeccable in both respects, whereas Hausman's exposition is much less comprehensive. However, Hausman has in personal conversation supplied valuable information on several points where the original study is vague.

It should also be noted that MaCurdy, in addition to his main data set, also presents an alternative set. However, this alternative data set generated estimates that closely resembled the estimates generated by the main data set. In the following we therefore focus primarily on his main data set.

3.1 Selection

Despite the fact that both Hausman and MaCurdy investigate the labor supply of prime aged married men, they nevertheless report data sets characterized by substantially different sample sizes and variable means. While Hausman presents a sample of 1,085 observations, the data set used by MaCurdy does only include 1,017 observations. Since the inconsistency between sample sizes is nontrivial we obviously need to construct two separate data sets, based on separate sample selection criteria, in our effort to replicate as closely as possible the samples originally selected by Hausman and MaCurdy.

Both Hausman and MaCurdy report that they include married men, between 25 and 55 years of age, either employed, temporarily laid off or unemployed. Due to extraordinary labor supply activities they exclude self-employed, farmers, severely disabled and observations which were part of the low income survey. In order to avoid troublesome interpretations of the tax system they also omitt households which had a change in family composition or lived outside the U.S. during 1975. Finally, they leave out families with a female head of household. Applying these criteria to PSID-76 we obtain a sample of 1,084 observations.[5] Noticing that we arrive close to the sample size of 1,085 reported by Hausman, this sample will be referred to as data set ‘H’.

MaCurdy reports that, since some of the questions put in the PSID-76 refer to the respondents status in 1976 rather than the status in 1975, he considers selected responses in the PSID-75 as well. Specifically, MaCurdy requires that observations pass the selection criteria concerning employment status, self-employment and farming when checked against PSID-75 as well as PSID-76. Applying these extended sample selection criteria causes the omission of another 66 observations. This leaves us with a sample of 1,018 observations - in the following referred to as data set ‘M’ - which is as close as we can get to MaCurdy’s sample size of 1,017.

Table 2 displays the effects in terms of excluded observations of applying the sample selection criteria in the given order. The first column represents the selection process underlying our data set ‘H’ (i.e., our best approximation of Hausman’s original sample). Unfortunately, Hausman’s study contains no comparable information. Thus, even though our final sample size matches Hausman’s quite closely, we have not been able to ascertain whether the samples actually contain the identical observations. The second column summarizes the selection process underlying our data set ‘M’. Comparing these figures to those presented by MaCurdy (table A1 in his Appendix A) we conclude that we have an almost perfect match between MaCurdy’s sample and our data set ‘M’.

5 Four additional observations were excluded due to missing values on the available hours of work-variables.

Table 2. Deleted observations

Selection criteria	data set 'H'	data set 'M'	Selection criteria (*cont.*)	data set 'H'	data set 'M'
# observations	5,862	5,862	Employment status 1975	-	15
Non-random sample	2,544	2,544	Self-employed 1975	-	202
Invalid marriage	1,071	1,071	Self-employed 1976	215	68
Family comp change	185	185	Farmer 1975	-	5
Under 25	229	229	Farmer 1976	11	6
Over 55	464	464			
Empl status 1976	44	44	# observations omitted	4,774	4,844
Outside USA	4	4	Deleted (missing hours)	4	-
Female head	7	7			
			Remaining sample size	**1,084**	**1,018**

1 MaCurdy reports having deleted one observation due to "...unreasonably high number of hours worked." We could not find any such observation.

3.2 Variables

Having established how to select our samples we now turn to the variables used in the empirical analysis. We need to define and measure the key economic variables - the hourly wage rate and nonlabor income (taxable and nontaxable) - used to construct individuals' budget sets. We also have to measure the dependent variable, i.e., annual hours of work, in order to perform the estimations. Besides these variables we also need information on a set of socioeconomic variables, such as the number of household members and the respondent's age, which are to be included as explanatory variables (representing observed preference heterogeneity) in the labor supply function.

Table 3 summarizes the averages of the key variables as reported by Hausman and MaCurdy. As will be shown below, the poor match between sample means is not due to differences in sample selection only, but also the result of Hausman and MaCurdy adopting widely divergent variable measures. In this subsection we try to pinpoint what we believe are the crucial differences between Hausman and MaCurdy in this respect.

Table 3. Characteristics of Hausman's and MaCurdy's key variables

Variable	Hausman	MaCurdy
No of observations	1,085	1,017
Hours of work	2,123	2,236
Nonlabor income	1,266	3,714
Wage rate	6.18	6.89

i) Hours of work

There are at least two ways we can measure annual hours of work for money in the PSID-76. The first is annual hours of work in 1975 as the question is put in the PSID-76. Another possibility is to construct a measure from responses to questions concerning average hours of work per week and the numbers of working weeks in 1975.[6] One would perhaps expect the two measures to coincide, but this is not the case; the measures are highly correlated (≈0.9), but certainly not identical.

Using our data set 'M' and measuring working hours according to the first definition, i.e., by using the single annual hours of work variable, we obtain an average for annual working hours of 2,236, which is the exact number reported by MaCurdy. Applying the same measure to data set 'H', we obtain a similar sample mean. In contrast, Hausman reports a considerably smaller mean for his annual hours of work variable, 2,123 hours, which suggests that this is in fact not the measure adopted by Hausman. Unfortunately, we have not been able to ascertain the exact way Hausman measured this variable. However, if we use our data set 'H' and construct a measure of annual hours of work by multiplying the respondent's average hours of work per week by the number of working weeks in 1975, we obtain a mean of 2,148 hours. This takes us reasonably close, but not as close as we might wish, to the mean of 2,123 hours reported by Hausman.

6 The questions are put with reference to the respondent's time spent on both the main job and on extra jobs. Preceding these questions are some inquiries about absence from work during 1975. This could imply that the second measure is more precise than the first.

We should also note that in data set ‘M’ all individuals report a positive number of annual working hours, while in data set ‘H’ there are 4 observations with zero hours of work.[7] This raises the additional difficulty of having to account for unobserved wage rates for these 4 individuals in the ‘H’ set (see subsection iii. below). A quick inspection reveals that, since both measures indicate zero hours of work for these 4 observations, the difference between samples with respect to the number of nonworkers is in fact due to variations in sample selection criteria rather than the different measures for hours worked.

ii) Nonlabor income

Persons recieve income from many sources other than employment, and this poses various difficulties for the empirical analysis of individual labor supply. The fundamental question is what should be included in the measure of nonlabor, or property, income. In principle, it seems correct to include not only money income, but also the money value of the stream of nonmoney services recieved from physical assets such as housing and durables. While conceptually relatively straightforward, in applied work the analyst often has to settle with incomplete or poor data on the various components of nonlabor income. In particular, since the economic variables in the PSID originate from questionnaire rather than register data, there might be reason to suspect that they are poorly measured.

One issue that deserves some further attention relates to the fact that various components of nonlabor income, such as the earnings of other family members, may be endogenous to labor supply. The conventional empirical model of individual labor supply is badly suited for taking account of more intricate household interactions, and researchers have in general adopted one of the following simplifying assumptions for couples filing jointly: 1) The labor supply decisions of husbands are made independently of the wives’ hours of work and earnings, implying that earnings of the wife should not enter as part of the husband’s nonlabor income. Consequently,

7 Hausman reports that there are approximately 0.5% ($\approx$5 observations) with zero hours of work in his sample.

wives are treated as secondary workers who face a ‘marriage tax’ in the sense that the first dollar of their earnings is taxed at the rate applicable to the last dollar of the husband’s earnings. 2) The labor supply decisions of wives are made independently of the husbands’ hours of work and earnings, and husbands are treated as secondary workers facing the marriage tax. This implies that we should include the wife’s earnings as taxable nonlabor income for the husband.[8]

MaCurdy constructs the nonlabor income variable by subtracting total labor earnings of the husband, defined as the sum of the respondent’s income from wages, the labor part of income from farming and rooming, income from bonuses, overtime and commissions etc., from the total 1975 taxable income of the household. This construction leaves asset income such as rent, interest, dividends etc. and the untaxed earnings of the spouse as the relevant measure of pretax nonlabor income for the husband. It should be noted that MaCurdy treats the husband as the secondary worker in the household, and he makes no attempt to incorporate the implicit income of housing and other physical assets in his measure of nonlabor income. Applying MaCurdy’s definition of taxable nonlabor income using our data set ‘M’ we obtain an average of $3,717, which is almost the exact number reported by MaCurdy.[9]

The corresponding sample mean reported by Hausman is merely one third of the amount presented by MaCurdy. Hausman states that he constructs the nonlabor income variable by

---

8 Suppose that the notion of primary and secondary workers is a correct description of the true data generating process, but we have no a priori knowledge of whether the husband or the wife should be treated as secondary. In particular, suppose wives are secondary workers, but we mistakenly treat them as primary. We would then falsely include the pretax earnings of the wife as taxable nonlabor income for the husband. This will cause not only a measurement error in taxable nonlabor income, but also create an endogeneity problem since there will be a correlation between the size of this measurement error and the husband’s labor supply. On the other hand, suppose husbands are secondary workers, but we mistakenly treat them as primary. We would then falsely exclude the pretax earnings of the wife as taxable nonlabor income for the husband. This will cause a measurement error in taxable nonlabor income, but there will be no correlation between the size of this measurement error and the husband’s labor supply since the wife as primary worker is making her labor supply decision independently of the husband’s hours of work and earnings. Thus, while different analysts may present different views on whether husbands or wives were secondary workers in the mid-seventies, it might perhaps be argued that by excluding the pretax earnings of the wife as taxable nonlabor income for the husband we at least avoid introducing a simultaneous equation bias in estimation.

9 MaCurdy reports that he excludes the untaxed earnings of the spouse in his alternative data set. This alternative nonlabor income measure then consists of the household’s income from rent, interest and dividends. The mean of the new income variable is $736 (1,100 observations), still not even close to the mean of $1,266 reported by Hausman (1,084 observations).

attributing an 8 percent return to financial assets. Thus, in contrast to MaCurdy, Hausman does not include the pretax income of the spouse, which implies that he treats the wife as the secondary worker in the household. Neither does his definition incorporate the PSID measurement of income from rent, interest and dividends included in the MaCurdy definition. Since information about equity in owner occupied homes is the only available data on asset holdings in the PSID, our inference is that Hausman attempts to impute the value of home ownership by taking an 8 percent return on the amount of equity the family have in their house (defined as house value minus the remaining mortgage). Using our data set 'H' and applying this measure we construct a nonlabor income variable with a mean of $1,262, which is reasonably close to the $1,266 reported by Hausman. In personal conversation, Hausman has confirmed that he in fact applied this definition of nonlabor income.[10]

We conclude that Hausman and MaCurdy use widely differing measures for the nonlabor income variable. Further, as a direct consequence of the different variable definitions, MaCurdy treats the nonlabor income as taxable income, whereas according to Hausman's definition the nonlabor income is to be treated as nontaxable income. Variations in nontaxable nonlabor income affect the vertical location of the budget constraint, but for each value of hours of work the slope will be unchanged. On the other hand, variations in taxable nonlabor income will affect both the intercept of the vertical axis of the budget constraint and the location of the kink points, which implies that the slope of the budget constraint will differ at a given value for hours of work. Consequently, the budget sets constructed by Hausman and MaCurdy might differ substantially for a certain individual, and this can presumably explain part of the diverging

10 The value of home ownership is presumably the most important asset income for many individuals who own their house. MaCurdy's motivation for choosing not to impute the implicit income of owner occupied homes is unclear, but for purposes of replicating Hausman's study (which is a clearly stated objective) it seems peculiar that MaCurdy actually fails to identify Hausman's definition of nonlabor income. The fact that Hausman refers to the amount of equity that families have in their houses as 'financial' assets might perhaps have distracted MaCurdy. However, since information on asset holdings in the PSID is confined to housing equity, one would expect that MaCurdy at least tried this possibility. Moreover, since he uses the housing equity variable as a taste shifter in the labor supply function, MaCurdy is certainly not unaware of its existence in the data.

empirical results. This is also confirmed by our results in sections 4.2-4.3 below, where we observe some significant changes in parameters estimates in response to the use of different nonlabor income measures.

iii) The hourly wage rate

The last of the key economic variables to be discussed is the hourly wage rate. One common way to measure the wage rate is to ask about it in a direct survey question. A possible alternative is to measure the hourly wage rate as annual labor earnings divided by annual hours of work. However, it is well known that the use of average hourly earnings can result in a serious measurement error in the wage rate: any error in the measurement of hours worked will be duplicated in the constructed wage measure, which will give rise to a spurious negative correlation between wage rates and hours of work. Of course, wage rates obtained by survey questions might also suffer from measurement error, but there is no compelling reason to suspect that this error should be correlated with measurement errors in hours of work. Thus, if both wage measures are available it is probably the case that average hourly earnings should be avoided.

In the PSID-76 both these measures for the hourly wage rate are in principle available to the investigator. However, there are some additional difficulties with respect to incomplete data that need to be considered if one chooses to use the response to an explicit question concerning the respondent's regular wage rate on the main job.[11] Firstly, the measure is truncated for wages above $9.98. Slightly less than 15 percent of the male respondents in PSID-76 fall into this category. Secondly, the measure contains missing values (zeros) for individuals who are not paid by the hour or salaried. Slightly more than 5 percent of the respondents fall into this category.

MaCurdy reports that he uses the construction of average hourly earnings in 1975 for all observations. Applying this measure to our data set 'M', using MaCurdy's measures of annual labor income and annual hours of work discussed in the previous subsections, we obtain an average hourly wage rate of \$6.89, which is the same mean as reported by MaCurdy. One convenient consequence of choosing this measure is that the wage is observed for all individuals in the sample; that is, there are no missing values or truncations. In addition, since the selection criteria for data set ‘M’ generate a sample without corner observations at zero hours of work, we need not consider how to handle unobserved wages for nonworkers.

In contrast, Hausman reported a mean hourly wage of \$6.18, which leads us to believe that the wage variable used there was not average hourly earnings.[12] In personal conversation Hausman has informed us that he used the directly reported hourly wage rates and estimated a wage equation to impute hourly wages for observations with unobserved or truncated wages.[13] The exact way in which this was carried out is however unknown to us. We have tried many different wage equations, but the results obtained in estimation of the complete labor supply model (discussed in the following section) appear to be quite robust to these variations. To simplify matters as far as possible we have settled with the following. We estimate a standard Tobit model by regressing observed, untruncated wages and truncated wages on a constant term, age, years of schooling, college degree, reading difficulty and family size. We then use the estimated coefficients to obtain predicted wages for truncated observations (149 in data set ‘H’), individuals who work but the wage variable is missing (87 observations) and for nonworkers (4 observations). Applying this procedure and using our data set ‘H’, we obtain an average hourly wage rate of \$6.21, which is reasonably close to the \$6.18 reported by Hausman.[14]

---

11 For salaried workers there is a separate question used to impute the average wage rate. Both questions relate to the current (1976) wage rate.

12 Applying the average hourly earnings measure to data set ‘H’ yields a sample mean of \$6.97.

13 MaCurdy reported that he used the directly reported hourly wage in his alternative data set. However, it is unclear how he accounted for unobserved/truncated wages. Since there is no mentioning of a wage equation procedure in MaCurdy’s paper, our inference is that he imputed the wage by taking the average hourly earnings for these observations. Also, the alternative set (1,103 observations) consisted of 3 nonworkers with unobserved wages. MaCurdy simply deleted these 3 observations from the sample.

14 As noted in footnote 11, the relevant question in the PSID asks about the wage at the time of the interview in early 1976 (most interviews took place in March/April) rather than the wage in 1975, which is the year in which hours of work are observed. Our inference is that Hausman did not attempt to discount the wage rate.

Again it is clearly the case that Hausman and MaCurdy use different measures for one of the key economic variables. Our results in sections 4.2-4.3 below strongly suggest that important aspects of the divergent empirical results reported by Hausman and MaCurdy can be attributed to their different wage measures. Furthermore, as noted in a previous paragraph, each of the measures can be expected to suffer from various defects. The Monte Carlo simulations in section 4.4 demonstrate that measurement errors in the wage rate can cause severely biased parameter estimates, and that certain types of errors are more damaging than others. These findings give rise to some interesting interpretations of the estimation results.

Finally, in addition to the economic variables, both Hausman and MaCurdy let the following socioeconomic variables represent observed preference heterogeneity in the labor supply function: the respondent's age (AGE45)[15], the number of children under six years of age (KIDSU6), family size (FAMSIZ), house equity - measured as house value minus remaining mortgage - (HOUSEQ) and a 0/1 dummy variable which takes a value of one if the respondent reported having a health condition that limited the amount of work he could do (BHLTH). Averages for the variables included in our data sets 'H' and 'M' are summarized in Table 4.

To sum up the results of this section, our main objective has been to reconstruct the data sets used by Hausman and MaCurdy. We believe we have generated a reasonably close approximation of Hausman's original data set and that we have successfully replicated the MaCurdy sample. MaCurdy concluded that the discrepancies between his estimation results and those reported by Hausman seemed "...perplexing in the light of the facts that both data sets are drawn from the same source and that the estimation approaches are the same" (p.481). Since Hausman's original results could not be replicated by MaCurdy, there now appears to be a widespread suspicion that Hausman's findings might be unreliable. In this section we have shown that there are several striking dissimilarities between the data sets constructed by

Hausman and MaCurdy. In fact, not only did they apply separate sample selection criteria, they also used widely differing measures for all the key economic variables. It therefore seems remarkable that the replicability of Hausman's results has been doubted on the basis of a comparison to the results reported by MaCurdy. In section 4 below it will become apparent that different strategies for measuring the economic variables play an important role in estimation.

Table 4. Descriptive statistics of data set 'H' and 'M'

Variable	data set 'H'	data set 'M'
Hours of work	2,148	2,236
Hourly wage rate	6.21	6.89
Non-labor income	1,262	3,717
KIDSU6	0.50	0.49
FAMSIZ	3.77	3.78
HOUSEQ	17,321	17,895
AGE45	1.32	1.37
BHLTH	0.06	0.05[1]
Number of obs.	1,084	1,018

1 MaCurdy reports a mean of 0.6 for the BHLTH variable. This must be a misprint, since it would mean that 60% of the respondents suffer from a severe health problem

4 Results

This section is organized as follows. Section 4.1 contains a brief description of the empirical labor supply model used both by Hausman and MaCurdy. In section 4.2 we discuss the empirical results originally reported by Hausman and MaCurdy, respectively, and the results we arrive at trying to replicate these two studies. In section 4.3 we study in greater detail the consequences of applying different sample selection criteria and variable definitions in estimation. Finally, in section 4.4 we perform a set of Monte Carlo simulations to study the bias associated with various types of measurement errors in the pretax hourly wage rate.

4.1 Model and econometric specification

This section briefly describes the empirical labor supply model used both by Hausman (1981) and MaCurdy et al. (1990). Individuals' preferences are represented by a strictly quasi-concave direct utility function $u(c,h;q)$, where c denotes consumption, h hours of work, and q is

15 The age variable is constructed by taking the respondent's age minus 45 if he was older than 45 in 1975.

a vector of preference parameters and individual characteristics. With no constraints on hours of work other than $0 \le h \le \overline{H}$, where $\overline{H}$ is an upper physical feasibility limit, the individual's utility maximization problem becomes

$$\max u(c,h;q) \quad \text{s.t.} \quad c = w^G h + y^G - T(w^G h, y^G; t) \quad \text{and} \quad 0 \le h \le \overline{H} \tag{1}$$

where w^G is the gross wage rate, y^G before tax nonlabor income, and $T(\cdot)$ a tax function with corresponding parameters t. Given a convex budget set and globally strictly convex preferences, the unique solution to the optimization problem can be written $h^* = h^*(w^G, y^G; q, t)$. Thus, the functional form of $h^*(\cdot)$ depends both on the utility function and the tax-transfer function.

For purposes of estimating the parameters of the utility function, let $b(w_j, y_j; q)$ be the the supply function generated by a linear budget constraint $c = y_j + w_j h$, where w_j is the marginal wage rate (the slope) of the j:th segment and y_j the virtual income (the intercept of the extended j:th segment) corresponding to the marginal wage rate w_j. Then we know that the individual's global optimum is located on the j:th segment of the budget constraint, i.e., $h^* = b(w_j, y_j; q)$ if $H_{j-1} < b(w_j, y_j; q) < H_j$, where H_j is the j:th kink of the budget constraint in terms of hours of work. Similarly, it must hold that desired hours of work fall at the j:th interior kink point if $b(w_{j+1}, y_{j+1}; q) \le H_j \le b(w_j, y_j; q)$. By the same argument, $h^* = 0$ if $b(w_1, y_1; q) < 0$ and $h^* = \overline{H}$ if $b(w_m, y_m; q) > \overline{H}$, where m is the total number of linear budget segments.

The labor supply function generated by a linear budget constraint is specified as

$$b(w, y; q) = cst + a\, w + b\, y + gz \tag{2}$$

where cst is a constant term and z a vector of observable personal attributes with corresponding vector of parameters g The uncompensated wage effect, a, is assumed to be constant across individuals. Unobserved preference heterogeneity enters through the individual specific random preference term b, which allows income effects to vary over the population.

Otherwise the variable is set equal to zero.

Assuming that leisure is a normal good, b is specified as a random draw from a normal distribution with mean m_b, variance s_b^2 and an upper truncation at zero.

The hours of work observed in the data are usually assumed to differ from the the utility maximizing quantity. We let a random term e ~ NID$(0, s_e^2)$ represent errors that contaminate the observation on hours of work for individuals who work. The assumed data generating process for observed hours of work $(\hat{h})$ can then be summarized by the following generalized Tobit model:

$$\hat{h} = \begin{cases} 0 & \text{if } h^* = 0 \text{ or } h^* + e \le 0 \\ h^* + e & \text{if } h^* > 0 \text{ and } 0 < h^* + e < \overline{H} \\ \overline{H} & \text{if } h^* > 0 \text{ and } h^* + e \ge \overline{H} \end{cases} \tag{3}$$

The appropriate maximum likelihood procedure for this model has been described in some detail in Hausman (1981), MaCurdy et al. (1990) and several other studies, and need not be repeated in this paper.

Suppose that data are actually generated by utility maximization with globally convex preferences as described by equations (1)-(3), and that individuals' true budget sets are measured without error. We then know that the maximum likelihood method produces estimates that are both consistent and asymptotically efficient. If the estimated parameters are inconsistent with the assumed data generating process we have an incoherent model, and we do not know how to interpret the results. In particular, we cannot interpret the parameters as preference parameters in a labor supply model.[16] The usual way to remedy the incoherency is to maximize the likelihood function subject to constraints that force the estimated preferences to be globally convex. For the linear supply function in (2) this implies satisfaction of the Slutsky inequality a-bh≥0, which

16 There are of course several reasons why we might obtain parameter values that are inconsistent with the utility maximization hypothesis. One is that data might not be generated by utility maximization. However, it must be realized that, due to bad choice of functional form, measurement errors in variables and/or sampling error, data might sometimes falsely refute the utility maximization hypothesis.

effectively rules out a negative uncompensated wage effect.[17] Of course, the results are of limited interest in the event that the optimum occurs at parameter values where the Slutsky condition represents a binding restriction; that the parameters can be interpreted as preference parameters in a labor supply model is then due only to the fact that preferences have been forced to satisfy the Slutsky condition.

At this point we should also comment on a potential numerical difficulty that might arise in the evaluation of the log-likelihood function. Using the truncated normal distribution as a description of the variation in b requires the evaluation of the term $1/\Phi[(0-m_b)/s_b]$, where $\Phi(\cdot)$ is the standard normal c.d.f. Standard approximations of $\Phi(\cdot)$ based on the erf function typically breaks down when the mean is more than six times the standard deviation. Obviously, since truncation of the normal distribution may occur quite far out in the tail, we need an algorithm that is capable of evaluating the standard normal c.d.f. with high precision out in the left tail of the distribution.

As described in the following section, MaCurdy reported that he encountered serious problems attempting to evaluate cumulative probabilities far out in the left tail of the normal distribution. MaCurdy did not investigate the use of an alternative algorithm. Instead, sticking to the standard routines he found it necessary to impose the restriction $m_b/s_b<6$ in order to be able to evaluate the relevant probabilities. In practice, he appears to have set $s_b=m_b/6$, which not only determines the size of the ratio m_b/s_b, but also the sign of m_b since s_b cannot take on a negative value. This approach might of course affect the estimated b distribution. In section 4.3 below we apply an algorithm that is capable of evaluating standard normal probabilities with high precision out in the left tail of the distribution. Specifically, when the ratio m_b/s_b exceeds six we switch to

17 We take this approach throughout this paper, i.e., in estimation we check continuously that $\alpha\geq0$ and signal an error return to discourage further search in that direction if the optimization routine attempts to evaluate the function for negative α. MaCurdy applies a slightly different method to achieve coherency, namely that the parameters are constrained to represent locally convex preferences at all kink points for all individuals in the

an asymptotic expansion of the error function as described by Abramowitz & Stegun (1965). We are then in a position to study the consequences of imposing unnecessary constraints on m_b and s_b.

4.2 Estimation results

Recall that our data sets 'H' and 'M' were designed to replicate as closely as possible the data sets used by Hausman and MaCurdy. Selected parameter estimates obtained for the 'H' and 'M' data sets are displayed in table 5, which also reproduces (in italics) the original Hausman and MaCurdy results. We will primarily focus on the estimated wage and income effects, and we therefore relegate the full set of parameter estimates to Appendix A1. In addition to the wage rate coefficient (a) and the parameters of the untruncated normal distribution (m_b, s_b), table 5 also reports the mean and the implied percentiles at the 1 percent, 25 percent, 50 percent (the median), 75 percent and the 99 percent level for the truncated distribution (i.e., the b distribution), and the labor supply elasticities implied by the estimates.

Table 5. Estimation results[1] (the original Hausman and MaCurdy results in italics)

sample. In practice, since the number of observations and kink points are large, both approaches virtually amount to imposing globally convex preferences.

Parameters	Data set ‘H’	*Hausman*	Data set ‘M’	*MaCurdy*
a	16.4	*0.200*	0.000	*0.000*
	(6.1)	*(9.0)*	(--)[2]	*(--)[2]*
E^{w}	0.036	*0.0004*	0.000	*0.000*
e^{w}	0.20	*0.55*	0.027	*0.024*
The untruncated b distribution:				
m_b	2.105	*2.037*	0.323	*0.296*
	(0.092)	*(0.073)*	(0.116)	*(0.321)*
s_b	0.336	*0.624*	0.054	*0.049*
	(0.015)	*(0.023)*	(--)[3]	*(--)[3]*
The truncated b distribution:				
99%	-0.001	*-0.002*	-0.000	*-0.000*
75%	-0.015	*-0.050*	-0.003	*-0.002*
50%	-0.036	*-0.120*	-0.006	*-0.006*
25%	-0.071	*-0.234*	-0.012	*-0.011*
1%	-0.229	*-0.707*	-0.038	*-0.035*
mean	-0.051	*-0.166*	-0.009	*-0.008*
E^{y}	-0.038	*-0.127*	-0.010	*-0.009*
Sample size	1,084	*1,085*	1,018	*1,017*

1 Annual hours of work. Wages and income are measured in dollars. Standard errors in paranthesis. Standard errors are not reported when the estimate encountered a binding inequality restriction. The uncompensated and compensated wage elasticities (E^{w} and e^{w}, respectively) implied by the estimates are computed using the median b and the median of net wage rates in our data sets ‘H’ and ‘M’ evaluated at 2,100 hours of work. Similarly, income elasticities (E^{y}) are computed using the median b and the median virtual income. Note that we use data sets ‘H’ and ‘M’ to compute the elasticities implied by Hausman’s and MaCurdy’s original estimates, since the elasticities were not reported in the two studies.
2 The nonnegativity constraint is binding.
3 The restriction s_b=m_b/6 is imposed.

4.2.1 The original Hausman and MaCurdy results

Consider first the results originally reported by Hausman and MaCurdy, respectively. The Hausman results indicated an uncompensated wage effect (a) close to zero. It should be noted that the estimate of a was not forced to take on a nonnegative value. Thus, the estimated parameters implied coherency in the sense that the Slutsky condition was satisfied for all observations in the sample, which suggests that data could have been generated by utility maximization with globally convex preferences. The main difference compared to previous studies was that the results indicated a substantial income effect (b); Hausman reported -0.12 as the median for his estimated b distribution. Except for family size, the coefficients for the socioeconomic variables were insignificant and estimated with poor precision (see Appendix A1).

Turning to the results obtained by MaCurdy, the wage rate coefficient had to be constrained to a nonnegative value in order not to violate the Slutsky condition (this is the reason why no standard error was reported). That is, the parameter estimate would have ended up negative unless the nonnegativity constraint was imposed. Another problem related to the estimates of m_b and s_b. Since the results implied that truncation of the normal distribution occured very far out in the tail, MaCurdy found it necessary to impose the restriction s_b=m_b/6 in order to be able to evaluate the relevant probabilities. The resulting estimates of m_b and s_b implied a tightly concentrated heterogeneity distribution. In fact, the estimates implied a negligible median income effect in the order of -0.006, which suggested a b distribution stacked up just below zero. Another striking difference concerned the estimate of the standard deviation for the error term e, which was almost twice as large as the standard deviation reported by Hausman (see Appendix A1).

The differences in the empirical results reported by Hausman and MaCurdy have, for several reasons, attracted widespread attention in recent years. First, it has been noted that these studies produced results that represent quite different views on the work disincentive effects and the dead weight losses associated with the progressive income tax system. The discrepancy for the estimated income effects imply sharply dissimilar responses to income changes and widely differing compensated substitution elasticities. In particular, MaCurdy's findings of simultaneously negligible wage and income responses imply small work disincentive effects and trivial welfare losses induced by progressive taxation. In contrast, Hausman calculated the welfare cost of labor supply distortions to be around 22 percent of tax revenues.

Second, as discussed in section 3 above, the fact that Hausman's results could not be replicated by MaCurdy has raised some concern about the reliability of Hausman's findings. However, our exposition on the construction of data sets has already shown that it makes no sense to judge the replicability of Hausman's results on the basis of the results reported by MaCurdy.

Third, the fact that MaCurdy encountered a binding nonnegativity constraint for the wage rate coefficient has important implications for how the results should be interpreted. One interpretation is that data correctly refute the utility maximization hypothesis as it is formulated in section 4.1. That the parameters can be interpreted as preference parameters in a labor supply model is due only to the fact that preferences have been forced to satisfy the Slutsky condition globally.[18] Of course, that estimated preferences sometimes might violate the convexity condition does not necessarily imply that preferences are nonconvex; Slutsky violations might for instance be due to misspecifications of the model or measurement errors in variables. Thus, a second possible interpretation of MaCurdy's results is that he used poorly measured variables in estimation such that data falsely refute the basic utility maximization hypothesis described in section 4.1.

4.2.2 Replications

Turning next to our replications of the empirical results reported by Hausman and MaCurdy, the results in table 5 and Appendix A1 suggest that we have successfully replicated MaCurdy's original results.[19] This is of course not very surprising, since it merely confirms that our 'M' data set constitutes a successful replication of the selection criteria and variable measures underlying MaCurdy's data set.[20] We have encountered greater problems attempting to replicate the results originally reported by Hausman. This is perhaps what might be expected, since reconstructing Hausman's original data set proved more difficult (section 3 above). Recall that, even though our sample size matched Hausman's quite closely, we were unable to ascertain

---

18 This would obviously imply that, for the labor supply specification under consideration, Hausman's results falsely support the utility maximization hypothesis. If data are generated by some mechanism other than utility maximization with convex preferences, it is of course fully conceivable that misspecifications or measurement errors in variables might produce results such that data falsely support the utility maximization hypothesis.

19 The GQOPT package of numerical optimization algorithms (version 6.00, see Goldfeld & Quandt 1972) was employed to find the optima for the log-likelihood functions. We have tried a large number of starting points to assure that global optima have been obtained. The variance-covariance matrix for the estimated parameter vector is calculated as the inverse of the Hessian of the log-likelihood function evaluated at the estimated parameter vector. We use numerically calculated second-order derivatives to construct the Hessian. It should be noted that we generally obtain significantly smaller standard errors than reported by MaCurdy et al. There is no information in their study on how standard errors were constructed. However, if they used analytic derivatives, then it might be the case that they obtained more reliable standard errors than those generated by our numerically calculated derivatives.

20 The wage rate coefficient moved quickly to the nonnegativity constraint and stayed there throughout the optimization. There was a steady increase in the ratio m_b/s_b . Eventually, the standard routines for evaluating the

whether the samples contained the identical observations. Similarly, there was some uncertainty concerning several of the variable definitions applied by Hausman. However, we nevertheless capture the distinctive features of Hausman's results.

Using our 'H' data set, the results in table 5 suggest that the wage rate coefficient is strictly positive, and we need not constrain the coefficient to a positive value in order to satisfy the Slutsky condition. The size of a implies a moderate uncompensated wage rate elasticity of around 0.036. It should be noted that the coefficient is estimated with poor precision. In fact, we obtained approximately the same log-likelihood value for a whole range of positive values for the wage rate coefficient, and the stopping point of each run varied with the positions of the starting values.

The median of -0.036 for the estimated b distribution is smaller (in absolute value) than Hausman's estimate of -0.12, but significantly larger than the median of -0.006 reported by MaCurdy. The b distribution is tighter than Hausman's, but significantly wider than the distribution estimated by MaCurdy. We have noted that the parameters of the untruncated normal distribution, m_b and s_b, are far from stable, i.e., there is a large number of combinations of m_b and s_b that generate approximately the same likelihood value, but the estimates appear to be correlated in such a way that the truncated distributions often are quite similar. However, we have also observed local optima for combinations of m_b and s_b similar to the one reported by Hausman. Since our 'H' data set differs in some respects from the data set used by Hausman, one should of course not misconstrue the latter result to conclude that Hausman's parameter estimates corresponded to a local optimum for his log-likelihood function. Rather, our inference is that small discrepancies between our data set and Hausman's might have been sufficient to induce a shift in the estimated b distribution.

---

normal c.d.f. could no longer provide meaningful evaluations. Following MaCurdy, we therefore imposed the restriction $s_b = m_b/6$.

To summarize, the results obtained for our data sets ‘H’ and ‘M’ highlight the sensitivity of parameter estimates to variations in variable definitions and sample selection criteria. Specifically, the results clearly demonstrate how different strategies may generate sharply dissimilar views on the work disincentive effects induced by progressive taxation and the consistency of data with the economic theory of consumer choice. In the following section we explore this issue in greater detail in order to pinpoint the driving forces behind the empirical results reported by Hausman and MaCurdy, respectively.

4.3 Why do results differ between studies?

As discussed in section 3 above, Hausman and MaCurdy apparently applied different sample selection criteria and widely divergent measures for the key economic variables. The parameter estimates obtained for our ‘M’ and ‘H’ data set in the previous subsection suggest that this might explain their seemingly irreconcilable empirical results. The objective of this section is to study in greater detail the consequences of applying the different selection criteria and variable definitions in estimation. We hope to shed some light on why MaCurdy, in contrast to Hausman, encounters a binding nonnegativity constraint for the wage rate coefficient (a). Similarly, we can hopefully provide an explanation to the conflicting views on the dispersion of income effects (b).

At this point it might be useful to recapitulate the principal differences between our data sets ‘H’ and ‘M’, which were designed to replicate as closely as possible the data sets used by Hausman and MaCurdy. The differences are summarized in table 6.

Table 6. Principal differences between the data sets ‘H’ and ‘M’

	Data set ‘H’	Data set ‘M’
Sample selection criteria	No criteria taken from the 1975 interview (n=1,084, including 4 observations with zero annual hours of work).	The criteria concerning employment status, self-employment and farming are checked against both PSID-75 and PSID-76 (n=1,017, zero nonworkers).

The wage measure	The reported hourly wage rate and a wage equation to impute hourly wages for those with unobserved or truncated wages.	Annual labor earnings divided by annual hours of work (for all observations).
The nonlabor income measure	An imputed income variable based on attributing an 8 percent return to house equity.	Labor income of the husband is subtracted from the taxable income of the household (leaving asset income and untaxed earnings of the spouse as the relevant measure).
The hours of work measure	Average hours per week are multiplied by the number of working weeks in 1975.	A measure based on a direct survey question on annual hours of work.

In the analysis below we break up our constructed data sets ‘H’ and ‘M’ and reestimate the model for all possible combinations of sample selection criteria, wage measures, definitions of nonlabor income and measures of annual hours worked. However, since our results suggest that variations in the hours of work variable produce only minor changes in the parameter estimates, and hence provide no useful insights as to the divergent results reported by Hausman and MaCurdy, we confine attention to the results where we apply MaCurdy’s measure of annual hours worked. It should also be noted that we apply the algorithm described in Abramowitz & Stegun (1965) throughout this secton in order to avoid imposing unnecessary constraints on m_b and s_b in those instances where the optimization routine attempts to evaluate cumulative probabilities far out in the tail of the untruncated normal distribution. The effects of the restriction s_b=m_b/6 imposed by MaCurdy will then become immediately apparent.

To simplify the exposition we analyze the results for wage and income effects separately in subsection 4.3.1 and 4.3.2, respectively.[21] We use the following shorthand notation: W_H represents the wage measure in data set ‘H’ (i.e., our approximation of Hausman’s wage variable) and W_M the measure in data set ‘M’ (MaCurdy’s wage variable). Similarly, Y_H and Y_M denote the measures of nonlabor income in data set ‘H’ and ‘M’, respectively. Finally, S_H and S_M represent the set of sample selection criteria associated with our two different data sets.

21 An Appendix containing the full set of parameter estimates and standard errors is available upon request from the authors.

4.3.1 The wage effect (a)

Table 7 displays estimates of the net wage rate coefficient (a) and the uncompensated wage elasticities[22] (E^{w}) implied by the estimates, for all possible combinations of sample selection criteria, wage and nonlabor income measures. The column labeled '*base case*' refers to MaCurdy's selection criteria and variable definitions (i.e., the combination S_M, W_M, Y_M). The other columns display the results obtained as we substitute one or more of Hausman procedures for the corresponding MaCurdy procedures. For example, the column 'W_H' shows what happens compared to the base case when we substitute W_H for W_M. The table is organized so that the upper panel shows the results for all combinations that still include W_M, whereas the lower panel shows the results for all combinations that include W_H.

Table 7. The estimated wage effect (a)

	base case	Y_H	S_H	Y_H, S_H
a	0.000	0.000	0.000	0.000
E^{w}	0.000	0.000	0.000	0.000
	W_H	W_H, S_H	W_H, Y_H	W_H, Y_H, S_H
a	10.3	18.6	26.5	26.9
E^{w}	0.021	0.038	0.058	0.058

Note: The S_H selection generates a sample that includes 4 nonworkers with unobserved wages. The imputation procedure for the case where S_H appears in combination with W_H was discussed in section 3.2 above. When S_H appears in combination with W_M we estimate a conventional linear wage equation (using the set of explanatory variables described in section 3.2) in order to impute wages for these 4 observations. It should also be noted that, since all results in the table are based on MaCurdy's hours of work measure, the estimate for the combination (W_H, Y_H, S_H) is not the same as for the Hausman replication in table 5.

The results convey a quite suggestive picture of how the estimated wage effect varies with our choice of wage measure. MaCurdy's wage measure, W_M, causes the estimate of a to run into the nonnegativity constraint for all combinations of nonlabor income definitions and sample selection criteria.[23] In contrast, Hausman's wage measure, W_H, yields a positive wage effect for

22 Uncompensated wage elasticities, E^{w}, are computed using the median net wage rate evaluated at 2,100 hours.

23 The results for the MaCurdy base case in table 7 are obtained without imposing a constraint on m_b and s_b. Similar to the results in table 5 where our replication of MaCurdy's original results was carried out by imposing

all combinations of nonlabor income definitions and sample selection criteria. If we apply S_M and Hausman's wage measure, then Y_H yields a larger uncompensated wage elasticity (0.058) than generated by Y_M (0.021). The choice of selection criteria has virtually no effect on a if we apply Hausman's Y. If we use MaCurdy's Y, then S_H yields a slightly larger uncompensated wage elasticity (0.038) compared to the elasticity generated by S_M (0.021).

Thus, the estimates in table 7 suggest that the main difference between MaCurdy's and Hausman's original results with respect to the wage coefficient - i.e., MaCurdy's estimate of the wage coefficient encountered a binding nonnegativity constraint, whereas Hausman did not experience this in his study - most likely can be attributed to the use of different wage measures in these two studies. In fact, MaCurdy would have obtained a positive wage coefficient if he had used W_H rather than W_M to construct individuals' budget sets.[24] Suppose data are in fact generated by utility maximization with globally convex preferences as described by eq. (1)-(3). Then W_H generates a wage coefficient of the correct sign, whereas W_M gives rise to a downward biased coefficient such that data falsely refute the utility maximization hypothesis. Alternatively, suppose data are generated by some other mechanism. Then W_M generates a wage coefficient that correctly indicates violation of the Slutsky condition, whereas W_H yields a coefficient such that the results falsely indicate that data could have been generated by utility maximization with globally convex preferences.

Of course, there is no way we can establish with certainty which interpretation is correct. As discussed in section 3.2 above, what we do know is that there is potentially damaging division bias associated with W_M (MaCurdy's construction of average hourly earnings). In particular,

---

the constraint $s_b=m_b/6$ suggested by MaCurdy, the estimate of a encounters the nonnegativity constraint. Thus, the restriction $s_b=m_b/6$ imposed by MaCurdy appears to have no influence on the estimated wage effect.

24 MaCurdy reported that he used the directly reported hourly wage in his alternative data set (see footnote 13 above) and still encountered a binding nonnegativity constraint for the wage rate coefficient. However, since there is no mentioning of a wage equation procedure in MaCurdy's paper, our inference is that he accounted for unobserved/truncated wages (approximately 20% of the sample) by taking the average hourly earnings for these observations. Thus, this alternative measure is not equal to the WH measure. We have tried what we believe is

since the estimation results in table 5 indicated a substantial variance for the error *e*, it might seem odd that MaCurdy did not even reflect on this being a potential problem. On the other hand, W_H (the directly reported hourly wage rate used by Hausman) can also be expected to suffer from measurement error, which might cause an inconsistent estimate of the wage rate coefficient. In section 4.4 below we perform a set of Monte Carlo simulations which hopefully can shed some light on the seriousness and the direction of the bias associated with various types of measurement errors in the pretax hourly wage rate.

4.3.2 The income effect (b)

Table 8 displays estimates of the untruncated normal distribution, the mean and the implied percentiles at the 1 percent, 25 percent, 50 percent (the median), the 75 percent and the 99 percent level for the truncated *b* distribution, and the income elasticities (E^y) implied by the estimates.[25] We have reestimated the model for all possible combinations of sample selection criteria, wage and nonlabor income measures. However, in order to simplify the exposition and conserve on space, we confine attention to a subset of the estimations that conveys essentially all the relevant information. Specifically, the column labeled '*base case*' in table 8 refers to MaCurdy's selection criteria and variable definitions, whereas the other columns display the results when we substitute one of Hausman's procedures for the corresponding MaCurdy procedure.

Table 8. The estimated income effect (*b*)

MaCurdy's alternative wage measure, and the estimate of *a* still runs into the nonnegativity constraint for all combinations of nonlabor income measures and selection criteria.

25 Income elasticities, E^y, are computed using the median *b* and the median virtual income evaluated at 2,100 hours.

	base case	W_H	Y_H	S_H
The untruncated *b* distribution:				
m_b	-0.011	0.179	-0.079	0.893
s_b	0.002	0.026	0.019	0.143
The truncated *b* distribution:				
99%	-0.007	-0.000	-0.034	-0.000
75%	-0.010	-0.001	-0.066	-0.006
50%	-0.011	-0.003	-0.079	-0.015
25%	-0.012	-0.005	-0.092	-0.031
1%	-0.015	-0.017	-0.124	-0.098
mean	-0.011	-0.004	-0.079	-0.022
E^y	-0.018	-0.004	-0.089	-0.025

Note: The S_H selection generates a sample that includes 4 nonworkers with unobserved wages. The imputation procedure for the case where S_H appears in combination with W_H was discussed in section 3.2 above. When S_H appears in combination with W_M we estimate a conventional linear wage equation (using the set of explanatory variables described in section 3.2) in order to impute wages for these 4 observations

The results in table 8 suggest that our choice of nonlabor income measure, Y_M or Y_H, has a considerable influence on the estimated *b* distribution. Compared to MaCurdy's base case, the truncated distribution is much wider, and both the mean and median *b* become considerably more negative if we apply Hausman's rather than MaCurdy's definition of nonlabor income; we observe a change in the mean of the *b* distribution from -0.011 to -0.079, an increase in the interquartile range from 0.002 to 0.026 and a change in the income elasticity from -0.018 to -0.089 as we shift from Y_M to Y_H. Thus, MaCurdy would have obtained a *b* distribution much more in line with the one estimated by Hausman if he had used Y_H rather than Y_M.

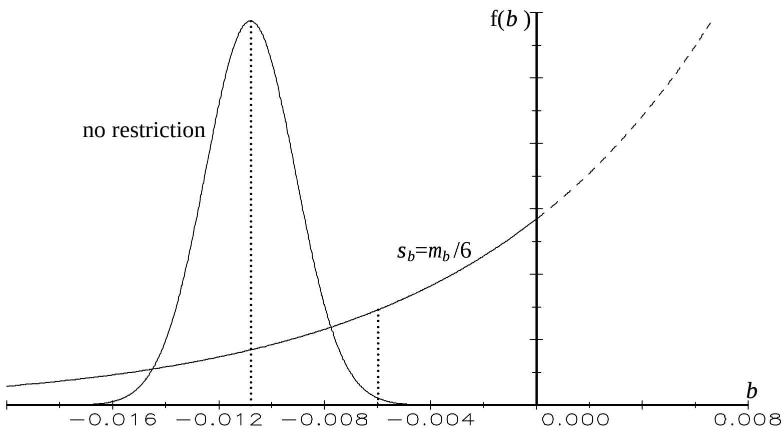
Compared to MaCurdy's base case, the results in table 8 suggest that our choice of selection criteria, S_M or S_H, has some influence on the estimated *b* distribution. Not only do the parameters of the untruncated normal distribution change as we change selection criteria, there are also differences between the implied truncated distributions. Compared to MaCurdy's base case, the truncated distribution is wider, and both the mean and median *b* become more negative if we apply S_H rather than S_M; we observe a change in the mean of the *b* distribution from -0.011 to -0.022, an increase in the interquartile range from 0.002 to 0.025 and a change in the income elasticity from -0.018 to -0.025 as we shift from S_M to S_H. Thus, MaCurdy would have obtained a *b* distribution slightly more in line with the one estimated by Hausman if he had applied the

Hausman selection rather than his actual sample selection criteria. Intuitively, this result can surely be attributed to the fact that the S_H selection generates a sample that includes 4 nonworkers, whereas the S_M selection generates no such observations. To the extent that the nonworkers included in the S_H sample are true outliers, the heterogeneity distribution will have to be stretched out in order to explain why we observe zero hours of work for these observations. Consequently, we might expect a wider distribution if nonworkers are included in the sample.

A shift from W_M to W_H has some effect on the parameters of the untruncated normal distribution. However, the truncated distribution implied by the estimates is relatively similar to the base case, except that W_H generates a slightly wider distribution. Thus, in contrast to the striking effect on the wage rate coefficient discussed in the previous subsection, our choice of wage measure appears to have a limited effect on the estimated b distribution. In particular, the fact that the wage rate coefficient hits the nonnegativity constraint in one case and is strictly positive in the other appears to be of little significance for the estimated b distribution.

Finally, consider the effects on MaCurdy's original results (table 5) of imposing the restriction $s_b = m_b/6$. As shown in figure 1, we obtain a considerably more spiked untruncated normal distribution when the restriction is relaxed. The mean (m_b) shifts from 0.3 to a negative value of approximately -0.01, and there is virtually no probability mass above zero. This, of course, implies a distribution for b that essentially coincides with the untruncated normal distribution. Compared to MaCurdy's original results, we observe a change in the median of the b distribution from -0.006 to -0.011, a decrease in the interquartile range from 0.009 to 0.002, a decrease (in absolute value) in the first percentile from -0.035 to -0.015 and a change in the income elasticity from -0.009 to -0.018 as we relax the restriction $s_b = m_b/6$. Thus, although the results still indicate that income effects are tightly distributed and moderate in size, the restriction has a substantial influence on the general shape of the estimated b distribution.

Figure 1. The effect of imposing the constraint s_b=m_b/6



To summarize, the sensitivity analysis performed in this section conveys a clear picture of the driving forces behind the divergent empirical results reported by Hausman and MaCurdy. Our findings suggest that differences with respect to the estimated wage effect can be attributed to the use of competing wage measures. MaCurdy's wage measure causes the estimated wage effect to run into the nonnegativity constraint for all combinations of nonlabor income definitions and sample selection criteria. In contrast, applying the W_H wage measure we obtain a strictly positive wage effect for all combinations of nonlabor income measures and sample selection criteria. Similarly, the choice of nonlabor income measure, Y_M or Y_H, has a crucial influence on the estimated distribution of income effects.

4.4 Monte Carlo simulations

The results of the previous section suggested that the main difference between MaCurdy's and Hausman's original results with respect to the wage coefficient - i.e., MaCurdy's estimate of the wage coefficient encountered a binding nonnegativity constraint, whereas Hausman did not experience this in his study - can be attributed to the use of differing wage measures in the two studies. One interpretation of MaCurdy's results is that data correctly refute the utility maximization hypothesis as it is formulated in section 4.1. However, an alternative interpretation is that data have actually been generated by utility maximization with globally convex preferences as described by eq. (1)-(3), but MaCurdy's wage measure causes a downward biased wage coefficient such that data falsely refute the utility maximization hypothesis.

As discussed in section 3.2, it is well known that there is potentially damaging division bias associated with MaCurdy's imputed wage measure, i.e., the construction of average hourly earnings. On the other hand, the directly reported hourly wage rate used by Hausman (including predictions by means of an estimated wage equation for observations with unobserved or truncated wages) can also be expected to measure the true wage with some error. In this section we perform a set of Monte Carlo simulations which hopefully can shed some light on the seriousness of the bias associated with various types of measurement errors in the pretax hourly wage rate.[26]

The data generating process (DGP) is designed as follows. The income tax system is the same as in actual estimation. Individuals' true budget sets are constructed on the basis of the reported hourly wage rate in the PSID-76 and MaCurdy's nonlabor income measure. The latter is observed without error in estimation. Applying MaCurdy's sample selection criteria, and excluding observations for which the reported hourly wage rate is missing or truncated, the sample consists of 818 observations throughout the simulations. Observed hours of work are generated by utility maximization with globally convex preferences according to the observational scheme in (3). The labor supply function has the same form as in actual estimation, but we exclude the influence of socioeconomic variables. The parameters are given the following values: a=20, m_b=1, s_b=0.3 and constant=2,200.[27]

---

26 Given the way the estimated income effects varied with the nonlabor income measure in the previous section, it would of course also be of interest to study how uncertainty about the appropriate definition of nonlabor income affects parameter estimates. For example, what happens if we falsely exclude the implicit income of housing and other physical assets in the measure of nonlabor income when it is in fact part of the data generating process? Similarly, what happens if we misspecify the way the earnings of the spouse enter the data generating process? However, since it is difficult to make a convincing case for a particular nonlabor income measure, and hence for the true data generating process, such experiments would be of limited interest for our study.

27 The combination of m_b and s_b implies a median b of -0.056. The income elasticity for the median virtual income and median b is in the order of -0.075 when evaluated at 2,100 hours of work. The uncompensated wage elasticity for the median net wage rate is approximately 0.04. It should be noted that there are no observations with zero or $\overline{H}$ hours of work in any of our simulations.

In the DGP individuals are confronted with budget constraints based on their true wage rate (w^*), but in estimation the observed wage ($\hat{w}$) may differ from the true wage rate. For example, if we ask about it in a direct survey question, the true wage may be measured with error. Alternatively, if data contain no such direct question, or, if for some reason the quality of the existing survey data is deemed unacceptable, we may have to settle with an imputed wage rate. We perform six different experiments, and we generate 100 samples for each experiment. The distribution of true wages and nonlabor income is the same in all samples, but we generate different realizations for the random variables (unobserved preference heterogeneity and various measurement errors) in each sample. The percentage bias and standard deviation for the estimated parameters are displayed in Appendix A2 (for the b distribution we confine attention to the median implied by the parameters m_b and s_b). For the wage rate coefficient (a) we also display the fraction of estimates for which the nonnegativity constraint is binding.

In *simulation 1* the true wage rate is observed and measured without error for all observations (i.e., we estimate the true DGP). The simulation is repeated for s_e=300, s_e=400 and s_e=500. The results in Appendix A2 suggest that there is a moderate small sample bias towards zero for the wage rate coefficient (a), whereas the median b is slightly biased away from zero. The small sample bias appears to be quite insensitive to the size of s_e.[28]

In *simulation 2* we assume that the wage used in estimation measures the true wage with error.[29] Specifically, we assume that the error is additive and normally distributed with mean zero; that is, $\hat{w}=w^*+v$, v~N.I.D(0, s_v). This might be a reasonable description of the nature of measurement errors for the reported hourly wages in the PSID. The simulation is repeated for s_v =0.25, s_v =0.5 and s_v =1.0, with s_e set equal to 300 (the standard deviations for v amount to approximately 5%, 10% and 20%, respectively, of the average w^* in the DGP). The results

28 We have also performed simulations with larger sample sizes. To increase the sample size we use multiples of the sample of 818 observations. For s_e=300, the bias in a is virtually zero already for a sample size of 1,636.

suggest that the estimation method is quite robust to additive measurement error in the hourly wage rate. In particular, for s_v =1.0 the percentage bias in a is only -6.6%. Further, only in 1% of the generated samples do estimates hit the nonnegativity constraint. The median for the estimated b distribution appears also to be virtually unaffected by this type of measurement error in the hourly wage rate.

In *simulations 3* and *4* we assume that there are no reported hourly wages in the data (or, equivalently, that reported hourly wages for some reason are deemed unacceptable). Instead, in estimation we construct a measure based on the individual's average hourly earnings. This approach corresponds to the way MaCurdy imputed the hourly wage rate for all observations in his main data set. In *simulation 3* we assume that annual labor earnings, w^*h^*, are measured *without* error; that is, $\hat{w}=w^*h^*/\hat{h}$. The simulation is repeated for s_e=100, s_e=300 and s_e=500. The results indicate that this type of error might be extremely damaging. The wage rate coefficient is heavily downward biased towards zero, -57.4%, already for small measurement errors in hours of work. The bias increases rapidly and approaches -100% as s_e increases. For s_e equal to 300 we encounter a binding nonnegativity constraint for 100% of the estimates. The median for the estimated b distribution is also affected, but to a lesser extent. For s_e=300 there is a bias (away from zero) of approximately -17.5% for the estimated median. The bias increases with the size of s_e and approaches -50% as s_e moves towards 500.

Simulation 4 is the same as simulation 3, except that we now assume that reported annual labor earnings measure the true earnings *with* error. If the economic variables originate from questionnaire rather than register data, which is the case for the PSID, this presumably offers a more realistic description of the quality of actual data. Specifically, we assume that the error is additive and normally distributed with mean zero; that is, $\hat{w}= (w^*h^*+u)/\hat{h}$, where u~N.I.D(0, s_u).

---

29 In order to avoid generating observed wages less or equal to zero we truncate observed wages at $.59, which is the minimum wage used in the data generating process. We apply this procedure throughout the simulations.

The simulation is repeated for s_u=500, s_u=1,000 and s_u=2,000, with s_e set equal to 100 (the standard deviations for *u* amount to approximately 5%, 10% and 20%, respectively, of the average w^*h^* in the DGP). The results are similar to those obtained for s_e=100 in simulation 3, where labor earnings were measured *without* error (i.e., s_u=0). Thus, the results suggest that the estimation method is quite robust to additive measurement errors in labor earnings.

The DGP for *simulations 5* and *6* is the same as for simulations 1-4, but we now assume that in estimation the true wage is unobserved for 20 percent of the sample. We assume that there are no measurement errors in reported labor earnings or in reported hourly wages for the remaining 80 percent. For each sample we generate a new set of observations for which the wage is missing.

In *simulation 5* we construct average hourly earnings to impute a measure for observations with unobserved wages; that is, $\hat{w}=w^*h^*/\hat{h}$. This approach corresponds to the way we believe MaCurdy handled unobserved hourly wages in his alternative data set (see footnote 13). The results provide another strong warning against the use of average hourly earnings. The wage rate coefficient is slightly downward biased towards zero, -13.5%, for small measurement errors in hours of work. However, the bias increases rapidly and approaches -100% as s_e increases towards 500. For s_e equal to 300 the percentage bias is approximately -85%, and we encounter a binding nonnegativity constraint in almost 60% of the generated samples. The median for the estimated *b* distribution is also affected, but to a much lesser extent; for s_e=300 there is a bias (away from zero) of approximately -5% for the estimated median.

In *simulation 6* we predict wages for 20% of the sample by means of a simple linear wage equation.[30] This approach corresponds to the way Hausman handled unobserved hourly wages. The wage rate coefficient and the median *b* are estimated about as well as when the true wage is

30 We use the same set of explanatory variables as in actual estimation; that is, a constant term, age, years of schooling, college degree, reading difficulty and family size.

observed without error for all observations (simulation 1), which clearly suggests that this imputation procedure is preferred to the construction of average hourly earnings.

The simulations performed in this section demonstrate that measurement errors in the hourly wage rate can cause severely biased parameter estimates, and that certain types of errors are much more damaging than others. In particular, for sufficiently large measurement errors in hours of work, measuring the hourly wage rate by average hourly earnings might be extremely damaging. This is true even if we apply this construction for a subset of only 20% of the full sample. Considering that the estimation results in table 5 indicated a substantial standard deviation for the error ***e*** [31], the simulations give rise to some interesting interpretations of the estimation results reported by MaCurdy (generated by the main or the alternative data set). The results of the analysis in section 4.3.1 stressed the sensitivity of the estimated wage rate coefficient to the use of different wage measures. In particular, using MaCurdy's wage measure we encountered a binding nonnegativity constraint for all combinations of nonlabor income definitions and sample selection criteria. The simulation results in this section are consistent with the interpretation that MaCurdy's wage measure in fact causes a downward biased wage rate coefficient such that data falsely refute the utility maximization hypothesis.

5 Summary and Conclusion

The divergent empirical results reported by Hausman (1981) and MaCurdy et al. (1990) have attracted widespread attention in recent years. The studies represent sharply contradicting views on the size of the work disincentive effects induced by progressive taxation and the consistency of data with the economic theory of consumer choice. This might, at first sight, seem puzzling, since they applied the same model specification, the same estimation technique and

---

31 There is, of course, no way we can establish the separate contributions to the estimated error distribution from measurement errors in hours of work and from general specification errors. However, since the annual hours of work-variable originates from questionnaire rather than register data, there might be reason to suspect that measurement/reporting errors are substantial in size.

also collected data from the same source. In this paper we seek to uncover the driving forces behind the seemingly irreconcilable results.

In section 3 we set out to reconstruct the data sets used by Hausman and MaCurdy. A careful reading of the two studies reveals that there are several conspicuous inconsistencies between their data sets. First, despite the fact that both Hausman and MaCurdy et al. investigated the labor supply in 1975 of prime aged married men, Hausman presented a sample of 1,085 observations (including 0.5% nonworkers), whereas the sample used by MaCurdy et al. included only 1,017 observations (zero nonworkers). Concluding that they obviously applied different sample selection criteria, our exposition in section 3.1 shows that Hausman presumably did not apply any criteria taken from the 1975 interview, whereas MaCurdy et al. checked the criteria concerning employment status, self-employment and farming against both PSID-75 and PSID-76.

Second, they also used divergent measures for the key economic variables. Hausman used the reported hourly wage rate and estimated a wage equation to impute hourly wages for observations with unobserved or truncated wages, whereas MaCurdy et al. measured the hourly wage rate as annual labor earnings divided by annual hours of work. As for nonlabor income, Hausman applied an imputed income variable based on attributing an 8 percent return to house equity. In contrast, MaCurdy et al. used asset income such as rent, interest, dividends etc. and the untaxed earnings of the spouse as the relevant measure of pretax nonlabor income. Consequently, the budget sets constructed by Hausman and MaCurdy et al. might have differed substantially for a certain individual. In addition, they appear to have used different measures for the dependent variable annual hours of work. Thus, even if the studies relied on a common data source and empirical methodology, we see no reason to expect a close match between the estimation results in the two studies. In particular, it makes little sense to judge the replicability of Hausman’s results on the basis of a comparison to the results reported by MaCurdy et al.

The results presented in section 4.2 suggest that we have successfully replicated the empirical results reported by MaCurdy et al. We encountered greater problems attempting to replicate the results originally reported by Hausman, but we nevertheless capture the distinctive features of Hausman's results. In section 4.3 we study in greater detail the consequences of applying the different selection criteria and constructions of variables in estimation. The results suggest that important aspects of the divergent empirical results reported by Hausman and MaCurdy et al. can be attributed to their differing wage measures. The measure adopted by MaCurdy et al. causes the wage rate coefficient to run into the nonnegativity constraint for all combinations of nonlabor income definitions and sample selection criteria. In contrast, our approximation of Hausman's wage measure yields a positive wage effect for all combinations of nonlabor income definitions and sample selection criteria. Similarly, the choice of nonlabor income measure has a major influence on the estimated distribution of income effects. MaCurdy et al. would have estimated a distribution much more in line with the one estimated by Hausman if, *ceteris paribus*, they had used the nonlabor income measure applied by Hausman.

The imputed wage measure adopted by MaCurdy et al. (the construction of average hourly earnings) suffers from a well known, potentially damaging, division bias. On the other hand, the reported hourly wage rate used by Hausman (including predictions by means of an estimated wage equation for observations with unobserved or truncated wages) can also be expected to measure the true wage with error. The Monte Carlo experiments in section 4.4 demonstrate that, for sufficiently large measurement errors in hours of work, measuring the hourly wage rate by average hourly earnings might cause a severe bias towards zero for the wage rate coefficient. In contrast, the estimation method is quite robust to the type of errors which presumably contaminate Hausman's wage measure (additive errors in the reported hourly wage rate and errors caused by predicting wages by means of a simple wage equation). Thus, the simulation results are consistent with the interpretation that the wage measure adopted by MaCurdy et al.

causes a downward biased wage rate coefficient such that data falsely refute the utility maximization hypothesis.

The main contribution of the present paper is perhaps not a more reliable set of parameter estimates than provided by Hausman or MaCurdy et al. Much more, our results demonstrate how different strategies for measuring variables may generate sharply dissimilar views on the work disincentive effects induced by progressive taxation and on the consistency of data with the basic notion of utility maximization. It should be noted that the sensitivity of parameter estimates to measurement errors in the explanatory variables is by no means unique to the maximum likelihood estimation method. Simulations in Blomquist (1996) show that simpler methods, such as linearization techniques with instrumental variables, often are even more sensitive to measurement errors in the independent variables. Until more robust estimation techniques are available, it is essential that empirical studies are accompanied by rigourous sensitivity analyses in those instances where variable definitions are not self-evident. Similarly, Monte Carlo simulations to study the effects of measurement errors in variables should be a matter of routine. Precautionary steps like these would presumably serve at least two important purposes. First, they may preclude unnecessary misconceptions and mitigate the need for studies such as ours. Second, they would hopefully draw attention to variable constructions which could have serious effects on the performance of the estimation method.

APPENDIX A

A1. Estimation results[1] (the original Hausman and MaCurdy results in italics)

Parameters	Data set 'H'	*Hausman*	Data set 'M'	*MaCurdy*
CONSTANT	2,1703	*2,4195*	2,3191	*2,3191*
	(35.4)	*(58.9)*	(28.4)	*(111.9)*
KIDSU6	33.6	*-3.9*	8.89	*11.6*
	(14.8)	*(25.5)*	(5.8)	*(49.9)*
FAMSIZ	-14.7	*34.1*	-10.2	*-11.2*
	(4.9)	*(17.0)*	(3.6)	*(20.6)*
AGE45	-20.6	*-1.1*	-23.4	*-22.5*
	(5.8)	*(10.8)*	(5.7)	*(14.8)*
BHLTH	-180.1	*-138.7*	-153.3	*-153.2*
	(68.2)	*(143.6)*	(71.8)	*(173.4)*
HOUSEQ	0.0066	*0.0026*	0.0019	*0.0018*
	(0.0011)	*(0.0009)*	(0.0006)	*(0.0013)*
a	16.6	*0.2*	0.0	*0.0*
	(6.1)	*(9.0)*	(-)[2]	*(-)[2]*
m_b	2.105	*2.037*	0.323	*0.296*
	(0.092)	*(0.073)*	(0.116)	*(0.321)*
s_b	0.336	*0.624*	0.054	*0.049*
	(0.015)	*(0.023)*	(-)[3]	*(-)[3]*
s_e	556.4	*279.4*	536.4	*536.4*
	(12.5)	*(17.8)*	(12.0)	*(35.7)*
sample size	1,084	*1,085*	1,018	*1,017*

1 Annual hours of work. Wages and income are measured in dollars. Standard errors in paranthesis.
Standard errors are not reported when the estimate encountered a binding inequality restriction.
2 The nonnegativity constraint is binding.
3 The restriction $s_b = m_b/6$ is imposed.

A2. Simulation results

	Sim.1: No errors in reported wages ($\hat{w}=w^*$).			Sim.2: $\hat{w}=w^*+v, v$~NID(0, s_v), σ_e=300.		
Parameter	s_e=300	s_e=400	s_e=500	s_v=0.25	s_v=0.5	s_v=1.0
a:						
bias (%)	-8.5	-8.9	-10.3	-9.0	-5.0	-6.6
st.dev.	9.6	11.8	14.2	9.3	9.7	9.0
% constrained	6.0	7.0	14.0	5.0	5.0	1.0
median b:						
bias (%)	-4.0	-5.4	-7.2	-3.9	-3.5	-3.2
st.dev.	0.006	0.007	0.008	0.006	0.006	0.006
constant:						
bias (%)	0.5	0.5	0.4	0.5	0.3	0.3
st.dev.	42.3	51.0	60.9	40.4	42.6	40.0
s_e:						
bias (%)	-0.2	0.1	0.8	-0.2	0.0	0.1
st.dev.	11.0	14.4	18.1	11.0	10.9	11.2

Parameter	Sim.3: $\hat{w}=w^*h^*/\hat{h}$.			Sim.4: $\hat{w}=(w^*h^*+u)/\hat{h}$, u~N.I.D(0,s_u), s_e=100.		
	s_e=100	s_e=300	s_e=500	s_u=500	s_u=1,000	s_u=2,000
a:						
bias (%)	-57.4	-100.0	-100.0	-55.7	-54.9	-49.3
st.dev.	4.7	0	0	4.9	4.8	4.8
% constrained	4.0	100.0	100.0	7.0	5.0	4.0
median b:						
bias (%)	-1.1	-17.5	-48.6	-0.5	0.4	2.0
st.dev.	0.004	0.007	0.008	0.004	0.004	0.004
constant:						
bias (%)	2.1	4.9	7.5	2.0	1.9	1.6
st.dev.	22.0	19.9	30.6	23.6	24.3	24.8
s_e:						
bias (%)	-0.2	-2.8	-3.8	0.5	1.8	4.1
st.dev.	6.3	12.0	18.4	6.0	6.2	5.9

Parameter	Sim.5: $\hat{w}=w^*h^*/\hat{h}$ for 20% of the sample			Sim.6: Predicted wages for 20% of the sample.		
	s_e=100	s_e=300	s_e=500	s_e=300	s_e=400	s_e=500
a:						
bias (%)	-13.5	-83.8	-100.0	-9.8	-7.9	-7.1
st.dev.	4.5	5.3	0	10.4	12.1	14.2
% constrained	0.0	58.0	100.0	4.0	9.0	12.0
median b:						
bias (%)	-1.6	-4.7	-12.3	-2.9	-3.8	-5.7
st.dev.	0.003	0.006	0.009	0.006	0.007	0.008
constant:						
bias (%)	0.5	3.2	4.1	0.4	0.3	0.2
st.dev.	19.0	29.2	31.8	42.9	51.0	60.8
s_e:						
bias (%)	-0.1	-0.2	0.1	0.1	0.3	0.8
st.dev.	6.0	11.5	18.6	11.0	15.5	19.8

REFERENCES

Abramowitz, M. & Stegun, I.A. (1965), *Handbook of Mathematical Functions,* Washington, DC: US Government Printing Office

Blomquist, N. S. (1995), "Restrictions in Labor Supply Estimation: Is the MaCurdy Critique Correct?", *Economics Letters* 47, 229-235.

Blomquist, N. S. (1996), "Estimation Methods for Male Labor Supply Functions: How to Take Account of Taxes", *Journal of Econometrics* 70, 383-405.

Burtless, G. & Hausman, J. A. (1978), "The Effect of Taxation on Labor Supply", *Journal of Political Economy* 86, 1103-1130.

Goldfeld, S. M. & Quandt, R. E. (1972), "GQOPT: A Package of Numerical Optimization of Functions", Department of Economics, Princeton University.

Hausman, J. A. (1981), "Labor Supply" in *How Taxes Affect Economic Behavior*, (eds. Aaron, H. & Pechman, J.), Washington D.C: Brookings Institution.

Hausman, J. A. (1985), "Taxes and Labor Supply" in *Handbook of Public Economics* (eds. Auerbach, A. & Feldstein, M.), 213-263, Amsterdam: North Holland.

MaCurdy, T., Green, D. & Paarsch, H. (1990), "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply", *Journal of Human Resources* 25, 415-490.

Pencavel, J. (1986), "Labor Supply of Men", in *Handbook of Labor Economics* (eds. Aschenfelter, O. & Layard, R.), 3-102, Amsterdam: North Holland.