TABLE 2—CORRELATION MATRICES

A. *Identical Twins*

Variable	Y_1	Y_2	S_1^1	S_1^2	S_2^2	S_2^1	E_F^1	E_F^2	E_M^1	E_M^2
Y_1	1.000									
Y_2	0.563	1.000								
S_1^1	0.382	0.168	1.000							
S_1^2	0.375	0.140	0.920	1.000						
S_2^2	0.267	0.272	0.658	0.697	1.000					
S_2^1	0.248	0.247	0.700	0.643	0.877	1.000				
Father's education (E_F^1)	0.155	0.088	0.345	0.266	0.361	0.416	1.000			
Father's education (E_F^2)	0.159	0.091	0.357	0.278	0.320	0.389	0.857	1.000		
Mother's education (E_M^1)	0.102	0.088	0.348	0.343	0.392	0.410	0.614	0.644	1.000	
Mother's education (E_M^2)	0.126	0.087	0.316	0.321	0.322	0.337	0.503	0.579	0.837	1.000

B. *Fraternal Twins*

Variable	Y_1	Y_2	S_1^1	S_1^2	S_2^2	S_2^1	E_F^1	E_F^2	E_M^1	E_M^2
Y_1	1.000									
Y_2	0.364	1.000								
S_1^1	0.142	0.233	1.000							
S_1^2	0.128	0.256	0.869	1.000						
S_2^2	0.140	0.367	0.543	0.535	1.000					
S_2^1	0.136	0.387	0.621	0.565	0.951	1.000				
Father's education (E_F^1)	0.109	0.028	0.332	0.408	0.353	0.407	1.000			
Father's education (E_F^2)	0.025	−0.107	0.259	0.392	0.230	0.253	0.803	1.000		
Mother's education (E_M^1)	0.147	−0.117	0.025	0.127	0.244	0.244	0.547	0.458	1.000	
Mother's education (E_M^2)	−0.065	−0.178	0.180	0.216	0.109	0.180	0.587	0.600	0.742	1.000

Note: Y_1 and Y_2 represent sibling 1's and sibling 2's log hourly wage rate, respectively.

sibling-reported) education levels, and father's and mother's education levels for our sample of twins. In all our analyses we have randomly selected one twin as the first in each pair. We write S_1^1 for the self-reported education level of the first twin, S_1^2 for the sibling-reported education level of the first twin, S_2^2 for the self-reported education level of the second twin, and S_2^1 for the sibling-reported education level of the second twin. (That is, S_n^m, $m, n = 1,2$, refers to the education level of the nth twin as reported by the mth twin.) All six of the possible correlations are reported in the table. It is apparent that the independent measures of education levels are highly correlated. There are, of course, two measures of the father's and mother's education levels, and we have

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	−0.007 (0.015)	−0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷100)	−0.087 (0.023)	−0.089 (0.028)	−0.090 (0.029)	−0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	−0.410 (0.127)	−0.417 (0.143)	−0.424 (0.144)	−0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

[a]Own education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

cause the own-reports contain a common measurement-error component that the cross-sibling reports do not contain. In contrast, in the presence of classical measurement error these correlations would be identical. In fact, the correlations in Table 2 are consistent with the hypothesis of positively correlated measurement error in the siblings' reports.

In the presence of correlated measurement errors the instrumental-variables estimators of equation (4), (5), or (6) will be inconsistent. For example, instrumental variables used to obtain the fixed-effects estimator in (6) leads to

$$\operatorname{plim}\hat{\beta}_{\mathrm{FEIV}} = \beta \big/ \{1 - 2\rho_\nu[\mathrm{Var}(\nu)/\mathrm{Var}(\Delta S)]\}.$$

A straightforward consistent estimator of equation (6) may be obtained by instrumental-variables estimation of

$$(8)\quad y_{1i} - y_{2i} = \beta\left(S_1^1 - S_2^1\right) + \varepsilon_{1i} - \varepsilon_{2i}$$
$$= \beta \Delta S^* + \Delta\varepsilon$$

in which $\Delta S^{**} = S_1^2 - S_2^2$ is used as an instrument for ΔS^*, and we report this estimate below.[5]

C. *The Basic Empirical Results*

Table 3 contains simple estimates of the effect of schooling on earnings that control only for demographic variables (that may be considered strictly exogenous). In columns (i) and (ii) we report the results of stacking equations (1) and (2) and fitting them by least squares and generalized least squares (the seemingly-unrelated-regression method due to Arnold Zellner [1962]). The results in columns (i) and (ii) are comparable to most of the estimates that have appeared in the literature which ignore the potential correlation between schooling level and

[5]Note that the estimates using averages of the schooling differences will be inconsistent in the presence of correlated measurement errors, but as in the classical case, the inconsistency will be reduced by averaging.

TABLE 5—GLS, IV, AND FIXED-EFFECTS ESTIMATES OF AUGMENTED LOG-WAGE EQUATIONS FOR IDENTICAL TWINS

Variable	GLS (i)	GLS (ii)	IV[a] (iii)	First difference (iv)	First difference by IV (v)
Own education	0.105	0.105	0.147	0.091	0.179
	(0.016)	(0.016)	(0.034)	(0.022)	(0.041)
Sibling's education	—	−0.008	−0.062	—	—
			(0.016)	(0.035)	
Age	0.082	0.082	0.082	—	—
	(0.023)	(0.023)	(0.019)		
Age squared (÷100)	−0.094	−0.094	−0.092	—	—
	(0.029)	(0.029)	(0.024)		
Male	0.147	0.149	0.139	—	—
	(0.080)	(0.081)	(0.066)		
White	−0.472	−0.482	−0.506	—	—
	(0.143)	(0.144)	(0.130)		
Covered by union	0.115	0.118	0.153	0.063	0.095
	(0.072)	(0.072)	(0.081)	(0.090)	(0.095)
Married	0.089	0.086	0.051	0.142	0.140
	(0.065)	(0.065)	(0.073)	(0.081)	(0.086)
Years of tenure	0.025	0.024	0.020	0.028	0.028
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Father's education	0.001	0.001	0.006	—	—
	(0.014)	(0.014)	(0.013)		
Mother's education	0.013	0.015	0.019	—	—
	(0.017)	(0.018)	(0.017)		
Sample size:	284	284	284	147	147
R^2:	0.320	0.320	—	0.257	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

[a]Own education and sibling's education are instrumented using sibling's report of the other sibling's education as instruments.

indicates that white workers earn less than nonwhite workers. It seems possible that this result is due to selection in the relatively small sample of nonwhites who attended the twins festival and turned up in our sample. We have, therefore, computed the results in Tables 4 and 5 deleting the sample of nonwhite workers. The results of these regressions for white workers do not differ in any material way from those already reported. (The effect of schooling on wage rates is slightly higher for white twin pairs than for the group as a whole, but this difference is not statistically significant.)

Finally, we implement an instrumental-variables approach that is consistent in the presence of measurement errors that are correlated between the twins' reports of their own schooling and of their siblings' schooling. Specifically, we include $\Delta S^* = S_1^1 - S_2^1$ in the first-differenced wage equations, and use $\Delta S^{**} = S_1^2 - S_2^2$ as an instrument for ΔS^*. These instrumental-variables first-difference estimates, along with least-squares first-difference estimates, are reported in Table 6. When no other covariates are included, the instrumental-variable estimate that is robust to correlated measurement errors is 0.129, which is 20 percent greater than the OLS estimate of 0.107. Similar results hold when other variables are added to the regression [see columns

Table 1. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings (standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)	.063 (.000)	.060 (.029)
F (excluded instruments)		13.486		4.747		1.613
Partial R^2 (excluded instruments, $\times 100$)		.012		.043		.014
F (overidentification)		.932		.775		.725
Age Control Variables						
Age, Age^2	x	x			x	x
9 Year of birth dummies			x	x	x	x
Excluded Instruments						
Quarter of birth		x		x		x
Quarter of birth $\times$ year of birth				x		x
Number of excluded instruments		3		30		28

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), and 8 Regional dummies as control variables. F (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basmann (1960).

finite-sample bias. Because quarter of birth is related, by definition, to age measured in quarters within a single year of birth, and because age is an important determinant of earnings, we find the specification using within-year age controls [column (6)] to be more sensible than the specification that does not [column (4)]. The F statistic on the excluded instruments in column (6) indicates that quantitatively important finite-sample biases may affect the estimate. Comparing the partial R^2 in columns (2) and (6) shows that adding 25 instruments does not change the explanatory power of the excluded instruments by very much, explaining why the F statistic deteriorates so much between the two specifications.

Compulsory attendance laws, and the degree to which these laws are enforced, vary by state. In AK-91 the authors used this cross-state variation to help identify the coefficient on education by including state of birth $\times$ quarter of birth interactions as instruments in some of their specifications. Besides improving the precision of the estimates, using variation across state of birth should mitigate problems of multicollinearity between age and quarter of birth. In Table 2 we report replications of AK-91's Table VII, columns (5) through (8). These models use quarter of birth $\times$ state of birth interactions in addition to quarter of birth and quarter of birth $\times$ year of birth interactions as instruments for educational attainment.

Including the state of birth $\times$ quarter of birth interactions reduces the standard errors on the IV results by more than a factor of two and stabilizes the point estimates considerably. The F statistics on the excluded instruments in the first stage of IV do not improve, however. These F statistics indicate that although including state of birth $\times$ quarter of birth interactions improves the precision and reduces the instability of the estimates, the possibility that small-sample bias may be a problem remains.

To illustrate that second-stage results do not give us any indication of the existence of quantitatively important finite-sample biases, we reestimated Table 1, columns (4) and (6), and Table 2, columns (2) and (4), using randomly generated information in place of the actual quarter of birth, following a suggestion by Alan Krueger. The means of the estimated standard errors reported in the last row are quite close to the actual standard deviations of the 500 estimates for each model. Moreover, the distribution of the estimates appears to be quite symmetric. In these cases, therefore, the asymptotic standard errors give reasonably accurate information on the sampling variability of the IV estimator. This is specific to these cases, however. Nelson and Startz (1990a) showed, in the context of a different example, that asymptotic standard errors can give very misleading information about the actual sampling distribution of the IV estimator when the correlation between the instrument and the endogenous variable is weak.

Table 2. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Controlling for State of Birth (standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.083 (.009)	.063 (.000)	.081 (.011)
F (excluded instruments)		2.428		1.869
Partial R^2 (excluded instruments, $\times 100$)		.133		.101
F (overidentification)		.919		.917
Age Control Variables				
Age, Age^2			x	x
9 Year of birth dummies	x	x	x	x
Excluded Instruments				
Quarter of birth		x		x
Quarter of birth $\times$ year of birth		x		x
Quarter of birth $\times$ state of birth		x		x
Number of excluded instruments		180		178

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), 8 Regional dummies, and 50 State of Birth dummies as control variables. F (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basmann (1960).