Proxying Ability by Family Background in Returns to Schooling Estimations is Generally a Bad Idea*

Erik Mellander
Institute of Labour Market Policy Evaluation (IFAU), SE-751 20 Uppsala, Sweden
erik.mellander@ifau.uu.se

Sofia Sandgren-Massih
Uppsala University, SE-751 20 Uppsala, Sweden
sofia.sandgren@nek.uu.se

Abstract

When schooling is measured with error and data on ability are lacking, return to schooling estimates will be subject to positive omitted variable bias (OVB) and negative measurement error bias (MEB). We investigate how these biases are affected when ability is proxied by family background variables. We show that the effect on OVB is uncertain, while MEB invariably increases in magnitude. Our empirical analysis demonstrates that MEB generally dominates OVB. With more background variables or increased measurement error, the total bias rapidly becomes negative and increasing in magnitude, thereby driving the return estimate further and further away from the true value.

Keywords: Missing data; proxy variables; measurement error; consistent estimates of omitted variable bias and measurement error bias

JEL classification: C13; C20; C52; J31

I. Introduction

In estimation of the return to schooling, many practitioners have noted that the return estimates tend to fall when, for want of ability measures, family background variables are included in the earnings equation. Could this be a general property, i.e., is it possible to demonstrate analytically that it holds under a large variety of circumstances?

Lam and Schoeni (1993) claim that this is indeed the case. Referring to Welch (1975) and Griliches (1977), they note that if there is measurement

* Financial support from the Swedish Agency of Innovation Systems and the Swedish Research Council is gratefully acknowledged. Thanks to Colm Harmon for comments on an early version of this paper. Seminar participants at the Institute for Labour Market Policy Evaluation, in particular Björn Öckert, provided constructive criticism on a later version. The paper was considerably improved by useful comments from two anonymous referees. We also thank Anna Sjögren for help with a graphical illustration.

© The editors of the Scandinavian Journal of Economics 2008. Published by Blackwell Publishing. 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.
error in the schooling variable, the inclusion of a variable that is correlated with a worker’s schooling may increase the measurement error bias as well as reduce the omitted variable bias. To take this into account, they provide equations for the asymptotic bias in the estimated return to schooling where the total bias is additively decomposed into omitted variable bias and measurement error bias, before and after the inclusion of a family background variable. Without proof, Lam and Schoeni (1993, p. 719) claim that “…under plausible assumptions…”: (i) the omitted variable bias is positive in both cases but smaller after the inclusion of the family background variable and (ii) the measurement error bias is negative in both cases but larger in magnitude when the family background variable is included. The addition of the family background variable is thus claimed to affect the omitted variable bias and the measurement error bias in the same direction: both changes lower the estimated return to schooling.

In a subsequent empirical analysis, Lam and Schoeni (1993) implicitly extend these theoretical conclusions, drawn in the context of a single family background variable, to the case with many family background variables.

The purpose of this paper is threefold. The first is to correct an error in Lam and Schoeni’s theoretical analysis of the effects of including a single family background variable in the earnings regression. In Section II we demonstrate that while their conclusion (ii) is right, conclusion (i) is in general incorrect. It is shown, however, that there are restrictive conditions under which (i) does hold true. We also consider related results that have been established earlier in another literature, which focuses specifically on omitted variable bias.

The second purpose is to extend the theoretical analysis to an arbitrary number (K) of family background variables. It is shown in Section III that the conditions under which the omitted variable bias is reduced towards zero do not carry over to the case with several family background variables. The measurement error bias results continue to hold in the general case, however.

The third purpose of the paper is to provide an empirical assessment of the effects of proxying ability by means of family background variables. How are the omitted variable bias and the measurement error bias affected? What does this imply for the total bias?

To answer these empirical questions we need data on ability, i.e., the information whose absence is the very reason for the problem considered. It might seem odd to address the problem of proxying ability when data on ability are actually available, but only in this way can conclusions be drawn about the consequences of proxying when ability data are not at hand. Moreover, as schooling is treated as predetermined in the theoretical analysis we need either information substantiating this assumption or, otherwise, an instrument for observed schooling.
The unique dataset that we employ satisfies these requirements. It is a panel dataset, covering 555 males, born in 1928 in the city of Malmö in southern Sweden; cf. Section V for details. In Section IV we derive estimates of the omitted variable bias and the measurement error bias conditional on a given ratio of the measurement error variance to the total variance in schooling. Explicit information on the amount of measurement error in the schooling variable is thus not a prerequisite for the empirical analysis in Section VI. Concluding comments are provided in Section VII.

II. The Case with One Family Background Variable

To facilitate comparison with Lam and Schoeni (1993) we derive our basic results in the context of their stylized model. We then show how the analysis can be extended to accommodate an arbitrary number of control variables.

Correction of the Results in Lam and Schoeni (1993)

The starting point in Lam and Schoeni (henceforth LS) is the following equation, giving the “true” relation between (log) earnings, \( Y \), schooling, \( S \), and (unknown) ability, \( A \), for the \( i \)th individual:

\[
Y_i = \beta_0 + \beta_s S_i + \beta_a A_i + u_i, \quad \text{where } \beta_s, \beta_a > 0. \tag{1}
\]

Further, \( u_i \) is a random disturbance, viewed as realizations of the random variable \( u \), characterized by \( E(u) = 0 \) and \( \text{Var}(u) = \sigma_u^2 \). In addition, it is assumed that the schooling variable is measured with error, such that observed schooling, \( S^* \), can be expressed according to

\[
S^*_i = S_i + w_i, \tag{2}
\]

where \( w \) is pure measurement error uncorrelated with \( S \), i.e., \( E(w) = 0 \), \( \text{Var}(w) = \sigma_w^2 \), and \( \text{Cov}(S, w) = 0 \). LS also implicitly take \( w \) to be uncorrelated with \( A \) and \( u \), and \( u \) and \( w \) to be uncorrelated with the family background variable, \( F \). Thus:

\[
\text{Cov}(w, S) = \text{Cov}(w, A) = \text{Cov}(w, u) = \text{Cov}(w, F) = \text{Cov}(u, F) = 0. \tag{3}
\]

Another assumption implicitly made by LS is that schooling can be treated as a predetermined variable. In general, this is a strong assumption; see e.g. Card (1999). However, in the present context it is merely a simplifying assumption which allows us to focus on the problems of omitted variable bias and measurement error bias.\(^2\)

\(^1\) Lam and Schoeni do not explicitly state the positivity constraints on \( \beta_s \) and \( \beta_a \). They consistently use these restrictions in their analysis, however.

\(^2\) It is always possible to think of \( S \) as an instrument for schooling—rather than schooling itself—and \( w \) as an associated random error. Thus, if schooling is endogenous, our results can be applied once an instrument has been substituted for the schooling variable.
LS first consider the case when $Y$ is simply regressed on $S^*$, i.e., when the unobserved ability variable is disregarded and the measurement error in schooling ignored. The probability limit of the estimated return to education is then given by

$$\text{plim} \hat{\beta}_S = \beta_s - \beta_s \lambda + \beta_a \hat{\beta}_{AS} (1 - \lambda),$$

(4)

where $\lambda$ is the noise-to-signal ratio, i.e.,

$$\lambda \equiv \frac{\text{Var}(w)}{\text{Var}(S^*)}, \quad 0 \leq \lambda < 1,$$

(5)

and $\hat{\beta}_{AS}$ is the coefficient from a hypothetical regression of ability on true schooling:

$$\hat{\beta}_{AS} \equiv \frac{\text{Cov}(A, S)}{\text{Var}(S)}, \quad \hat{\beta}_{AS} > 0.$$  

(6)

The second term on the RHS of (4) is the measurement error bias and the third the omitted variable bias. It can be seen that the measurement error bias is negative, and increasing in magnitude with the variance of the measurement error. As schooling and ability are assumed to be positively correlated, cf. (6), the omitted variable bias is positive. Note that, in general, one cannot rule out the possibility that the measurement error bias dominates the omitted variable bias, thus making the total bias negative.

LS consider how $\text{plim} \hat{\beta}_S$ is affected if a measure of family background, $F$, is added to the regression. Their result for this case contains an error, however. The correct expression is provided in Proposition 1. LS’s equation is considered immediately after the proposition. Three corollaries are then given. The last of these provides a bridge between the general result in Proposition 1 and the equation suggested by LS.

**Proposition 1.** Given equations (1), (2) and (3), OLS regression of $Y$ on $S^*$ and a family background measure, $F$, yields an estimate of $\beta_s$ whose probability limit is given by

$$\text{plim} \hat{\beta}_{S,F} = \beta_s - \beta_s \frac{\lambda}{1 - R_{S^*F}^2} + \beta_a \hat{\beta}_{AS} (1 - \lambda) \left(1 - \theta \cdot \rho_{AF,S^*}^2\right),$$

(7)

where $\lambda$ and $\hat{\beta}_{AS}$ are defined by (5) and (6), respectively. Further, $R_{S^*F}^2(< 1)$ is the squared correlation of $S^*$ and $F$, and $\rho_{AF,S^*}^2$ is the squared partial correlation of ability and the family background measure when one controls for schooling, i.e.,

$$\rho_{AF,S^*}^2 = \left(\frac{\rho_{AF} - \rho_{AS^*} \cdot \rho_{S^*F}}{\sqrt{1 - \rho_{AS^*}^2} \sqrt{1 - \rho_{S^*F}^2}}\right)^2,$$

while θ is defined according to
\[
\theta = \frac{\rho_{SF}/\rho_{AS} - \rho_{AS} \cdot \rho_{SF}}{\rho_{AF} - \rho_{AS} \cdot \rho_{SF}}; \quad \rho_{AF} - \rho_{AS} \cdot \rho_{SF} \neq 0,
\]
where \( \rho_{SF}, \rho_{AS} \) and \( \rho_{AF} \) denote bivariate correlations.

**Proof**: Follows as a special case of the proof of Proposition 2.3

The equation provided by LS, equation (8) in their paper, is

\[
\text{plim } \beta_{SF} = \beta_s - \beta_s \frac{\lambda}{1 - R^2_{SF}} + \beta_a \beta_{AS} (1 - \lambda) (1 - \rho^2_{SF}) \cdot \lambda.
\]

This equation differs from equation (7) with respect to the final term, the expression for the omitted variable bias. Specifically, the last parenthesis reads \((1 - \rho^2_{AF,S^*})\) instead of \((1 - \theta \cdot \rho^2_{AF,S^*})\) in (7). Since \(\rho^2_{AF,S^*} \in [0, 1]\) by construction, and \(\rho^2_{AF,S^*} \in [0, 1]\) by assumption, (8) implies that the omitted variable bias invariably decreases towards zero upon the inclusion of a family background variable. Corollary 1 shows, however, that the omitted variable bias may well increase, driving the estimate of \(\beta_s\) upward, away from zero.

**Corollary 1.** If schooling and the family background variable are correlated, i.e., if \(R^2_{SF} > 0\), then the inclusion of the family background variable unambiguously increases the measurement error bias, compared to when no family background variable is included. If \(R^2_{SF} = 0\) the measurement error bias will be unchanged. The effect on the omitted variable bias cannot be determined a priori; the omitted variable bias may decrease or it may increase. This is true also in the absence of measurement error.

**Proof**: The first part of the corollary follows trivially from the facts that, by construction, \(R^2_{SF} \in [0, 1]\) and, by assumption, \(R^2_{SF} < 1\). The second part follows from the fact that \(\theta\) may be both negative and positive. Moreover, \(\theta\) is not bounded, either from below or from above. If \(\theta\) is negative, the omitted variable bias increases with certainty. If \(\theta\) is positive, the omitted variable bias may decrease—if \(\theta\) is small enough to ascertain that \(\theta \cdot \rho^2_{AF,S^*} < 1\). But it might also increase—if \(\theta\) is large enough to yield \(\theta \cdot \rho^2_{AF,S^*} > 1\).\(^4\) Comparison of (4) and Proposition 1 shows that the effect on the omitted variable bias is manifested in the factor \((1 - \theta \cdot \rho^2_{AF,S^*})\); it is independent of the factor \((1 - \lambda)\), i.e., the extent of measurement error.

---

\(^3\) A proof of Proposition 1 is provided in Mellander and Sandgren-Massih (2008).

\(^4\) In principle, \(\theta = 0\) is also a possibility. However, that requires \(\rho_{AS} = 1\), which is a pathological case in the sense that it implies that the parameters in equation (1) cannot be identified.
To provide some intuition for Corollary 1, note that controlling for $F$ means purging the schooling variable of "$A$-factors" and of other factors as well. At some point, the latter, undesired, effect will outweigh the former, desired, effect.

In Corollary 2 we discuss a special case of the general situation considered in Corollary 1. A constraint on $\theta$ is provided which ascertains that the omitted variable bias stays positive and is reduced towards zero. A condition is also given which is necessary, but not sufficient, for this constraint to be satisfied.

**Corollary 2.** If, and only if, $0 < \theta \leq 1/\rho_{AF,S^*}^2$, then the positive omitted variable bias will remain positive and be reduced towards zero when the family background variable is added to the earnings equation. A necessary, but not sufficient, condition for these inequalities to hold is that $\text{sign}(\rho_{AF}) = \text{sign}(\rho_{S^*F})$.

**Proof:** That the constraint implies that the omitted variable is reduced while staying positive follows directly from the fact that the change in the bias is determined by $(1 - \theta \cdot \rho_{AF,S^*}^2)$ where $\rho_{AF,S^*}^2 \in [0, 1]$. For values of $\theta$ above the value of $\rho_{AF,S^*}^2$, the omitted variable bias changes sign. For $\theta = 0$ the omitted variable bias is unaffected and thus not reduced. For $\theta < 0$ the omitted variable bias increases.

To prove the necessary condition, first consider the case when $\rho_{S^*F} > 0$. In this case the numerator of $\theta$ is unambiguously positive; cf. the definition of $\theta$ in Proposition 1 and recall that $\rho_{AS^*} = \rho_{AS} \in [0, 1]$, by assumption. A necessary requirement for $\theta$ to be positive, which in turn is necessary for $\theta$ to belong to $[0, 1/\rho_{AF,S^*}^2]$, is thus that the denominator of $\theta$ is positive, too. This requires $\rho_{AF} > 0$. But it may be that $0 < \rho_{AF} < \rho_{AS^*} \cdot \rho_{S^*F}$, in which case $\theta < 0$. Hence, for $\rho_{S^*F} > 0$ the condition $\rho_{AF} > 0$ is necessary but not sufficient for the omitted variable bias to remain positive and be reduced towards zero. In a perfectly analogous way it can be shown that if $\rho_{S^*F} < 0$, then $\rho_{AF} < 0$ is a necessary but not sufficient condition for keeping the omitted variable bias positive and decreasing towards zero. The case $\rho_{S^*F} = 0$ can be disregarded because it implies $\theta = 0$. Together, the results for $\rho_{S^*F} > 0$ and $\rho_{S^*F} < 0$ yield the necessary but not sufficient condition stated in the corollary.

We next consider “a special case of the special case” characterized in Corollary 2, namely when $\theta = 1$, the constraint implicitly imposed by LS. Corollary 3 provides an interpretation of this constraint, in terms of the correlation between schooling and family background, conditional on ability.
Corollary 3. If the correlation between schooling and family background is equal to zero when ability is controlled for, i.e., if $\rho_{S^*F} = 0$, then $\theta = 1$. This is a sufficient, but not necessary, condition for the positive omitted variable bias to decrease towards zero when one family background variable is included in the earnings regression.

Proof: Note that, by definition,

$$
\rho_{S^*F} = \frac{\rho_{S^*F} - \rho_{AS^*} \cdot \rho_{AF}}{\sqrt{1 - \rho_{AS^*}^2} \sqrt{1 - \rho_{AF}^2}}. \quad (9)
$$

Thus, if $\rho_{S^*F} = 0$ then $(\rho_{S^*F} / \rho_{AS^*}) = \rho_{AF}$ and, consequently, $\theta = 1$. Given $\theta = 1$, the second part follows directly from Corollary 2.

At this point, a relevant question is whether the assumption $\theta = 1$ is not implicit in the model which provides the starting point of LS’s analysis, i.e., (1)–(3)? The answer is no. The best way to see this is to note that just as (1)–(3) allow $F$ to be an instrument for $A$, they also allow $F$ to be an instrument for $S^*$. That $F$ is an admissible instrument for $A$ follows from the properties $\text{Cov}(u, F) = 0$, given in (3), and $\rho_{AF, S^*} \neq 0$, ascertained in Proposition 1. Likewise, $\text{Cov}(u, F) = 0$ and $\rho_{S^*F} = 0$ together imply that $F$ is a valid instrument for $S^*$. The latter possibility can only be excluded by explicitly specifying that $\rho_{S^*F} = 0$, which is certainly not done by LS.\(^5\)

An alternative to the conditions provided in Corollaries 2 and 3 is to impose constraints on $F$ that imply $\rho_{S^*F} = 0$. One (trivial) example is provided by McCallum (1972) and Wickens (1972) in the context of omitted variable bias only. Applied to the present context, they assume that $F = A + \eta$ where $\eta$ is a random error which is uncorrelated with all of the model’s explanatory variables and all of its random terms.\(^6\) Then $\rho_{S^*F} = \rho_{AS^*}$ and $\rho_{AF} = 1$, implying that $\rho_{S^*F} = 0$. Yet another alternative is to consider whether the assumption $\rho_{S^*F} = 0$ can be justified empirically. Again, the answer is no. On the contrary, it is quite natural to think of family background as reflecting both nature (ability) and nurture (schooling). To the extent that this view is justified, $\rho_{S^*F} = 0$ is most likely to be non-zero. To illustrate this, we provide three examples of $\rho_{S^*F}$ at the end of Section V.

\(^5\)It should be noted that $\rho_{S^*F}$ is not affected by the fact that $F$ is not included in equation (1). Leaving out $F$ amounts to assuming $\rho_{YF,S^*} = 0$, which is equivalent to $\rho_{UF} = 0$. The latter implies $\text{Cov}(u, F) = 0$ which we have explicitly accounted for in (3).

\(^6\)That is, $\text{Cov}(\eta, S^*) = \text{Cov}(\eta, A) = \text{Cov}(\eta, u) = \text{Cov}(\eta, w) = 0$. 

Allowing for Control Variables

Frost (1979) provides (without proof) results that are qualitatively equivalent to Proposition 1, in the context of omitted variable bias only. In so doing, he allows for an arbitrary number of control variables, \(X_1, X_2, \ldots, X_M\). Frost’s approach amounts to preceding the above analysis by purging the RHS variables from the influences of the control variables. To this end, consider estimating the following regressions by OLS:

\[
S_i = \gamma_0 + \gamma_1 X_{1i} + \cdots + \gamma_M X_{Mi} + e_{Si}, \\
A_i = \delta_0 + \delta_1 X_{1i} + \cdots + \delta_M X_{Mi} + e_{Ai}, \\
F_i = \kappa_0 + \kappa_1 X_{1i} + \cdots + \kappa_M X_{Mi} + e_{Fi}.
\] (10)

What Frost’s analysis shows is that for \(\lambda = 0\), the estimated residuals \(\hat{e}_S, \hat{e}_A\), and \(\hat{e}_F\) can be substituted for \(S_i, A_i\) and \(F_i\), respectively, in Proposition 1 and Corollaries 1–3, and the results will still go through—although they must, of course, be reinterpreted as “control-variables-corrected”.

Fortunately, allowing for \(\lambda \neq 0\) is unproblematic. In this case, the first regression is replaced by

\[
S_i^* = \gamma_0 + \gamma_1 X_{1i} + \cdots + \gamma_M X_{Mi} + \left(e_{Si} + w_i\right),
\] (11)

and \(\hat{e}_{S_i^*} = \hat{e}_{Si} + w_i\) is substituted for \(S_i^*\). As \(\text{Cov}(e_{Si}, w_i) = 0\), by (3), we have \(\text{Var}(e_{Si} + w_i) = \text{Var}(e_{Si}) + \text{Var}(w_i)\). Thus, given an assumption about \(\lambda\) (cf. Section VI), yielding an estimate of \(\text{Var}(w_i)\), an estimate of \(\text{Var}(e_{Si})\) can also be determined, which is needed in the “control-variables-corrected” version of (16) in Section IV.

III. Several Family Background Variables

The number of family background variables is now taken to be equal to \(K \geq 1\). For analytical simplicity we abstract from control variables; from the previous sub-section it should be clear that we can do so without loss of generality. The \(K\)-variable counterpart to Proposition 1 is given by the following proposition.

**Proposition 2.** Given equations (1), (2) and (3), OLS regression of \(Y\) on \(S^*\) and a \(K \times 1\) vector \(F\) of family background variables yields an estimate
of $\beta_s$ whose probability limit is given by:

$$
\text{plim} \hat{\beta}_{S,F} = \beta_s - \beta_s \cdot \frac{\lambda}{1 - R^2_{S,F}} + \beta_a \hat{\beta}_{AS} \frac{(1 - \lambda)}{1 - R^2_{S,F}} \left[ 1 - \sum_{j=1}^{K} \rho_{AFj} \sqrt{\frac{\text{Var}(F_j)}{\text{Var}(S^*)}} \text{plim}(\hat{\alpha}_j) \right],
$$

where $\lambda$ and $\hat{\beta}_{AS}$ are defined by (5) and (6), respectively, and $\hat{\alpha}_j$ is the OLS estimate of the coefficient for $F_j$ in the linear regression of $S^*$ on $F$.

**Proof**: See the Appendix.

Two features of Proposition 2 are worth noting. First, the result for the measurement error bias is a straightforward extension of the result in the case with one family background variable. Inclusion of family background variables will always increase the negative measurement bias, thus driving the estimate of $\beta_s$ downward.

Second, like the measurement error bias, the omitted variable bias is inversely related to $1 - R^2_{S,F}$. Thus, the larger the part of the variance in $S^*$ explained by the family background variables, the higher is the probability that the omitted variable bias increases, compared to when family background variables are disregarded. This tendency may be balanced by the sum within square brackets but, in general, it is impossible to say anything about the relative weights of these opposing forces.$^8$

**IV. Estimation of the Omitted Variable Bias and the Measurement Error Bias**

The analysis in this section corresponds to the following hypothetical experiment. Assume that, initially, the econometrician does not have access to information about ability and, therefore, tries to proxy ability by means of family background variables. At a later stage, (s)he gets access to a measure of ability and wants to use this information to estimate the omitted variable and the measurement error biases associated with the initial estimates. We take the initial stage as given here, i.e., we assume that we have results from regressions of $Y$ on $S^*$ and of $Y$ on $S^*$ plus family background variables.

From (4) and Proposition 1 we know that two unknown parameters involved in the omitted variable bias and the measurement error bias are $\beta_s$

$^8$ To illustrate how difficult it is to say anything *a priori* about how the omitted variable bias is affected, it is instructive to consider the case $K = 2$. This is done in Example 1 in Mellander and Sandgren-Massih (2008) which is omitted here.

and $\beta_a$. Given data on true schooling, $S$, and ability, $A$, these parameters could be estimated by application of OLS to equation (1). However, while we know $A$ we lack information about $S$; what we have is $S^*$. Accordingly, we can run a regression which is very close to equation (1), namely:

$$Y_i = \beta_0^* + \beta_s^*S_i^* + \beta_a^*A_i + u_i^*, \quad \text{where } \beta_{s^*}, \beta_{a^*} > 0. \quad (12)$$

The only difference between equations (12) and (1) is that in (12), observed schooling, $S_i^*$, replaces true schooling, $S_i$. This changes the model’s parameters and stochastic disturbance, compared to equation (1); to reflect this we attach an asterisk (*) to them.

Intuitively, it seems likely that OLS estimates of the parameters in (12) can be useful in the construction of estimates of the omitted variable bias and the measurement error bias. Proposition 3 and Corollary 4 substantiate this intuition.

**Proposition 3.** The OLS estimates of the parameters $\beta_{s^*}$ and $\beta_{a^*}$ in (12) have the following probability limits:

$$\text{plim } \hat{\beta}_{s^*} = \beta_s - \beta_s \frac{\lambda}{1 - \rho_{AS}^2},$$

$$\text{plim } \hat{\beta}_{a^*} = \beta_a + \beta_s \frac{\lambda \cdot \hat{\beta}_{S^*A}}{1 - \rho_{AS}^2},$$

where

$$\hat{\beta}_{S^*A} = \frac{\text{Cov}(A, S^*)}{\text{Var}(A)} \quad (14)$$

is the coefficient from a regression of observed schooling on ability.

**Proof:** Follows directly from the results in Carroll, Ruppert and Stefanski (1995, Ch. 2.2.3).

**Corollary 4.** Conditional on a consistent estimate of $\lambda$, denoted $\lambda^\dagger$, consistent estimates of $\beta_s$ and $\beta_a$ in equation (1) can be constructed as follows:

$$\hat{\beta}_{s|\lambda^\dagger} = \hat{\beta}_{s^*} \cdot \left(1 - \frac{\lambda^\dagger}{1 - \rho_{AS}^2}\right)^{-1},$$

$$\hat{\beta}_{a|\lambda^\dagger} = \hat{\beta}_{a^*} - \hat{\beta}_{s|\lambda^\dagger} \cdot \frac{\lambda^\dagger \cdot \hat{\beta}_{S^*A}}{1 - \rho_{AS}^2}. \quad (15)$$

**Proof:** Implied by Proposition 3 and the properties of the plim operator.

By means of (4), Proposition 2 and Corollary 4, we can construct consistent estimates of the omitted variable bias conditional on $\lambda^\dagger$, without and
with family background variables, according to:

\[ \hat{OVB}_{S^*|\lambda^\dagger} = \hat{\beta}_{a|\lambda^\dagger} \hat{\beta}_{AS|\lambda^\dagger} (1 - \hat{\lambda}^\dagger), \]

\[ \hat{OVB}_{S^*,F|\lambda^\dagger} = \hat{\beta}_{a|\lambda^\dagger} \hat{\beta}_{AS|\lambda^\dagger} (1 - \hat{\lambda}^\dagger) \]

\[ \times \frac{1}{1 - R_{S^*F}^2} \left[ 1 - \sum_{j=1}^{K} \frac{\rho_{AF_j} \sqrt{\text{Var}(F_j)}}{\rho_{AS^*} \sqrt{\text{Var}(S^*)}} (\hat{\lambda}_j) \right]. \]  

(15)

Note that in the last row of (15) the consistent estimate \( \hat{\lambda}_j \) has been substituted for the corresponding probability limit in Proposition 2. This is admissible because consistency of \( \hat{\lambda}_j \) is sufficient to ascertain consistency of \( \hat{OVB}_{S^*,F|\lambda^\dagger} \), *ceteris paribus*. It still remains, however, to determine the parameter \( \hat{\beta}_{AS|\lambda^\dagger} \). While the original parameter \( \hat{\beta}_{AS} \) in (6) cannot be computed unless true schooling, \( S \), is known, this is, fortunately, not the case when we condition on \( \lambda^\dagger \). We have:

\[ \hat{\beta}_{AS|\lambda^\dagger} \equiv \frac{\text{Cov}(A, S)}{\text{Var}(S)} \bigg|_{\lambda^\dagger} = \frac{\text{Cov}(A, S^*)}{\text{Var}(S^*) - \hat{\lambda}^\dagger \cdot \text{Var}(S^*)} \bigg|_{\lambda^\dagger} \]

\[ = \frac{\text{Cov}(A, S^*)}{\text{Var}(S^*)(1 - \hat{\lambda}^\dagger)} \bigg|_{\lambda^\dagger} = \rho_{AS^*} \frac{\sqrt{\text{Var}(A)}}{\sqrt{\text{Var}(S^*)}} \frac{1}{(1 - \hat{\lambda}^\dagger)} \bigg|_{\lambda^\dagger}, \]  

(16)

where the first equality follows from (2) and (3) and the third equality follows directly from the definition of the coefficient of correlation.

Finally, we get the following consistent estimates of the measurement error bias before and after the inclusion of family background variables, again combining Corollary 4 with (4) and Proposition 2:

\[ \hat{MEB}_{S^*|\lambda^\dagger} = -\hat{\beta}_{s|\lambda^\dagger} \hat{\lambda}^\dagger, \]

\[ \hat{MEB}_{S^*,F|\lambda^\dagger} = -\hat{\beta}_{s|\lambda^\dagger} \frac{\hat{\lambda}^\dagger}{1 - R_{S^*F}^2}. \]  

(17)

V. Data and Variable Specifications

Our dataset stems from a unique longitudinal survey, initiated in 1938 in Malmö, Sweden. For the purpose of disentangling the effects of cognitive ability and social background on student achievement, a Swedish sociologist conducted a survey involving all of the city’s third-grade pupils. As the common school starting age at the time was seven, the children were generally 10 years old when they were interviewed in the spring of 1938. Altogether, 1,542 children were surveyed. These individuals were followed up and recurrently interviewed until 1993, when they were 65 years old,
the Swedish retirement age. Presumably, this makes the Malmö survey one of the longest panels in the world.9

With respect to schooling, the original data contain information about the type and level of education, and, in many cases, about whether the respondent completed the education or not; altogether 42 alternatives are specified. Because of the large number of detailed alternatives, the likelihood that individuals have been wrongly categorized is small. A measurement error arises, however, when we transform the categorical data to number of years of schooling. This transformation is based on information about the stipulated number of years for each level and type combination (compared to the corresponding next lower level).

Specifically, individuals with completed educations have been assigned the stipulated number of years. To individuals reporting incomplete educations we have assigned the stipulated years minus 1. As the difference in stipulated years between two subsequent educational levels is at least two years, this means that the number of years we assign to reported incomplete educations correspond to at least half of the stipulated completion times. By doing so, we (strongly) increase the probability that the continuous measurement error that we generate is symmetric around zero.10

In accordance with the preferred practice in the literature, we use IQ test results to measure ability. This test contained four themes: “Word opposites”, “Sentence completion”, “Perception of identical figures” and “Disarranged sentences”.11

Regarding the age at which the test was conducted, it has been argued by several authors that it is important to measure IQ at a young age, to avoid contaminating the ability measure by schooling.12 On the other hand, constructing reliable tests for very young children is difficult. Age 10 strikes a balance between these opposing considerations. Moreover, to the extent that education does influence test results, we want the influence to be similar across individuals. This is an additional argument for conducting the test at age 10 because, at the time, education was comprehensive only up to the fourth grade.

9 For information about the Malmö survey, see Fagerlind (1975) and Furu (2000).
10 Essentially, the reason is that within a given level and type combination we know that all individuals have had a strictly positive amount of education, irrespective of whether they completed it or not. This implies that a consistent estimate of the number of years of schooling must exceed half of the corresponding stipulated years; see Mellander and Sandgren-Massih (2008) for a more detailed discussion of this issue.
11 A pilot test had been carried out on third-graders in municipalities in the neighborhood of Malmö in the year preceding the survey. Moreover, when the real test was conducted, considerable effort was made to ascertain that the test conditions were the same for all students.
12 See e.g. Hansen, Heckman and Mullen (2004), and the references therein.
Proxying ability by family background is generally a bad idea

Table 1. Descriptive statistics for the males in the Malmö study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample used (N = 555)</th>
<th>Original sample (N = 834)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
</tr>
<tr>
<td>$Y$: ln of gross income 1968, in 1,000 SEK</td>
<td>10.29</td>
<td>0.664</td>
</tr>
<tr>
<td>$S^*$: observed years of schooling, 1964</td>
<td>8.867</td>
<td>2.571</td>
</tr>
<tr>
<td>$A$: IQ, measured at age 10, i.e., 1938</td>
<td>98.26</td>
<td>15.75</td>
</tr>
<tr>
<td>$F_1$: father’s education, in years, 1938</td>
<td>7.690</td>
<td>1.492</td>
</tr>
<tr>
<td>$F_2$: family income 1937, in SEK</td>
<td>3,653</td>
<td>5,780</td>
</tr>
<tr>
<td>$F_3$: family size in 1938</td>
<td>4,500</td>
<td>1,557</td>
</tr>
</tbody>
</table>

The data include extensive information about family background. We have chosen information about the father’s education, family income and family size for two reasons. First, they are measured very closely in time to the IQ test. This should be an advantage as we want to use the family background variables as proxies for ability/IQ. Second, these variables are standard choices in the literature.

Information about work experience is missing in our data. However, since we are considering a cohort of male individuals, differences in work experience will, for given schooling, be determined by unemployment spells only. As the period we consider was characterized by almost full employment in Sweden, neglecting differences in unemployment across individuals is likely to be of minor importance.13

It can be seen in Table 1 that the average length of schooling was close to nine years for the 1928 cohort that we study. As the school-starting age was seven this means that, on average, the individuals finished their studies in 1944. The time span between completing school and the point in time when income is measured, 1968, is thus 24 years. With this extensive time lag, schooling is likely to be a predetermined variable. This conjecture is supported by Sandgren (2005) who, using very similar data from the Malmö survey, cannot reject the hypothesis that schooling is exogenous in an earnings regression.

Owing to attrition and incomplete data for some individuals, the sample we use comprises two-thirds of the original sample. According to Table 1, however, the loss of observations has barely affected means and standard deviations.

Coefficients of correlations are provided in Table 2. The table also includes examples of a partial correlation which plays a crucial role in Section II, the partial correlation between schooling and family background, while controlling for ability; cf. Corollary 3. Here, for simplicity, we only

13 Regarding military service, it was compulsory in Sweden at the time. The length of duty varied somewhat across individuals, but not much.

consider one background variable at a time. These examples clearly show that our data do not satisfy $\rho_{S^*F} = 0$, which would ascertain a reduction in omitted variable bias when a family background variable is used as proxy for ability in the earnings equation.

VI. Empirical Analysis

We now apply the results derived in Section IV. As all our empirical results are conditional on the unknown noise-to-signal ratio, we begin by considering the relevant range for this parameter. We then compute the estimates of the return to schooling and ability. Finally, we estimate the omitted variable bias, the measurement error bias and the total bias when earnings are regressed on (i) observed schooling only or (ii) observed schooling and various constellations of our three family background variables.

The Noise-to-Signal Ratio

An estimate of the noise-to-signal ratio, i.e., $\lambda$, in Swedish data is provided by Isacsson (1999). He estimates $\lambda$ to be 0.12 for imputed years of schooling, subject to a continuous (classical) measurement error of the type (2).

For our purposes, a precise estimate of $\lambda$ is not important. As we can estimate the omitted variable bias (OVB) and the measurement error bias (MEB) associated with any $\lambda$, we only need some idea about its range. Thus, it does not matter that our method of imputing years of schooling differs somewhat from Isacsson’s.

We report estimates of OVB and MEB for three values on $\lambda$: 0.08, 0.13 and 0.18.\(^\text{14}\) For the total bias, which is of primary interest, we report estimates when $\lambda$ varies continuously between 0 and 0.20; cf. Figure 1.

\(^\text{14}\) That we have chosen $\lambda = 0.13$ as our middle estimate, rather than Isacsson’s 0.12 estimate, is for expository reasons.
Proxying ability by family background is generally a bad idea

Table 3. OLS estimates of the parameters in equation (12)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. error</th>
<th>t-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0^*$</td>
<td>8.7910</td>
<td>0.1580</td>
</tr>
<tr>
<td>$\beta_s^*$</td>
<td>0.0961</td>
<td>0.0111</td>
</tr>
<tr>
<td>$\beta_a^*$</td>
<td>0.0066</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

$N = 555$  $R^2 = 0.22$

Table 4. Estimates of $\beta_s$ and $\beta_a$, conditional on $\lambda^l$

| $\lambda^l$ | $\tilde{\beta}_s|\lambda^l$ | $\tilde{\beta}_a|\lambda^l$ |
|------------|-------------------------------|------------------------------|
| $\lambda^l = 0.08$ | 0.1074 | 0.0057 |
| $\lambda^l = 0.13$ | 0.1159 | 0.0050 |
| $\lambda^l = 0.18$ | 0.1259 | 0.0042 |

Estimates of the Return to Schooling and Ability

The OLS estimates corresponding to equation (12) are reported in Table 3. The estimated return is 0.096. For comparison, Björklund and Kjellström (2002) report an estimate of 0.087 for an earnings equation for Swedish males, based on cross-section data for 1968 and using as regressors years of schooling and years of work experience.\footnote{The difference between the two estimates is not statistically significant.}

To construct consistent estimates of $\beta_s$ and $\beta_a$ from Table 3, we first use Tables 1 and 2 to obtain $\rho_{SA}^2 = 0.2395$ and $\tilde{\beta}_{S-A} = 0.0799$; cf. (14). Next, Corollary 4 yields the estimates in Table 4. Comparing Tables 3 and 4, we see that, as expected, the OLS estimate of the return to schooling is biased downwards, due to the measurement error in schooling. Given that the return estimate is biased downwards, the OLS coefficient for ability must be biased upwards, just as we find it to be.

Conditional Estimates of Omitted Variable Bias, Measurement Error Bias and Total Bias

Table 5 provides estimates of the OVB and the MEB, corresponding to (15) and (17), respectively. Regarding OVB, the most important aspect to note is that for a given noise-to-signal ratio, inclusion of family background variables has a very small effect on the OVB. Moreover, the bias is not monotonically decreasing in the number of family background variables included. Of course, this is not surprising, given our theoretical results. With increasing measurement error in schooling OVB decreases somewhat,
Table 5. Estimates of omitted variable bias (OVB) and measurement error bias (MEB), conditional on \( \lambda^\dagger \), for different sets of family background variables proxying for ability

<table>
<thead>
<tr>
<th>( \lambda^\dagger )</th>
<th>No ( F_i )'s</th>
<th>( F_1 = \text{father’s education} )</th>
<th>( F_2 = \text{family income} )</th>
<th>( F_3 = \text{family size} )</th>
<th>( F_1 &amp; F_2 )</th>
<th>( F_1 &amp; F_3 )</th>
<th>( F_2 &amp; F_3 )</th>
<th>( F_1 &amp; F_2 &amp; F_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda^\dagger = 0.08 )</td>
<td>0.0170</td>
<td>-0.0086</td>
<td>0.0150</td>
<td>-0.0151</td>
<td>0.0126</td>
<td>-0.0227</td>
<td>0.0170</td>
<td>-0.0107</td>
</tr>
<tr>
<td>( \lambda^\dagger = 0.13 )</td>
<td>0.0150</td>
<td>-0.0151</td>
<td>0.0126</td>
<td>-0.0227</td>
<td>0.0170</td>
<td>-0.0107</td>
<td>0.0150</td>
<td>-0.0188</td>
</tr>
<tr>
<td>( \lambda^\dagger = 0.18 )</td>
<td>0.0126</td>
<td>-0.0227</td>
<td>0.0170</td>
<td>-0.0107</td>
<td>0.0150</td>
<td>-0.0188</td>
<td>0.0126</td>
<td>-0.0282</td>
</tr>
</tbody>
</table>

Note: The estimates have been computed using (15) and (17).

but not very much. For example, more than doubling \( \lambda^\dagger \) (going from 0.08 to 0.18), decreases OVB less than 0.5 percentage points, or slightly more than a quarter of the initial bias (at \( \lambda^\dagger = 0.08 \)).

The MEB estimates are larger (in absolute value) than the OVB estimates, except when the measurement error is small (\( \lambda^\dagger = 0.08 \)). Inclusion of family background variables sizably increases the MEB. For instance, when \( \lambda^\dagger = 0.13 \) and all three of the family background variables are included, the (absolute value of the) MEB increases by almost 0.5 percentage points compared to when no background variables are included. The OVB is virtually unaffected by the same operation. Moreover, in contrast to OVB, the MEB is monotonically related to the number of background variables. The MEB is also much more sensitive to \( \lambda^\dagger \) than is the OVB. Going from \( \lambda^\dagger = 0.08 \) to \( \lambda^\dagger = 0.18 \) increases the MEB by more than 250 percent, i.e., the relative change in the MEB is larger than the relative change in the noise-to-signal ratio.

The total bias, TB, is equal to OVB + MEB. Since the MEB dominates the OVB, the TB has the same general properties as the MEB. In particular, additional family background variables will always make the TB fall. Referring back to our Introduction, this means that Lam and Schoeni (1993) were right in claiming that the estimated return to schooling can always be made to decrease, by inclusion of family background variables.

A decrease in TB can be a good thing if the TB is positive to begin with. But it is a bad thing if the TB is negative initially. In this case the change implies a move further away from the true return to schooling. This case is illustrated in Table 5 for \( \lambda^\dagger = 0.13 \). When no family background variables are included the TB is negative, but very small—virtually zero—because the OVB and the MEB cancel out. Adding family background variables
Proxying ability by family background is generally a bad idea

\[ \text{ABS}(TB(\lambda)|_{S^F}) - \text{ABS}(TB(\lambda)|_{S^*}) \]

\[ \lambda \]

\[ \text{Percent} \]

\[ \text{Fig. 1. Total bias, in percent of } \tilde{\beta}_{s|\lambda}, \text{ as a function of } \lambda, \text{ in the absence of family background variables, } TB(\lambda)|_{S^*}, \text{ and when } F_1, F_2 \text{ and } F_3 \text{ are included, } TB(\lambda)|_{S^*F} \]

merely has the effect of making a very good return estimate increasingly more biased.

To see when the use of family background variables turns from a useful to a wasteful practice, we need to consider the total bias as a continuous function of \( \lambda^\dagger \). This is done in Figure 1, which shows the estimated TB, in percent of \( \tilde{\beta}_{s|\lambda} \), in two cases. In the first case, no family background variables are included in the regression and, in the second case, \( F_1, F_2 \) and \( F_3 \) are included. The figure also shows the difference between the absolute values of the relative TBs in the two cases. The vertical dotted line delimits the areas where the inclusion of the family background variables decreases the TB—to the left of the line—and where the bias increases—to the right.

First, note how little the inclusion of the background variables decreases the relative TB when \( \lambda = 0 \). The reduction equals \( 20.5 - 19.6 = 0.9 \) percentage points of the return estimate. With \( \lambda \) strictly positive, the family background variables at first yield an increasing advantage in terms of relative TB. But this advantage is present only when \( \lambda < 0.117 \). Moreover, it is quite small; only when \( \lambda \) varies between 0.04 and 0.11 is the decrease at least 2 percent of the corresponding return estimate. The gain is largest when the bias is small to begin with. The maximum reduction occurs when the bias without family background variables is less than 4 percent of the estimated true return (at \( \lambda \approx 0.105 \)). If \( \lambda \) grows beyond 0.12 the relative TB rapidly becomes much larger with family background variables than without. For instance, when \( \lambda = 0.18 \), the relative TB is 6 percentage points larger with family background variables included, than when they are left out.

Qualitatively, the results in Figure 1 are the same if, instead, only one or two family background variables are included. The quantitative difference is that the gains/losses in the total bias are smaller (in absolute terms) if the number of background variables is reduced.

VII. Concluding Comments

This study was inspired by a remarkable claim, implied by an analysis in Lam and Schoeni (1993): using family background variables as proxies for unobserved ability in earnings regressions, the estimated return to schooling can be driven down to arbitrarily low levels. A compact way to summarize our findings is that Lam and Schoeni were right—but for the wrong reasons.

We have shown that Lam and Schoeni’s assertion that inclusion of family background variables will reduce the positive omitted variable bias (OVB) and increase the negative measurement error bias (MEB) is partly incorrect. The OVB may increase, as well as decrease. We also demonstrate, though, that in the context of a single family background variable, the OVB will indeed decrease for sure if the correlation between schooling and family background is identically zero when one controls for ability. But this a restrictive assumption; in our empirics we find that when ability is controlled for, the correlation between schooling and family background is far from zero.

Unlike Lam and Schoeni, we also conduct a theoretical analysis of the case with several family background variables. We show that the MEB result in the one-variable case can be extended to the $K$-variable case. The indeterminacy of the OVB, on the other hand, becomes even larger; the conditions which ascertain that the OVB is reduced in the one-variable case cannot be extended to the case when $K \geq 2$.

Theoretically, the reduction in the estimated return induced by increased MEB can thus be counteracted by increases in OVB. The effect on the total bias, i.e., $\text{MEB} + \text{OVB}$, hence becomes an empirical matter.

For the empirical analysis, we derive OVB and MEB estimates that are consistent, conditional on the ratio of measurement error variance to total variance in observed schooling ($\lambda$). The estimators are applied to a unique Swedish dataset that is extremely well suited to our analysis.

Our empirical results yield three conclusions. First, to the extent that inclusion of family background variables leads to increases in OVB, these increases are very small. Second, all changes in the OVB are very small, irrespective of whether they are positive or negative. Third, except for small values on $\lambda$—below 0.13 in our application—the OVB is dominated by the MEB, thereby implying a negative total bias. Furthermore, the MEB is much more sensitive than the OVB to changes in the number of family background variables and in $\lambda$. Both additional family background variables and
variables and increases in $\lambda$ monotonically increase the magnitude of the MEB.

Proxying ability by family background is thus generally a bad idea because the bias that one wants to influence—the OVB—is barely affected, while the side-effect—an increase in the MEB—is substantial and generally makes the total bias larger than in the absence of the background variables. For example, if $\lambda = 0.13$, the total bias when family background variables are excluded is estimated in our data to be virtually zero: $-0.08$ percent of the estimated true return. Adding three family background variables increases the relative bias to $-4.6$ percent. Moreover, when the inclusion of family background variables does indeed reduce the total bias, the reduction is largest when the bias is small to begin with, i.e., when a reduction is not very important.

It might be argued that our analysis is too stylized to provide useful insights about how omitted variables and measurement error affect estimates of the rate of return to education. We do not think so. Essentially, there are two potential problems: endogeneity of schooling and the lack of variables besides schooling and ability. Regarding the lack of variables, we have shown how an arbitrary number of control variables can be accounted for; cf. Section II.

Concerning the endogeneity problem, assume that schooling, $S$, is endogenous. This means that $S$ is a stochastic variable that is correlated with the stochastic disturbance in the wage equation. The standard remedy to this problem is to find an instrument for $S$. Such an instrument is another stochastic variable that is correlated with $S$ but not with the disturbance term. Our $S^*$ variable can be interpreted in precisely this way. Accordingly, our analysis can either be interpreted as concerning the case where the measure of schooling is predetermined or as taking place after an instrument has been found for the endogenous schooling variable.

**Appendix. Proof of Proposition 2**

Let the $(K + 1)$ square matrix $Q_{xx}$ be defined as

$$Q_{xx} = \begin{pmatrix} q_{s^*s^*} & q_{s^*f} \\ (1 \times 1) & (1 \times K) \\ q_{f^*} & Q_{ff} \\ (K \times 1) & (K \times K) \end{pmatrix}, \quad (A1)$$

where

$$q_{s^*s^*} = \frac{1}{N} \sum_{i=1}^{N} (S_i^* - \bar{S}^*)^2, \quad (A2)$$

and the typical elements of the vector $q_{f^*} = (q_{f^*, r})$ and the matrix $Q_{ff}$ are

$$q_{f^*} = \left( q_{f^*, r} \right) = \left[ \frac{1}{N} \sum_{i=1}^{N} (F_{ij} - \bar{F}_j)(S_i^* - \bar{S}^*) \right],$$

and

$$Q_{ff} = \left( q_{f^*, f^*} \right) = \left[ \frac{1}{N} \sum_{i=1}^{N} (F_{ik} - \bar{F}_k)(F_{ij} - \bar{F}_j) \right],$$

respectively. Similarly, denote by $q_{x}$ the $(K + 1) \times 1$ vector whose first element is

$$q_{s^* y} = \frac{1}{N} \sum_{i=1}^{N} (S_i^* - \bar{S}^*)(Y_i - \bar{Y})$$

and whose following elements are

$$q_{f^* y} = \frac{1}{N} \sum_{i=1}^{N} (F_{ij} - \bar{F}_j)(Y_i - \bar{Y}), \quad j = 1, \ldots, K.$$ (A6)

The OLS estimate of $\beta_s$ is given by the first element of $(K + 1)$ vector

$$Q_{xx}^{-1} q_x = \frac{1}{\det(Q_{xx})} \text{adj}(Q_{xx}) q_x,$$ (A7)

where

$$\text{adj}(Q_{xx}) = \begin{pmatrix} C_{s^* s^*} & C_{f^* s^*} & \ldots & C_{f^* f^*} \\ C_{f^* s^*} & C_{f^* f_1} & \ldots & C_{f^* f_K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{f^* f^*} & C_{f^* f_K} & \ldots & C_{f^* f_{f_K}} \end{pmatrix}$$ (A8)

is the transpose of the matrix of co-factors of $Q_{xx}$. Thus, $C_{s^* s^*} = \det(Q_{ff})$ and, for example, $C_{f^* s^*}$ is $(-1)$ times the determinant of the matrix obtained by deleting the first column and the fourth row of $Q_{xx}$. Accordingly,

$$\text{plim} \hat{\beta}_{S,F} = \frac{\text{plim}(q_{s^* y}) \cdot \text{plim}(C_{s^* s^*}) + \sum_{j=1}^{K} \text{plim}(q_{f^* y}) \cdot \text{plim}(C_{f^* s^*})}{\text{plim} \det(Q_{xx})}. $$ (A9)

To simplify (A9), note that by equations (1) to (3),

$$\text{plim}(q_{s^* y}) = \beta_s \text{Var}(S) + \beta_a \text{Cov}(A, S^*) = \beta_s \text{Var}(S) + \beta_a \text{Cov}(A, S),$$

(A10)

and

$$\text{plim}(q_{s^* s^*}) = \text{Var}(S^*) = \text{Var}(S) + \text{Var}(w)$$ (A11)

and

$$\text{plim}(q_{f^* y}) = \beta_s \text{Cov}(S, F_j) + \beta_a \text{Cov}(A, F_j) = \beta_s \text{Cov}(S^*, F_j) + \beta_a \text{Cov}(A, F_j).$$ (A12)
By means of (A10) to (A12), (A9) can be reformulated as

$$\begin{align*}
\text{plim } \hat{\beta}_{S\cdot F} = & \beta_s \frac{[\text{Var}(S^*) - \text{Var}(w)] \text{plim}(C_{s^*s^*}) + \sum_{j=1}^{K} \text{Cov}(S^* F_j) \text{plim}(C_{f_j s^*})}{\text{plim}[\det(Q_{xx})]} \\
& + \beta_a \text{Cov}(A, S) \cdot \text{plim}(C_{s^*s^*}) + \sum_{j=1}^{K} \text{Cov}(A, F_j) \text{plim}(C_{f_j s^*}) \quad \text{plim}[\det(Q_{xx})].
\end{align*}$$

(A13)

We now further simplify the two terms in (A13) in turn. Concerning the first term, note that in accordance with the rules for Laplace expansions of determinants

$$\begin{align*}
\text{Var}(S^*) \text{plim}(C_{s^*s^*}) + \sum_{j=1}^{K} \text{Cov}(S^*, F_j) \text{plim}(C_{f_j s^*}) = \text{plim}[\det(Q_{xx})].
\end{align*}$$

(A14)

Thus, by (A14), (5) and (A7):

$$\begin{align*}
\beta_s \frac{[\text{Var}(S^*) - \text{Var}(w)] \text{plim}(C_{s^*s^*}) + \sum_{j=1}^{K} \text{Cov}(S^*, F_j) \text{plim}(C_{f_j s^*})}{\text{plim}[\det(Q_{xx})]} = \beta_s - \beta_s \lambda \text{Var}(S^*) \text{plim}(Q_{s^*s^*}^{-1}),
\end{align*}$$

(A15)

where $Q_{s^*s^*}^{-1}$ denotes the first element in the first row of $Q_{xx}^{-1}$, i.e.,

$$Q_{s^*s^*}^{-1} = C_{s^*s^*}/\text{det}(Q_{xx}) = \text{det}(Q_{ff})/\text{det}(Q_{xx}).$$

(A16)

It remains to show that $\text{Var}(S^*) \text{plim}(Q_{s^*s^*}^{-1}) = (1 - R_{S^* F}^2)^{-1}$. Using (A14) and the rules for the plim operator we get:

$$\begin{align*}
\text{Var}(S^*) \text{plim}(Q_{s^*s^*}^{-1}) &= \left[1 - \frac{\sum_{j=1}^{K} \text{Cov}(S^*, F_j) \text{plim} \left[ - \frac{C_{f_j s^*}}{\text{det}(Q_{ff})} \right]}{\text{Var}(S^*)} \right]^{-1}.
\end{align*}$$

(A17)

The (asymptotic) $R_{S^* F}^2$ can be written:

$$R_{S^* F}^2 = \frac{\sum_{k=1}^{K} \text{Cov}(S^*, F_j) \text{plim } \hat{\alpha}_j}{\text{Var}(S^*)},$$

(A18)

where $\hat{\alpha}_j$ denotes the OLS estimate of the $j$th slope coefficient in the regression of $S^*$ on $F$; cf. Maddala (1977, p. 107). The final step then amounts to demonstrating that $[-C_{f_j s^*}/\text{det}(Q_{ff})] = \hat{\alpha}_j$. To this end, write the minor of the element $q_{s^*f_j}$ in $Q_{xx}$ as $\text{det}(M_{s^*f_j})$ and denote by $\Psi_j$ the matrix obtained by replacing the $j$th column of $Q_{ff}$
by the column vector \( q_{fs^*} \). Then:

\[
- \frac{C_{fs^*} \text{ det}(Q_{ff})}{\text{ det}(Q_{ff})} = \frac{-[(−1)^{(j+1)\text{ det}(M_{s^*f_j})}}}{\text{ det}(Q_{ff})} = \frac{-[(−1)^{(j+1)\text{ det}(\Psi_j)}]}{\text{ det}(Q_{ff})} = \frac{\text{ det}(\Psi_j)}{\text{ det}(Q_{ff})} = \hat{\alpha}_j. \tag{A19}
\]

The first equality follows from the definition of the co-factor \( C_{fs^*} \). The second equality is due to the fact that \( \Psi_j \) can be obtained by \((j−1)\) interchanges of the columns in \( M_{s^*f_j} \), each of which results in the associated determinant being multiplied by \((-1)\). The third equality follows because \((-1)^2(j+1) = 1\), for all \( j \). The final equality is an application of Cramer’s rule to the system \( Q_{ff} \hat{\alpha} = q_{fs^*} \). Substituting (A19) in (A17), using (A18) and inserting the result in (A15) we get the two first terms in Proposition 2.

To rewrite the second term in (A13) first note that, by (2), (5) and (6):

\[
\frac{\text{ Cov}(A, S)}{\text{ Var}(S^*)} = \hat{\beta}_{AS} (1 − \lambda). \tag{A20}
\]

Next, use (A16), and the equality

\[
\text{ Var}(S^*) \text{ plim}(Q_{s^*s^*}^{-1}) = (1 − R_{S^*F}^2)^{-1}, \tag{A21}
\]

implied by (A17) to (A19) to get:

\[
\beta_a \frac{\text{ Cov}(A, S) \cdot \text{ plim}(C_{s^*s^*}) + \sum_{j=1}^{K} \text{ Cov}(A, F_j) \text{ plim}(C_{f_j s^*})}{\text{ plim}[\text{ det}(Q_{xx})]} = \beta_a \hat{\beta}_{AS} \frac{(1 − \lambda)}{1 − R_{S^*F}^2} [1 + \Phi]. \tag{A22}
\]

The variable \( \Phi \) is given by

\[
\Phi = \sum_{j=1}^{K} \frac{\text{ Cov}(A, F_j)}{\text{ Cov}(A, S^*)} \frac{\text{ plim}[C_{f_j s^*}/\text{ det}(Q_{xx})]}{\text{ plim}(Q_{s^*s^*}^{-1})} = - \sum_{j=1}^{K} \frac{\text{ Cov}(A, F_j)}{\text{ Cov}(A, S^*)} \text{ plim} \left[ - \frac{C_{f_j s^*}}{\text{ det}(Q_{ff})} \right], \tag{A23}
\]

where \( \text{ Cov}(A, S) = \text{ Cov}(A, S^*) \) has been used to obtain the first equality. To get the second equality, (A16) has been employed and the sign of \( C_{f_j s^*} \) has been changed, whereupon the whole expression has been multiplied by \(-1\). By (A19), the term within brackets is equal to \( \hat{\alpha}_j \). Finally, some straightforward manipulations yield:

\[
\frac{\text{ Cov}(A, F_j)}{\text{ Cov}(A, S^*)} = \frac{\rho_{AF_j}}{\rho_{AS^*}} \frac{\sqrt{\text{ Var}(F_j)}}{\sqrt{\text{ Var}(S^*)}}. \tag{A24}
\]

Together, (A22) to (A24) yield the last term in Proposition 2. ■
References


