On Omitted Variable Bias and Measurement Error in Returns to Schooling Estimates

Erik Mellander

November 13, 1997

Abstract. Lam and Schoeni (1993) consider a regression where earnings are explained by schooling and ability. They assume that data on ability are lacking and that schooling is measured with error. The estimate obtained by regressing earnings on schooling thus contains both omitted variable bias, which is positive, and measurement error bias, which is negative. Adding a family background variable is claimed to: i) decrease the omitted variable bias towards, but not below, zero and ii) decrease the measurement error bias even further. This note claims that while ii) is true, even in the context of multiple family background variables, i) is in general incorrect. The omitted variable bias may decrease or increase in magnitude as well as change with respect to sign. Conditions are provided under which i) holds. A simulation procedure is suggested that will yield consistent estimates of the total bias and its components, conditional upon values on the true return and the measurement error variance.

1. Introduction

Many practitioners estimating the return to schooling have noted the tendency for the return estimates 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, although this is not the main point in their article. Their analysis is primarily empirical and reports on a study of earnings and returns to education for a large sample of prime-aged Brazilian males. By means of a model of assortative mating they motivate a large number of family background variables, intended as proxies for ability in the earnings regression.

When the family background variables are included, the estimated returns to schooling fall from around 18 percent to approximately 12 percent. In discussing this finding, Lam and Schoeni (henceforth LS) point out that if there is measurement error in the schooling variable then this 6 percent reduction cannot be taken to imply that traditional estimates of the return to schooling (excluding family background variables) are one-third omitted variable bias. Referring to Welch (1975) and

Griliches (1977), they note that inclusion of variables that are correlated with a worker's schooling may increase the measurement error bias as well as reduce the omitted variable bias. The addition of family background variables will thus not necessarily generate estimates of the return to schooling that are closer to the true return than the original estimates.

LS attempt to assess how much of the decrease in the estimated return to education that can be attributed to increased measurement error bias. Their approach can be described as follows. They first provide formulas for the asymptotic bias in the estimated return to schooling, before and after the inclusion of a single family background variable in the earnings regression. In these equations, the total biases are additively decomposed into omitted variable bias and measurement error bias, respectively. LS claim, without proof, that "....under plausible assumptions ...." (op.cit., p. 719) 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.

Secondly, they note that given knowledge about the true return to schooling and the noise-to-signal-ratio, i.e. the variance of the measurement error in schooling divided by the total variance in schooling, their analytical results imply that one can numerically compute the measurement error bias. They use this finding in a simulation analysis to compute how much of the change in the estimated return to schooling that can be attributed to a change in the measurement error bias. The simulations are conducted both with a single family background variable and *several* family background variables. That is to say, in the application they implicitly extend their theoretical conclusions, drawn in the context of a single family background variable, to the case with many family background variables.

The purpose of this note is threefold. The first purpose is to correct an error in LS's theoretical analysis of the effects of including a single family background variable in the earnings regression. In Section 2 it will be demonstrated that, in general, the conclusion i) is incorrect. It is shown, however, that there are conditions under which the claim i) is true. Finally, it is demonstrated that the formula suggested by LS corresponds to a special case of these conditions. The implicit constraint upon which LS's expression for the omitted variable bias is based is shown to be equivalent to assuming that the correlation between schooling and the family background variable disappears completely when one controls for ability.

The second purpose is to extend the theoretical analysis to an arbitrary number (K) of family background variables. This is done in Section 3 where a general asymptotic bias equation is derived. It turns out that with respect to the measurement error bias the result derived in the context of one family background variable holds in the $K-$ variable case, too. The indeterminacy of the omitted variable bias is even larger in the general case than in the case with $K = 1$; the conditions which ascertain that the omitted variable bias is reduced and unchanged in the one family background variable case cannot readily be extended to the case when $K \geq 2$. It is shown, however, that there is a multivariate counterpart to LS' implicit constraint which is quite useful.

The final purpose is to suggest an alternative to the simulation procedure employed by LS to study to what extent a change in the estimated return to schooling, brought about by the inclusion of family background variables, is due to increased measurement error bias. The LS procedure has the undesirable property that the *share* of the change in the estimated return that they attribute to a change in the measurement error bias can exceed 100 percent. It is shown that using the same information as LS do one can avoid this troublesome feature and, moreover, obtain more information about the biases in the returns to schooling estimates. Concluding comments are given in Section 5.

2. The Case with One Family Background Variable

LS's starting point is the following equation, giving the "true" relation between income, Y, schooling, S, and unobserved "ability", A, for the ith individual

$$Y_i = \beta_0 + \beta_s S_i + \beta_a A_i + u_i\,, \qquad \text{where} \qquad \beta_s\;,\, \beta_a > 0, \tag{1}$$

and u_i is a random disturbance with zero mean and constant variance.[1] For simplicity, the individual observations will be treated as random draws from one and the same underlying population. The u_i are thus viewed as realizations of the random variable u, characterized by $E\,(u) = 0$ and $Var\,(u) = \sigma_u^2$.[2] In addition, they assume that the schooling variable is measured with error, such that observed schooling, S^*, can be

[1] This is LS's equation (1), except that the subindex h for "husband" has been suppressed. Also, LS do not explicitly state the positivity constraints on β_s and β_a in connection with their equation (1). They consistently use these restrictions in their discussion about omitted variable bias and measurement error bias, however.

[2] The individual observations could equivalently be regarded as corresponding to independent but identical distributions. The disadvantage of this formulation is that the stochastic assumption has to be written in the following, lengthier, way

$$E\,(u_i) = 0, \quad E\,(u_i u_j) = \begin{cases} \sigma_u^2 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \quad \text{for all } i.$$

expressed according to

$$S_i^* = S_i + w_i \ , \tag{2}$$

where w represents pure measurement error uncorrelated with S, i.e. $E(w) = 0$, $Var(w) = \sigma_w^2$, and $Cov(S, w) = 0$. Finally, LS implicitly take w to be uncorrelated with A and u, as well, and both u and w to be uncorrelated with the family background variable, F. Collecting all these zero covariances we have

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

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[3]

$$\text{plim}\, \hat{\beta}_S = \beta_s - \beta_s \lambda + \beta_a \hat{\beta}_{AS} (1 - \lambda) \ , \tag{4}$$

where λ is the noise-to-signal ratio, i.e.

$$\lambda \equiv \frac{Var(w)}{Var(S^*)} \ , \qquad 0 \leq \lambda < 1, \tag{5}$$

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

$$\hat{\beta}_{AS} \equiv \frac{Cov(A, S)}{Var(S)} \ , \qquad \hat{\beta}_{AS} > 0. \tag{6}$$

The second term on the RHS of (4) is the measurement error bias and the third term is 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. Since we are assuming, like LS, that schooling and ability are positively correlated so that $\hat{\beta}_{AS} > 0$ the omitted variable bias is positive. It should be noted that, in general, one cannot rule out the possibility that the measurement error bias dominates the omitted variable bias, in which case the total bias is negative.

Given (4), LS consider how the probability limit of the estimate $\hat{\beta}_S$ 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 to Proposition 1 are then given. The last of these provides an interpretation between the general result in Proposition 1 and the equation suggested by LS.

[3] Equation (4) is equal to LS's equation (7), which they give without proof. Since the result can be obtained as a special case of Proposition 1 below the proof is omitted here as well.

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

$$\operatorname{plim} \hat{\beta}_{S \cdot F} = \beta_s - \beta_s \frac{\lambda}{1 - R^2_{S^* F}} + \beta_a \hat{\beta}_{AS} (1 - \lambda) \left(1 - \theta \cdot \rho^2_{AF \cdot S^*}\right) \tag{7}$$

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

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

while θ is defined according to

$$\theta = \frac{(\rho_{S^* F} / \rho_{AS^*}) - \rho_{AS^*} \cdot \rho_{S^* F}}{\rho_{AF} - \rho_{AS^*} \cdot \rho_{S^* F}} ; \qquad \rho_{AF} - \rho_{AS^*} \cdot \rho_{S^* F} \neq 0 ,$$

where $\rho_{S^* F}$, ρ_{AS^*} , and ρ_{AF} denote bivariate correlations.

Proof. See Appendix.

The equation provided by LS [eq. (8) in their paper] is

$$\operatorname{plim} \hat{\beta}_{S \cdot F} = \beta_s - \beta_s \frac{\lambda}{1 - R^2_{S^* \cdot F}} + \beta_a \hat{\beta}_{AS} (1 - \lambda) \left(1 - \rho^2_{AF \cdot S^*}\right) .$$

This equation differs from equation (7) with respect to the final term, i.e. the expression for the omitted variable bias. More specifically, the last parenthesis reads $(1 - \rho^2_{AF \cdot S^*})$ instead of $(1 - \theta \cdot \rho^2_{AF \cdot S^*})$ in (7). Since $\rho^2_{AF \cdot S^*} \in [0, 1]$ by construction, and $\rho^2_{AF \cdot S^*} \in]0, 1[$ by assumption, LS's formulation implies the omitted variable bias invariably decreases towards zero upon the inclusion of a family background variable. Corollary 2 shows, however, that the omitted variable bias may well increase, thus driving the estimate of β_s upward, rather than downward. Furthermore, Corollary 2 demonstrates that the inclusion of the family background variable may change the sign of the omitted variable bias. This possibility was, for obvious reasons, overlooked by LS.

Corollary 2. *If schooling and the family background variable are correlated, i.e. if $R^2_{S*F} > 0$, then the inclusion of the family background variable unambiguously increases the measurement error bias, compared to when no family background variable is included. If $R^2_{S*F} = 0$ the measurement error bias will be unchanged. The effect on the omitted variable bias cannot be determined a priori with respect either to magnitude or to sign. 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, $R^2_{S*F} \in [0,1]$ and, by assumption, $R^2_{S*F} < 1$. The second part is easily established by means of examples; cf. Table 1.

Table 1: The effect on omitted variable bias from including a family background variable in the earnings equation: four examples.

$(\rho_{S^*F},\ \rho_{AF},\ \rho_{AS^*})$	θ	$(1-\theta\cdot\rho^2_{AF\cdot S^*})$	Omitted variable bias
(0.4, 0.3, 0.5)	6.000	0.905	decreases
(0.5 , 0.2, 0.6)	−5.333	1.111	increases
(0.7, 0.7, 0.4)	3.500	−0.441	changes sign, decreases in abs. value
(0.7 , 0.8, 0.3)	3.599	−1.699	changes sign, increases in abs. value

The examples in Table 1 clearly demonstrate that virtually anything can happen to the omitted variable bias when the family background variable is included in the earnings equation. Finally, as shown in the table, the effects on the omitted variable bias are manifested in the factor $(1-\theta\cdot\rho^2_{AF\cdot S^*})$; they are independent of the factor $(1-\lambda)$, i.e. the extent of measurement error. Q.E.D.

It is rather difficult to argue that the example given by row 2 in Table 1 is implausible. Yet it has the effect of increasing the omitted variable bias, thus driving the estimate of β_s upward, contrary to the claim in LS. Also, LS did not consider the possibility that the change in the omitted variable bias may lower the estimate of β_s through an alteration in sign, from positive to negative. As indicated by rows 3 and 4 in Table 1, examples of such cases are not hard to construct.[4]

[4]In studying Table 1, the careful reader might ask if there isn't a connection between the three bivariate correlations in the first column and, if so, has this connection been taken into consideration? The answer is yes, on both questions. The fact that $\rho^2_{AF\cdot S^*} \in [0,1]$ does indeed impose constraints on its bivariate components. The latter have accordingly been chosen such that they are consistent with the property $\rho^2_{AF\cdot S^*} \in [0,1]$.

In Corollary 3 we discuss a special case of the general situation considered in Corollary 2. A constraint on θ is provided which ascertains that the omitted variable bias stays positive *and* is reduced towards zero, i.e. the situation exemplified by row 1 in Table 1. A condition is also given which is necessary, but not sufficient, for this constraint to be satisfied.

Corollary 3. *If, and only if,* $0 < \theta \leq 1/\rho^2_{AF \cdot S^*}$ *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* $\operatorname{sign}(\rho_{AF}) = \operatorname{sign}(\rho_{S^*F})$.

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 $(1-\theta \cdot \rho^2_{AF \cdot S^*})$ where $\rho^2_{AF \cdot S^*} \in \,]0,1]$. For values on θ above $\rho^2_{AF \cdot S^*}$ 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 biases increases.

To prove the necessary condition, first consider the case when $\rho_{S^*F} > 0$. In this case the numerator of θ is unambiguously positive; cf. the definition of θ in Proposition 1 and remember that $\rho_{AS^*} = \rho_{AS} \in \,]0,1[$, by assumption. A necessary requirement for θ to be positive, which in turn is necessary for θ to belong to $]0, 1/\rho^2_{AF \cdot S^*}]$, is thus that the denominator of θ 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 maintaining the omitted variable bias positive and decreasing it towards zero. The case $\rho_{S^*F} = 0$ can be disregarded because it implies $\theta = 0$. Putting the results for the cases $\rho_{S^*F} > 0$ and $\rho_{S^*F} < 0$ together one obtains the necessary and condition stated in the corollary. Q.E.D.

In Corollary 4 we proceed to a special case of the special case characterized in Corollary 3, namely when $\theta = 1$, the constraint implicitly imposed by LS. Corollary 4 provides an interpretation of this constraint, in terms of the correlation between schooling and family background, conditional on ability.

Corollary 4. *If the correlation between schooling and family background is equal to zero when ability is controlled for, i.e. if* $\rho_{S^*F \cdot A} = 0$, *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,

$$\rho_{S^*F\cdot A} = \frac{\rho_{S^*F} - \rho_{AS^*} \cdot \rho_{AF}}{\sqrt{1-\rho_{AS^*}^2}\sqrt{1-\rho_{AF}^2}} \,. \tag{8}$$

Accordingly, if $\rho_{S^*F\cdot A} = 0$ then $(\rho_{S^*F} / \rho_{AS^*}) = \rho_{AF}$ and, consequently, $\theta = 1$. Given that $\theta = 1$, the second part of the corollary follows immediately from Corollary 2. Q.E.D.

In summary, the above results show that LS draw too strong conclusions about how the probability limit of the estimated return to schooling is affected when a family background variable is used as a proxy for "ability" and the schooling variable is subject to measurement error. They claim that the positive omitted variable bias and the negative measurement error bias, which together make up the total bias, are unchanged with respect to sign but decrease and increase in magnitude, respectively. As shown by Corollary 2 the claim is correct only with respect to the measurement error bias. The omitted variable bias may change sign and/or increase in magnitude. Accordingly, contrary to what LS maintain the inclusion of the family background variable can increase the probability limit of the estimated return to schooling.

3. The General Case

In this section the number of family background variables will be taken to be equal to K $(K \geq 1)$. The K-variable counterpart to Proposition 1 is given by the following proposition.

Proposition 5. *Given (1), (2), and (3), OLS regression of Y on S^* and a $K \times 1$ vector $\mathbf{F}$ of family background variables yields an estimate of β_s whose probability limit is given by*

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

where λ 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 $\mathbf{F}$.

Proof. See Appendix.

There are several features of Proposition 2 that are worth noting. The first and most important is that the result for the measurement error bias in the case with one family background variable extends to the case with an arbitrary number of family background variables. Inclusion of family background variables as proxies for ability in a wage equation where schooling is measured with error will always increase the negative measurement error bias, thus driving the estimate of β_s downward.

The second interesting property is that, just like the measurement error bias, the omitted variable bias is inversely related to $1 - R^2_{S^*\mathbf{F}}$. That is to say, 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 the case when family background variables are disregarded. However, this tendency is balanced by the sum within brackets. For example, if the coefficients in the regression of S^* on $\mathbf{F}$ are all positive and if ability is negatively correlated with the family background variables then the sum will be positive, as $\rho_{AS^*} > 0$ by assumption. This will create a downward pressure on the omitted variable bias. In general, it is impossible to say anything about the relative weights of these opposing forces.

To illustrate how difficult it is to say anything a priori about how the omitted variable bias is affected in the general case, it is instructive to consider the case $K = 2$. This is done in Example 1, below. The example also enables a simple, albeit non-stringent, demonstration of the equivalence between Proposition 1 and Proposition 2 when $K = 1$.

Example 6. *The omitted variable bias for $K = 2$.*

By means of standard results, the probability limits of the coefficients in the regression of S^* on F_1 and F_2 can be expressed as

$$\text{plim}\left(\hat{\alpha}_1\right) = \frac{\left(\rho_{S^*F_1} - \rho_{F_1F_2} \cdot \rho_{S^*F_2}\right)}{\left(1 - \rho^2_{F_1F_2}\right)} \frac{\sqrt{Var\left(S^*\right)}}{\sqrt{Var\left(F_1\right)}}$$

and

$$\text{plim}\left(\hat{\alpha}_2\right) = \frac{\left(\rho_{S^*F_2} - \rho_{F_1F_2} \cdot \rho_{S^*F_1}\right)}{\left(1 - \rho^2_{F_1F_2}\right)} \frac{\sqrt{Var\left(S^*\right)}}{\sqrt{Var\left(F_2\right)}}.$$

By Proposition 2, the omitted variable bias thus equals

$$\beta_a \hat{\beta}_{AS} \frac{(1-\lambda)}{1 - R^2_{S^*\mathbf{F}}} \left\{ 1 - \left[\frac{\rho_{AF_1}\left(\rho_{S^*F_1} - \rho_{F_1F_2} \cdot \rho_{S^*F_2}\right) + \rho_{AF_2}\left(\rho_{S^*F_2} - \rho_{F_1F_2} \cdot \rho_{S^*F_1}\right)}{\rho_{AS^*} \cdot \left(1 - \rho^2_{F_1F_2}\right)} \right] \right\}.$$

Concentrating on the ratio within brackets we see that the denominator is unambiguously positive, as $\rho_{AS^*} \in]0,1[$, by assumption. A necessary, but not sufficient, *requirement to create a downward pressure on the omitted variable bias is thus that the numerator is positive, too. However, the sign of the numerator depends to a large extent on the signs and relative magnitudes of* ρ_{AF_1} *and* ρ_{AF_2}*, both of which are unknown and can lie anywhere in the closed interval [-1,1].*

Example 1 enables us to check that Proposition 1 and Proposition 2 yield the same results for $K = 1$. To this end, set $\rho_{F_1F_2} = \rho_{S^*F_2} = \rho_{AF_2} = 0$ and remember that, for $K = 1$, $R^2_{S^*\mathbf{F}}$ is equal to $\rho^2_{S^*F_1}$. This reduces the expression for the omitted variable bias in Example 1 to

$$\beta_a \hat{\beta}_{AS} (1 - \lambda) \times \frac{1}{1 - \rho^2_{S^*F_1}} \left(1 - \frac{\rho_{AF_1} \cdot \rho_{S^*F_1}}{\rho_{AS^*}}\right).$$

By means of Table 1 it can be seen that numerical evaluation of the factor after " $\times$ " yields the same result as evaluation of $(1 - \theta \cdot \rho^2_{AF \cdot S^*})$ which demonstrates that (for the examples in the table) the two propositions are equivalent.

We next give a corollary which is a multivariate extension of Corollary 3.

Corollary 7. *For* $K \geq 2$ *assume that the correlations between schooling and all the family background variables are zero when ability is controlled for, i.e.* $\rho_{S^*F_j \cdot A} = 0$ *for* $j = 1, ..., K$*. Then*

$$L \equiv \sum_{j=1}^{K} \rho_{S^*F_j} \cdot \frac{\sqrt{Var(F_j)}}{\sqrt{Var(S^*)}} \operatorname{plim}(\hat{\alpha}_j) \begin{cases} > \ 0 & \textit{is a necessary, but not sufficient,} \\ \geq \ R^2_{S^*\mathbf{F}} & \textit{is a sufficient, but not necessary,} \end{cases}$$

condition for the omitted variable bias to decrease when two or more family background variables are included in the earnings equation. The omitted variable bias may decrease below zero under both the necessary and the sufficient conditions. However, given either of these conditions, $\left(1/\sqrt{L}\right) > \rho_{AS^*}$ *is both necessary and sufficient for the omitted variable bias to decrease towards but not below zero.*

Proof. If $\rho_{S^*F_j \cdot A} = 0$ for $j = 1, ..., K$ then $\rho_{AF_j} = \rho_{S^*F_j} / \rho_{AS^*}$, by (8). Thus

$$-\sum_{j=1}^{K} \frac{\rho_{AF_j}}{\rho_{AS^*}} \frac{\sqrt{Var(F_j)}}{\sqrt{Var(S^*)}} \operatorname{plim}(\hat{\alpha}_j) = -\frac{1}{\rho^2_{AS^*}} \sum_{j=1}^{K} \rho_{S^*F_j} \cdot \frac{\sqrt{Var(F_j)}}{\sqrt{Var(S^*)}} \operatorname{plim}(\hat{\alpha}_j).$$

Using the definition of L and Proposition 2 it is clear that the omitted variable bias will only decrease if

$$1 - \frac{1}{\rho^2_{AS^*}} \cdot L < 1 - R^2_{S^*\mathbf{F}} \quad \Leftrightarrow \quad \frac{L}{R^2_{S^*\mathbf{F}}} > \rho^2_{AS^*},$$

from which both the necessary and the sufficient conditions follow immediately because $\rho^2_{AS^*} \in]0,1[$. Given that L is positive, the omitted variable bias will decrease and stay positive only if

$$L < \frac{1}{\rho^2_{AS^*}} \quad \Leftrightarrow \quad \frac{1}{\sqrt{L}} > \rho_{AS^*} .$$

This proves the last claim. Q.E.D.

Corollary 4 shows that the property that Corollary 3 cannot be readily extended to the multivariate case. With the given additional conditions, the omitted variable bias is certain to *decrease*, however. A nice feature is that these additional conditions can easily be checked; it is straightforward to compute L by means of the data that are assumed to be available. Moreover, the corollary provides an upper limit on the unknown correlation between schooling and ability, i.e. ρ_{AS^*}, which ascertains that the omitted variable bias both decreases and stays positive.

Finally, we consider what the results imply for the effects on the total bias. As we have seen, the probability limit of the estimated return to schooling may not necessarily decrease when family background variables are added to the earnings equation.[5] Moreover, as noted in Section 2, it cannot be taken for granted that the total bias is positive before the family background variables are included. In trying to assess what happens to the total bias when the family background variables are added to the regression we thus have four different possibilities to consider, cf. Table 2.[6]

Table 2: The effect on the total bias from including family background variables, as a function of the initial total bias and the change in the estimated return

	Initial total bias > 0	Initial total bias < 0
$\text{plim}\left(\hat{\beta}_{S\cdot\mathbf{F}} - \beta_s\right) < 0$	?	Total bias ↑ (in abs. value)
$\text{plim}\left(\hat{\beta}_{S\cdot\mathbf{F}} - \beta_s\right) > 0$	Total bias ↑	?

The reason for the question marks in the diagonal entries in the table is the possibility of "over-shooting" – i.e. a change in the estimated return in the right

[5] Admittedly, this section contains no proof or example showing that for $K \geq 2$ the omitted variable bias may actually increase so much that $\text{plim}\,\hat{\beta}_{S\cdot\mathbf{F}} > \text{plim}\,\hat{\beta}_S$. In Section 2 it has been shown to be a real possibility in the case $K = 1$, however, and that allowing for several family background variables would eliminate this possibility seems rather far-fetched. Given data on schooling and family background variables one could compute examples similar to those in Table 1.

[6] For simplicity, Table 2 abstracts from the theoretically possible but in practice highly unlikely situations where the initial bias and/or the change in the parameter estimates are exactly zero.

direction but by a too large amount. Thus, the only cases that can be unambiguously characterized are those where the inclusion of family background variables leave the researcher worse off than if he/she had disregarded this information. This a quite discouraging result, of course.

4. Simulating the bias in the estimated returns

What, then, are the implications for the simulation procedure suggested by LS? As the above results show, to compute the *share* of the change in the estimated return that is attributable to changed measurement error is in general not justified. The change in the measurement error may be larger than the change in the estimated return, thus invalidating the share interpretation.

However, if we condition upon β_s and λ, as LS do in their simulations, we can actually generate much more information than they do, without adding any further assumptions. Specifically, consistent estimates can be constructed of both the total bias and the omitted variable bias, before and after the inclusion of the family background variables.

Denote the total bias by TB. For a given value of β_s the consistent estimates of the total biases are given by

$$\widehat{TB}_S\Big|_{\beta_s} = \hat{\beta}_S - \beta_s \tag{9}$$

and

$$\widehat{TB}_{S\cdot\mathbf{F}}\Big|_{\beta_s} = \hat{\beta}_{S\cdot\mathbf{F}} - \beta_s\,. \tag{10}$$

From (9) and (10) it is clear that the change in the estimated returns considered by LS equals the change in the total biases. However, the difference $\widehat{TB}_{S\cdot\mathbf{F}} - \widehat{TB}_S = \hat{\beta}_{S\cdot\mathbf{F}} - \hat{\beta}_S$ does not give any information about whether the family background variable has brought the estimated return closer to the (presumed) true return, β_s.

Next, denote the measurement error bias by MEB. Conditional upon β_s and λ, *exact* measures of the measurement error biases are given by

$$MEB_S|_{\beta_s,\lambda} = \beta_s\lambda \tag{11}$$

and

$$MEB_{S\cdot\mathbf{F}}|_{\beta_s,\lambda} = \beta_s\frac{\lambda}{1-R^2_{S*\mathbf{F}}}\,. \tag{12}$$

Finally, denote the omitted variable bias by OVB. Using (4) and (7), consistent estimates of the omitted variable biases are simply obtained according to

$$\widehat{OVB}_S\Big|_{\beta_s,\lambda} = \widehat{TB}_S\Big|_{\beta_s} - MEB_S|_{\beta_s,\lambda} \tag{13}$$

and

$$\widehat{OVB}_{S\cdot\mathbf{F}}\Big|_{\beta_s,\lambda} = \widehat{TB}_{S\cdot\mathbf{F}}\Big|_{\beta_s} - MEB_{S\cdot\mathbf{F}}|_{\beta_s,\lambda}\ . \tag{14}$$

The formulas (9) – (14) are always applicable, irrespective of how the family background variable affects the omitted variable bias.

5. Concluding comments

Using previous work by Lam and Schoeni (1993) as the starting point, the intention of this note has been to shed further light on how the common practice of using family background variables as proxies for ability affects the bias in returns to schooling estimates, when the measure of schooling is subject to random error. The bias can then be additively, but not independently, decomposed into omitted variable bias, arising because of the lacking ability measure, and measurement error bias, due to the random error in schooling.

Lam and Schoeni (LS) conduct their theoretical analysis for the case with one family background only and implicitly assume that it can be extended to the case with several family background variables. This study has first examined whether LS results hold true in the one variable case and then considered the generalization to the case with an arbitrary number of variables. The results are partly discouraging and partly reassuring.

The discouraging result is that the analysis of how the omitted variable bias is affected seems to be considerably more complicated than claimed by LS. LS maintained that i) the omitted variable bias is always positive and bounded from below by zero and ii) is driven towards its lower bound when family background variables are introduced in the earnings equation. This has been shown to be true only under special conditions in the one variable case and under even more restrictive conditions in the case with several family background variables. In general, whereas the omitted variable bias is positive in the absence of the family background variables, including the family background variables in the wage equation may reduce the omitted variable bias below zero, i.e. alter its sign. Alternatively, the bias may increase, rather than decrease. Moreover, if the bias changes sign it is impossible on a priori grounds to determine whether it is reduced in the sense of becoming smaller in absolute value.

The reassuring result concerns the measurement error bias. LS claimed that this bias in the estimated return to schooling is invariably negative and increases in magnitude with the inclusion of family background variables in the earnings equation. This quite strong claim has been shown to be true not only in the case with one family background variable but in the general case as well. Specifically, the downward pressure on the estimated return increases monotonically with the number of family background variables.

A somewhat surprising finding is that, qualitatively, the interdependence between the two types of biases studied does not matter very much for the results. The existence of measurement error affects the *level* of the omitted variable bias – larger measurement error implies smaller omitted variable bias – but it has nothing to do with whether the omitted variable bias increases or changes sign when the family background variables are introduced. These possibilities are intrinsic to the omitted variable bias and thus do exist even if schooling is measured without error.

Given the results for omitted variable bias and the measurement error bias, what can be said about their sum, i.e. the total bias? It has been shown that a priori very little can be said, indeed. In general, we cannot be certain whether the total bias is negative or positive before the introduction of the family background variables. Moreover, the estimated return to schooling may either decrease or increase. This yields four different outcomes and of these the only ones that can be unambiguously characterized are those where the total bias is larger after the inclusion of the background variables than before. This happens when the initial total bias is positive and the estimated return increases or when the initial bias is negative and the estimated return decreases. In the other cases, i.e. when the initial bias and the change in the estimated return have different signs, the total bias may decrease but it may also increase (in absolute value). The latter case may arise because of "over-shooting", i.e. a change in the right direction but by a too large amount.

Faced with the difficulties to establish analytical results it is natural to turn to simulations. LS employed a simulation procedure to determine the share of the change in the estimated return that can be attributed to increased measurement error bias. The present analysis shows that LS's approach is not justified in general, because the change in the measurement error bias can exceed the change in the total bias. However, it is demonstrated that with minor alterations LS's procedure can still be applied. Conditional on given values on the true return to schooling and the measurement error variance, consistent estimates can be obtained of the levels of the total bias and its components, before and after the introduction of family background variables.

The analysis has been based on a very stylized "true" earnings equation; earnings are explained by schooling and ability. The question thus arises if a richer specification would yield qualitatively different results. Including control variables like age, sex, etc. probably merely adds only algebra and no content. However, if family background were to have not only an indirect effect on earnings, as in the present model, but a direct effect as well the picture might change. Another issue that should be considered in future research is the practical importance of the analytical results derived here. Simulations on the LS data set could perhaps provide an answer to this question.

6. References

Griliches, Zvi, 1977, "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, Vol. 45, pp. 1–22.

Lam, David, and Robert F. Schoeni, 1993, "Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil", *Journal of Political Economy*, Vol. 101, pp. 710–740.

Maddala, G.S., 1977, *Econometrics*, McGraw–Hill, New York.

Welch, Finis, 1975, "Human Capital Theory: Education, Discrimination, and Life Cycles", *American Economic Review Papers and Proceedings*, Vol. 65, pp. 63–73.

A. Appendix: Proofs of Propositions

Proof of Proposition 1. Denote mean sums of squares and cross-products according to

$$q_{s^*s^*} = \frac{1}{N}\sum_{i=1}^{N}\left(S_i^* - \bar{S}^*\right)^2 \qquad \text{and} \qquad q_{s^*f} = \frac{1}{N}\sum_{i=1}^{N}\left(S_i^* - \bar{S}^*\right)\left(F_i - \bar{F}\right) .$$

Then, by standard results for regressions involving two regressors and in accordance with well-known properties of probability limits,

$$\text{plim } \hat{\beta}_{S \cdot F} = \frac{\text{plim}\left(q_{ff}\right)\text{plim}\left(q_{s*y}\right) - \text{plim}\left(q_{s^*f}\right)\text{plim}\left(q_{fy}\right)}{\text{plim}\left(q_{s^*s^*}\right)\text{plim}\left(q_{ff}\right) - \left[\text{plim}\left(q_{s^*f}\right)\right]^2} .$$

To evaluate this expression, first note that

$$\text{plim}\left(q_{ff}\right) = Var\left(F\right) . \tag{A.1}$$

Next, by (2) and (3),

$$\text{plim}\left(q_{s^*s^*}\right) = Var\left(S^*\right) = Var\left(S\right) + Var\left(w\right) . \tag{A.2}$$

$$\text{plim}\left(q_{s^*f}\right) = Cov\left(S^*, F\right) = Cov\left(S, F\right) \tag{A.3}$$

Further, by (1)and (3)

$$\text{plim}\left(q_{fy}\right) = \beta_s Cov\left(S, F\right) + \beta_a Cov\left(A, F\right) = \beta_s Cov\left(S^*, F\right) + \beta_a Cov\left(A, F\right), \tag{A.4}$$

where the last equality follows from (A.3). Finally, by (1) - (3)

$$\text{plim}\left(q_{s*y}\right) = \beta_s Var\left(S\right) + \beta_a Cov\left(A, S^*\right) = \beta_s Var\left(S\right) + \beta_a Cov\left(A, S\right) . \tag{A.5}$$

Collecting results and rearranging we get

$$\begin{aligned} \text{plim } \hat{\beta}_{S \cdot F} &= \frac{\beta_s\left\{Var\left(S\right)Var\left(F\right) - \left[Cov\left(S^*, F\right)\right]^2\right\}}{Var\left(S^*\right)Var\left(F\right) - \left[Cov\left(S^*, F\right)\right]^2} \\ &+ \frac{\beta_a\left[Var\left(F\right)Cov\left(A, S\right) - Cov\left(S^*, F\right)Cov\left(A, F\right)\right]}{Var\left(S^*\right)Var\left(F\right) - \left[Cov\left(S^*, F\right)\right]^2} . \end{aligned} \tag{A.6}$$

To simplify the first term in (A.6), first substitute $[Var(S^*) - Var(w)]$ for $Var(S)$, then divide the numerator and the denominator by $[Var(S^*)Var(F)]$, and, finally, use (5). This yields

$$\frac{\beta_s\left\{Var(S)Var(F) - [Cov(S^*,F)]^2\right\}}{Var(S^*)Var(F) - [Cov(S^*,F)]^2} = \beta_s - \beta_s\frac{\lambda}{1-\rho^2_{S^*F}} \ . \tag{A.7}$$

To rewrite the second term in (A.6) first divide the numerator and the denominator by $[Var(S^*)Var(F)]$ and note that, by (2)

$$\frac{Cov(A,S)}{Var(S^*)} = \hat{\beta}_{AS}(1-\lambda) \ , \tag{A.8}$$

where $\hat{\beta}_{AS}$ and λ are defined by (6) and (5), respectively. This yields

$$\frac{\beta_a[Var(F)Cov(A,S) - Cov(S^*,F)Cov(A,F)]}{Var(S^*)Var(F) - [Cov(S^*,F)]^2} = \beta_a\hat{\beta}_{AS}(1-\lambda)\cdot\zeta \tag{A.9}$$

where

$$\zeta \equiv \frac{1-\rho^2_{S^*F}\frac{Cov(A,F)Var(S^*)}{Cov(S^*,F)\cdot Cov(A,S)}}{1-\rho^2_{S^*F}} \ . \tag{A.10}$$

Using the equality $Cov(A,S) = Cov(A,S^*)$ and rearranging one can rewrite ζ according to

$$\zeta = 1 - \frac{\rho^2_{S^*F}\dfrac{\rho_{AF} - \rho_{S^*F}\cdot\rho_{AS^*}}{\rho_{S^*F}\cdot\rho_{AS^*}}}{1-\rho^2_{S^*F}} \ . \tag{A.11}$$

It remains to prove that the second term on the RHS of (A.11) is equal to the product $\theta\cdot\rho^2_{AF\cdot S^*}$. Multiplication of the numerator and the denominator by $(1-\rho^2_{AS^*})$ yields

$$\begin{aligned}\frac{\rho^2_{S^*F}\dfrac{\rho_{AF} - \rho_{S^*F}\cdot\rho_{AS^*}}{\rho_{S^*F}\cdot\rho_{AS^*}}}{1-\rho^2_{S^*F}} &= \frac{[(\rho_{S^*F}/\rho_{AS^*}) - \rho_{S^*F}\cdot\rho_{AS^*}](\rho_{AF} - \rho_{S^*F}\cdot\rho_{AS^*})}{(1-\rho^2_{S^*F})(1-\rho^2_{AS^*})} \\ &= \theta\cdot\rho^2_{AF\cdot S^*},\end{aligned}$$

where θ and $\rho^2_{AF\cdot S^*}$ are defined in Proposition 1. Now, substitute this equality in (A.11) and, subsequently, (A.11) in (A.9). The resulting expression and (A.7) can then be used in (A.6). Finally, to get (7), note that in the case with only one family background variable $\rho^2_{S^*F} = R^2_{S^*F}$. Q.E.D.

Proof of Proposition 2. Let the $(K+1)$ square matrix $\mathbf{Q_{xx}}$ be defined as

$$\mathbf{Q_{xx}} = \begin{pmatrix} q_{s^*s^*} & \mathbf{q}_{s^*\mathbf{f}} \\ (1 \times 1) & (1 \times K) \\ \mathbf{q}_{\mathbf{f}s^*} & \mathbf{Q_{ff}} \\ (K \times 1) & (K \times K) \end{pmatrix}$$

where

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

and the typical elements of the vector $\mathbf{q}_{\mathbf{f}s^*}$ $\left(= \mathbf{q}'_{s^*\mathbf{f}}\right)$ and the matrix $\mathbf{Q_{ff}}$ are

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

and

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

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

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

and whose following elements are

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

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

$$\mathbf{Q}_{\mathbf{xx}}^{-1} \mathbf{q}_{\mathbf{x}y} = \frac{1}{\det\left(\mathbf{Q_{xx}}\right)} \operatorname{adj}\left(\mathbf{Q_{xx}}\right) \mathbf{q}_{\mathbf{x}y}\ . \tag{A.12}$$

where

$$\operatorname{adj}\left(\mathbf{Q_{xx}}\right) = \begin{pmatrix} C_{s^*s^*} & C_{f_1 s^*} & \cdots & C_{f_K s^*} \\ C_{f_1 s^*} & C_{f_1 f_1} & \cdots & C_{f_1 f_K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{f_K s^*} & C_{f_1 f_K} & \cdots & C_{f_K f_K} \end{pmatrix}$$

is the transpose of the matrix of cofactors of $\mathbf{Q}_{\mathbf{xx}}$. Thus, $C_{s^*s^*} = \det(\mathbf{Q}_{\mathbf{ff}})$ and, e.g., $C_{f_3s^*}$ is (-1) times the determinant of the matrix obtained by deleting the first column and the fourth row of $\mathbf{Q}_{\mathbf{xx}}$. Accordingly,

$$\text{plim}\,\hat{\beta}_{S\cdot\mathbf{F}} = \frac{\text{plim}\,(q_{s^*y})\cdot\text{plim}\,(C_{s^*s^*}) + \sum_{j=1}^{K}\text{plim}\left(q_{f_jy}\right)\text{plim}\left(C_{f_js^*}\right)}{\text{plim}\,[\det(\mathbf{Q}_{\mathbf{xx}})]}.$$

To simplify this expression first use $(A.5)$, (A.2), and $(A.4)$ to get

$$\begin{aligned}\text{plim}\,\hat{\beta}_{S\cdot\mathbf{F}} &= \beta_s\frac{[Var(S^*)-Var(w)]\,\text{plim}(C_{s^*s^*})+\sum_{j=1}^{K}Cov(S^*,F_j)\,\text{plim}\left(C_{f_js^*}\right)}{\text{plim}[\det(\mathbf{Q}_{\mathbf{xx}})]} \\ &\quad+\beta_a\frac{Cov(A,S)\cdot\text{plim}(C_{s^*s^*})+\sum_{j=1}^{K}Cov(A,F_j)\,\text{plim}\left(C_{f_js^*}\right)}{\text{plim}[\det(\mathbf{Q}_{\mathbf{xx}})]}.\end{aligned} \tag{A.13}$$

We now further simplify the two terms in (A.13) in turn.

Concerning the first term in (A.13), note that in accordance with the rules for Laplace expansions of determinants

$$Var\,(S^*)\,\text{plim}\,(C_{s^*s^*}) + \sum_{j=1}^{K}Cov\,(S^*,F_j)\,\text{plim}\left(C_{f_js^*}\right) = \text{plim}\,[\det(\mathbf{Q}_{\mathbf{xx}})]\,. \tag{A.14}$$

Thus, by (A.14), (5), and (A.12)

$$\beta_s\frac{[Var(S^*)-Var(w)]\,\text{plim}(C_{s^*s^*})+\sum_{j=1}^{K}Cov(S^*,F_j)\,\text{plim}\left(C_{f_js^*}\right)}{\text{plim}[\det(\mathbf{Q}_{\mathbf{xx}})]} = \beta_s - \beta_s\lambda Var\,(S^*)\,\text{plim}\left(Q^{-1}_{s^*s^*}\right) \tag{A.15}$$

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

$$Q^{-1}_{s^*s^*} = C_{s^*s^*}/\det(\mathbf{Q}_{\mathbf{xx}}) = \det(\mathbf{Q}_{\mathbf{ff}})\,/\det(\mathbf{Q}_{\mathbf{xx}})\,. \tag{A.16}$$

It remains to show that $Var\,(S^*)\,\text{plim}\left(Q^{-1}_{s^*s^*}\right) = (1-R^2_{S^*\cdot\mathbf{F}})^{-1}$. Using (A.14) and the rules for the plim operator we get

$$Var\,(S^*)\,\text{plim}\left(Q^{-1}_{s^*s^*}\right) = \left[1 - \frac{\sum_{j=1}^{K}Cov\,(S^*,F_j)\,\text{plim}\left[\frac{-C_{f_js^*}}{\det(\mathbf{Q}_{\mathbf{ff}})}\right]}{Var\,(S^*)}\right]^{-1}. \tag{A.17}$$

As the (asymptotic) $R^2_{S^*\cdot\mathbf{F}}$ can be written

$$R^2_{S^*\cdot\mathbf{F}} = \frac{\sum_{j=1}^{K}Cov\,(S^*,F_j)\,\text{plim}\,\hat{\alpha}_j}{Var\,(S^*)}, \tag{A.18}$$

where $\hat{\alpha}_j$ denotes the OLS estimate of the jth slope coefficient in the regression of S^* on $\mathbf{F}$ [cf. Maddala (1977, p. 107)], the final step amounts to demonstrating that $\left[-C_{f_j s^*}/\det\left(\mathbf{Q_{ff}}\right)\right] = \hat{\alpha}_j$. To this end, write the minor of the element $q_{s^* f_j}$ in $\mathbf{Q_{xx}}$ as $\det\left(\mathbf{M}_{s^* f_j}\right)$ and denote by $\mathbf{\Psi}_j$ the matrix obtained by replacing the jth column of $\mathbf{Q_{ff}}$ by the column vector $\mathbf{q}_{\mathbf{f}s^*}$. Then

$$\begin{aligned} \frac{-C_{f_j s^*}}{\det(\mathbf{Q_{ff}})} &= \frac{-\left[(-1)^{(j+1)+1}\det\left(\mathbf{M}_{s^* f_j}\right)\right]}{\det(\mathbf{Q_{ff}})} = \frac{-\left[(-1)^{(j+1)+1}(-1)^{j-1}\det(\mathbf{\Psi}_j)\right]}{\det(\mathbf{Q_{ff}})} \\ &= \frac{\det(\mathbf{\Psi}_j)}{\det(\mathbf{Q_{ff}})} = \hat{\alpha}_j\,. \end{aligned} \tag{A.19}$$

The first equality follows directly from the definition of the cofactor $C_{f_j s^*}$. The second equality is due to the fact that $\mathbf{\Psi}_j$ can be obtained by $(j-1)$ interchanges of the columns in $\mathbf{M}_{s^* f_j}$, each of which results in the associated determinant being multiplied by (-1). The third equality follows because $(-1)^{2(j+1)} = 1\ \forall j$. The final equality is just an application of Cramer's rule to the system $\mathbf{Q_{ff}}\boldsymbol{\alpha} \hat{=} \mathbf{q_{fs^*}}$.

To rewrite the second term in (A.13) first use (A.8), (A.16), and the equality

$$Var\left(S^*\right)\operatorname{plim}\left(Q^{-1}_{s^*s^*}\right) = \left(1 - R^2_{S^*\cdot\mathbf{F}}\right)^{-1}$$

implied by (A.17) – (A.19) to get

$$\beta_a \frac{Cov\left(A,S\right)\cdot\operatorname{plim}\left(C_{s^*s^*}\right) + \sum_{j=1}^{K} Cov\left(A,F_j\right)\operatorname{plim}\left(C_{f_j s^*}\right)}{\operatorname{plim}\left[\det\left(\mathbf{Q_{xx}}\right)\right]} = \beta_a\hat{\beta}_{AS}\frac{(1-\lambda)}{1-R^2_{S^*\mathbf{F}}}\left[1+\Phi\right] \tag{A.20}$$

The variable Φ is given by

$$\Phi = \sum_{j=1}^{K}\frac{Cov\left(A,F_j\right)}{Cov\left(A,S^*\right)}\frac{\operatorname{plim}\left[C_{f_j s^*}/\det\left(\mathbf{Q_{xx}}\right)\right]}{\operatorname{plim}\left(Q^{-1}_{s^*s^*}\right)} = -\sum_{j=1}^{K}\frac{Cov\left(A,F_j\right)}{Cov\left(A,S^*\right)}\operatorname{plim}\left[\frac{-C_{f_j s^*}}{\det\left(\mathbf{Q_{ff}}\right)}\right] \tag{A.21}$$

where $Cov\left(A,S\right) = Cov\left(A,S^*\right)$ has been used to obtain the first equality. To get the second equality, (A.16) has been employed and the sign of $C_{f_j s^*}$ has been changed, whereupon the whole expression has been multiplied by -1. By (A.19), the term within brackets is equal to $\hat{\alpha}_j$. Finally, some straightforward manipulations yield

$$\frac{Cov\left(A,F_j\right)}{Cov\left(A,S^*\right)} = \frac{\rho_{AF_j}}{\rho_{AS^*}}\frac{\sqrt{Var\left(F_j\right)}}{\sqrt{Var\left(S^*\right)}}\,. \tag{A.22}$$

Collecting results we thus have

$$\Phi = -\sum_{j=1}^{K} \frac{\rho_{AF_j}}{\rho_{AS^*}} \frac{\sqrt{Var\left(F_j\right)}}{\sqrt{Var\left(S^*\right)}} \operatorname{plim}\left(\hat{\alpha}_j\right). \tag{A.23}$$

Substitution of (A.23) in (A.20) yields the omitted variable bias term in Proposition 2. This completes the proof. Q.E.D.