WORKING PAPER #234
INDUSTRIAL RELATIONS SECTION
PRINCETON UNIVERSITY
JULY 1988

Grouped Data Estimation and Testing in Simple Labor Supply Models

by

Joshua D. Angrist¹

Industrial Relations Section
Princeton University

July 1988

¹I thank O. Ashenfelter, D. Benjamin, D. Card, C. Goldin, A. Krueger and W. Newey for many helpful comments and conversations. Remaining errors are, of course, my own. I also thank D. Card for the use of his PSID extract and the Industrial Relations Section and the Princeton University Department of Economics for financial support.

Abstract

Labor supply research has not yet produced a clear statement of the size of the labor supply elasticity nor how it should be measured. Measurement error in hourly wage data and the use of inappropriate identifying assumptions can account for the poor performance of some empirical labor supply models. I propose here a generalization of Wald's method of fitting straight lines that is robust to measurement error, imposes mild testable identifying assumptions, and is useful for the estimation of life-cycle labor supply models with panel data. A convenient Two-Stage Least Squares (TSLS) equivalent of the generalized Wald estimator is presented and a TSLS over-identification test statistic is shown to be the test statistic for equality of alternative Wald estimates of the same parameter. These results are applied to labor supply models using a sample of continuously employed prime-age males. Labor supply elasticities from the two best-fitting models that pass tests of over-identifying restrictions range from 0.6 to 0.8 . A test for measurement error based on the difference between generalized Wald and Analysis of Covariance estimators is also proposed. Application of the test indicates that measurement error can account for low or negative Analysis of Covariance estimates of labor supply elasticities.

I. Introduction and overview

Discussions of economic issues ranging from welfare to tax reform generally involve assertions about the individual behavioral response to a change in wage rates.² Yet, this behavioral response, usually summarized by the elasticity of labor supply, has proven frustratingly difficult to measure. Surveys by Killingsworth (1983) and Pencavel (1986) catalog scores of estimates that vary widely and often fail to satisfy theoretical restrictions. Pencavel goes so far as to conclude that the "canonical labor supply model" may be of little empirical relevance. In studies using panel data, Ashenfelter (1984), Card and Abowd (1986, 1987), and Card (1987) present results that suggest many of the models used to estimate labor supply parameters are not internally consistent. Abowd and Card fit a number of intertemporal substitution and contracting models, nearly all of which fail tests of over-identifying restrictions.

A variety of economic and econometric explanations for the shortcomings of empirical labor supply models have been proposed. I focus here on a solution to two of the most cited econometric problems - measurement error in explanatory variables and the imposition of inappropriate identifying restrictions on the labor supply equation. Identifying restrictions are required because labor supply theory suggests that components of the error term in the labor supply equation, such as the discounted present value of lifetime wealth, are correlated with the hourly wage. When correlated error components are time-invariant, they may be eliminated by transformations like differencing or deviations from means (Analysis of Covariance). In practice, however, the problem of omitted variables is complicated by the fact that differencing and deviations from means tend to aggravate measurement error bias (Griliches and Hausman 1986). It is now

²A recent example is Blinder's (1987) discussion of the "Laffer Curve."

generally recognized that measurement error may account for a substantial fraction of the observed variation in earnings and hours worked in many of the data sets used for labor supply research (Abowd and Card 1987, Altonji 1986, Bound and Krueger 1987).

To overcome problems of measurement error and omitted variables without imposing strong assumptions, Ashenfelter (1984) suggested that Wald's (1940) method of fitting straight lines be used to estimate life-cycle labor supply models. Wald's method, motivated by the measurement error problem, is robust to measurement error. Furthermore, in the context of labor supply estimation, Wald's method imposes a fairly mild and testable identifying assumption: the absence of aggregate year effects in the labor supply equation. Finally, Wald estimates are easy to compute and intuitively appealing. The data are divided into two groups in a manner statistically independent of the unobservables in the equation to be estimated. The Wald slope estimate is the difference in dependent variable group means divided by the difference in independent variable group means.

The principal shortcoming of Wald's method in the context of labor supply estimation is that if year effects are omitted variables in equations grouped into annual averages, then Wald estimates will be inconsistent. But there is generally no theoretical reason for aggregate period-specific demand shocks to be in the labor supply equation, although they will appear in the reduced form for wages.³ In fact, it is the non-stationarity of wage rates over time which provides the identifying information used by the Wald estimator.

³Examples of labor supply models without year effects include Ashenfelter and Ham (1979), Heckman and MaCurdy (1980) and some of the equations estimated by MaCurdy (1981), Browning, Deaton and Irish (1985), Altonji (1986) and Ham (1986). MaCurdy (1981) and Altonji (1986) also use year effects as instrumental variables. On the other hand, Browning Deaton and Irish (1985) find year effects to be important in some specifications and Abowd and Card (1986, 1987) work exclusively with data which has had the time means removed.

A further useful feature of Wald's method is that more than one Wald estimate of a single regression coefficient may be computed whenever the data can be divided into more than two groups. Comparison of alternative Wald estimates of the same parameter provides a simple over-identification test for the labor supply model. For example, the difference between annual averages from the first two years of a panel can be used to form one Wald estimate and the difference between annual averages from the last two years can be used to form another. The presence of aggregate year effects in the labor supply equation, as well as other forms of misspecification, should be signalled by a statistically significant difference between the two estimates.

Alternative Wald estimates of the same parameter may be combined into a single efficient estimate when the equality restrictions on the estimates are found to be acceptable. In the next section, an efficient linear combination of multiple Wald estimators is shown to be the Generalized Least Squares (GLS) estimator for grouped data. A convenient Two-Stage Least Squares (TSLS) equivalent of the GLS estimator is also presented. The Generalized Method of Moments (GMM) over-identification test statistic associated with this TSLS estimator is shown to be a test statistic for equality of alternative Wald estimates. In Section III, the GMM test statistic is used to confirm Ashenfelter's informal rejection of over-identifying restrictions in labor supply models estimated using the Panel Survey of Income Dynamics (PSID). The restrictions are found to be acceptable, however, when taste parameters in the models include either a linear or quadratic trend. Labor supply elasticities in the range 0.6 to 0.8 are estimated from the two best-fitting models that pass tests of over-identifying restrictions.

The fourth section of the paper discusses a test for measurement error based on the difference between Efficient Wald and Analysis of Covariance estimates.

In the absence of measurement error, Analysis of Covariance is the minimum variance estimator for linear models with unobserved individual effects. Consequently, the covariance matrix of the difference between an Efficient Wald estimate and an Analysis of Covariance estimate is the difference in their covariance matrices. This fact allows construction of a computationally convenient Hausman (1978) test statistic. Application of the Hausman test indicates that measurement error imparts substantial downward bias into Analysis of Covariance estimates constructed from regressions on PSID wage data.

II. Wald estimators and specification tests

Grouped Data and Instrumental Variables

Wald's method is based on grouping the data in a manner independent of the unobservables in the equation to be estimated. Thus, in labor supply models with a time-invariant fixed effect, time is an instrumental variable that may be used to group the data. In principle, however, any categorical instrument may be used to construct a grouped data estimator. When the instrument takes on more values than the number of parameters to be estimated, these parameters will be over-identified. Not surprisingly, the estimator which efficiently combines alternative Wald estimates is a form of Two-Stage Least Squares.

The equation of interest is assumed to be a bivariate regression model, with regressor X and an error term, u :

$$y = \alpha + X\beta + u, \quad E u u' = \Omega.$$

Even though X may be correlated with u , the researcher is assumed to have access

to an instrumental variable, R , which is independent of u and correlated with X .⁴ Furthermore, R is assumed to take on at least r distinct values, with $r \geq 2$, and the data are assumed to be ordered by the values of R . For example, in a panel data set, data from different periods should be stacked. It is also useful to establish notation for a matrix Z , the columns of which are a set of indicators for each value of R . When the data are ordered by the values of R , Z has the form

$$\begin{array}{cccccccc} \ell_1 & 0 & 0 & . & . & . & 0 \\ 0 & \ell_2 & 0 & . & . & . & 0 \\ 0 & 0 & . & & & & . \\ . & . & & & & & . \\ . & . & & & \ell_j & . & . \\ . & . & & & . & . & . \\ 0 & . & . & . & . & . & \ell_r \end{array}$$

where ℓ_j is a vector of n_j 1's and n_j is the number of individuals having the j th value of R .

Premultiplying by $(Z'Z)^{-1}Z'$ transforms the regression into a model for data grouped by the values of R . The structure of the ungrouped residual covariance matrix, Ω , is restricted here so that the covariance matrix of the grouped residual is proportional to a diagonal matrix with the reciprocal of the group sizes on the diagonal. That is, $E[(Z'Z)^{-1}Z'uu'Z(Z'Z)^{-1}]$ is assumed to be $(Z'Z)^{-1}\sigma^2$, for some scalar, σ^2 . Clearly if $\Omega = \sigma^2 I$, this requirement will be satisfied. Note, however, that this restriction includes the most commonly encountered variance components (fixed effect) model for panel data, grouped into annual averages. The usual variance components model for a balanced panel is, for $i = 1, \dots, N$ and $t = 1, \dots, T$:

$$\begin{aligned} y_{it} &= \alpha + X_{it}\beta + u_{it}, \quad u_{it} = \kappa_i + v_{it} \\ E\kappa_i v_{it} &= 0 \text{ for all } i \text{ and } t, \end{aligned}$$

⁴Independence is required because the indicator variables used to group the data are nonlinear functions of R .

$$E\kappa_i^2 = \tau^2, E\kappa_i\kappa_j = 0 \text{ for } i \neq j,$$

$$Evv' = \sigma_v^2 I_{NT}, \text{ where } v \text{ is the vector of } v_{it}.$$

The fixed effect, κ_i , may be viewed as a random variable that is correlated with X_{it} . Grouping the model by time gives

$$\bar{y}_t = [\alpha + \bar{\kappa}] + \bar{x}_t\beta + \bar{v}_t,$$

where $\bar{y}_t$, $\bar{x}_t$, $\bar{v}_t$ and $\bar{\kappa}$ denote time means. Because $\bar{\kappa}$ does not vary over time, it may be absorbed into the intercept. The vector of $\bar{v}_t$ has covariance matrix $(Z'Z)^{-1}\sigma_v^2$, which is the required form.⁵

A pair of grouped observations may always be used to construct a single Wald estimate. I will refer to a Wald estimate constructed from two groups as a "Pairwise Wald estimate." In general, however, when the data are divided into r groups, there are $(r-1)$ linearly independent differences that may be used to construct distinct Wald estimates of a regression coefficient. The GLS slope estimate constructed from r grouped observations is a minimum variance linear combination of $(r-1)$ Pairwise Wald estimates computed from a full set of linearly independent differences. This result is established by observing that the GLS grouped data estimator and any linear combination of Pairwise Wald estimators are both linear combinations of the grouped observations on the dependent variable. Because GLS is minimum variance in this class of linear estimators, there can be

⁵The fixed effect model for an unbalanced panel is also grouped into annual averages with a residual covariance matrix proportional to $(Z'Z)^{-1}$. Suppose there are n_t individuals in each period, m_t of whom appear only in period t . Let $f_t = (n_t - m_t)/n_t$. Then $\bar{\kappa}_t = f_t\kappa^* + (1-f_t)\bar{\kappa}_t^t$, where κ^* is the mean of κ_i for individuals appearing in all periods and $\bar{\kappa}_t^t$ is the mean of κ_i for those appearing only in period t . Assume that the mean of $(1-f_t)\bar{\kappa}_t^t$ has been included in α . The grouped model may be written as

$$\bar{y}_t = \alpha + f_t\kappa^* + \bar{x}_t\beta + \{(1-f_t)\bar{\kappa}_t^t + \bar{v}_t\},$$

where f_t is treated as a regressor, κ^* is a parameter, and the term in curly brackets is a compound error term with variance $(\tau^2 + \sigma_v^2)/n_t$. Because $\bar{\kappa}_t^t$ is an average of individuals appearing only in t , it has a diagonal covariance matrix. Therefore, the compound error variance is proportional to $(Z'Z)^{-1}$.

no more efficient linear combination of Pairwise Wald estimates than the GLS estimate. In the first section of Appendix A, I derive the general form of linear combinations of Pairwise Wald estimates that gives the GLS estimate.

Assuming that the grouped disturbances have a covariance matrix proportional to $(Z'Z)^{-1}$, the GLS minimand for the grouped data may be written as

$$(\bar{y} - \alpha - \bar{x}\beta)' [(Z'Z)] (\bar{y} - \alpha - \bar{x}\beta) / \sigma^2,$$

where $\bar{y}$ and $\bar{x}$ denote the vectors of r group means. Now, the Two-Stage Least Squares estimator of α and β that uses Z as the matrix of instruments minimizes

$$\mu = (y - \alpha - X\beta)' P_Z (y - \alpha - X\beta) / \sigma^2,$$

where $P_Z = Z(Z'Z)^{-1}Z'$. The TSLS and GLS minimands are easily seen to be identical and so the Two-Stage Least Squares estimates that minimize μ are identical to the GLS grouped data estimates. Because it is an efficient linear combination of Pairwise Wald estimators, I refer to the Two-Stage Least Squares estimator as an Efficient Wald estimator. The term Efficient Wald estimator will also be used for multivariate regression models that minimize a quadratic form in P_Z . Clearly the relationship between grouped data GLS and TSLS applies to multivariate models as well.⁶

⁶Durbin (1954) appears to have been the first to observe that Wald's method is an Instrumental Variables (IV) estimator. Malinvaud (1966, p. 359-63) notes (with reference to Wald and to a 1959 application published in French) that when there are more groups than parameters, models with mismeasured regressors may be consistently estimated by OLS on group means. He also speculates on the relationship between grouping and instrumental variables (p. 361-2): "Clearly the method now considered and that which consists of using the [the variable whose values are used to classify the data into groups] as instruments are closely related . . . The results must be virtually identical whenever the number of groups is sufficiently great for the values of the [classification variable] not to vary too much within groups." This speculation is not quite right, however, because the grouping estimator actually uses separate indicators for each value of the classification variable as instruments. Bowden and Turkington (1984, p. 59-60) also discuss the IV interpretation of Wald estimators, mentioning that OLS on grouped data is related to the use of categorical dummies as instruments. They refer to Friedman's (1957) use of group means as an application of this principle, although Friedman makes no explicit mention of the IV interpretation of his estimation strategy.

Deaton (1985, p. 113) also observes that running least squares on grouped data is an instrumental variables procedure, although he does not discuss the exact relationship between grouped data GLS and Two-Stage Least Squares. Much of Deaton's paper is concerned with the bias introduced because the sample group means are "error-ridden" observations on the population group means. It is apparent from the analogy to Two-Stage Least Squares, however, that this "measurement-error bias" is only relevant in small-samples and that the GLS grouping estimator will be consistent (as group sizes get large) whenever the columns of Z are uncorrelated with the unobservables in the regression. This is simply the requirement that the grouping be done in a manner statistically independent of the unobservables.⁷ On the other hand, if one is concerned with the consistency of estimates as the number of groups gets large when the group size is fixed, Deaton's correction for measurement error may be appropriate. In practice, an applied researcher must decide which dimension of her sample is large enough to justify the use of asymptotic theory.

An Over-Identification Test for Grouped Data

Let $\hat{\mu}$ denote the Two-Stage Least Squares minimand evaluated at the parameter estimates. The fact that the Efficient Wald estimator is a form of Two-Stage Least Squares is useful because $\hat{\mu}$ is also a Generalized Method of Moments (GMM) over-identification test statistic. In a bivariate regression, there are r

⁷ A requirement emphasized by Wald (1940, p. 294) in his original paper: "...the formulas and statements . . . remain exactly valid for any arbitrary subdivision of the observations into two equal groups, provided that the subdivision is defined independently of the errors . . .". Deaton suggested that the data be grouped on the basis of "cohorts" which can be followed through a time series of cross-sections and that the cohort means of the unobservables be removed by differencing the grouped data. Yet, if the data are grouped in a manner independent of the unobservables, differencing is no longer necessary.

instruments available to estimate the regression slope and intercept and so $\hat{\mu}$ is distributed as chi-square with $(r-2)$ degrees of freedom. Now, $(r-2)$ is also the number of equality restrictions on the $(r-1)$ linearly independent Pairwise Wald estimates of β that can be constructed using the grouped data. In fact, the test statistic for these $(r-2)$ equality restrictions is identical to the GMM over-identification test statistic associated with the Efficient Wald estimator. Proof is outlined below only for the case where the data have been divided into equal size groups. Missing details and a proof for the more general case are relegated to Appendix A (Sections 2, 3 and 4).

Pairwise Wald estimates based on $(r-1)$ first differences may be constructed by solving for the parameters in the following system:

$$\begin{array}{l} \bar{y}_2 - \bar{y}_1 \\ \bar{y}_3 - \bar{y}_2 \\ \vdots \\ \bar{y}_r - \bar{y}_{r-1} \end{array} = \begin{bmatrix} \bar{x}_2 - \bar{x}_1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & \bar{x}_3 - \bar{x}_2 & 0 & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdot & \cdot & \cdot & \bar{x}_r - \bar{x}_{r-1} & \cdot \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_r \end{bmatrix}$$

Denote a column vector of first differences by the prefix Δ so that this system may be written compactly as $\Delta \bar{y} = H\beta$, where $H = \text{diag}(\Delta \bar{x})$. The hypothesis, $H_0: \beta_2 = \beta_3 = \dots = \beta_r$, can be tested by computing the Wald test statistic (W) for equality of the unrestricted estimates, or by computing the Lagrange Multiplier (LM) statistic in restricted estimation. The argument for equality of W and $\hat{\mu}$ turns on the fact that the W and LM statistics for linear hypotheses in linear models differ only by the estimator used for the residual variance (Newey and West 1987). I assume that the same residual variance estimator is used to construct both statistics so that $W = LM$. It is straightforward to show that LM is equal to $\hat{\mu}$.

The linear hypothesis H_0 may be written

$$B = \ell_{r-1} \beta,$$

where ℓ_{r-1} is a vector of $(r-1)$ 1's and β is the scalar regression coefficient to which the elements of B are restricted to be equal. Then, following the notation in Newey and West (1987), the LM test statistic is

$$\begin{aligned} \text{LM} &= \min_B (\Delta\bar{y} - H\bar{B})' \hat{V} (\Delta\bar{y} - H\bar{B}) \text{ s.t. } B = \ell_{r-1}\beta \\ &= \min_{\beta} (\Delta\bar{y} - \Delta\bar{x}\beta)' \hat{V} (\Delta\bar{y} - \Delta\bar{x}\beta) \end{aligned}$$

where $\hat{V}$ is a consistent estimate of the GLS weighting matrix. Let the matrix which transforms r observations into $(r-1)$ first differences be A_1 . Then $\hat{V} = (A_1 A_1')^{-1} / \hat{\sigma}^2$. Now, in Appendix A (Sections 2 and 3), $A_w' (A_w A_w')^{-1} A_w$ is shown to be equal to A_g , where A_w is any matrix which converts an r -vector into $r-1$ linearly independent differences and A_g is a matrix which deviates an r -vector from its mean. Therefore, $A_g = A_1' (A_1 A_1')^{-1} A_1$. Let $\bar{y}^*$ and $\bar{x}^*$ be the mean of $\bar{y}$ and $\bar{x}$ and let $\tilde{y}^*$ and $\tilde{x}^*$ denote $\bar{y}$ and $\bar{x}$ in deviations from means form. The relationship between A_g and A_1 may be used to write

$$\begin{aligned} \text{LM} &= \min_{\beta} (\Delta\bar{y} - \Delta\bar{x}\beta)' \hat{V} (\Delta\bar{y} - \Delta\bar{x}\beta) = \min_{\beta} (\bar{y} - \bar{x}\beta)' A_1' (A_1 A_1')^{-1} A_1 (\bar{y} - \bar{x}\beta) / \hat{\sigma}^2 \\ &= \min_{\beta} (\bar{y} - \bar{x}\beta)' A_g (\bar{y} - \bar{x}\beta) / \hat{\sigma}^2 = (\bar{y}^* - \bar{x}^*\beta)' (\bar{y}^* - \bar{x}^*\beta) / \hat{\sigma}^2, \end{aligned}$$

where the last equality is a consequence of the fact that A_g is idempotent.

$\hat{\mu}$ reduces to the ordinary least squares minimand for grouped data when the groups are of equal size. Let $\hat{\alpha}$ and $\hat{\beta}$ be the least squares estimates of α and β . Then,

$$\hat{\mu} = \min_{\alpha, \beta} (\bar{y} - \alpha - \bar{x}\beta)' (\bar{y} - \alpha - \bar{x}\beta) / \hat{\sigma}^2 = (\bar{y} - \hat{\alpha} - \bar{x}\hat{\beta})' (\bar{y} - \hat{\alpha} - \bar{x}\hat{\beta}) / \hat{\sigma}^2.$$

Equality of $\hat{\mu}$ and LM follows immediately by noting that $\bar{y} - \hat{\alpha} - \bar{x}\hat{\beta}$ may be written as $(\tilde{y}^* - \tilde{x}^*\hat{\beta}) + (\bar{y}^* - \hat{\alpha} - \bar{x}^*\hat{\beta})$ and that $\hat{\alpha}$ satisfies $\bar{y}^* - \hat{\alpha} - \bar{x}^*\hat{\beta} = 0$.

Procedures for grouped data estimation and testing are also useful for multivariate regression models. When X has more than one column and β is a vector, it no longer makes sense to speak of a test statistic for the equality

of Pairwise Wald estimates. However, TSLS estimators which minimize a quadratic form in P_Z are consistent for parameters in multivariate models with mismeasured regressors. Furthermore, $\hat{\mu}$ remains a valid over-identification test statistic. In particular, the orthogonality of the columns of Z and the regression error is tested by $\hat{\mu}$ in both bivariate and multivariate models.

III. Efficient Wald estimates of labor supply elasticities from the PSID

Simple Labor Supply Models with Fixed effects

In labor supply models with intertemporally additive preferences in a certain environment, current period labor supply or labor earnings depend solely on contemporaneous variables and the time-invariant discounted utility of lifetime wealth. Three labor supply equations of this type are estimated here. The absence of time-varying regressors other than the wage rate allows the use of these equations to compute Pairwise Wald estimates, as well as Efficient Wald estimates.

The first model estimated is the Stone-Geary labor earnings equation:

$$(1) \quad w_{it} h_{it} = \gamma w_{it} - \phi \lambda_i^{-1}.$$

w_{it} and h_{it} are individual i 's wages and hours in period t , so that $w_{it} h_{it}$ are i 's earnings in t ; λ_i is i 's (unobserved) discounted present value of lifetime wealth. The parameter γ has the interpretation of being the maximum feasible hours of work and the Stone-Geary intertemporal labor supply elasticity is $\theta_1 = (\gamma/h_{it}) - 1$. To allow for the fact that λ_i may be correlated with w_{it} , λ_i can be treated as a "fixed effect" and, if there is no measurement error in w_{it} , γ may

be consistently estimated using first differences or Analysis of Covariance.⁸ Alternatively, Pairwise Wald estimates of γ may be constructed by replacing population means with sample means in the expression

$$(2) \quad [E(w_{it}h_{it}) - E(w_{it-j}h_{it-j})]/[E(w_{it}) - E(w_{it-j})] = \gamma.$$

Efficient Wald estimates are constructed by running GLS on annual averages when there are more than two periods.

The second model is a log-linear equation for hours, derived from the utility function discussed by Heckman and MaCurdy (1980). When the rate of time preference equals the interest rate, the Heckman and MaCurdy hours equation has the form

$$(3) \quad \ln h_{it} = A(1+\theta_2) + \theta_2 \ln w_{it} + \theta_2 \ln \lambda_i,$$

where λ_i is the time invariant discounted present value of wealth, 'A' is a taste parameter which shifts the utility of leisure and θ_2 is the Heckman and MaCurdy labor supply elasticity. The Pairwise Wald estimator of θ_2 replaces expectations by sample means in the expression

$$(4) \quad [E(\ln h_{it}) - E(\ln h_{it-j})]/[E(\ln w_{it}) - E(\ln w_{it-j})] = \theta_2.$$

When the rate of time preference and the interest rate are unequal, equation (3) includes a linear trend, the coefficient on which is interpreted as the difference between the two rates.

The Heckman and MaCurdy hours equation appears to generate a relationship between log hours and log earnings which may be used to construct a third Pairwise Wald estimate of the labor supply elasticity. Let the theoretical counterpart of this estimate be denoted by θ_3 . Adding $\theta_3 \ln h_{it}$ to both sides of (4) and dividing by $(1+\theta_3)$ gives

⁸ Ashenfelter and Ham (1979) estimate the Stone-Geary earnings equation by Analysis of Covariance and Heckman and MaCurdy (1980) treat λ_i as a fixed effect in a nonlinear hours equation. See also MaCurdy (1981), Browning, Deaton and Irish (1985), Altonji (1986) and Ham (1986), each of whom estimate hours equations in first differences.

$$(5) \quad \ln h_{it} = A + [\theta_3/(1+\theta_3)] \ln(w_{it} h_{it}) + [\theta_3/(1+\theta_3)] \ln \lambda_i,$$

so that, in principle, Pairwise Wald estimates of θ_3 may be computed from the sample analog of

$$(6) \quad [E(\ln h_{it}) - E(\ln h_{it-j})] / [E(\ln w_{it} h_{it}) - E(\ln w_{it-j} h_{it-j})] = \theta_3 / (1 + \theta_3).$$

In many data sets, however, including the PSID, w_{it} is constructed by dividing annual earnings by annual hours. When wage data are computed in this manner, Pairwise Wald estimates of θ_2 will be numerically equal to Pairwise Wald estimates of θ_3 .⁹ The only reason Ashenfelter (1984) got different estimates of θ_2 and θ_3 using the PSID is because he computed the sample analog of

$$[\ln E(h_{it}) - \ln E(h_{it-1})] / [\ln E(w_{it}) - \ln E(w_{it-1})] = \theta_2'$$

and of

$$[\ln E(h_{it}) - \ln E(h_{it-1})] / [\ln E(w_{it-1} h_{it-1}) - \ln E(w_{it-1} h_{it-1})] = \theta_3' / (1 + \theta_3').$$

Because the mean of the log is not equal to the log of the mean, Pairwise Wald estimates of θ_2' and θ_3' will generally be different from Pairwise Wald estimates of θ_2 and θ_3 ; Pairwise Wald estimates of θ_2' will also differ from Pairwise Wald estimates of θ_3' .

Data and Results

The PSID sample used to demonstrate the Efficient Wald estimator consists of 1437 male household heads, aged 21 to 64, with positive hours and earnings in every year from 1969 to 1979 (waves III through XII). This sample is similar to that used by Abowd and Card (1986, 1987) and to that used for some of the grouped

⁹ Estimates of θ_2 and θ_3 will not be equal when equations (3) and (5) are estimated using procedures more general than Wald's method. Thus, MaCurdy (1981) got different estimates of θ_2 and θ_3 when fitting (3) and (5) in first differences using instrumental variables. In general, the Efficient Wald procedure also produces different estimates of θ_2 and θ_3 .

data examples in Ashenfelter (1984).¹⁰ Mean earnings, wages and hours in 1967 dollars are shown in the first three columns of Table 1. The wage variable is computed by dividing annual earnings by annual hours.

The results of estimating labor supply models with and without a trend are discussed separately, below.

1. Pairwise and Efficient Wald estimates of models without a trend

The first set of results, also reported in Table 1, consists of all Pairwise Wald estimates of θ_1 , θ_2 , θ_2' , θ_3 and θ_3' that may be computed from first differences. As noted above, when the wage variable is computed by dividing annual earnings by annual hours, Pairwise Wald estimates of θ_2 and θ_3 will be equal. Pairwise Wald estimates computed using the log of the mean (θ_2' and θ_3'), however, will be different from θ_2 and θ_3 and from each other.¹¹ The results in Table 1 indicate that, in some years, the differences between estimates of θ_2 and θ_2' and between estimates of θ_3 and θ_3' are substantial. Also, while they have a common pattern, the alternative Pairwise Wald estimates vary widely by both year and model specification. The variability of elasticities from year to year suggests that neither equations (1), (3), or (5) provide an internally consistent description of the data. In addition, negative estimates of θ_1 violate the concavity restrictions on the underlying Stone-Geary utility function.

Efficient Wald estimates and the associated GMM over-identification test

¹⁰The sample, described in Appendix B, excludes men with real hourly wages in excess of \$100 in any year and men with hours in excess of 4680 in any year. The Abowd and Card sample (1986, 1987) is constructed using these selection criteria as well.

¹¹The mean of the log and the log of the mean of PSID wages, earnings, and hours are reported in Appendix B. Because $E[\ln(x)] \approx \ln[E(x)] - (1/2)[E(x)]^{-2}\sigma_x^2$, the mean of the log is always less than the log of the mean.

statistic are presented in the upper panel of Table 2, labeled "No Trend." The GMM test statistic provides a formal check of whether instability in the Pairwise Wald estimates is large relative to sampling variance. Estimates and test statistics are constructed by running TSLS on equations (1), (3), and (5), using 10 year dummies plus an intercept as instruments.¹² The results show that the over-identifying restrictions on each equation are strongly rejected.¹³

2. Efficient Wald estimates of models with a trend

The time invariance of all labor supply parameters is a strong restriction, especially in view of the fact that the data are observations on an aging cohort. A modest relaxation of time-invariance is to allow the maximum number of work hours in the Stone-Geary specification (γ in equation 1), and the leisure preference parameter in the Heckman and MaCurdy specifications ('A' in equations 3 and 5), to be a function of time. Inclusion of a (linear) trend in the Heckman and MaCurdy specifications is also justified whenever the rate of time preference is not equal to the interest rate. Finally, inclusion of a trend may be justified by human capital accumulation over the life-cycle. The concavity of life-cycle earnings is a well-known theoretical property of human capital earnings functions and a concave trend in life-cycle hours and earnings is a "stylized fact" which has been well documented (e.g., Ghez and Becker 1975).

The assumption that γ has a trend leads to interaction terms in the Stone-Geary earnings equation. The trend in the Heckman and MaCurdy taste parameter is additive. Therefore, the estimating equations for models with a quadratic

¹² This instrument list is equivalent to a full set of 11 year dummies.

¹³ The GMM test statistic is computed as N times the R^2 from a regression of the TSLS residuals on the instruments (Newey 1985).

trend are

$$(1') \quad w_{it}h_{it} = \gamma_0 w_{it} + \gamma_1 t w_{it} + \gamma_2 t^2 w_{it} - \phi \lambda_i^{-1}$$

$$(3') \quad \ln h_{it} = A_0(1+\theta_2) + A_1(1+\theta_2)t + A_2(1+\theta_2)t^2 + \theta_2 \ln w_{it} + \theta_2 \ln \lambda_i$$

$$(5') \quad \ln h_{it} = A_0 + A_1 t + A_2 t^2 + [\theta_3/(1+\theta_3)] \ln(w_{it}h_{it}) + [\theta_3/(1+\theta_3)] \ln \lambda_i.$$

Models with a linear trend are estimated by setting A_2 or γ_2 to zero.

The identity of grouped data GLS and TSLS discussed in the previous section applies equally well to the equations above, even though these equations contain more than one time-varying regressor. Also, while the analogy to a test for equality of Pairwise Wald estimators is no longer appropriate, the TSLS minimand is still a GMM test statistic for the validity of group indicators as instruments in the multivariate model. If the addition of one or two trend parameters does not adequately capture all the time effects in the labor supply equation, the GMM over-identification test should provide a signal of misspecification.

Labor supply estimates computed with a trend and using year dummies as instruments are presented in the two lower panels of Table 2, labeled "Quadratic Trend" and "Linear Trend." Because the estimate of γ is allowed to vary over time, Table 2 shows the value of γ calculated at the mean of trend. The complete set of estimates, including the actual trend coefficients, are reported in Appendix C. All the linear trend coefficients indicate a decline in hours worked over the sample period.

In equations with a quadratic trend, the over-identifying restrictions are rejected with a prob-value around .10 for θ_1 and are found to be acceptable for θ_2 and θ_3 . In specifications (3') and (5'), the quadratic term in the trend is not significantly different from zero (see Appendix C). Restricting the trend to be linear does not appreciably affect the estimates from these specifications,

although the Stone-Geary elasticity, θ_1 , becomes positive. With a linear trend, the prob-value for the GMM test in specification (1') is again around .10, while the tests for (3') and (5') provide no evidence of misspecification. Thus, at conventional levels of significance, both the Stone-Geary and Heckman-MaCurdy equations with either a quadratic or linear trend pass the test of over-identifying restrictions. However, estimates of θ_2 and θ_3 computed with a linear or quadratic trend appear to be the most plausible, and, according to the GMM test statistic, the Heckman and MaCurdy equations give a substantially better fit than does the Stone-Geary. Elasticities computed from equations (3') and (5') range from 0.56 to 0.79.¹⁴

IV. A test for measurement error

The Efficient Wald estimator is only one of many solutions to the problem of measurement error in panel data. Thus, Griliches and Hausman (1986) use alternative assumptions about the serial correlation properties of the measurement error to construct instrumental variables of the form PX , where P is a transformation matrix which is orthogonal to the fixed effect. The Griliches and Hausman estimators are over-identified as long as the serial correlation in the measurement error is sufficiently restricted. In some applications, however, there may be no instrument list which passes tests of over-identifying restrictions.¹⁵ Furthermore, there is generally no theoretical reason for

¹⁴The Stone-Geary and Heckman-MaCurdy labor supply models are not nested and so classical theories of statistical inference do not justify preference for the model which produces the smallest $\hat{\mu}$. Selection on the basis of $\hat{\mu}$ may be rationalized, however, by a Bayesian posterior-probability criterion. See Schwarz (1978) and Chow (1983, Chapter 9).

¹⁵MaCurdy (1981) and Altonji (1986) also use instrumental variables to deal with measurement error in log-linear labor supply equations estimated in first differences. The instruments used include age and education, as well as year

preferring a particular set of assumptions about serially correlated measurement error. Consistency of Efficient Wald estimates requires only that serially correlated measurement error have time means that are asymptotically zero.

Describing what is now called a "Hausman test", Durbin (1954) suggested that ordinary least squares (OLS) estimates be tested for measurement error bias by comparing them to estimates that are consistent in the presence of measurement error. Following Durbin, it seems natural to compare Analysis of Covariance estimates, which aggravate measurement error bias, with Efficient Wald estimates.¹⁶ In the absence of measurement error, the Analysis of Covariance estimator is Best Linear Unbiased (Mundlak 1978) for linear models with unobserved individual effects. Consequently, the covariance matrix of the difference between Analysis of Covariance and Efficient Wald estimates will be the difference in their covariance matrices. The degrees of freedom in a Hausman (1978) test for this difference is equal to the number of unknown parameters in the coefficient vector when there is no misspecification (Newey 1985, Theorem 3). Let the Efficient Wald estimator of a k by 1 parameter vector, ξ , be ξ_w and let the Analysis of Covariance estimator be ξ_c . Then the Hausman test statistic for the difference between ξ_w and ξ_c is simply

$$h_1 = (\xi_w - \xi_c)' [\hat{v}_w - \hat{v}_c]^{-1} (\xi_w - \xi_c) \sim \chi^2_k,$$

where $\hat{v}_w$ and $\hat{v}_c$ are the appropriate covariance matrices.

dummies. Yet, it seems likely that education is correlated with the residual from a regression of earnings growth on wage growth.

¹⁶The bias of an Analysis of Covariance slope estimate in the presence of serially uncorrelated measurement error in a panel of length T is $[-(T-1)/T]\sigma_v^2/\sigma_x^2$ where σ_v^2 is the variance of the measurement error and σ_x^2 is the variance of the mismeasured regressor in deviations from person means form (Griliches and Hausman 1986). The measurement error bias of an OLS slope estimate is $-\sigma_v^2/\sigma_x^2$, where σ_x^2 is the variance of the levels of the mismeasured regressor. Analysis of Covariance tends to aggravate measurement error bias because $\sigma_x^2 \geq \sigma_x^2$. For similar reasons, differencing also aggravates measurement error bias.

Results from a comparison of Analysis of Covariance and Efficient Wald estimates of equations which include a trend are shown in Table 3. Ordinary Least Square estimates of each equation are also presented in the table. Measurement error tests for specifications without a trend are not presented, however, because Efficient Wald estimates do not pass the test of over-identifying restrictions in these specifications. The assumption that the Efficient Wald estimates are consistent underlies the measurement error test. This assumption appears to be satisfied only for specifications which include a trend.

The measurement error tests indicate that Analysis of Covariance estimates from equations (1') and (3') are significantly smaller than the Efficient Wald estimates.¹⁷ The Analysis of Covariance estimates, negative and statistically significant, are rejected in models (1') and (3') with both a linear and quadratic trend. The rejection is particularly powerful for the Stone-Geary equation, (1'). On the other hand, the Analysis of Covariance and Efficient Wald estimates of (5') are not significantly different. Thus, the results indicate that coefficients from regressions on average hourly earnings are biased by measurement error and that coefficients from regressions on annual earnings are not.¹⁸ Duncan and Hill's (1985) validation study of the PSID suggests that average hourly earnings are much more poorly measured than annual earnings. The fact that specifications (1') and (3') fail the measurement test, while estimation of (5') provide no evidence of measurement error, appears to

¹⁷ Each Hausman test is for the difference between all parameters in the model except the intercept.

¹⁸ A small amount of measurement error in earnings may account for the fact that the Analysis of Covariance estimates of parameters in (5') are slightly larger than the Efficient Wald estimates. If earnings equals the product of wages and hours, then measurement error in hours induces spurious positive correlation between hours and earnings.

corroborate the Duncan and Hill findings.

Ordinary Least Squares estimates of θ_1 and θ_2 are bracketed by the Efficient Wald and Analysis of Covariance estimates. This result suggests that, in regressions on average hourly earnings, measurement error may induce even more bias than omitted variables. In contrast, OLS estimates of θ_3 , significant and positive, are relatively close to the Efficient Wald estimates.

V. Conclusion

This paper shows how multiple Wald estimates of the same parameter may be conveniently combined to estimate and test life-cycle labor supply models. An efficient linear combination of multiple Wald estimates of the same parameter is the GLS grouped data estimator, or, equivalently, a Two-Stage Least Squares estimator which uses a set of group dummy variables as instruments. Like Wald's original estimator, the Efficient Wald estimator is robust to measurement error in the explanatory variables and imposes mild identifying assumptions. The Wald test for equality of a full set of linearly independent Wald estimates is shown to be equal to an over-identification test statistic associated with the Two-Stage Least Squares equivalent of the Efficient Wald estimator. These results are applied to labor supply models estimated using data from the PSID. The empirical work indicates that simple labor supply models provide an internally consistent description of the data when the models are augmented with a linear or quadratic time trend. Labor supply elasticities in the range 0.6 to 0.8 are obtained from the two best-fitting models that pass tests of over-identifying restrictions.

A test for measurement error based on the difference between an Efficient Wald estimator and an Analysis of Covariance estimator is also proposed.

Consistent with previous research, there is evidence of substantial measurement error bias in estimates constructed from regressions on average hourly earnings. Many of the regressions on average hourly earnings reported in the labor supply literature are estimated using transformations that tend to aggravate measurement error bias. The fact that previously reported elasticities are generally too small to explain much of the variance in annual hours may be partly a consequence of measurement error in the explanatory variable.

Table 1: Pairwise Elasticity Measures

Year	<u>Means</u>			(a)	(b)	(c)	(d)
	Earnings	Wages	Hours ($\bar{h}_t$)	θ_1	$\theta_2 - \theta_3$	θ_2'	θ_3'
1969	7994.8	3.6110	2254.5				
1970	8122.0	3.7470	2226.3	-0.583	-0.1321	-0.341	-0.444
1971	8368.8	3.8618	2241.9	-0.038	0.1049	0.231	0.304
1972	8880.8	4.0493	2247.8	0.216	0.2012	0.056	0.047
1973	9267.9	4.1605	2283.7	0.537	0.9603	0.585	0.590
1974	8891.5	4.1100	2206.7	2.323	3.3967	2.813	4.800
1975	8695.8	4.0552	2182.7	0.626	1.7867	0.812	0.961
1976	8989.4	4.1740	2200.4	0.128	0.4140	0.278	0.319
1977	9244.3	4.2637	2201.8	0.291	0.1106	0.032	0.025
1978	9249.0	4.4551	2205.3	-0.989	-0.9585	0.036	-1.492
1979	8901.4	4.4522	2155.0	55.075	2.3180	36.127	1.513

Notes: There are 1437 individuals in the sample.

$$\theta_1 = \gamma_t / \bar{h}_t - 1 \text{ where } \bar{h}_t = (\bar{h}_t + \bar{h}_{t-1})/2 \text{ and}$$

$$\gamma_t = [E(w_{it}h_{it}) - E(w_{it-1}h_{it-1})] / [E(w_{it}) - E(w_{it-1})].$$

$$\theta_2 = [E(\ln h_{it}) - E(\ln h_{it-1})] / [E(\ln w_{it}) - E(\ln w_{it-1})].$$

$$\theta_2' = [\ln E(h_{it}) - \ln E(h_{it-1})] / [\ln E(w_{it}) - \ln E(w_{it-1})].$$

$$\theta_3 = \eta / (1 - \eta) \text{ where}$$

$$\eta = \theta_3 / (1 + \theta_3) = [E(\ln h_{it}) - E(\ln h_{it-1})] / [E(\ln w_{it}h_{it}) - E(\ln w_{it-1}h_{it-1})].$$

$$\theta_3' = \eta' / (1 - \eta') \text{ where}$$

$$\eta' = [\ln E(h_{it}) - \ln E(h_{it-1})] / [\ln E(w_{it}h_{it}) - \ln E(w_{it-1}h_{it-1})].$$

Table 2

Equation	Efficient Wald Estimates (standard errors)			Elasticities (delta-method standard errors)		
	(1)	(3)	(5)	θ_1^{***}	θ_2	θ_3
	γ	θ_2	$\theta_3/(1+\theta_3)$			
<u>No Trend</u>						
estimate	1473.69 (117.7)	-.132 (.051)	.031 (.053)	-.336 (.053)	-.132 (.051)	.032 (.056)
GMM test (χ^2_9)	44.25	53.74	63.23			
<u>Quadratic Trend</u> *						
	(1')	(3')	(5')			
estimate	2121.11 (436.0)	.634 (.269)	.442 (.099)	-.044 (.197)	.634 (.269)	.792 (.318)
GMM test (χ^2_7)	12.65	3.16	3.16			
<u>Linear Trend</u> **						
estimate	2754.87 (501.4)	.556 (.160)	.385 (.065)	.242 (.226)	.556 (.160)	.626 (.172)
GMM test (χ^2_8)	14.23	4.74	3.16			
Notes: * γ and standard error evaluated using $\bar{\gamma} = \gamma_0 + \gamma_1 E(t) + \gamma_2 E(t^2)$.						
** γ and standard error evaluated using $\bar{\gamma} = \gamma_0 + \gamma_1 E(t)$.						
*** Mean hours ($\bar{h}$) for all years = 2218.73, $\theta_1 = \bar{\gamma}/\bar{h} - 1$.						
delta-method standard errors for θ_1 are computed ignoring the correlation between γ and $\bar{h}$.						

Table 3

	Parameter Estimates (standard errors)			Elasticities (delta-method standard errors)		
	(1')	(3')	(5')			
<u>Quadratic Trend</u> *	γ	θ_2	$\theta_3/(1+\theta_3)$	θ_1^{***}	θ_2	θ_3
Efficient Wald	2121.11 (436.0)	.634 (.269)	.442 (.099)	-.044 (.197)	.634 (.269)	.792 (.318)
Analysis of Covariance	699.82 (10.66)	-.270 (.008)	.474 (.006)	-.685 (.005)	-.270 (.008)	.901 (.022)
Hausman Test(χ^2_3)	209.1	11.27	.105			
Ordinary Least Squares	1526.56 (9.92)	-.063 (.005)	.257 (.004)	-.312 (.005)	-.063 (.005)	.346 (.007)

Linear Trend**

Efficient Wald	2754.87 (501.4)	.556 (.160)	.385 (.065)	.242 (.226)	.556 (.160)	.626 (.172)
Analysis of Covariance	643.12 (10.75)	-.267 (.008)	.473 (.006)	-.710 (.005)	-.267 (.008)	.898 (.022)
Hausman Test(χ^2_2)	103.8	26.53	1.84			
Ordinary Least Squares	1490.56 (10.16)	-.063 (.005)	.258 (.004)	-.328 (.005)	-.063 (.005)	.348 (.007)

Notes: * γ and standard error are evaluated using $\bar{\gamma} = \gamma_0 + \gamma_1 E(t) + \gamma_2 E(t^2)$.

** γ and standard error are evaluated using $\bar{\gamma} = \gamma_0 + \gamma_1 E(t)$.

*** Mean hours ($\bar{h}$) for all years = 2218.73, $\theta_1 = \bar{\gamma}/\bar{h} - 1$.

delta-method standard errors for θ_1 are computed ignoring the correlation between γ and $\bar{h}$.

Hausman tests are for the differences between Analysis of Covariance and efficient Wald estimates of all the parameters in specifications (1'), (3') and (5') except the intercept.

Appendix A

The Algebra of Wald Estimators

1. Construction of grouped data GLS from a linear combination of Pairwise Wald estimators.

(r-1) linearly independent differences may be computed from r grouped observations. The complete set of Pairwise Wald estimators associated with (r-1) linearly independent differences is defined here to be a "full set of linearly independent Pairwise Wald estimators." Note that such a set is not unique. For example, Pairwise Wald estimators can be constructed by taking first differences or by differencing each observation from the first observation.

Let $\bar{y}$ denote the vector of $\bar{y}_i$ and denote the $r \times r$ matrix which transforms $\bar{y}$ into deviations from means by A_g and the $(r-1) \times r$ matrix which transforms $\bar{y}$ into first differences by A_1 . The GLS estimator is a linear combination of $A_g \bar{y}$, while any linear combination of a full set of Wald estimators constructed from first differences is a linear combination of $A_1 \bar{y}$. For example, if ϕ_1 denotes an arbitrary linear combination of Pairwise Wald estimators constructed from first differences, then ϕ_1 may be written

$$\phi_1 = \delta_1' C_1^{-1} A_1 \bar{y}$$

where δ_1 is any $(r-1) \times 1$ vector and C_1 is an $(r-1) \times (r-1)$ matrix with $\bar{x}_i - \bar{x}_{i-1}$ on the diagonal and zeros on the off-diagonal. In general, any linear combination of a full set of linearly independent Wald estimators may be written as

$$\phi_w = \delta_w' C_w^{-1} A_w \bar{y}$$

where A_w is the $(r-1) \times r$ matrix which transforms $\bar{y}$ into (r-1) linearly independent pairwise differences, C_w is an $(r-1) \times (r-1)$ diagonal matrix containing the "denominator" of the Wald estimators and δ_w is any $(r-1) \times 1$ vector. Now, let $\hat{x}$ be the mean of the ungrouped regressor, X . That is, $\hat{x} = (1/\sum n_j) \sum n_j \bar{x}_j$. Then we may

also write

$$\phi_g = \delta_g' C_g^{-1} A_g \bar{y}$$

for the GLS grouped data estimator, where δ_g is an r by 1 vector with elements equal to $(\bar{x}_j - \hat{x})n_j$ and C_g is an $r \times r$ matrix with $\Sigma(\bar{x}_j - \hat{x})^2$ on the diagonal and zeros on the off-diagonal.

For any full set of linearly independent Pairwise Wald estimators, defined by choice of A_w and C_w , it is possible to find a vector δ_w' such that $\phi_w = \phi_g$. This result is established by applying a theorem on differencing transformations, discussion of which begins with some definitions:

Definition 1 (Ijiri 1965). An incidence matrix is any $m \times n$ matrix with no zero rows and with columns which each contain exactly two non-zero entries equal to 1 and -1.

Definition 2 (Ijiri 1965). A connected incidence matrix is an incidence matrix where any two rows of the matrix are either directly connected or indirectly connected. Row i and row j of an incidence matrix are said to be directly connected if there is a column which has non-zero entries in both rows. Row i and row j are said to be indirectly connected if there is a sequence of rows, beginning with row i and ending with row j , in which every pair of adjacent rows are directly connected.

These definitions are related to Pairwise Wald estimates by the following lemma.

Lemma 1. The transpose of any matrix, A_w , which transforms a $1 \times r$ vector into a full set of $(r-1)$ linearly independent pairwise differences is a connected incidence matrix.

proof. First, note that the $(r-1) \times r$ matrix, A_1 , which transforms an r -vector into $(r-1)$ first differences has the form

$$A_1 = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & 1 & -1 \end{bmatrix}$$

and that A_1' is clearly a connected incidence matrix. However, it is obvious that any desired set of differences can be constructed from an $(r-1) \times r$ matrix (A_w) whose elements are exclusively 1, -1 and 0, where the placement of these elements will differ from the placement of elements in A_1 only by the distance between the 1 and -1 in each row. Also, since every observation must be included in at least one pairwise difference for the matrix to generate a full set of differences, there will be no zero columns in A_w and therefore, no zero rows in A_w' . Therefore, A_w' is an incidence matrix.

Proof of connectedness is established by observing that if A_w' were not connected, then it would be possible to write it as a block diagonal matrix with connected matrices on the diagonal (Ijiri 1965, p. 827-8). It is sufficient to consider the case wherein A_w' can be written as

$$A_w' = \begin{pmatrix} B_1' & O_1 \\ O_2 & B_2' \end{pmatrix}$$

and where B_1 is $r_1 \times n_1$, B_2 is $r_2 \times n_2$, O_1 and O_2 are conformable matrices of zeros and $r_1 + r_2 = r$. It is apparent that B_1 transforms the first r_1 observations into n_1 differences, that B_2 transforms the remaining r_2 observations into n_2 differences and that none of the first r_1 observations are differenced by A_w from the remaining r_2 observations. However, for the first n_1 differences to be linearly independent, it must true that $n_1 \leq r_1 - 1$; likewise, $n_2 \leq r_2 - 1$. But then the total number of differences generated by A_w will be $n_1 + n_2 \leq r - 2$. Therefore, because there are $(r-1)$ differences in a full set, the matrix which generates a full set must be connected.

The fact that A_w' is a connected incidence matrix enables application of the following theorem, which relates A_w to A_g :

Theorem 1 (Ijiri 1965). Let M be any $(r-1) \times r$ incidence matrix and let M^+ be the Moore-Penrose generalized inverse of M . Then we have,

$$I_r - MM^+ = \ell_r \ell_r' / r,$$

where ℓ_r is a vector of r 1's and I_r is the $r \times r$ identity matrix.

Applying theorem 1, it is immediate that

$$A_w' [A_w']^+ = I_r - \ell_r \ell_r' / r = A_g.$$

and, because A_g is symmetric, $([A_w']^+)' A_w = A_w^+ A_w = A_g$. This relationship is now used to construct δ_w' :

Lemma 2. Let $\delta_w' = \delta_g' C_g^{-1} A_w^+ C_w$. Then $\phi_w = \delta_w' C_w^{-1} A_w \bar{y} = \delta_g' C_g^{-1} A_g \bar{y} = \phi_g$.

Proof. Substitute for δ_w' in the expression for ϕ_w and use the fact that $A_w^+ A_w = A_g$.

We have have shown how to construct the GLS grouped data estimator from a linear combination of any full set of linearly independent Pairwise Wald estimators. As an example of how the GLS grouped data estimator may be

constructed from a particular set, consider the $(r-1)$ estimators which may be computed by differencing each group mean from the first group mean:

$$\hat{\beta}_2 = (\bar{y}_2 - \bar{y}_1)/(\bar{x}_2 - \bar{x}_1), \hat{\beta}_3 = (\bar{y}_3 - \bar{y}_1)/(\bar{x}_3 - \bar{x}_1), \dots, \hat{\beta}_r = (\bar{y}_r - \bar{y}_1)/(\bar{x}_r - \bar{x}_1).$$

Recall that GLS on grouped data is simply ordinary least squares weighted by the square root of the group size (Prais and Aitchison 1954). Then it is easy to verify that the GLS slope estimator may be constructed from the Pairwise Wald estimators above by taking

$$\sum_{j=2}^r a_j \hat{\beta}_j, \text{ where } a_j = [(\bar{x}_j - \bar{x}) (\bar{x}_j - \bar{x}_1) n_j] / \left[\sum_{j=1}^r (\bar{x}_j - \bar{x})^2 n_j \right].$$

2. The relationship between two full sets of linearly independent Pairwise Wald estimators.

The transpose of any connected incidence matrix, A_w' , has the property that $A_w A_g = A_w$. This is because $A_w A_g = A_w [A_w^+ A_w] = A_w$. Also note that A_g , A_w and A_w^+ all have rank $(r-1)$. These facts may be used to show that full sets of linearly independent Pairwise Wald estimators are related by a non-singular transformation.

Lemma 3. Let $\Psi_1 = C_1^{-1} A_1 \bar{y}$ and $\Psi_2 = C_2^{-1} A_2 \bar{y}$ be two distinct full sets of $(r-1)$ linearly independent Pairwise Wald estimators constructed from r groups. Then $\Psi_1 = G \Psi_2$ where G is a non-singular $(r-1)$ by $(r-1)$ matrix.

proof. Let $G = C_1^{-1} A_1 A_2^+ C_2$. G is a nonsingular $(r-1)$ by $(r-1)$ matrix because A_1 and A_2^+ are of rank $(r-1)$, and C_1 and C_2 are both diagonal, non-singular $(r-1)$ by $(r-1)$ matrices. $\Psi_1 = G \Psi_2$ may be verified by substituting and using the fact that $A_1 A_2^+ A_2 = A_1 A_g = A_1$.

3. GLS Quadratic forms for models in differences and deviations from means.

A regression equation may be transformed into a model for linearly independent differences by premultiplying by the transpose of some connected incidence matrix, A_w' . As shown in the text, the LM statistic for the restrictions implied by equality of a full set of linearly independent Pairwise Wald estimators is the GLS minimand for the differenced model. If the undifferenced error has covariance matrix $\sigma^2 I$, the residual in the transformed model has covariance matrix $A_w A_w' \sigma^2$. GLS estimates of the transformed model therefore minimize

$$m_1 = (\bar{y} - \bar{x}\beta)' A_w' (A_w A_w')^{-1} A_w (\bar{y} - \bar{x}\beta) / \hat{\sigma}^2.$$

The constant may be dropped because A_w is orthogonal to a vector of 1's. Let A_g be the idempotent matrix which deviates an r-vector from its mean. To show that the minimized value of m_1 is the OLS minimand for grouped data, and hence the GMM test statistic for the case of equal grouped sizes, it is sufficient to establish that $A_w' (A_w A_w')^{-1} A_w = A_g$. Let A^- denote any generalized inverse (g-inverse) of a matrix A and let $\rho(A)$ denote the rank of A. Then,

Lemma 4 (Rao and Mitra 1971, Lemma 2.2.5c). If $\rho(QAP) = \rho(A)$ then $P(QAP)^- Q$ is a g-inverse of A.

Application of this lemma gives the desired result.

Lemma 5. $A_w' (A_w A_w')^{-1} A_w = A_g$ for any connected incidence matrix, A_w' .
proof. $A_w A_g = A_w$ and A_g is idempotent so $A_w' (A_w A_w')^{-1} A_w = A_w' (A_w A_g A_w')^{-1} A_w$. Also, $\rho(A_w) = \rho(A_g) = \rho(A_w' (A_w A_g A_w')^{-1} A_w)$, so $A_w' (A_w A_w')^{-1} A_w$ is a g-inverse of A_g by lemma 4. But if $A_w' (A_w A_w')^{-1} A_w$ is a g-inverse of A_g , then

$$A_g A_g^- A_g = A_g [A_w' (A_w A_w')^{-1} A_w] A_g = A_g.$$

Finally, because $A_w A_g = A_w$ and $A_w' = A_g A_w'$, we have $A_w' (A_w A_w')^{-1} A_w = A_g$.

4. GMM over-identification test statistics and the equality of Pairwise Wald estimates when the Wald estimates are based on groups of unequal size.

The argument for the general case begins exactly as in the special case discussed in Section 3, above. A regression equation may be grouped and transformed into a model for linearly independent differences by premultiplying by $A_w(Z'Z)^{-1}Z'$, where A_w' is any $(r-1)$ by r connected incidence matrix. As shown in the text, the LM statistic for the restrictions implied by the equality of a full set of linearly independent Pairwise Wald estimators is the the GLS minimand for the differenced model. If the original regression error covariance matrix is spherical, the transformed error covariance matrix is $[A_w(Z'Z)^{-1}A_w']\sigma^2$.

Therefore, GLS estimates of the transformed model minimize

$$\begin{aligned} m_2 &= (\bar{y} - \bar{x}\beta)' A_w' [A_w(Z'Z)^{-1}A_w']^{-1} A_w (\bar{y} - \bar{x}\beta) / \hat{\sigma}^2 \\ &= (y - \alpha - X\beta)' Z(Z'Z)^{-1} A_w' [A_w(Z'Z)^{-1}A_w']^{-1} A_w (Z'Z)^{-1} Z' (y - \alpha - X\beta) / \hat{\sigma}^2. \end{aligned}$$

The constant may be dropped because A_w is orthogonal to a vector of 1's; likewise the minimand is unaffected by writing the model in deviations from means form.

Let $F = (Z'Z)^{-1}A_w' [A_w(Z'Z)^{-1}A_w']^{-1}A_w(Z'Z)^{-1}$. Then the LM test statistic is

$$m_2 = (\bar{y} - \bar{x}\beta)' ZFZ' (\bar{y} - \bar{x}\beta) / \hat{\sigma}^2,$$

where $\bar{\cdot}$ denotes deviations from means.

Because $(Z'Z)$ has rank r , while A_w has rank $r-1$, $A_w' [A_w(Z'Z)^{-1}A_w']^{-1}A_w$ is not a g -inverse of $(Z'Z)^{-1}$. The key algebraic fact that relates m_2 to $\hat{\mu}$ is that the matrix $D = F - (Z'Z)^{-1}$ has columns which are in the null space of A_g . This result is established below.

Lemma 6 (Rao and Mitra 1971, Lemma 2.2.6c). $(A'VA)(A'VA)^-A' = A'$ for any matrices A and V such that $\rho(A'VA) = \rho(A)$.

Lemma 7. Let D be defined as above and let A_g be the r by r symmetric idempotent matrix which deviates an r -vector from it's mean. Then $A_g D = [0]_{r \times r}$, where $[0]_{r \times r}$ is an $r \times r$ matrix of zeros.

proof. First, note that $A_w' [A_w(Z'Z)^{-1}A_w']^{-1}A_w = A_g A_w' [A_w A_g (Z'Z)^{-1}A_g A_w']^{-1}A_g A_g =$

$A_g [A_g (Z'Z)^{-1} A_g']^{-1} A_g$ by Lemma 4, above. Therefore,

$$D = (Z'Z)^{-1} A_g' [A_g (Z'Z)^{-1} A_g']^{-1} A_g (Z'Z)^{-1} - (Z'Z)^{-1}$$

and

$$A_g D = A_g (Z'Z)^{-1} (Z'Z) (Z'Z)^{-1} A_g' [A_g (Z'Z)^{-1} (Z'Z) (Z'Z)^{-1} A_g']^{-1} A_g (Z'Z)^{-1} - A_g (Z'Z)^{-1}.$$

By setting $A' = A_g (Z'Z)^{-1}$ and $V = Z'Z$, the conclusion of Lemma 7 follows immediately from Lemma 6.

Using the fact that $F = (Z'Z)^{-1} + D$, m_2 may be written as

$$\begin{aligned} m_2 &= (\tilde{y} - \tilde{x}\beta)' Z [(Z'Z)^{-1} + D] Z' (\tilde{y} - \tilde{x}\beta) / \hat{\sigma}^2 \\ &= (\tilde{y} - \tilde{x}\beta)' Z (Z'Z)^{-1} Z' (\tilde{y} - \tilde{x}\beta) / \hat{\sigma}^2 + (\tilde{y} - \tilde{x}\beta)' Z D Z' (\tilde{y} - \tilde{x}\beta) / \hat{\sigma}^2. \end{aligned}$$

Because the minimized value of the GMM test statistic is the same whether the model is in deviations from means or in levels,¹⁹ equality of m_2 (evaluated at the parameter estimates) and the GMM statistic, $\hat{\mu}$, are established if the second term in the above expression for m_2 is zero. Now, the null space of A_g has dimension 1 (because the rank of A_g is $r-1$ and A_g is $r \times r$) and is spanned by an r -vector of 1's. Therefore, each column of D is proportional to an r -vector of 1's. Let a typical column of D be denoted by d_i . Because Z is a full set of indicators, the vector constructed by adding the columns of Z is a column of N 1's. We may write $Z d_i = \ell_N k$, where k is a scalar constant of proportionality and ℓ_N is a column vector of N 1's. But both $\tilde{y}' \ell_N$ and $\tilde{x}' \ell_N$ are equal to zero because $\tilde{x}$ and $\tilde{y}$ are vectors in deviations from means. Therefore, the minimized value of m_2 is indeed equal to $\hat{\mu}$.

The algebra of this section applies equally well to models with more than one regressor, although the analogy to a test statistic for equality of Pairwise Wald estimates is no longer appropriate. Provided that the number of groups exceeds the number of regressors (including the constant), the LM statistic for equality of slope coefficients on grouped and differenced data is the same as the GMM test statistic for the use of the columns of Z as instruments.

¹⁹The Efficient Wald estimator is equivalent to a Two-Stage Least Squares estimator for an equation with an intercept. The estimated parameters therefore fit the means of the data exactly (Theil 1971, Chapter 9).

Appendix B

Data Sources and Sample Selection

1. Data Sources

Earnings, average hourly earnings (wages) and hours are taken from the Michigan Panel Survey of Income Dynamics, Wave 13 Individual Family Tape.

Annual Earnings variables are V1196, V1897, V2498, V3051, V3463, V3863, V5031, V5627, V6174, V6767, V7413.

Annual Hours of work variables are V1138, V1839, V2439, V3027, V3423, V3823, V4332, V5232, V5731, V6336, V6934.

Average hourly earnings equals Annual Earnings divided by Annual Hours.

Nominal data were deflated by the CPI (all items), taken from the Economic Report of the President 1987, p. 307.

2. Sample Selection

Male heads of household aged 21-64 over the period 1969-79. Only household heads in households which experienced no change in head were included.

Sample restricted to heads with positive hours and earnings in each year from 1969-79.

Sample restricted to heads with average hourly earnings (in 1967\$) less than 100\$ in each year and with annual hours less than 4680 in each year.

There are 1437 men in the sample.

Mean Logs and Log Means of PSID Wages, Earnings and Hours

Year	Earnings	Wages	Hours	Mean Ln Earnings	Ln Mean Earnings	Mean Ln Wages	Ln Mean Wages	Mean Ln Hours	Ln Mean Hours
1969	7994.81	3.61102	2254.49	8.78937	8.98655	1.12021	1.28399	7.66916	7.72068
1970	8122.01	3.74702	2226.27	8.83115	9.00233	1.16835	1.32096	7.66280	7.70808
1971	8368.80	3.86179	2241.87	8.86199	9.03227	1.19626	1.35113	7.66573	7.71507
1972	8880.81	4.04932	2247.83	8.92320	9.09165	1.24722	1.39855	7.67598	7.71772
1973	9267.94	4.16048	2283.72	8.97366	9.13432	1.27296	1.42563	7.70070	7.73356
1974	8891.49	4.11003	2206.68	8.92422	9.09285	1.26171	1.41343	7.66250	7.69924
1975	8695.81	4.05521	2182.74	8.88971	9.07060	1.24933	1.40000	7.64038	7.68834
1976	8989.35	4.17397	2200.35	8.91672	9.10380	1.26843	1.42887	7.64829	7.69637
1977	9244.25	4.26366	2201.83	8.95189	9.13176	1.30010	1.45013	7.65179	7.69704
1978	9249.03	4.45507	2205.28	8.95217	9.13227	1.30677	1.49404	7.64540	7.69861
1979	8901.44	4.45222	2155.00	8.89821	9.09397	1.29051	1.49340	7.60770	7.67555

Appendix C

Parameter Estimates for Models (1'), (3') and (5') (standard errors)

(1')	<u>Quadratic Trend</u>				<u>Linear Trend</u>		
	Intercept	γ_1	γ_2	γ_0	Intercept	γ_1	γ_0
Efficient Wald	126.42 (1775.3)	27.82 (18.14)	-3.47 (1.04)	2113.9 (498.32)	-2452.01 (2041.72)	-26.82 (9.88)	2915.8 (557.6)
ANACOVA		135.53 (5.55)	-12.17 (.43)	446.3 (18.59)		-16.37 (1.30)	741.34 (15.8)
OLS	2579.05 (48.91)	188.48 (8.12)	-19.81 (.63)	1306.82 (24.14)	2739.9 (50.15)	-61.9 (1.81)	1861.9 (17.1)
(3')	$A_0(1+\theta_2)$	$A_1(1+\theta_2)$	$A_2(1+\theta_2)$	θ_2	$A_0(1+\theta_2)$	$A_1(1+\theta_2)$	θ_2
Efficient Wald	6.96 (.294)	-.018 (.013)	.0003 (.0008)	.634 (.269)	7.05 (.184)	-.014 (.003)	.556 (.160)
ANACOVA		.020 (.003)	-.002 (.0003)	-.270 (.008)		-.0007 (.0007)	-.267 (.008)
OLS	7.73 (.012)	.011 (.004)	-.0013 (.0003)	-.063 (.005)	7.76 (.008)	-.0039 (.0009)	-.063 (.005)
(5')	A_0	A_1	A_2	$\theta_3/(1+\theta_3)$	A_0	A_1	$\theta_3/(1+\theta_3)$
Efficient Wald	3.79 (.870)	-.014 (.006)	.00035 (.00045)	.442 (.099)	4.29 (.578)	-.009 (.001)	.385 (.065)
ANACOVA		-.016 (.003)	.0005 (.0002)	.474 (.006)		-.010 (.0006)	.473 (.006)
OLS	5.41 (.035)	-.005 (.004)	-.0003 (.0003)	.257 (.004)	5.41 (.035)	-.008 (.0008)	.258 (.004)

References

- Altonji, J.G. (1986). "Intertemporal Substitution in Labor Supply: Evidence From Micro Data," Journal of Political Economy 94 (Vol. 3, Part 2): S176-S215.
- Abowd, J.M. and D. Card (1986). "On the Covariance Structure of Earnings and Hours Changes," National Bureau of Economic Research Working Paper No. 1832, February 1986.
- Abowd, J.M. and D. Card (1987). "Intertemporal Labor Supply and Long Term Employment Contracts," American Economic Review 77: 50-68.
- Ashenfelter, O.C. (1984). "Macroeconomic Analyses and Microeconomic Analyses of Labor Supply," Carnegie-Rochester Conference on Public Policy 21: 117-156.
- Ashenfelter, O.C. and J. Ham. "Education, Unemployment and Earnings," Journal of Political Economy 87: S99-S116.
- Blinder, A. (1987). Hard Heads, Soft Hearts: Tough-Minded Economics for a Just Society, Reading, Mass.: Addison-Wesley Publishing Co., 1987.
- Bound, J. and A.B. Krueger (1987). "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right ?," Princeton University, Industrial Relations Section, Mimeo, November 1987.
- Bowden, R. and D. Turkington (1984). Instrumental Variables, Cambridge: Cambridge University Press, 1984.
- Browning, M., A. Deaton and M. Irish (1985). "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," Econometrica 53: 503-543.
- Card, D. (1987). "Supply and Demand in the Labor Market," Industrial relations Section Working Paper No. 228, November 1987.
- Chow, G.C. (1983). Econometrics, New York: McGraw Hill, Inc., 1983.
- Deaton, A. (1985). "Panel Data from a Time Series of Cross-Sections," Journal of Econometrics 30: 109-126.
- Duncan, G.J. and D.H. Hill (1985). "An Investigation into the Extent and Consequences of Measurement Error in Labor-economic Survey Data," Journal of Labor Economics 3: 508-532.
- Durbin, J. (1954). "Errors in Variables," Review of the International Statistical Institute 22: 23-32.
- Friedman, M. (1957). A Theory of the Consumption Function, Princeton: Princeton University Press, 1957.
- Ghez, G.R and G.S. Becker (1975). The Allocation of Time and goods Over the Life Cycle, New York: National Bureau of Economic Research, 1975.
- Griliches, Z. and J.A. Hausman (1986). "Errors in Variables in Panel Data," Journal of Econometrics 31: 93-118.

- Ham, J.C. (1986). "Testing Whether Unemployment Represents Intertemporal Labor Supply," Review of Economic Studies 53: 559-578.
- Hausman, J.A. (1978). "Specification Tests in Econometrics," Econometrica 46: 1251-1271.
- Heckman, J.J. and T.E. MaCurdy (1980). "A Life-Cycle Model of Female Labor Supply," The Review of Economic Studies 47: 47-74.
- Ijiri, Yuji (1965). "On the Generalized Inverse of an Incidence Matrix," Journal of the Society of Industrial and Applied Mathematics 13: 827-836.
- Killingsworth, M. (1983). Labor Supply, Cambridge: Cambridge University Press, 1983.
- MaCurdy, T.E. (1981). "An Empirical Model of Labor Supply in a Life Cycle Setting," Journal of Political Economy 89: 1059-1086.
- Malinvaud, E. (1966). Statistical Methods of Econometrics, Chicago: Rand McNally and Company, 1966.
- Mundlak, Y. (1978). "On the Pooling of Time Series and Cross-Section Data," Econometrica 46: 69-85.
- Newey, W.K. (1985). "Generalized Method of Moments Estimation and Testing," Journal of Econometrics 29: 229-256.
- Newey, W.K. and K.D. West (1987). "Hypothesis Testing with Efficient Method of Moments Estimation," International Economic Review 28: 777-787.
- Pencavel, J. (1986). "Labor Supply of Men: A Survey," In O. Ashenfelter and R. Layard, eds. The Handbook of Labor Economics Volume I, Amsterdam: North Holland, 1986.
- Prais, S.J. and J. Aitchison (1954). "The Grouping of Observations in Regression Analysis," Review of the International Statistical Institute 22: 1-22.
- Rao, C.R. and S.K. Mitra (1971). Generalized Inverse of Matrices and its Applications, New York: John Wiley and Sons, 1971.
- Schwarz, G. (1978). "Estimating The Dimensions of a Model," Annals of Mathematical Statistics 6: 461-464.
- Theil, H. (1971). Principles of Econometrics, New York: John Wiley and Sons, 1971.
- Wald, A. (1940). "The Fitting of Straight Lines if Both Variables are Subject to Error," Annals of Mathematical Statistics 11: 284-300.