Efficient Estimation With Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators

James P. Ziliak
Department of Economics, University of Oregon, Eugene, OR 97403-1285

I examine the empirical performance of instrumental variables estimators with predetermined instruments in an application to life-cycle labor supply under uncertainty. The estimators studied are two-stage least squares, generalized method-of-moments (GMM), forward filter, independently weighted GMM, and split-sample instrumental variables. I compare the bias/efficiency trade-off for the estimators using bootstrap algorithms suggested by Freedman and by Brown and Newey. Results indicate that the downward bias in GMM is quite severe as the number of moment conditions expands, outweighing the gains in efficiency. The forward-filter estimator, however, has lower bias and is more efficient than two-stage least squares.

KEY WORDS: Bootstrap; Life-cycle labor supply; Overidentifying restrictions; Split samples.

The panel-data literature offers little guidance on the relative empirical performance of instrumental variables (IV) estimators when applied to samples of the size typically encountered in practice. For example, when the number of observations is large (say, greater than 500), are the small-sample concerns of a bias/efficiency trade-off in generalized method-of-moments (GMM) raised by Tauchen (1986) and Altonji and Segal (1994) binding? Moreover, if bias is present, is it due to a correlation between the sample moments and the sample weight matrix (Altonji and Segal 1994) or to instruments weakly correlated with the endogenous regressor as recently discussed by Bound, Jaeger, and Baker (1995) and Angrist and Krueger (1995)? The latter issue is crucial for efficiency considerations because the optimal number of moments may contain instruments dated far into the past, possibly weakening the correlation between the instruments and the endogenous regressor(s). In this article, I examine the sample properties of several panel-data IV estimators first by applying them to a well-known life-cycle labor-supply model and then comparing the estimators in terms of the bias/efficiency trade-off via a bootstrap Monte Carlo.

For many panel-data applications, GMM is the obvious estimator of choice: It does not require a full specification of the stochastic process (Hansen and Singleton 1982), it is consistent asymptotically under a variety of situations including when the only instruments available are predetermined rather than strictly exogenous (Anderson and Hsiao 1982; Arellano and Bond 1991), and it attains the efficiency bound in the class of IV estimators (Chamberlain 1987; Ahn and Schmidt 1995; Arellano and Bover 1995). Tauchen (1986) demonstrated, however, that in samples typically encountered in time series applications ($N = 50$ or 75) GMM is biased as the number of moment conditions expands, leading to a bias/efficiency trade-off, and thus he recommended the use of "suboptimal" instrument sets.

Altonji and Segal (1994) extended the small-sample analysis of Tauchen to least squares optimal minimum-distance (OMD) estimation of covariance structures. They demonstrated via Monte Carlo simulation that the bias in OMD is quite severe, the source of which, like feasible general least squares, is due to a correlation between the sample moments and the estimated weight matrix used in optimally minimizing the distance between population and sample moments. They attempted to correct for the bias by developing a split-sample estimator, called independently weighted OMD (IW-OMD) but found that both OMD and IW-OMD are dominated in terms of lower bias and root mean squared error (RMSE) by equally weighted minimum distance. Although the findings of bias in GMM by Tauchen and by Altonji and Segal seem conclusive, they are for small samples, and there is no a priori expectation that comparable results exist in a large panel-data setting. If comparable negative results against GMM do exist, then a case could be made for less efficient estimators such as two-stage least squares (2SLS) or Keane and Runkle's (1992) forward-filter (FF) estimator.

Parallel to the time series research on IV estimators of Altonji and Segal (1994), Nelson and Startz (1990), and Tauchen (1986) is cross-sectional research on the properties of IV estimators when the correlation between the instruments and the endogenous regressor is weak (Bekker 1994; Staiger and Stock 1994; Angrist and Krueger 1995; Bound et al. 1995). As pointed out by Bound et al., a weak correlation between the instruments and the endogenous regressor can lead to (1) a large standard error, (2) a bias in IV even if a weak correlation exists between the instruments and the structural error, and (3) a bias in IV toward ordinary least squares (OLS) as the explanatory power of the instruments approaches 0. They demonstrated that the returns-to-schooling results from the overidentified models reported by Angrist and Krueger (1991) are biased toward OLS due to weakly correlated instruments and, worse still,

Journal of Business & Economic Statistics
October 1997, Vol. 15, No. 4

that similar IV results are found by using instruments from a uniform random-number generator.

Addressing the criticisms of Bound et al., Angrist and Krueger (1995) developed a computationally convenient split-sample IV (SSIV) estimator as an alternative to 2SLS. SSIV is not biased toward OLS; however, the estimator is biased toward 0, so they proposed to "inflate" the SSIV estimates with a bias-correction factor, giving unbiased SSIV (USSIV). They concluded that USSIV gives results comparable to 2SLS but at a substantial loss of efficiency due to the smaller samples used in estimation. Because panel data with a long time series present the opportunity for many (possibly weakly correlated) instruments, SSIV and USSIV may be viable alternatives to standard 2SLS or GMM.

In this article, I focus on panel-data IV estimators that are consistent asymptotically when only predetermined instruments are available. Applications that fall within this class of estimators include dynamic models, rational-expectations models, and simultaneous-equations models. Predetermined instruments complicate the estimation of such models because certain transformations that eliminate the model's temporally persistent latent heterogeneity, such as deviations from time-means (the "within" estimator), are inconsistent when instruments are predetermined (Keane and Runkle 1992). The estimators considered here include 2SLS, GMM, FF, SSIV, USSIV, and an IV analog to IW-OMD that I refer to as IW-GMM.

I study the sample properties of the estimators in a realistic setting using data from the Panel Study of Income Dynamics (PSID). The empirical model employed is MaCurdy's (1985) life-cycle labor-supply model under uncertainty. MaCurdy's model is pertinent to the issues studied here because uncertainty suggests a rational-expectations solution to the consumer's problem, thereby making the instrument set predetermined. In addition, the model produces estimates of the intertemporal substitution elasticity, a key parameter used in understanding the comovements in earnings and hours over the business cycle. For each estimator, I sequentially build up the moment matrix by adding extra years as instruments, and I test the specification with Sargan's test of the overidentifying restrictions (Godfrey 1988).

I complement the empirical investigation with a bootstrap Monte Carlo as developed by Efron (1979) and extended to IV by Freedman (1984) and Freedman and Peters (1984). The bootstrap is used to compare the estimators in terms of bias, efficiency, RMSE, median absolute error (MAE), and asymptotic coverage rates. The strength of the bootstrap over standard Monte Carlo analysis lies in the fact that, like the IV estimators themselves, the researcher simply approximates the empirical distribution of the estimator with a nonparametric estimate of the underlying error distribution. Although Freedman's method is asymptotically valid for bootstrapping estimators from overidentified models (Hahn in press), Brown and Newey (1995) and Hall and Horowitz (in press) showed that the Freedman algorithm does not yield an improvement in terms of coverage rates over asymptotic theory, and more importantly, it gives the wrong size asymptotically for the overidentifying restrictions test. Because each algorithm is asymptotically valid for the estimator, I compare bootstrap results from the approaches of both Freedman and Brown and Newey.

From a 10-year balanced panel of men, I find the following results of note. The downward bias in GMM is quite severe as the number of moment conditions expands, outweighing the gains in efficiency. The bias is due to a correlation between the sample moments used in estimation and the estimated weight matrix. The IW-GMM estimator is generally successful at eliminating the bias in GMM parameter estimates using Freedman's bootstrap algorithm; however, the standard errors from asymptotic theory seem to understate the true sampling variation, and the overidentifying-restrictions test tends to overreject. The levels distortion in the overidentifying-restrictions test persists in models with many moments, even after recentering the distribution using Brown and Newey's (1995) algorithm. Finally, the bias in FF parameter estimates is less than GMM and 2SLS, and it is more efficient than 2SLS.

1. ESTIMATION AND PREDETERMINED INSTRUMENTS

I begin with a brief overview of the IV estimators used in the empirical application and bootstrap simulations. Consider the linear regression for individual $i\,(i=1,\ldots,N)$ in time $t\,(t=1,\ldots,T)$

$$y_{it}=\alpha_i+x_{it}\beta+\varepsilon_{it}, \tag{1}$$

where α_i represents fixed latent heterogeneity, x_{it} is a $(1\times K)$ vector of predetermined explanatory variables, β is a $(K\times 1)$ vector of parameters to estimate, and ε_{it} is a random error that varies over i and t and is assumed distributed iid $(0,\sigma_\varepsilon^2)$. Under the assumption of fixed effects, the latent heterogeneity is correlated with the explanatory variables for all periods; that is, $E[\alpha_i|x_{it}]\neq 0$ for all t.

A common practice in panel data is to eliminate the fixed effect by taking deviations from the individual's time series means, known as the within transformation. In IV estimation with predetermined instruments, however, the within transformation is inconsistent. In particular, for the within transformation to be consistent it is necessary for the instrument set to be strictly exogenous to the model's error term for all periods, $E[\varepsilon_{is}|W_{it}]=0$ for all s,t; however, predetermined instruments only guarantee weak exogeneity, $E[\varepsilon_{is}|W_{it}]=0$ for all $s\geq t$. This inconsistency carries over into the class of endogenous random-effects estimators studied by Hausman and Taylor (1981). The first-difference (Anderson and Hsiao 1982; Keane and Runkle 1992; Schmidt, Ahn, and Wyhowski 1992) and orthogonal-deviations (Arellano and Bover 1995) transformations are consistent when applied with lagged levels of predetermined (or endogenous) regressors as instruments. Orthogonal deviations may offer efficiency gains over first differences because differencing exacerbates measurement errors-in-variables (Maeshiro and Vali 1988). Each transformation will be applied to the estimators.

1.1 Method-of-Moments Estimators

The first estimator I consider is 2SLS, which minimizes the distance between the sample moments and the population moments, giving equal weight to each observation. The 2SLS estimator produces consistent parameter estimates for either the first-difference or orthogonal-deviations transformations and is given as

$$\hat{\beta}_{2SLS} = (X'P(W)X)^{-1}(X'P(W)y), \tag{2}$$

where W is an $(N(T-1) \times L)$ matrix of instruments, $P(W) = W(W'W)^{-1}W'$ is the projection matrix of instruments, ε is the stacked $(N(T-1) \times 1)$ vector of residuals, X is the stacked $(N(T-1) \times K)$ matrix of regressors, and y is the stacked $(N(T-1) \times 1)$ dependent variable. Under the assumption of conditional homoscedasticity, $E[\varepsilon^2|W] = 0$, inference for the estimated 2SLS parameters is conducted with the variance–covariance matrix

$$\begin{aligned} \text{var}(\hat{\beta}_{2SLS}) &= \hat{\sigma}_\varepsilon^2 (X'P(W)X)^{-1} \\ \hat{\sigma}_\varepsilon^2 &= \left(\frac{1}{NT-N-K}\right)(y - X\hat{\beta}_{2SLS})^2. \end{aligned} \tag{3}$$

As noted by White (1982), 2SLS standard errors are inconsistent when the conditional homoscedasticity assumption is violated; thus, he proposed a robust covariance matrix estimated as

$$\begin{aligned} \text{var}(\hat{\beta}_{2SLS}) = (X'P(W)X)^{-1}X'P(W) \\ \times \hat{\Omega}P(W)X(X'P(W)X)^{-1}, \end{aligned} \tag{4}$$

where $\hat{\Omega}$ is a diagonal matrix of squared residuals.

Because the number of instruments typically exceeds the number of parameters estimated $(L > K)$, one can test the overidentifying restrictions with the Sargan test as $\varepsilon(\hat{\beta}_{2SLS})'P(W)\varepsilon(\hat{\beta}_{2SLS})/\hat{\sigma}_\varepsilon^2$, which is asymptotically distributed χ^2 with $L - K$ df (Godfrey 1988). In the case of conditional heteroscedasticity, Hansen's (1982) robust variant of the overidentifying-restrictions test given later is necessary for consistent inference.

1.1.1 Generalized Method-of-Moments. Hansen (1982) and White (1982) showed that improvements in efficiency over 2SLS are possible by optimally weighting the distance between the sample and population moments, with the weight being the inverse of the covariance matrix of sample moments. The ensuing GMM estimator typically relies on residuals from the 2SLS estimator for an initial consistent estimate of the covariance matrix. The GMM estimator is

$$\hat{\beta}_{gmm} = (X'P(W(\hat{S}))X)^{-1}(X'P(W(\hat{S}))y), \tag{5}$$

where $P(W(\hat{S})) = W(\hat{S})^{-1}W'$ is the projection matrix of instruments and $\hat{S} = W'\hat{\Omega}W$ is the optimal weight matrix that permits both conditional heteroscedasticity and autocorrelation in the covariance matrix $\hat{\Omega}$. The variance–covariance matrix for the GMM estimator is

$$\text{var}(\hat{\beta}_{gmm}) = (X'P(W(\hat{S}))X)^{-1}, \tag{6}$$

and Hansen's version of the overidentifying-restrictions test is $\varepsilon(\hat{\beta}_{gmm})'P(W(\hat{S}))\varepsilon(\hat{\beta}_{gmm})$, which is distributed asymptotically χ^2 with $(L-K)$ df.

1.1.2 Forward Filtering. As an alternative to 2SLS and GMM, Keane and Runkle (1992) proposed an estimator called the forward-filter (FF) estimator. The FF estimator eliminates all forms of serial correlation while still maintaining orthogonality between the initial instrument set, which contains lagged values of predetermined/endogenous variables, and the structural error. There is a similarity between forward-filtering and the orthogonal-deviations transformation developed by Arellano and Bover (1995) in that both methods demean the variables with only current and future values; however, filtering is likely to be superior because it eliminates all forms of serial correlation. Although Schmidt et al. (1992) argued that filtering is irrelevant if one exploits all sample moments during estimation, filtering may be a desirable alternative to GMM in practice. First, because the dimension of the GMM moment matrix grows exponentially as the number of time periods and regressors expands, it can be computationally intractable and the overidentifying restrictions are less likely to be satisfied, possibly due to a weak correlation between instruments and endogenous regressors. Second, if the small-sample evidence from Tauchen (1986) and Altonji and Segal (1994) carries over to the panel-data setting, a bias/efficiency trade-off will arise with the optimal GMM estimator.

Similar to GMM the FF estimator is a two-step estimator. In the first step, the first-differenced equation is estimated by 2SLS. The $((T-1) \times 1)$ vector of first-difference residuals for individual i from the 2SLS regression, $\hat{\varepsilon}_{i,2SLS}$, are used in constructing a $(T-1) \times (T-1)$ covariance matrix, $\hat{\Sigma}_{FD} = \frac{1}{N}\sum_{i=1}^{N} \hat{\varepsilon}_{i,2SLS}\hat{\varepsilon}'_{i,2SLS}$. The inverse of the covariance matrix is then filtered by a Cholesky decomposition, $\hat{C}_{FF} = \text{Chol}(\hat{\Sigma}_{FD}^{-1})$, that eliminates serial correlation in the differenced errors. The second step involves transforming the original stacked $N(T-1)$ first-difference observations by $\hat{Q}_{FF} = I_N \otimes \hat{C}_{FF}$, leading to the FF estimator

$$\hat{\beta}_{FF} = (X'\hat{Q}_{FF}P(W)\hat{Q}_{FF}X)^{-1}(X'\hat{Q}_{FF}P(W)\hat{Q}_{FF}y), \tag{7}$$

with variance–covariance matrix $\text{var}(\hat{\beta}_{FF}) = \hat{\sigma}_\varepsilon^2(X'\hat{Q}_{FF} P(W)\hat{Q}_{FF}X)^{-1}$. As noted by Hayashi (1992), inference with the FF estimator is inconsistent if the conditional homoscedasticity assumption is violated; however, heteroscedasticity-robust variants of the variance and the overidentifying restrictions test are easy to compute.

1.1.3 Independently Weighted Generalized Method-of-Moments. In an attempt to mitigate the finite-sample bias in OMD, Altonji and Segal (1994) developed a new estimator called independently weighted OMD. I extend their idea to the case of IV with panel data. The motivation behind IW-GMM is to break the correlation between the sample moments used in estimation, $(1/N)[W'\varepsilon]$, and the estimated weight matrix, $\hat{S} = W'\hat{\Omega}W$, that is constructed with the same data. The procedure is to randomly split the sample into independent groups (g), say two $(g = 1, 2)$, with group 1 used in constructing the weight matrix, $\hat{S}_1 = W_1'\hat{\Omega}_1 W_1$, and group 2 used in constructing the sample moments to estimate, $(1/N)[W_2'\varepsilon_2]$. The sample split must occur on the cross-sectional dimension of the data because each cross-sectional unit retains its own time series for instruments. Each group is used alternately in constructing the weight matrix and the sample moments so that the resulting IW-GMM estimator is the average of the independent estimations:

$$\hat{\beta}_{\text{IW-GMM}} = \frac{1}{G}\sum_{g=1}^{G} (X_g'W_g(W_{-g}'\hat{\Omega}_{-g}W_{-g})^{-1}W_g'X_g)^{-1} \times (X_g'W_g(W_{-g}'\hat{\Omega}_{-g}W_{-g})^{-1}W_g'y_g), \quad (8)$$

where $-g$ refers to the excluded group. Because the sampling errors of the population moments and weight matrix are independent, the average is a consistent estimator. The variance of the estimated average is constructed as $\text{var}(\hat{\beta}_{\text{IW-GMM}}) = (1/G^2)[\text{var}(\hat{\beta}_1) + \cdots + \text{var}(\hat{\beta}_G)]$, where the covariance between ith and jth sample split is 0 by construction. The IW-GMM estimator is expected to be less efficient than GMM because of the loss in degrees of freedom. In the application to follow, I fix $G = 2$ for the IW-GMM estimator.

1.1.4 Split-Sample Instrumental Variables. Angrist and Krueger (1995) confronted the problem of a weak correlation between the instruments and regressors in their earlier work by developing the split-sample IV (SSIV) estimator. Recall that if the instruments and the endogenous regressor(s) are weakly correlated and there exists a (weak) correlation between the error from the first-stage fitted value and the structural error, then IV is biased. SSIV breaks the correlation between the two errors by randomly splitting the sample in half and using one-half to estimate the first-stage equation and the other half to estimate the structural parameters. Let sample 2 estimate the first-stage equation and combine that with W_1 to form the fitted value for X_1, which is then regressed on y_1, yielding the SSIV estimator

$$\hat{\beta}_{\text{SSIV}} = (X_2'W_2(W_2'W_2)^{-1}W_1'W_1(W_2'W_2)^{-1}W_2'X_2)^{-1} \times X_2'W_2(W_2'W_2)^{-1}W_1'y_1, \quad (9)$$

which under conditional homoscedasticity has an asymptotic covariance matrix of $\text{var}(\hat{\beta}_{\text{SSIV}}) = \hat{\sigma}_\varepsilon^2(X_2'W_2(W_2'W_2)^{-1}W_2'X_2)^{-1}$. A loss of efficiency relative to 2SLS is expected with SSIV because only one-half of the observations are used in estimation. This divergence in efficiency may be exacerbated when heteroscedasticity-robust covariance matrices of Equation (4) are employed.

Because SSIV is biased toward 0, Angrist and Krueger (1995) inflated the SSIV estimator with a bias-correction factor, resulting in a just-identified 2SLS estimator in sample 1. USSIV is consistent under group asymptotics and is given by the formula

$$\hat{\beta}_{\text{USSIV}} = (X_1'W_{21}'(W_{21}'W_{21})^{-1}W_{21}'X_1)^{-1} \times X_1'W_{21}'(W_{21}'W_{21})^{-1}W_{21}y_1, \quad (10)$$

where $W_{21} = W_1(W_2'W_2)^{-1}W_2'X_2$. The asymptotic covariance matrix takes the usual form as 2SLS under conditional homoscedasticity as $\hat{\sigma}_{\varepsilon_1}^2(X_1'W_{21}(W_{21}'W_{21})^{-1}W_{21}'X_1)^{-1}$ and the robust covariance matrix is the same as in Equation (4) with W_{21} replacing W. For obvious reasons, there is no test of the overidentifying restrictions in this case.

1.2 Choice of Instrument Set

To this point, little attention has been given to the specification of the instrument set other than the overidentifying-restrictions test. When the matrix of explanatory variables is predetermined, the first-difference and orthogonal-deviations transformations make instruments dated $t - 1$ and earlier valid for estimation. In the standard IV estimator, choosing one-period lagged instruments, $W = X_{(t-1)}$, leaves the system just identified ($L = K$), whereas choosing instruments dated $t-1$ and $t-2$ imposes $p = 2K - K$ overidentifying restrictions but sacrifices an extra period of data for each observation.

Schmidt et al. (1992) argued that efficiency gains are possible if, instead of the usual instrument set, one exploits all of the linear moments available as implied by the orthogonality conditions $E[W_i'\varepsilon_{i,d}] = 0$, where W_i is the matrix of instruments for individual i and $\varepsilon_{i,d}$ is the vector of first-difference or orthogonal deviations residuals. The Schmidt et al. approach uses levels of instruments from different time periods for different observations. The instrument set for individual i when all the regressors are predetermined is constructed as

$$W_i = \begin{bmatrix} x_1 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & \cdot & 0 & 0 \\ 0 & x_1 & x_2 & 0 & \cdot & \cdot & 0 & 0 & \cdot & 0 & 0 \\ \cdot & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & x_1 & x_2 & \cdot & \cdot & x_{T-1} \end{bmatrix}, \quad (11)$$

which has dimension $(T-1) \times (T)(T-1)(K)/2$ and where $x_t(t = 1, \ldots, T-1)$ are the lagged levels of the explanatory variables. For example, in period 2 variables from period 1 are valid instruments, in period 3 variables from both periods 1 and 2 are valid, and so on until period T, where variables from periods 1 to $T-1$ are valid instruments. When T and K are both large, say $T = 15$ and $K = 10$, there are 1,040 overidentifying restrictions, highlighting the computational burden of the Schmidt et al. approach. In the application to follow, I compare the efficiency of the standard instrument set to the stacked instrument set in Equation (11).

2. AN APPLICATION TO LIFE-CYCLE LABOR SUPPLY UNDER UNCERTAINTY

A focal point of interest in the labor-supply literature is consistent and efficient estimation of the intertemporal substitution elasticity (ISE). The ISE measures intertemporal changes in hours of work due to an anticipated change in the real wage and aids in understanding comovements in earnings and hours over the business cycle. Consequently, reliable estimates of the ISE are of import to public policy in labor markets. The life-cycle labor-supply model developed by MaCurdy (1985) and Altonji (1986) forms the basis of the empirical exercise.

Based on a utility function that is additively separable between consumption and leisure, the double-log life-cycle labor-supply function for individual i $(i = 1, \ldots, N)$ in time t $(t = 1, \ldots, T)$ is

$$\ln h_{it} = \lambda_{i0} + \delta \ln \omega_{it} + z_{it}\gamma + \varepsilon_{it}, \qquad (12)$$

where ln is the natural log operator; h_{it} is annual hours of work; λ_{i0} is the marginal utility of initial wealth that is a function of all future wages, assets, prices, and tastes and is correlated with the explanatory variables; z_{it} is a vector of time-varying demographics; and δ is the ISE parameter. The assumptions between the regressors and structural random error are $E[\varepsilon_{is} | \ln \omega_{it}] = 0$ for all $s > t$ and $E[\varepsilon_{is} | z_{it}] = 0$ for all $s \geq t$. Wages are assumed to be endogenous rather than predetermined to account for the possible presence of nonlinear income taxes and/or human capital considerations, or possibly measurement error.

Estimation proceeds by taking first differences or orthogonal deviations on Equation (12) to eliminate the unobserved marginal utility of wealth and then using the estimators outlined in Section 1. Because of the endogeneity of wages and the desire to use lagged values of wages as instruments, an extra year of data is lost; consequently, there are a maximum of $(T - 2) \times (T - 2)[(K - 1)(T + 1) + T]/2$ linear moment conditions to exploit in estimation. The stacked instrument matrix in Equation (11) for the labor-supply model contains, in period 3, wages from period 1 along with demographics in periods 1 and 2 as instruments, wages from periods 1 and 2 along with demographics from periods 1 to 3 are instruments in period 4, and so on until period T when there are $(T - 2)$ lagged wages and $(K - 1)(T - 1)$ lagged demographic variables as instruments.

2.1 Data

The data used to estimate the life-cycle labor-supply parameters come from Waves XII–XXI (calendar years 1978–1987) of the PSID. The sample is selected on many dimensions and is similar to other research studying life-cycle models of labor supply. The sample is restricted to continuously married, continuously working, prime-age men aged 22–51 in 1978 from the Survey Research Center random subsample of the PSID. In addition the individual must either be paid an hourly wage rate or must be salaried, and he cannot be a piece-rate worker or self-employed. This selection process resulted in a balanced panel of 532 men over 10 years or 5,320 observations. The real wage rate, ω_{it}, is the hourly wage reported by the panel participant rather than the average wage (annual earnings over annual hours) to minimize division bias (Borjas 1981). The predetermined time-varying taste shifters, z_{it}, include a quadratic in age, the number of children in the household, and a dummy variable for bad health.

2.2 Results

In Table 1 I report the results from 101 regressions, 55 based on the first-difference transformation and 46 based on the orthogonal-deviations transformation. The FF estimator is not included under orthogonal deviations because filtering would be redundant in this case. I begin with the standard instrument set with z_{it}'s from $t-1$ and $t-2$, along with lagged wages from $t-2$. I then sequentially build up the stacked instrument set from Equation (11). Because the ISE is the parameter of primary interest, I only present results for the ISE in Table 1. The efficiency pattern in the demographic variables with changes in moment conditions is similar to the pattern in the ISE. The weighting matrix for the first-difference GMM estimator that corrected for a first-order moving average error failed to be positive definite in the stacked-moment models, even with modified Bartlett weights; hence, the results reported only correct for conditional heteroscedasticity. As a point of departure, I present OLS estimates at the bottom of Table 1. The OLS first-difference ISE is .111 with a heteroscedasticity-robust standard error of .079, and the orthogonal-deviations ISE is .176 with a heteroscedasticity-robust standard error of .074.

2.2.1 Base-Case Results. In the base case with demographic variables from $t-1$ and $t-2$ and wages from $t-2$ as instruments for the contemporaneous period, the first-difference 2SLS and GMM estimates of the ISE are .21 and .52, respectively; however, neither estimate is significant even at the 10% level. As a benchmark to judge the first-difference base-case results, Altonji (1986) estimated a range of values from .01 to .45, most of which are imprecisely estimated. The orthogonal-deviations 2SLS estimates are more efficient than their first-difference counterparts, possibly due to reduced measurement error relative to differencing; however, there is no general pattern of efficiency gains across the other orthogonal-deviations estimators. The IW-GMM estimates vary across the first-difference and orthogonal-deviations transformations in that the former inflate the corresponding GMM estimate by about 40% and the latter deflate it by about 40%. Of course, the objective of IW-GMM is to correct for bias in GMM, whether it be positive or negative bias. On the contrary, SSIV and USSIV uniformly inflate their 2SLS counterparts across specifications, although there is a substantial loss in efficiency due to smaller sample sizes. Interestingly, under conditional homoscedasticity, the FF estimator performs best on efficiency grounds in the base case, even surpassing GMM; however, once one accounts for heteroscedasticity, GMM and IW-GMM improve on the FF estimates.

The results from the Sargan overidentifying-restrictions tests are mixed across the models. The conditionally homoscedastic p values of .02 and .01 from the first-difference 2SLS and FF estimators do not lend much support to the choice of instruments; however, the GMM and orthogonal-deviations versions do not reject the overidentifying restrictions. The test rejects the IW-GMM specifications under first differences but does not reject SSIV models. Two important findings that emerge in the base case and are magnified in the stacked-moment results to follow are that (1) once the Sargan test is modified for the presence of heteroscedasticity, the overidentifying restrictions are not rejected in any of the 2SLS or FF specifications and (2) as

Table 1. Intertemporal Substitution Elasticities Under Alternative Moment-Condition Estimators

	First difference						*Orthogonal deviations*				
Moments	*2SLS*	*GMM*	*FF*	*IW-GMM*	*SSIV*	*USSIV*	*2SLS*	*GMM*	*IW-GMM*	*SSIV*	*USSIV*
Base case:											
$T-1, T-2$	.2091	.5192	.1350	.7161	.7158	1.7491	.2552	.4506	.2603	.8931	1.1224
	(.4155)	(.3638)	(.3163)	(.3498)	(.6183)	(2.0709)	(.3773)	(.4011)	(.4044)	(.6427)	(.7987)
	(.4231)	[.4482]	(.3857)	[.0059]	(.7109)	(1.8272)	(.4488)	[.6431]	[.0392]	(.8237)	(.8270)
{9}	[.0234]		[.0138]		[.3447]		[.1361]			[.5862]	
	[.3147]		[.1925]		[.4356]		[.5744]			[.8326]	
Stacked cases:											
$T-1$ to $T-2$	.5428	.3942	.5422	.3971	.6752	6.0887	.7130	.6158	.5069	.4289	8.2390
	(.1808)	(.1504)	(.1525)	(.0846)	(.2120)	(7.3626)	(.1719)	(.1586)	(.0908)	(.1764)	(19.205)
	(.2259)	[.4248]	(.2007)	[.0000]	(.4878)	(7.1342)	(.2119)	[.2040]	[.0000]	(.4880)	(18.444)
{72}	[.0700]		[.0284]		[.0000]		[.0697]			[.8188]	
	[.3837]		[.2778]		[.0002]		[.1185]			[.9315]	
$T-1$ to $T-3$	.5620	.3916	.5093	.2868	.3768	−3.4139	.5885	.3768	.3724	.1070	−2.4665
	(.1436)	(.1158)	(.1225)	(.0584)	(.1609)	(3.6721)	(.1353)	(.1156)	(.0626)	(.1342)	(6.1924)
	(.1817)	[.4560]	(.1779)	[.0000]	(.3564)	(4.2839)	(.1792)	[.2209]	[.0000]	(.3406)	(7.2017)
{107}	[.0632]		[.0089]		[.0017]		[.0109]			[.3946]	
	[.3932]		[.2096]		[.0065]		[.0959]			[.4759]	
$T-1$ to $T-4$	.3774	.1933	.4568	.1832	.0881	.6702	.5432	.2778	.2819	−.0489	.3272
	(.1279)	(.1013)	(.1118)	(.0462)	(.1168)	(.8119)	(.1242)	(.1031)	(.0497)	(.1064)	(.9251)
	(.1780)	[.3192]	(.1711)	[.0000]	(.2886)	(1.2464)	(.1733)	[.2363]	[.0000]	(.3007)	(1.4420)
{137}	[.0000]		[.0003]		[.0000]		[.0002]			[.0000]	
	[.2211]		[.2715]		[.0002]		[.0772]			[.0000]	
$T-1$ to $T-5$	.3123	.1186	.3108	.3170	.0218	−.1169	.4006	.1714	.2838	−.0898	.5920
	(.1179)	(.0878)	(.1008)	(.0346)	(.1044)	(.5247)	(.1133)	(.0898)	(.0412)	(.0972)	(.7368)
	(.1742)	[.1393]	(.1829)	[.0000]	(.2389)	(.7319)	(.2000)	[.0480]	[.0000]	(.2779)	(1.1203)
{162}	[.0000]		[.0927]		[.0000]		[.0000]			[.0000]	
	[.0705]		[.0824]		[.0000]		[.0125]			[.0000]	
$T-1$ to $T-6$	.3502	.1524	.3517	.3615	.0299	−.2473	.4051	.1448	.4031	−.0754	.7449
	(.1084)	(.0802)	(.0927)	(.0311)	(.0940)	(.6665)	(.1044)	(.0799)	(.0346)	(.0910)	(.9966)
	(.1772)	[.3509]	(.1778)	[.0000]	(.2242)	(.9609)	(.1860)	[.1303]	[.0000]	(.2550)	(1.4588)
{182}	[.0000]		[.0000]		[.0000]		[.0000]			[.0000]	
	[.1927]		[.1383]		[.0000]		[.0253]			[.0000]	
$T-1$ to $T-7$	.2763	.1017	.3069	.3444	.0158	−.2577	.3650	.1413	.3082	−.0654	.7629
	(.1028)	(.0741)	(.0893)	(.0247)	(.0849)	(.7653)	(.1005)	(.0743)	(.0287)	(.0850)	(1.1336)
	(.1758)	[.3510]	(.1677)	[.0000]	(.2109)	(1.0709)	(.1799)	[.1849]	[.0000]	(.2331)	(1.6145)
{197}	[.0000]		[.0000]		[.0000]		[.0000]			[.0000]	
	[.2051]		[.2126]		[.0000]		[.0505]			[.0000]	
$T-1$ to $T-8$	.2653	.0659	.2781	.3479	.0017	−.1355	.3408	.1051	.2781	−.0800	.8239
	(.1003)	(.0691)	(.0869)	(.0237)	(.0834)	(.7492)	(.0984)	(.0679)	(.0262)	(.0825)	(.9977)
	(.1686)	[.3562]	(.1592)	[.0000]	(.2035)	(.9927)	(.1738)	[.1709]	[.0000]	(.2270)	(1.3779)
{207}	[.0000]		[.0000]		[.0000]		[.0000]			[.0000]	
	[.1663]		[.1593]		[.0000]		[.0364]			[.0000]	
$T-1$ to $T-9$	.2814	.0931	.2961	.3173	.0046	−.1808	.3512	.1227	.2544	−.0824	.7825
	(.0987)	(.0674)	(.0856)	(.0228)	(.0827)	(.7238)	(.0971)	(.0661)	(.0255)	(.0822)	(.9308)
	(.1623)	[.3787]	(.1540)	[.0000]	(.1949)	(.9267)	(.1694)	[.1926]	[.0000]	(.2207)	(1.2813)
{212}	[.0000]		[.0000]		[.0000]		[.0000]			[.0000]	
	[.1875]		[.1793]		[.0000]		[.0431]			[.0000]	
OLS	.1115						.1755				
	(.0247)						(.0224)				
	(.0791)						(.0743)				

NOTE: Standard errors are in parentheses and p values for the null hypothesis of correct overidentifying restrictions are in square brackets. The second values in parentheses and square brackets for the 2SLS, FF, SSIV, and USSIV estimators refer to heteroscedasticity-robust standard errors and p values. $N = 532$ and $T = 8$. The numbers between { } are the number of instruments used in estimation.

the ratio of instruments to *cross-sectional* sample size rises, Hansen's test seems to overreject. This will become evident in Subsection 2.2.2 when one compares GMM to IW-GMM and 2SLS to SSIV.

2.2.2 Stacked Moment Results. Turning now to the stacked moment-condition results reveals a striking change in the efficiency of the ISE. With 72 moment conditions in the stacked $t-1$ and $t-2$ case, significant efficiency gains are achieved for all the estimators, with the average reduction in standard errors around 50%. The FF estimator outperforms 2SLS in efficiency in all cases, but GMM is more efficient than FF in all of the stacked-moment specifications. No clear efficiency pattern emerges between orthogonal deviations and first-difference GMM, nor with heteroscedasticity-robust 2SLS across the transformations. Quite remarkably, IW-GMM outperforms GMM across the

Table 2. First-Stage F and Wald Tests for Wage Changes

	First-difference		Orthogonal-deviations	
Moments	F test	Wald Test	F test	Wald Test
Base case:				
$T-1$, $T-2$	2.541	20.095	3.979	30.293
{9}	[.007]	[.017]	[.000]	[.000]
Stacked cases:				
$T-1$ to $T-2$	1.182	92.589	1.382	104.331
{72}	[.141]	[.052]	[.019]	[.008]
$T-1$ to $T-3$	1.197	129.862	1.303	150.497
{107}	[.084]	[.066]	[.021]	[.004]
$T-1$ to $T-4$	1.098	160.540	1.154	187.569
{137}	[.208]	[.083]	[.109]	[.003]
$T-1$ to $T-5$	1.074	210.043	1.092	245.218
{162}	[.251]	[.007]	[.205]	[.000]
$T-1$ to $T-6$	1.138	241.216	1.120	286.980
{182}	[.104]	[.002]	[.134]	[.000]
$T-1$ to $T-7$	1.149	271.128	1.097	309.148
{197}	[.080]	[.000]	[.175]	[.000]
$T-1$ to $T-8$	1.146	291.226	1.079	326.667
{207}	[.079]	[.000]	[.214]	[.000]
$T-1$ to $T-9$	1.160	325.604	1.080	362.342
{212}	[.061]	[.000]	[.210]	[.000]

NOTE: The null hypothesis is that the instruments jointly explain none of the variation in the wage changes. P values are given in brackets. For the F test, the numerator degrees of freedom are (# of moments − 1) and the denominator degrees of freedom are $NT - N -$ #moments, where $N = 532$ and $T = 8$. The degrees of freedom for the Wald test are the number of moment conditions.

board on efficiency grounds. This result is surprising given that IW-GMM only uses half of the sample for each estimation and may point to some small-sample advantages of IW-GMM over GMM. Likewise, SSIV standard errors under homoscedasticity dominate 2SLS and FF; however, 2SLS and FF are far superior once one controls for heteroscedasticity. Proceeding down the columns, note that, as additional moments are added to the instrument set, the ISE is estimated more precisely as predicted. For example, exploiting all moments (212) reduces the 2SLS and FF standard errors by at least 45% and the GMM standard errors by at least 55% compared to the suboptimal matrix with 72 moments.

The efficiency gains of extra instruments, nonetheless, come at a cost. A clear pattern of downward bias in the first-difference and orthogonal-deviations GMM and SSIV estimators emerges in the stacked-moment results of Table 1. With 72 moments imposed, the first-difference 2SLS and GMM ISE parameters differ by 27%, but the difference is 67% with 212 moments. The bias is even more severe under orthogonal deviations. Unlike GMM, the finding that the SSIV parameter estimates converge toward 0 as the number of overidentifying restrictions expands is an expected property of the estimator (Angrist and Krueger 1995). What is not expected is the erratic and puzzling behavior of the USSIV estimator. The estimated ISE ranges from 8.239 in the 72-moment orthogonal-deviations case to −3.414 in the 107-moment first-difference case. No obvious explanation emerges for the disappointing USSIV results, but the findings indicate that "forcing" exact identification through the bias-correction factor may be undesirable with many overidentifying restrictions. The FF estimator from Keane and Runkle (1992) does not appear to suffer from the same bias as in GMM and yet is more efficient than 2SLS.

2.2.3 Tests for Weak Instruments. An obvious question arises: Is the bias in GMM due to instruments weakly correlated with the wage or due to a correlation between the sample moments and the estimated weight matrix? Several corroborating facts suggest that a correlation between the sample moments and the estimated weight matrix is the source of the bias. In Table 2 I report first-stage F tests and Wald tests with their associated p values for the endogenous wage regressor. I present the Wald test because the F test is inconsistent when there is heteroscedasticity. As seen in Table 2, the F test frequently leads to the incorrect conclusion that there are weakly correlated instruments when the Wald test does not, suggesting that when heteroscedasticity is present the Wald test is the more appropriate test for the correlation between instruments and regressors. The conclusion that the instruments are not weak is strengthened by noting that the overidentifying-restrictions test does not reject the model under heteroscedasticity-robust GMM, 2SLS, or FF.

Potentially the strongest piece of evidence that the bias comes from the estimated weight matrix is the IW-GMM parameter estimates. Beginning with 162 moments, the IW-GMM estimates are at least twice as large as GMM on average and are comparable to 2SLS and FF estimates, a result that would not arise if there were no correlation between the GMM sample moments and estimated weight matrix or if the instruments are weak. Unfortunately, Hansen's overidentifying-restrictions test rejects all of the IW-GMM specifications (and all SSIV models beginning with 137 moments). This "overrejection" as the ratio of moments to cross-sectional sample size increases, which reaches a maximum of 212/266 under IW-GMM compared to 212/532 under GMM, may indicate a weakness in the test.

2.2.4 Summary. The results of Tables 1 and 2 suggest that GMM is biased downward relative to 2SLS and FF as the number of moment conditions expands because of a correlation between the estimated weight matrix and the sample moments. Moreover, IW-GMM is successful in correcting the bias in GMM; however, the IW-GMM parameter estimates are more variable than 2SLS and FF and the overidentifying restrictions are rejected in each specification. SSIV is biased toward 0 as the number of instruments expands and USSIV performs poorly. Finally, inference, whether it be standard errors, overidentifying restrictions, or first-stage F tests, can be quite misleading when the assumption of conditional homoscedasticity is incorrect.

3. BOOTSTRAPPING OVERIDENTIFIED MODELS

To investigate the sample properties of the estimators such as the potential bias/efficiency trade-off in GMM, more deeply, I now turn to the bootstrap. The bootstrap, recently surveyed by Efron and Tibshirani (1993) and Jeong and Maddala (1993), is a powerful statistical technique for the computation of measures of variability, confidence intervals, and bias of an estimator. In the current application, the bootstrap is a natural alternative to standard Monte Carlo analysis because it is based on a nonparametric estimate of the underlying error distribution. Recall that a strength of each of the IV estimators considered in Section 2 emanates from their lack of dependence on an a priori specified small-sample distribution. Moreover, the parametric Monte Carlo is difficult to implement because the labor-supply literature offers little help in the way of the joint small-sample distribution of wages and hours, making the choice of a distribution a dubious exercise at best. Consequently, the bootstrap maintains the semiparametric, empirical-based spirit of the article.

The typical regression-based bootstrap is a multistep procedure whereby the researcher resamples with replacement the estimated residuals, constructs a new dependent variable as the sum of the fitted value from the regression plus the bootstrapped residual, reestimates the model, and repeats the exercise B times ($b = 1, \ldots, B$). There are then B observations from which to compute measures of bias, variability, or confidence intervals. This approach is consistent only under the assumptions of conditional homoscedasticity, no serial dependence, and nonstochastic regressors.

When the regressors are stochastic or there is conditional heteroscedasticity as is typical in IV estimation, Freedman (1984) and Freedman and Peters (1984) suggested an alternative procedure. Instead of resampling the residuals, one resamples simultaneously the estimated residuals along with the regressors and instruments. More specifically, one resamples with replacement from $(\hat{\varepsilon}, X, W)$, where $\hat{\varepsilon}$ is the vector of estimated first-difference or orthogonal-deviations residuals, X is the matrix of first-difference or orthogonal-deviations regressors, and W is the matrix of instruments. Call the constructed "pseudodata" $(\hat{\varepsilon}^*, X^*, W^*)$. The new dependent variable is $\hat{y}^* = X^*\hat{\beta} + \hat{\varepsilon}^*$, which is regressed on X^* with W^* as instruments to generate a new $\hat{\beta}^*$, and the procedure is repeated B times. This approach, in which each observation has equal probability weight $1/N$ of being drawn from the discrete empirical distribution function, is an asymptotically valid method of bootstrapping an IV estimator, even when the model is overidentified, and provides asymptotic coverage rates equal to their nominal rates (Hahn in press).

Brown and Newey (1995) and Hall and Horowitz (in press), however, showed that the Freedman method does not yield an improvement in terms of coverage rates over first-order asymptotic theory, and more importantly, it gives the wrong size, even asymptotically, for the overidentifying-restrictions test. The problem lies in the fact that when the model is overidentified the sample moments are typically not 0, and thus one bootstraps from an empirical distribution that is a poor approximation to the true underlying distribution. Hall and Horowitz attempted to mitigate this problem by recentering the moments at their sample values, and Brown and Newey suggested recentering the empirical distribution by imposing the moment conditions on the data.

Because Brown and Newey showed that their approach is efficient in the class of bootstrap methods, I use their resampling algorithm as an alternative to Freedman's method. The idea is to resample from the multinomial distribution in which each observation receives a different probability weight, $\hat{p}_i$, such that the sample moments, $\hat{g}_i = g(\hat{\varepsilon}, X, W)$, are satisfied by construction, $\sum_{i=1}^{N} \hat{p}_i \hat{g}_i = 0$. In general, $\hat{p}_i$ is solved numerically from a linear programming problem; however, Brown and Newey suggested a closed-form solution for the probability given as $\hat{p}_i = (1 - \bar{g}'\hat{V}^{-1}\hat{g}_i)/[N * (1 - \bar{g}'\hat{V}^{-1}\bar{g})]$, where $\bar{g} = 1/N \sum_{i=1}^{N} \hat{g}_i$ and $\hat{V} = 1/N \sum_{i=1}^{N} \hat{g}_i\hat{g}_i'$. The estimated probabilities are expected to satisfy the usual regularity conditions such as $\hat{p}_i > 0$ and $\sum_{i=1}^{N} \hat{p}_i = 1$.

For each experiment, I compute the bias as $(1/B)\sum_{b=1}^{B} \hat{\delta}_b - \hat{\delta}$, where $\hat{\delta}_b$ is the bth bootstrap estimate of the ISE and $\hat{\delta}$ is the pseudotrue value of the ISE defined later, the average heteroscedasticity-consistent standard error constructed from the asymptotic formulas in Section 1 (SE-A), the bootstrap standard error as $[(\sum_{b=1}^{B}(\hat{\delta}_b - \sum_{b=1}^{B}\hat{\delta}_b/B)^2)/(B-1)]^{1/2}/\sqrt{B}$ (SE-B), the RMSE defined as the square root of the sum of bias squared and the average variance constructed from asymptotic formulas (RMSE-A), the RMSE defined as the square root of the sum of bias squared and boot SE squared (RMSE-B), and the MAE. In addition, I report a measure of reliability of asymptotic inference by constructing the coverage rates on a 95% two-tailed confidence interval as the fraction of times the pseudotrue value of the ISE falls within the interval based on asymptotic standard-error formulas and asymptotic critical values.

Because the t distribution does not adjust confidence intervals for the presence of skewness in the underlying population, I also construct bootstrap-t critical values that have the ability to account for skewness (Efron and Tibshirani 1993, pp. 159–162). The idea is to approximate the asymptotic pivotal t statistic from the bootstrap as $\hat{t}(b) = (\hat{\delta}_b - \hat{\delta})/\widehat{\mathrm{se}}(b)$, where $\widehat{\mathrm{se}}(b)$ is the bth estimated asymptotic standard error. The α and $1 - \alpha$ critical values are found by ordering the $\hat{t}(b)$'s for the entire bootstrap sample and are then used in constructing bootstrap-t confidence intervals. The bootstrap-t may offer asymptotic refinements over critical values from first-order asymptotic theory (Brown and Newey 1995; Hall and Horowitz in press). Finally, to gauge the potential gains in the Brown and Newey algorithm over Freedman's for correcting the size of the overidentifying-restrictions test, I report the fraction of times the bootstrap value of the overidentifying-restrictions test exceeds the asymptotic critical χ^2 value at the .05 level (Boot-J).

3.1 Results

To make relative comparisons of bias, RMSE, and MAE across the estimators I need to "prime" the bootstrap (i.e., construct $\hat{\varepsilon}$) with a common vector of parameters. This approach of choosing a common set of starting values mimics the parametric Monte Carlo and has been applied to the bootstrap by Freedman and Peters (1984). Because the focus is on the ISE, I use the estimated ISE most often encountered in male labor-supply studies as the pseudotrue value of δ. Pencavel (1986) noted that the ISE has a cen-

Table 3. Bootstrap Comparisons of First-Difference Moment-Condition Estimators Where Observations Are Drawn With Probability 1/N

Estimators	Bias	SE-A	SE-B	RMSE-A	RMSE-B	MAE	95% asym-t coverage	95% boot-t critical values	Boot-J rejection rate: 5%
					9 moments				
2SLS	−.108	.485	.474	.497	.486	.263	.99	−1.59, 1.71	.29
GMM	.174	.360	.330	.400	.373	.280	.95	−1.35, 2.15	.13
FF	−.097	.351	.309	.365	.324	.183	.98	−1.93, 1.29	.44
IW-GMM	.189	.459	.565	.497	.596	.336	.88	−1.48, 3.99	.66
SSIV	.208	.854	.909	.879	.932	.451	1.00	−1.63, 1.67	.29
USSIV	.828	15.621	6.082	15.643	6.138	1.683	.99	−1.04, 1.53	—
					72 moments				
2SLS	.128	.169	.199	.212	.236	.145	.85	−1.53, 3.27	.99
GMM	−.093	.097	.108	.134	.142	.108	.81	−3.34, .98	.99
FF	.055	.142	.156	.152	.168	.136	.87	−1.79, 2.63	.99
IW-GMM	.013	.070	.337	.071	.337	.149	.47	−7.28, 8.06	1.00
SSIV	.137	.310	.273	.339	.306	.160	.95	−2.13, 1.95	.63
USSIV	−1.107	3.4E2	42.527	3.4E2	42.541	4.965	1.00	−1.19, 1.69	—
					162 moments				
2SLS	.022	.131	.219	.132	.220	.147	.74	−3.73, 4.05	.99
GMM	−.169	.055	.077	.178	.186	.179	.21	−5.62, −.2	1.00
FF	−.046	.119	.128	.128	.136	.090	.87	−3.74, 1.49	1.00
IW-GMM	−.020	.029	.229	.035	.230	.156	.19	−15.1, 19.1	1.00
SSIV	−.216	.188	.124	.286	.249	.211	.84	−3.07, .31	.90
USSIV	−.932	11.598	5.707	11.636	5.782	.927	.96	−2.26, 1.79	—
					212 moments				
2SLS	−.014	.124	.166	.125	.166	.084	.79	−4.50, 3.12	1.00
GMM	−.175	.043	.047	.180	.181	.172	.04	−6.0, −1.74	1.00
FF	−.037	.100	.106	.107	.112	.069	.85	−3.69, 1.84	1.00
IW-GMM	.044	.018	.219	.047	.224	.144	.14	−27.4, 32.5	1.00
SSIV	−.209	.175	.096	.273	.231	.192	.83	−2.86, −.31	.95
USSIV	−.372	13.467	5.783	13.472	5.794	1.099	.99	−1.58, 1.41	—

NOTE: All calculations are based on 100 bootstrap replications. The pseudotrue value of the ISE is fixed at .21 for all calculations. Bias is computed as the difference between the average bootstrap estimate and the pseudotrue ISE, SE-A is the average standard error computed from asymptotic theory, SE-B is the average standard error constructed from the bootstrap, RMSE-A is the root mean squared error computed with the variance from asymptotic theory, RMSE-B is the root mean squared error computed with the variance from the bootstrap, MAE is the median absolute error, 95% asym-t coverage is the fraction of times the pseudotrue value of the ISE falls within the 95% confidence interval based on standard errors and critical values from asymptotic theory, 95% boot-t critical values are the .025 and .975 values from the bootstrap t statistic, and boot-J rejection rate is the fraction of times the bootstrap value of the overidentifying restrictions test exceeds the critical χ^2 value at the .05 level.

tral tendency of .20, which, coincidentally, is consistent with the 2SLS 9-moment model of Table 1; thus, the 2SLS 9-moment parameters form the starting values. I present 100 bootstrap replications for the 9-, 72-, 162-, and 212-moment cases for both the first-difference and orthogonal-deviations transformations. The small number of bootstrap Monte Carlo draws is sufficient for measures of bias and variability for most sample sizes but is sufficient for confidence intervals only in larger samples (Hall 1986; Efron and Tibshirani 1993, p. 161). All experiments were conducted in Gauss with a Pentium-90, where the models with 212 moments took at least 48 hours to compute. I present the results from the Freedman algorithm in Tables 3 and 4 and from Brown and Newey's algorithm in Tables 5 and 6. I focus on the first-difference results because the implications from orthogonal deviations are largely the same.

3.1.1 Results From Freedman Algorithm. Beginning with the base case of 9 moments in Table 3, 2SLS and FF tend to be biased toward OLS ($\hat{\delta}_{\text{OLS}} = .11$), which is consistent with the presence of a finite-sample bias in IV estimators. The absolute value of bias in GMM exceeds both 2SLS and FF, whereas GMM dominates all three of the split-sample estimators in terms of lower bias. Comparing the average asymptotic standard error (SE-A) to the bootstrap standard error (SE-B) suggests that there is little difference for 2SLS, FF, and GMM standard errors, but the asymptotic and bootstrap standard errors diverge for IW-GMM, SSIV, and USSIV. FF dominates all of the estimators in terms of lower RMSE and MAE, but GMM does better than 2SLS in terms of RMSE but not for MAE. Moreover, GMM dominates all of the split-sample estimators for lower RMSE and MAE. In terms of asym-t coverage rates, GMM performs best, 2SLS, FF, SSIV, and USSIV tend to under-reject, and IW-GMM overrejects. The bootstrap-t critical values suggest that there is a slight departure from symmetry for each of the estimators and that asymptotic critical values are too large. The boot-J rejection rates, however, reveal that there is a serious level distortion in the overidentifying-restrictions test computed with Freedman's algorithm. Overall, Keane and Runkle's FF estimator performs best in the base case.

The most obvious trend arising in the stacked-moment cases is the increasing bias in GMM as the number of moment conditions expands. The bias in the GMM ISE rises (i.e., becomes more negative) from −.093 under 72 moments to −.175 with 212 moments. This finding, reenforc-

Table 4. Bootstrap Comparisons of Orthogonal-Deviations Moment-Condition Estimators Where Observations Are Drawn With Probability 1/N

Estimators	Bias	SE-A	SE-B	RMSE-A	RMSE-B	MAE	95% asym-t coverage	95% boot-t critical values	Boot-J rejection rate: 5%
				9 moments					
2SLS	.099	.432	.440	.443	.452	.259	.96	−1.85, 2.16	.30
GMM	.303	.370	.366	.478	.475	.317	.87	−1.68, 2.67	.16
IW-GMM	.068	.444	.451	.450	.457	.285	.90	−1.72, 2.63	.73
SSIV	.419	.753	.647	.862	.771	.494	.99	−1.61, 1.69	.12
USSIV	.312	5.992	3.793	5.999	3.806	1.079	.97	−1.05, 2.11	—
				72 moments					
2SLS	.280	.165	.219	.325	.356	.270	.62	−.80, 4.40	1.00
GMM	.011	.096	.113	.097	.114	.080	.88	−2.36, 2.62	1.00
IW-GMM	.124	.069	.264	.142	.292	.160	.45	−8.95, 11.5	1.00
SSIV	.058	.292	.168	.298	.178	.130	.99	−1.78, 1.39	.42
USSIV	.355	31.512	12.082	31.512	12.087	3.825	1.00	−1.17, 1.28	—
				162 moments					
2SLS	.112	.141	.209	.180	.238	.160	.77	−2.24, 7.36	1.00
GMM	−.141	.053	.074	.150	.159	.140	.27	−5.21, .12	1.00
IW-GMM	.042	.031	.248	.052	.252	.184	.20	−17.3, 17.3	1.00
SSIV	−.252	.222	.134	.336	.286	.255	.83	−3.11, .07	.85
USSIV	−1.458	83.064	10.147	83.077	10.252	1.220	.91	−2.23, 2.32	—
				212 moments					
2SLS	.087	.123	.175	.151	.196	.121	.84	−2.49, 7.62	1.00
GMM	−.154	.041	.056	.159	.164	.150	.07	−6.62, −1.21	1.00
IW-GMM	.009	.019	.251	.021	.251	.155	.13	−23.7, 40.5	1.00
SSIV	−.239	.186	.101	.302	.259	.243	.71	−2.81, −.01	.90
USSIV	−.039	17.758	6.229	17.759	6.230	1.274	.98	−1.49, 2.32	—

NOTE: See note to Table 3.

ing the results of Table 1, extends Tauchen's (1986) and Altonji and Segal's (1994) finding of bias in GMM in small samples to the large-sample case of panel data. The 2SLS and FF estimators are not biased toward OLS as in the base case, while IW-GMM has negligible bias in each of the stacked-moment simulations. Consistent with the results of Table 1 is the bias toward 0 in SSIV and USSIV as the number of moment conditions increases. The USSIV estimator does not succeed in its primary function to "inflate" the SSIV estimates.

In terms of statistical inference, there appears to be little difference between the bootstrap standard error and the average asymptotic standard error for the FF and GMM estimators. On the contrary, the standard errors do diverge for the other estimators, especially IW-GMM and USSIV. This suggests that the small IW-GMM standard errors in Table 1 may understate the true sampling variability in the estimator, making the bootstrap useful for correctly estimating IW-GMM standard errors. Examining the 95% asym-t coverage rates indicates that all of the estimators except USSIV tend to overreject the event that the pseudotrue ISE lies in the confidence region, with the overrejection being particularly acute for both GMM and IW-GMM. In GMM, the asymptotic confidence region is distorted because of the bias in the parameter estimates toward 0, coupled with a tight standard error. IW-GMM, on the other hand, has little bias in the estimated ISE's relative to GMM, but because the asymptotic standard errors are estimated tightly, the confidence interval is short, excluding the pseudotrue ISE too frequently. Moreover, inspecting the boot-t critical values reveals that GMM does not even include 0 in the critical region in the 162- and 212-moment cases. This suggests that the bootstrap-t confidence interval is useful for GMM and IW-GMM to capture the skewness in the underlying distribution. The FF estimator continues to outperform 2SLS in coverage rates; meanwhile, given the bias in the estimated ISE, the SSIV and USSIV estimators give relatively good coverage.

Comparing the omnibus measures of estimator performance (RMSE and MAE) for the stacked-moment bootstrap simulations reveals that FF dominates 2SLS in all cases and both FF and 2SLS always dominate GMM except with 72 moments. Highlighting the problems noted by Brown and Newey (1995) and Hall and Horowitz (in press), however, the overidentifying-restrictions test is distorted severely, resulting in 100% rejection rates when the nominal rejection rate is 5%. It must be stressed that, as established by Hahn (in press), overrejection of the overidentifying restrictions does not invalidate the other bias/efficiency results in Tables 3 and 4—the estimators and overidentifying restrictions are separate.

3.1.2 Results From Brown and Newey Algorithm. Tables 5 and 6 contain the bootstrap results for the first-difference and orthogonal-deviations transformations using Brown and Newey's (1995) algorithm. The USSIV estimator is not presented because the motivation for Brown and Newey's method is to get a correct overidentifying-restrictions test, which by definition does not exist for the USSIV. Compared to Table 3, the most striking result in

Table 5. Bootstrap Comparisons of First-Difference Moment-Condition Estimators Where Observations Are Drawn With Probability $\hat{p}_i$

Estimators	Bias	SE-A	SE-B	RMSE-A	RMSE-B	MAE	95% asym-t coverage	95% boot-t critical values	Boot-J rejection rate: 5%
					9 moments				
2SLS	−.045	.383	.306	.385	.309	.197	.99	−1.86, 1.09	.03
GMM	−.034	.316	.268	.317	.270	.161	.96	−2.07, 1.42	.00
FF	−.034	.316	.323	.316	.325	.210	.95	−2.06, 2.03	.12
IW-GMM	−.041	.251	.264	.254	.267	.143	.96	−1.99, 2.17	.03
SSIV	−.043	.460	.351	.462	.353	.220	.97	−2.35, .99	.00
					72 moments				
2SLS	−.086	.117	.123	.145	.150	.121	.84	−2.96, 1.61	.45
GMM	−.120	.074	.093	.141	.151	.130	.57	−3.65, 1.08	.31
FF	−.115	.101	.105	.153	.156	.118	.73	−3.83, .84	.58
IW-GMM	−.141	.034	.188	.145	.235	.116	.24	−22.4, 2.95	.96
SSIV	−.081	.078	.082	.112	.115	.088	.81	−3.37, 1.12	.00
					162 moments				
2SLS	−.099	.061	.082	.116	.128	.099	.58	−4.99, .84	.45
GMM	−.108	.039	.049	.115	.119	.111	.25	−5.55, −.4	.71
FF	−.103	.051	.061	.115	.120	.111	.43	−4.91, .50	.84
IW-GMM	—	—	—	—	—	—	—	—	—
SSIV	−.127	.056	.054	.138	.138	.120	.33	−4.79, .29	.00
					212 moments				
2SLS	−.134	.051	.068	.143	.150	.142	.30	−5.39, .36	.93
GMM	−.131	.031	.040	.135	.137	.136	.03	−6.6, −1.83	.74
FF	−.125	.045	.049	.133	.134	.129	.23	−5.45, −.16	.96
IW-GMM	—	—	—	—	—	—	—	—	—
SSIV	−.144	.046	.035	.152	.149	.144	.08	−5.35, −1.33	.03

NOTE: See note to Table 3.

the base case of 9 moments is the dramatic improvement in boot-J rejection rates. This is exemplified by 2SLS in which the boot-J rejection rate falls from .29 to .03. The test still overrejects with the FF estimator but 70% fewer times than under the Freedman method. In addition, the Brown and Newey algorithm improves on the efficiency of the bootstrap estimates, especially for 2SLS, IW-GMM, and SSIV. In terms of bias, FF and GMM are identical, but GMM dominates 2SLS and FF in terms of lower MAE and RMSE-B. IW-GMM, however, performs best overall in the

Table 6. Bootstrap Comparisons of Orthogonal-Deviations Moment-Condition Estimators Where Observations Are Drawn With Probability $\hat{p}_i$

Estimators	Bias	SE-A	SE-B	RMSE-A	RMSE-B	MAE	95% asym-t coverage	95% boot-t critical values	Boot-J rejection rate: 5%
					9 moments				
2SLS	−.009	.396	.454	.396	.454	.319	.92	−2.11, 2.21	.10
GMM	−.044	.341	.338	.344	.341	.240	.96	−2.27, 1.68	.02
IW-GMM	−.081	.291	.390	.302	.399	.230	.86	−2.74, 2.45	.03
SSIV	.011	.438	.424	.438	.425	.254	.98	−1.94, 1.25	.00
					72 moments				
2SLS	−.101	.096	.112	.140	.151	.109	.76	−3.74, 1.46	.32
GMM	−.108	.071	.072	.129	.130	.109	.62	−3.72, .38	.20
IW-GMM	−.090	.032	.138	.095	.138	.089	.33	−9.86, 3.71	.77
SSIV	−.119	.098	.101	.155	.156	.142	.56	−4.16, 1.33	.03
					162 moments				
2SLS	−.130	.053	.068	.140	.147	.126	.41	−5.82, −.11	.32
GMM	−.122	.037	.042	.127	.129	.119	.13	−6.01, −.99	.75
IW-GMM	—	—	—	—	—	—	—	—	—
SSIV	−.142	.046	.041	.149	.147	.140	.12	−6.75, −.67	.00
					212 moments				
2SLS	−.138	.049	.058	.147	.149	.136	.30	−5.52, −.37	.94
GMM	−.136	.030	.037	.140	.141	.139	.05	−7.15, −1.56	.85
IW-GMM	—	—	—	—	—	—	—	—	—
SSIV	−.167	.048	.029	.174	.170	.171	.04	−6.38, −1.86	.03

NOTE: See note to Table 3.

base case in terms of lower RMSE and MAE. Because of the gains in efficiency and in levels of the boot-J test, the Brown and Newey method seems to be preferable to the Freedman method with few moments.

On the other hand, problems arise with the Brown and Newey algorithm, especially in the many-moment models. First, some of the bootstrap probabilities ($\hat{p}_i$), although summing to one overall, are negative. This problem increased with the number of moment conditions, ranging from a low of .5% of the observations with 9 moments in 2SLS and GMM to over 40% with IW-GMM with 212 moments. I considered several methods to deal with this, and the method I chose was to redistribute a fraction of the negative probabilities to each observation and then assign a small positive probability (1/100th of the smallest positive probability) to those observations with negative probabilities, making sure the probabilities sum to 1. The results are not sensitive to the method used. Second, because so much weight (i.e., a "high" probability) is given to a subset of the observations for the IW-GMM estimators with 162 and 212 moments, the estimators are singular. Recall that in sampling with replacement an observation can be drawn more than once, which can occur with greater probability under the Brown and Newey algorithm. This singularity problem persisted with every bootstrap sample over a four-day period and with different random sample splits of the original data as well. Consequently, I am unable to report on the IW-GMM estimator for these cases. Brown and Newey also reported difficulties due to a singularity problem in their empirical application.

With the exception of SSIV, the boot-J test overrejects in the stacked-moment cases just as in Tables 3 and 4. The distortion in levels, although less than the Freedman method, becomes more severe as additional moments are appended to the instrument set. It is important to note that both Hall and Horowitz, with their Monte Carlo, and Brown and Newey, with their empirical application, found that level distortions persist in the test even after recentering the distribution. In terms of bias, the estimators do not distinguish themselves as in the Freedman method, and each tends to center on the OLS estimate. However, 2SLS and FF continue to dominate GMM in terms of lower bias. Moreover, asym-t coverage is worse for each estimator, suggesting that boot-t critical values may improve on coverage over first-order asymptotics. The central conclusion from the previous bootstrap results remains the same though; namely, in comparing the estimators across all criteria, the FF estimator continues to be preferred in this application.

4. SUMMARY AND CONCLUSIONS

In this article I compared several IV estimators for panel-data models with predetermined instruments. The empirical results from the life-cycle labor-supply model, in conjunction with results from two separate bootstrap Monte Carlo experiments, are summarized as follows. First, and most important, the GMM estimator is biased downward relative to the 2SLS and FF estimators as the number of moment conditions approaches the optimal number of moments because of a correlation between the estimated weight matrix and the sample moments. This leads to poor coverage rates for confidence intervals based on asymptotic critical values but provides a clear role for the bootstrap-t confidence interval as a basis of inference under GMM. GMM performs reasonably well with suboptimal instruments but is not recommended for panel-data applications when all of the moments are exploited for estimation.

Second, the IW-GMM estimator is generally successful at eliminating the bias in GMM parameter estimates using Freedman's bootstrap algorithm; however, the standard errors from asymptotic theory seem to understate the true sampling variation and the overidentifying-restrictions test is biased toward rejection, possibly due to the high ratio of moments to cross-section sample size. Further research on both the asymptotic and small-sample properties of this estimator is needed. In the meantime, the bootstrap is likely to be useful in IW-GMM for the construction of both standard errors and confidence intervals. Third, the USSIV estimator tends to be highly unstable and inefficient, in contrast to the SSIV estimator which works fairly well in models with few overidentifying restrictions.

Fourth, caution is warranted when conducting inference under the assumption of conditional homoscedasticity. On several occasions, incorrectly assuming homoscedasticity led to the rejection of the overidentifying restrictions and to the conclusion of weakly correlated instruments. When heteroscedasticity is present, one should use robust variants of the test of overidentifying restrictions and the Wald test for first-stage tests of correlation between instruments and regressors. In addition, after adjusting for conditional heteroscedasticity, there is no obvious efficiency gain of using orthogonal deviations rather than first-differences.

Fifth, further research is needed on improving methods of bootstrapping estimators from overidentified models. The levels distortion in the overidentifying-restrictions test persisted in models with many moments, even after recentering the distribution using Brown and Newey's (1995) algorithm. Moreover, complications arose in the implementation of their algorithm, notably negative probabilities and singularity problems. Hence, if the model has numerous moment restrictions and the focus is solely on the estimator and consistent confidence intervals and not on the overidentifying-restrictions test, then Freedman's (1984) method of giving equal weight to each observation is recommended.

Finally, although 2SLS is still a reliable method of estimation, the FF estimator performs best in terms of the bias/efficiency trade-off. It offers more efficient estimation than 2SLS but does not have the problem of excess bias found in GMM parameter estimates. Consequently, the FF estimator is attractive relative to the other estimators considered here.

ACKNOWLEDGMENTS

I thank the editor, an associate editor, and an anonymous referee for many helpful comments that have improved this article substantially. I also have benefited from discussions with George Evans, Tom Kniesner, Larry Singell, and Wes

Wilson. Most especially, I am indebted to Carlos Martins-Filho and Mark Thoma for their insightful advice. I alone am responsible for any remaining errors.

[Received December 1994. Revised July 1996.]

REFERENCES

Ahn, S. C., and Schmidt, P. (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

Altonji, J. G. (1986), "Intertemporal Substitution in Labor Supply: Evidence From Micro Data," *Journal of Political Economy*, 94, S176–S215.

Altonji, J. G., and Segal, L. M. (1994), "Small Sample Bias in GMM Estimation of Covariance Structures," Technical Working Paper 156, National Bureau of Economic Research, Cambridge, MA.

Anderson, T. W., and Hsiao, C. (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Angrist, J. D., and Krueger, A. B. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

——— (1995), "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business & Economic Statistics*, 13, 225–235.

Arellano, M., and Bond, S. R. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Arellano, M., and Bover, O. (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

Bekker, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica*, 62, 657–682.

Borjas, G. J. (1981), "The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias," *Journal of Human Resources*, 15, 409–423.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995), "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

Brown, B. W., and Newey, W. K. (1995), "Bootstrapping for GMM," unpublished manuscript, Rice University, Dept. of Economics.

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Freedman, D. A. (1984), "On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models," *The Annals of Statistics*, 12, 827–842.

Freedman, D. A., and Peters, S. C. (1984), "Bootstrapping an Econometric Model: Some Empirical Results," *Journal of Business & Economic Statistics*, 2, 150–158.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics*, Cambridge, U.K.: Cambridge University Press.

Hahn, J. (in press), "A Note on Bootstrapping Generalized Method-of-Moments Estimators," *Econometric Theory*, 12.

Hall, P. (1986), "On the Bootstrap and Confidence Intervals," *The Annals of Statistics*, 14, 1431–1452.

Hall, P., and Horowitz, J. L. (in press), "Bootstrap Critical Values for Tests Based on Generalized Method-of-Moments Estimators," *Econometrica*, 65.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method-of-Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., and Singleton, K. J. (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Hausman, J. A., and Taylor, W. E. (1981), "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398.

Hayashi, F. (1992), Comment on "On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," by M. Keane and D. Runkle, *Journal of Business & Economic Statistics*, 10, 15–17.

Jeong, J., and Maddala, G. S. (1993), "A Perspective on Application of Bootstrap Methods in Econometrics," in *Handbook of Statistics: Econometrics* (Vol. 11), eds. G. S. Maddala, C. R. Rao, and H. D. Vinod, Amsterdam: North-Holland, pp. 573–610.

Keane, M., and Runkle, D. (1992), "On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business & Economic Statistics*, 10, 1–9.

MaCurdy, T. E. (1985), "Interpreting Empirical Models of Labor Supply in an Intertemporal Framework With Uncertainty," in *Longitudinal Analysis of Economic Data*, eds. J. Heckman and B. Singer, Cambridge, U.K.: Cambridge University Press, pp. 111–155.

Maeshiro, A., and Vali, S. (1988), "Pitfalls in the Estimation of a Differenced Model," *Journal of Business & Economic Statistics*, 6, 511–515.

Nelson, C., and Startz, R. (1990), "The Distribution of the Instrumental Variables Estimator and its t-ratio When the Instrument Is a Poor One," *Journal of Business*, 63, S125–S140.

Pencavel, J. (1986), "Labor Supply of Men," in *Handbook of Labor Economics* (Vol. 1), eds. O. Ashenfelter and R. Layard, Amsterdam: North-Holland, pp. 3–102.

Schmidt, P., Ahn, S., and Wyhowski, D. (1992), Comment on "On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," by M. Keane and D. Runkle, *Journal of Business & Economic Statistics*, 10, 10–14.

Staiger, D., and Stock, J. H. (1994), "Instrumental Variables Regression With Weak Instruments," Technical Working Paper 151, National Bureau of Economic Research, Cambridge, MA.

Tauchen, G. (1986), "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data," *Journal of Business & Economic Statistics*, 4, 397–416.

White, H. (1982), "Instrumental Variables Regression With Independent Observations," *Econometrica*, 50, 483–500.

LINKED CITATIONS

You have printed the following article:

Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators
James P. Ziliak
Journal of Business & Economic Statistics, Vol. 15, No. 4. (Oct., 1997), pp. 419-431.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199710%2915%3A4%3C419%3AEEWPDW%3E2.0.CO%3B2-U

References

Intertemporal Substitution in Labor Supply: Evidence from Micro Data
Joseph G. Altonji
The Journal of Political Economy, Vol. 94, No. 3, Part 2: Hoover Institution Labor Conference. (Jun., 1986), pp. S176-S215.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198606%2994%3A3%3CS176%3AISILSE%3E2.0.CO%3B2-6

Does Compulsory School Attendance Affect Schooling and Earnings?
Joshua D. Angrist; Alan B. Krueger
The Quarterly Journal of Economics, Vol. 106, No. 4. (Nov., 1991), pp. 979-1014.
Stable URL:
http://links.jstor.org/sici?sici=0033-5533%28199111%29106%3A4%3C979%3ADCSAAS%3E2.0.CO%3B2-P

Split-Sample Instrumental Variables Estimates of the Return to Schooling
Joshua D. Angrist; Alan B. Krueger
Journal of Business & Economic Statistics, Vol. 13, No. 2, JBES Symposium on Program and Policy Evaluation. (Apr., 1995), pp. 225-235.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199504%2913%3A2%3C225%3ASIVEOT%3E2.0.CO%3B2-1

LINKED CITATIONS

Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations
Manuel Arellano; Stephen Bond
The Review of Economic Studies, Vol. 58, No. 2. (Apr., 1991), pp. 277-297.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199104%2958%3A2%3C277%3ASTOSFP%3E2.0.CO%3B2-2

Alternative Approximations to the Distributions of Instrumental Variable Estimators
Paul A. Bekker
Econometrica, Vol. 62, No. 3. (May, 1994), pp. 657-681.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199405%2962%3A3%3C657%3AAATTDO%3E2.0.CO%3B2-E

The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias
George J. Borjas
The Journal of Human Resources, Vol. 15, No. 3. (Summer, 1980), pp. 409-423.
Stable URL:
http://links.jstor.org/sici?sici=0022-166X%28198022%2915%3A3%3C409%3ATRBWAW%3E2.0.CO%3B2-E

Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak
John Bound; David A. Jaeger; Regina M. Baker
Journal of the American Statistical Association, Vol. 90, No. 430. (Jun., 1995), pp. 443-450.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199506%2990%3A430%3C443%3APWIVEW%3E2.0.CO%3B2-U

Bootstrap Methods: Another Look at the Jackknife
B. Efron
The Annals of Statistics, Vol. 7, No. 1. (Jan., 1979), pp. 1-26.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28197901%297%3A1%3C1%3ABMALAT%3E2.0.CO%3B2-6

On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models
D. Freedman
The Annals of Statistics, Vol. 12, No. 3. (Sep., 1984), pp. 827-842.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198409%2912%3A3%3C827%3AOBTLEI%3E2.0.CO%3B2-T

LINKED CITATIONS

Bootstrapping an Econometric Model: Some Empirical Results
David A. Freedman; Stephen C. Peters
Journal of Business & Economic Statistics, Vol. 2, No. 2. (Apr., 1984), pp. 150-158.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28198404%292%3A2%3C150%3ABAEMSE%3E2.0.CO%3B2-Z

On the Bootstrap and Confidence Intervals
Peter Hall
The Annals of Statistics, Vol. 14, No. 4. (Dec., 1986), pp. 1431-1452.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198612%2914%3A4%3C1431%3AOTBACI%3E2.0.CO%3B2-V

Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators
Peter Hall; Joel L. Horowitz
Econometrica, Vol. 64, No. 4. (Jul., 1996), pp. 891-916.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199607%2964%3A4%3C891%3ABCVFTB%3E2.0.CO%3B2-A

Large Sample Properties of Generalized Method of Moments Estimators
Lars Peter Hansen
Econometrica, Vol. 50, No. 4. (Jul., 1982), pp. 1029-1054.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198207%2950%3A4%3C1029%3ALSPOGM%3E2.0.CO%3B2-O

Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models
Lars Peter Hansen; Kenneth J. Singleton
Econometrica, Vol. 50, No. 5. (Sep., 1982), pp. 1269-1286.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198209%2950%3A5%3C1269%3AGIVEON%3E2.0.CO%3B2-G

Panel Data and Unobservable Individual Effects
Jerry A. Hausman; William E. Taylor
Econometrica, Vol. 49, No. 6. (Nov., 1981), pp. 1377-1398.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198111%2949%3A6%3C1377%3APDAUIE%3E2.0.CO%3B2-3

[On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous]: Comment
Fumio Hayashi
Journal of Business & Economic Statistics, Vol. 10, No. 1. (Jan., 1992), pp. 15-17.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199201%2910%3A1%3C15%3A%5BTEOPM%3E2.0.CO%3B2-6

On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous
Michael P. Keane; David E. Runkle
Journal of Business & Economic Statistics, Vol. 10, No. 1. (Jan., 1992), pp. 1-9.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199201%2910%3A1%3C1%3AOTEOPM%3E2.0.CO%3B2-D

Pitfalls in the Estimation of a Differenced Model
Asatoshi Maeshiro; Shapoor Vali
Journal of Business & Economic Statistics, Vol. 6, No. 4. (Oct., 1988), pp. 511-515.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28198810%296%3A4%3C511%3APITEOA%3E2.0.CO%3B2-K

The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One
Charles R. Nelson; Richard Startz
The Journal of Business, Vol. 63, No. 1, Part 2: A Conference in Honor of Merton H. Miller's Contributions to Finance and Economics. (Jan., 1990), pp. S125-S140.
Stable URL:
http://links.jstor.org/sici?sici=0021-9398%28199001%2963%3A1%3CS125%3ATDOTIV%3E2.0.CO%3B2-6

[On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous]: Comment
Peter Schmidt; Seung C. Ahn; Donald Wyhowski
Journal of Business & Economic Statistics, Vol. 10, No. 1. (Jan., 1992), pp. 10-14.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199201%2910%3A1%3C10%3A%5BTEOPM%3E2.0.CO%3B2-B

LINKED CITATIONS

Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data

George Tauchen

Journal of Business & Economic Statistics, Vol. 4, No. 4. (Oct., 1986), pp. 397-416.

Stable URL:

http://links.jstor.org/sici?sici=0735-0015%28198610%294%3A4%3C397%3ASPOGME%3E2.0.CO%3B2-Z

Instrumental Variables Regression with Independent Observations

Halbert White

Econometrica, Vol. 50, No. 2. (Mar., 1982), pp. 483-500.

Stable URL:

http://links.jstor.org/sici?sici=0012-9682%28198203%2950%3A2%3C483%3AIVRWIO%3E2.0.CO%3B2-L