DE GRUYTER DOI 10.1515/jem-2012-0008 — J. Econom. Meth. 2015; 4(1): 101–122

Research Article

Christopher R. Bollinger* and Jenny Minier

On the Robustness of Coefficient Estimates to the Inclusion of Proxy Variables

Abstract: This paper considers the use of multiple proxy measures for an unobserved variable and contrasts the approach taken in the measurement error literature to that of the model specification literature. We find that including all available proxy variables in the regression minimizes the bias on coefficients of correctly measured variables in the regression. We derive a set of bounds for all parameters in the model, and compare these results to extreme bounds analysis. Monte Carlo simulations demonstrate the performance of our bounds relative to extreme bounds. We conclude with an empirical example from the cross-country growth literature in which human capital is measured through three proxy variables: literacy rates, and enrollment in primary and secondary school, and show that our approach yields results that contrast sharply with extreme bounds analysis.

Keywords: cross-country growth regressions; econometric bounds; latent variable; measurement error.

JEL Codes: C4; C51; O47.

***Corresponding author: Christopher R. Bollinger,** Department of Economics, University of Kentucky, Lexington, KY 40506, USA, E-mail: crboll@uky.edu
Jenny Minier: Department of Economics, University of Kentucky, Lexington, KY 40506, USA

1 Introduction

We consider estimation of a model with an unobserved explanatory variable, when a set of variables thought to proxy for the regressor is observed. A related problem has played a prominent role in the empirical growth literature, where it is generally interpreted as a model specification issue.[1] We examine the issue in the context of a measurement error model similar to that considered by Bollinger (2003) and Lubotsky and Wittenberg (2006), who provide examples from labor economics. We focus on how the use of these proxy variables affects the estimates of other coefficients in the model, and discuss how best to use combinations of proxy variables in order to minimize the inconsistency on these other coefficients. We derive analytical results which extend and combine those of Lubotsky and Wittenberg (2006) with those of Bollinger (2003), and perform simulations and an empirical example. Although the difference between our approach – seeing this as a measurement issue – and the model specification approach commonly taken in the growth literature is subtle, the approaches yield strikingly different results. When the issue is posed as a measurement issue, which we argue is often realistic, our results are preferable to the model specification approach.

This paper considers estimation of a model such as the following:

$$y_i = \boldsymbol{\alpha}'\mathbf{Z}_{1i} + \beta Z_{2i} + u_i, \tag{1}$$

1 This literature includes the extreme bounds approach of Levine and Renelt (1992), the distributional approach of Sala-i-Martin (1997), and Bayesian Model Averaging as in Sala-i-Martin, Doppelhofer, and Miller (2004) and Durlauf, Kourtellos, and Tan (2008). For additional discussion of the issue of model uncertainty in the macroeconomic literature, see Brock and Durlauf (2001) and Brock, Durlauf, and West (2003).

$E[u_i|\mathbf{Z}_{1i}, Z_{2i}]=0$ and $V(\mathbf{Z}_{1i}, Z_{2i})$ is full rank. The assumption that this is a conditional expectation is not necessary; the linear projection model is a sufficient assumption. The vector $\boldsymbol{\alpha}$ is a $k\times1$ vector of parameters, while $\mathbf{Z}_{1i}$ is a $k\times1$ vector of random variables. The parameter β and its corresponding variable Z_{2i} are both scalar. We assume that the structural model is well-specified; that is, the researcher is interested in the model above, and in particular, estimation of $\boldsymbol{\alpha}$. However, the researcher does not observe the variable Z_{2i}, but only observes a set of variables $\mathbf{X}_i$ (referred to as proxy variables) which are thought to be related to Z_{2i}.

Examples of this situation include the case where y_i is earnings and Z_{1i} is gender or race while Z_{2i} is human capital (see Bollinger 2003 for some results). A second example is the case considered by Lubotsky and Wittenberg (2006) where y_i is consumption and Z_{2i} is permanent income. In our example in Section 4, following the human capital-augmented Solow model, y_i is economic growth, while $\mathbf{Z}_{1i}$ is a vector including initial GDP per capita, investment in physical capital, and population growth, and Z_{2i} is aggregate human capital. The problem in all of these cases is that while conceptually (or theoretically) Z_{2i} exists and plays an important role in the model, it is difficult or impossible to actually measure. Griliches (1974, 976), whose examples include both human capital and permanent income, describes this type of unobservable variable as one that "do[es] not correspond directly to anything that is likely to be measured." What are often available are variables termed "proxy" variables, thought to be correlated with, but not perfectly related to, the underlying theoretical concept. In the case of human capital in the earnings literature, measures such as AFQT scores (see Neal and Johnson 1996; Bollinger 2003) are often used. In the case of permanent income, measures of multiple years of realized income are typically available (Lubotsky and Wittenberg 2006). In the case of measuring human capital in the cross-country growth literature, various measures of school attainment or completion rates are often used: the numerous studies employing enrollment rates include Mankiw, Romer, and Weil (1992), who use secondary enrollment rates, and Sala-i-Martin (1997), who uses primary school enrollment rates. The seminal empirical growth study of Barro (1991) uses both. Many other studies use outcome measures to proxy for the stock, rather than the flow, of human capital, such as literacy rates or the school attainment measures most recently updated in Barro and Lee (2001).

Our approach is related to the model specification literature, including extreme bounds analysis (EBA).[2] We will show, however, that our approach provides very different results. The main difference between our approach and the model specification approach common in the economic growth literature derives from the framing of the problem. We begin with a clear theoretical model stated as a linear equation. The problem stems from the fact that the model includes a variable (e.g., human capital) that is not measured directly. This is also how the issue is often initially presented in the model specification literature: for example, Sala-i-Martin, Doppelhofer, and Miller (2004) state, "The problem faced by empirical economists is that growth theories are not explicit enough about what variables…belong in the "true" regression." Thus the issue, as we frame it, is one of having multiple proxy variables for a variable in a well-specified model. While the model specification literature interprets this as choosing one model from many possible models, we argue that the issue is one of *measurement* rather than model specification. While there may be situations where the model of interest is not well specified, our approach is appropriate in many contexts, including many where the model specification literature has been employed.

An important caveat to this paper, however, is that the "proxy" (observed) variables in the model specification literature do not always proxy for only one unobserved variable. For example, Levine and Renelt (1992) describe their seven variables corresponding to our proxies as "fiscal, trade, monetary, uncertainty, and political-instability indicators." Although we do not extend our results in this paper to the case with multiple unobserved variables, Bollinger (2003) presents results for the case with k unobserved regressors and k proxies. Wittenberg (2007) provides a test examining whether the proxies represent one or more

2 Note that the model specification approach also encompasses an informal but very common practice in many fields of economics: asserting that results are "robust" to alternative measures of explanatory variables, generally after simply substituting one for another in the analysis.

latent variables.[3] Our generalized proxy bounds approach, however, starts from the premise that the correct specification cannot be estimated, since one of the variables ("human capital") is not observed; rather, the researcher has a set of proxies for the unobserved variable. In this paper, we develop a procedure for bounding coefficient estimates in the presence of proxies for an unobserved variable.

The two approaches yield very different results: in fact, we show that the bounds from the two approaches do not overlap, although they have one bound in common. Through our analytic results and simulations, we show that our generalized proxy bounds tend to outperform EBA, particularly in the most relevant cases. We then provide an empirical example from the growth literature.

The analytic sections draws heavily from Lubotsky and Wittenberg (2006) and Bollinger (2003) to arrive at three results. First, we extend the result of Lubotsky and Wittenberg (2006) (hereafter L-W) and focus on the inconsistency in estimation of $\boldsymbol{\alpha}$. These results are also related to Bollinger (2003), which considers the case where $\mathbf{X}_i$ is of the same dimension as Z_{2i}. We show that the minimum inconsistent estimates of the parameters $\boldsymbol{\alpha}$ can be achieved from the results of the regression that includes all proxy variables. (L-W showed the minimum inconsistency on β occurred with all proxy variables included, but did not explicitly examine the inconsistency on $\boldsymbol{\alpha}$.) L-W show that the minimum inconsistency estimate of β is only estimable if there is an element of $\mathbf{X}_i$ which has the same scale as Z_{2i}. This will be discussed more formally below, but essentially requires that one of the proxies be a classical additive white noise measurement error process. Our results for β require the same assumption. However, we show that the minimum inconsistent estimates for $\boldsymbol{\alpha}$ do not require this assumption. We further show that, as in Bollinger (2003), minimum inconsistent estimates of β/ρ_1 (where ρ_1 is a scaling parameter) are always estimable. As discussed in Bollinger (2003), this is simply a normalization of the scale of the unobserved variable Z_{2i}. Finally, we extend the results of Bollinger (2003) to derive a set of bounds for the parameters $(\boldsymbol{\alpha}, \beta)$. The minimum inconsistent estimates of $(\boldsymbol{\alpha}, \beta)$ form one bound, and a reverse regression (like that used in Klepper and Leamer 1984; Bollinger 2003) provides the other bound.

We compare these results to extreme bounds analysis, as an illustrative example of the model specification approach. What is important in this comparison is that regressions that include only a subset of the proxy variables have at least as large an inconsistency as the regression that includes all proxy variables, and the inconsistency is of the same sign. We show that the extreme bounds approach will not provide bounds that include the parameters $\boldsymbol{\alpha}$ (or β) in the model above, while our results do provide such bounds. In addition to demonstrating that the prescription of Lubotsky and Wittenberg (2006), to include all proxy variables in the regression, extends to regressions with additional regressors, the bounds results are important in that they establish both the direction and the potential magnitude of the inconsistency resulting from the use of proxy variables. Typically researchers include proxy variables with little understanding of the potential inconsistency on other parameters in the linear model. This paper establishes those results, links them to omitted variable inconsistency, and demonstrates that the direction of the inconsistency can be identified.

Next, we demonstrate our analytic results using Monte Carlo evidence and comparing these generalized proxy bounds to extreme bounds analysis. We find that in many cases, the extreme bounds analysis provides the wrong conclusion, while the proxy bounds yield the correct conclusion. The only cases where extreme bounds analysis appears to work well are cases where the unmeasured variable Z_{2i} is uncorrelated with the other variables in the model, and hence all estimates from extreme bounds analysis are consistent for the parameter of interest.

We conclude with an empirical example from the cross-country growth literature in which human capital is measured through three proxy variables: literacy rates, and enrollment in primary and secondary school. Following the growth literature, we use our bounds as a robustness test, and compare them to extreme bounds results. We find that the coefficient estimate on initial income is "robust" (i.e., consistently negative and statistically significant), as previous extreme bound analyses have concluded. However, in contrast to

3 Our intuition is that EBA and other model specification approaches would also have problems in the case with multiple unobserved variables, although bounds derived under EBA *may* contain the true estimate, depending on the extent to which bias from one unobserved variable offsets bias from another.

previous results, we find that the coefficient estimate on investment cannot be distinguished from zero, while that on population growth is robustly statistically different from zero.

2 Analytic Results

This section proceeds as follows. First, we extend the results of L-W and establish that there is a linear combination of proxy variables which simultaneously minimizes the inconsistency on all coefficients in the model. However, forming this linear combination requires knowledge of unidentified variances. Second, as with L-W, the OLS regression that includes all proxy variables provides coefficients on the observed variables that are equal to the coefficients that would be achieved by use of the inconsistency-minimizing linear combination of proxy variables. Following L-W, we also show that an available linear combination of the coefficients on the proxy variables achieves the minimum inconsistent estimate of the ratio of β/ρ_1 (where ρ_1 is a scaling parameter defined below that L-W assume to be 1), and that from the OLS results, the optimal linear combination of the proxy variables can be constructed. From this result we show that bounds on the coefficients can be achieved by applying results from Bollinger (2003).

Equation 1 provides the structural model of interest. We assume that the researcher's primary interest is estimation of the parameters $\boldsymbol{\alpha}$, the coefficients on $\mathbf{Z}_{1i}$. The relationship between the observed proxies and the variable of interest is

$$\mathbf{X}_i = \boldsymbol{\rho} Z_{2i} + \boldsymbol{\varepsilon}_i. \tag{2}$$

These are also sometimes referred to as multiple indicators (for example Goldberger and Jöreskog 1975 or Wooldridge 2010). The $l\times 1$ vector $\mathbf{X}_i$ contains multiple measures of the scalar variable Z_{2i}. The nuisance parameter $\boldsymbol{\rho}$ is a $l\times 1$ vector of unknown constants. We assume that

$$A1: Cov(\varepsilon_i, Z_{2i}) = 0$$
$$A2: Cov(\varepsilon_i, \mathbf{Z}_{1i}) = 0$$
$$A3: Cov(\varepsilon_i, u_i) = 0$$
$$A4: V(\varepsilon_i) = \Sigma \text{ is positive definite.}$$

The relationship expressed in equation 2 and assumption A1 is relatively benign and implies only that the linear projection of $\mathbf{X}_i$ on Z_i exists provided that both $\mathbf{X}_i$ and Z_i have finite first and second moments. Assumptions A2 and A3 do impose some important restrictions on the data generating process. Specifically, they state that, except as measures of Z_{2i}, there is no additional information about Y_y contained in these proxy variables and that the correlation between the proxy variables and the observed variables $\mathbf{Z}_{1i}$ is only through Z_{2i}. Since $V(\boldsymbol{\varepsilon}_i)$ is not diagonal, an instrumental variables approach is not available. We assume that the researcher observes (y_i, $\mathbf{Z}_{1i}$, $\mathbf{X}_i$).

Like L-W, we begin by considering the problem of choosing a linear combination of $\mathbf{X}_i$ to minimize the inconsistency on the resulting coefficient. That is, L-W are interested in the regression of y_i on $X_i^\delta = \boldsymbol{\delta}'\mathbf{X}_i$: the problem is to choose a $l\times 1$ vector $\boldsymbol{\delta}$ to minimize the inconsistency in estimation of β. (In their case, $\alpha=0$ and there are no other regressors.) We follow the same approach here, but include additional regressors. As noted in L-W,

$$X_i^\delta = \boldsymbol{\delta}'\mathbf{X}_i = \boldsymbol{\delta}'\boldsymbol{\rho} Z_{2i} + \boldsymbol{\delta}'\boldsymbol{\varepsilon}_i.$$

We can write this as

$$X_i^\delta = \gamma^\delta Z_{2i} + e_i^\delta, \tag{3}$$

which is a general measurement error specification as considered by Bollinger (2003). If the scalar $\gamma = \boldsymbol{\delta}'\boldsymbol{\rho} = 1$, then classical errors-in-variables results reveal that measurement error inconsistency from the regression of

y_i on $\mathbf{Z}_{1i}$ and X_i^δ is minimized when $V(e_i^\delta)$ is minimized. L-W examine this case. We extend this result beyond the results of L-W in two dimensions. First, we clarify the scaling issue with respect to the choice of $\gamma=\delta'\rho$. Second, we derive expressions for the inconsistency on $\boldsymbol{\alpha}$ and show that the result of L-W also minimizes the inconsistency on all regressors in the model.

Proposition 1 *Define $\gamma=\boldsymbol{\delta}'\boldsymbol{\rho}>0$ and $\theta=\beta/\gamma$. Let $(\mathbf{a}^\delta, t^\delta)$ be the coefficients from the regression of y_i on $\mathbf{Z}_{1i}, X_i^{\delta(\gamma)}$ for any δ and a given value of γ. Then, $\boldsymbol{\delta}=\gamma(\Sigma^{-1}\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho})^{-1}$ solves both $\min_\delta (t^\delta-\theta)^2$ and $\min_\delta \|a^\delta-\alpha\|$ for $\gamma=\boldsymbol{\delta}'\boldsymbol{\rho}$.*

Proof. By Lemma 1 (see Appendix A) we can write

$$(\mathbf{a}^\delta-\alpha)=(V_1-\mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\frac{V_2-\mathbf{C}'V_1^{-1}\mathbf{C}}{V_2}\right)\Omega\beta \tag{4}$$

and

$$(t^\delta-\theta)=-\Omega\theta \tag{5}$$

where $V_1=V(\mathbf{Z}_{1i})$, $V_2=V(Z_{2i})$, $\mathbf{C}=Cov(\mathbf{Z}_{1i}, Z_{2i})$, and

$$\Omega=\frac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})} \tag{6}$$

The common term Ω is a scalar. When Σ is positive definite, Ω is positive (see Lemma 2); hence the inconsistency on any coefficient (as measured by any norm) is minimized when this term is minimized. It is trivial to show that Ω is increasing in the term $(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$. Hence, choosing the inconsistency-minimizing $\boldsymbol{\delta}$ is equivalent to solving

$$\min_\delta(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}) \text{ subject to } \delta'\rho=\gamma,$$

the solution to which is $\boldsymbol{\delta}^*=\gamma(\Sigma^{-1}\boldsymbol{\rho})(\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho})^{-1}$. (See Appendix A.) ■

L-W derive the inconsistency expression for t^δ when $\delta'\rho=\gamma=1$, although they also discuss the more general case. The inconsistency minimization, relative to $\boldsymbol{\alpha}$ and θ, holds regardless of γ. When $\gamma=1$, the relationship between X_i^δ and Z_{2i} is a classical measurement error relationship. This provides a great deal of the intuition to these results. As is well known, the inconsistency from classical measurement error is determined by the variance of the error term, which in this case is $\mathrm{Var}(e_i^\delta)=\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}$. Hence the goal in combining X's is to choose a linear combination that minimizes the error variance. In the case where γ is some arbitrary constant, the intuition for the result can be found in Bollinger (2003), who shows that the model can be rescaled in terms of (β/γ) to be a classical measurement error model, and again, the inconsistency is minimized by choosing $\boldsymbol{\delta}$ to minimize the variance of e_i.

We next turn to the issue of the scaling γ. Unlike Bollinger (2003), γ is a choice variable (in the sense of the problem of choosing a linear combination of X). L-W focus on the case where $\gamma=1$ for a number of reasons. Their fundamentally important result shows a duality between the solution to choosing the optimal linear combination of X_i and the linear regression of y_i on $\mathbf{X}_i$. Another reason is that if there exists a $\boldsymbol{\delta}$ so that $\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}=0$, then the choice $\gamma=1$ achieves no measurement error inconsistency. We consider general implications of the choice of γ, and these will become important for the general result below which relaxes the assumption made by L-W that $\rho_1=1$ (the first element of $\boldsymbol{\rho}$, although any element can technically be chosen and the vector arranged appropriately).

Corollary 1 *If $\gamma=\boldsymbol{\delta}'\boldsymbol{\rho}=1+\dfrac{1}{(\rho'\Sigma^{-1}\rho)(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}$, then $(t-\beta)=0$: the OLS regression of y_i on $\mathbf{Z}_{1i}$ and X_i^δ provides consistent estimates of β.*

Corollary 2 *The inconsistency for α, as expressed by $(\mathbf{a}^\delta - \boldsymbol{\alpha})$, does not depend on γ. Even the choice above, which allows consistent estimation of β, does not provide consistent estimation of $\boldsymbol{\alpha}$.*

We leave the proof of corollary 1 to Appendix A and focus here on the proof of the second corollary.

Proof. As noted above,

$$(\mathbf{a}^\delta - \alpha) = (V_1 - \mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\frac{V_2 - \mathbf{C}'V_1^{-1}\mathbf{C}}{V_2}\right)\Omega\beta.$$

Substituting the optimal choice of δ (given a value of γ) from Proposition 1 into Ω (defined by (6)) results in

$$\begin{aligned}(\mathbf{a}^\delta - \alpha) &= (V_1 - \mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\frac{V_2 - \mathbf{C}'V_1^{-1}\mathbf{C}}{V_2}\right) \\ &\times \left(\frac{\gamma^2(\rho'\Sigma^{-1}\rho)^{-1}}{\gamma^2(V_2 - \mathbf{C}'V_1^{-1}\mathbf{C}) + \gamma^2(\rho'\Sigma^{-1}\rho)^{-1}}\right)\beta \\ &= (V_1 - \mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\frac{V_2 - \mathbf{C}'V_1^{-1}\mathbf{C}}{V_2}\right) \\ &\times \left(\frac{(\rho'\Sigma^{-1}\rho)^{-1}}{(V_2 - \mathbf{C}'V_1^{-1}\mathbf{C}) + (\rho'\Sigma^{-1}\rho)^{-1}}\right)\beta,\end{aligned} \tag{7}$$

which is not a function of γ. (See the proof of Corollary 1 in Appendix A for details.) Regardless of γ, the inconsistency on $\boldsymbol{\alpha}$ is determined by the underlying variance covariance structure of $(\mathbf{Z}_{1i}, Z_{2i}, \mathbf{X}_i)$. ■

The intuition is simple: unless $\Sigma=0$, measurement error exists and severs the relationship between $\mathbf{Z}_{1i}$ and Z_{2i}. Even though a rescaling of the X_i^δ variable exists, the OLS regression alone will not result in consistent estimation. To arrive at that, the variable $\mathbf{Z}_{1i}$ must be rescaled as well.

It also important here to note that since the term Ω is positive, the direction of the inconsistency on each coefficient is determined solely by the sign of $\mathbf{C}$ and β. Note that if $\mathbf{C}=0$, there is no inconsistency for $\boldsymbol{\alpha}^\delta$, and that the magnitude of the inconsistency is increasing in $\mathbf{C}$. The magnitude of the inconsistency for different linear combinations of $\mathbf{X}_i$ is determined solely by Ω. Hence, linear combinations of X_i that include only some subset of the proxies cannot be better than the optimal linear combination.

Clearly, however, both the optimal choice of δ and the choice of γ rely on information unavailable in typical applications: specifically Σ (the variance matrix of $\boldsymbol{\varepsilon}$), $\mathbf{C}$ (the covariance between $\mathbf{Z}_{1i}$ and Z_{2i}), and V_2 (the variance of Z_{2i}). The second key result of L-W is that the OLS regression of y_i on $\mathbf{Z}_{1i}$ and $\mathbf{X}_i$ provides slope coefficients on $\mathbf{Z}_{1i}$ equivalent to the coefficients from the regression of y_i on $\mathbf{Z}_{1i}$ and X^δ for the optimal choice of δ. Further, a linear combination of the coefficients on $\mathbf{X}_i$ can be combined to achieve the minimum inconsistency estimate of β for the case where $\gamma=1$. L-W showed this for the linear combination of the coefficients on X_i, even in the presence of additional regressors. We focus on the expression for the coefficients on additional regressors, which was not examined by L-W.

Proposition 2 *Let $(\mathbf{a}, \mathbf{b})$ be coefficients from the population least squares regression of y_i on $\mathbf{Z}_{1i}$ and $\mathbf{X}_i$. Then $\mathbf{a}=\mathbf{a}^\delta$ and $\boldsymbol{\rho}'\mathbf{b}=t^\delta$ for $\boldsymbol{\delta}=\Sigma^{-1}\rho/(\rho'\Sigma^{-1}\rho)$.*

The proof is provided in Appendix A and follows rather directly from Proposition 1 above and Proposition 3 in Lubotsky and Wittenberg (2006). It is important here to note the implication: the OLS regression that includes all proxy variables $(\mathbf{X}_i)$ achieves coefficients on all other regressors which have the minimum inconsistency achievable through any linear combination of regressors. Since the result does not depend on the number of proxy variables, using any subset of $\mathbf{X}_i$ is equivalent to using a linear combination of the full set of $X_i's$, and so necessarily has a larger inconsistency than using all $\mathbf{X}_i$. Thus the set of coefficients used in extreme bounds analysis (or other model specification approaches) represents coefficients where the inconsistency is larger as fewer and fewer $X_i's$ are included, and importantly, the inconsistency is always in the same direction. Indeed, if $a_j<\alpha_j$ (elements of **a** and $\boldsymbol{\alpha}$ respectively), then any coefficient $\tilde{a}_j$ from the regression of y_i on $\mathbf{Z}_{1i}$ and any subset of $\mathbf{X}_i$ will be less than or equal to $a_j:\tilde{a}_j<a_j<\alpha_j$. We state this formally in the following corollary:

Corollary 3 *Let a_j, α_j, and C_j be corresponding elements of* **a**, $\boldsymbol{\alpha}$, *and* **C**. *Let $\tilde{a}_j$ be the corresponding coefficient from any regression of y_i on Z_{1i} and any subset of elements of* $\mathbf{X}_i$. *Then if $C_j>0$, $\tilde{a}_j\geq a_j\geq\alpha$; if $C_j<0$ then $\alpha\geq a_j\geq\tilde{a}_j$.*

Lubotsky and Wittenberg (2006) show that X^δ can be formed empirically if ρ_1, the first element of $\boldsymbol{\rho}$, is known to be 1. We note that this is a desirable case, leading to inference on the parameter β. However, in practice, this assumption may be difficult to justify. Of importance here is the fact that information about the parameter vector α, can be obtained regardless of this assumption. L-W note that

$$Cov(X_{ij},y_i)=\rho_j\,Cov(Z_i,y_i)$$

for each element X_{ji} of the vector $\mathbf{X}_i$. Hence, if $\rho_1=1$, the terms in $\boldsymbol{\rho}$ are identified by $\frac{Cov(X_{ij},y_i)}{Cov(X_{i1},y_i)}$. We note that in general the ratio $\frac{Cov(X_{ij},y_i)}{Cov(X_{i1},y_i)}=\frac{\rho_j}{\rho_1}$, so the vector $\boldsymbol{\rho}^*=\frac{1}{\rho_1}\boldsymbol{\rho}$ is identified without further assumptions. Unlike L-W, this vector is actually overidentified, since the other regressors $\mathbf{Z}_{1i}$ can also be used in place of y_i: $\frac{Cov(X_{ij},Z_{li})}{Cov(X_{i1},Z_{li})}=\frac{\rho_j}{\rho_1}$. Next consider using $\rho^\star$ in place of ρ in the results for proposition 2: $\boldsymbol{\rho}^{\star\prime}\mathbf{b}=\frac{1}{\rho_1}\boldsymbol{\rho}'\mathbf{b}=\frac{1}{\rho_1}t^\delta$. Considering the results in corollary 1, this implies that use of $\rho^\star$ is equivalent to choosing $\boldsymbol{\delta}'\boldsymbol{\rho}=\rho_1$, rather than 1 as is the case in proposition 2. Note that this choice has no effect on the results for **a**. Therefore, $\boldsymbol{\rho}^{\star\prime}\mathbf{b}$ is the least inconsistent estimate of $\frac{\beta}{\rho_1}$.

L-W also consider construction of X^δ. They show that $(\mathbf{X}_i\mathbf{b})/\boldsymbol{\rho}'\mathbf{b})=X^\delta$ for the optimal $\boldsymbol{\delta}$ when $\boldsymbol{\delta}'\rho=1$. This result holds regardless of whether $\rho_1=1$. Similarly, $(\mathbf{X}_i'\mathbf{b})/(\boldsymbol{\rho}'\mathbf{b})=X^\delta$ for the optimal $\boldsymbol{\delta}$ when $\boldsymbol{\delta}'\boldsymbol{\rho}=\rho_1$. Thus, the regression of y_i on $\mathbf{Z}_{1i}$ and $(\mathbf{X}_i\mathbf{b})/(\boldsymbol{\rho}'\mathbf{b})$ will yield a slope coefficient of $\boldsymbol{\rho}'\mathbf{b}$.

We return now to the dual problem of linear combinations of $\mathbf{X}_i$. Let X^{ρ_1} be the optimal linear combination of $\mathbf{X}_i$ for the restriction that $\boldsymbol{\delta}'\boldsymbol{\rho}=\rho_1$. This implies that

$$X_i^{\rho_1}=\rho_1 Z_{2i}+v_i,$$

where $v_i=\boldsymbol{\rho}_1(\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho})^{-1}(\boldsymbol{\rho}'\Sigma^{-1})\boldsymbol{\varepsilon}$. Thus, $X_i^{\rho_1}$ is a mismeasured variable with a scaling coefficient. Bollinger (2003) considers this case and shows that the direct regression of y_i on $\mathbf{Z}_{1i}$ and X^{ρ_1} provides a lower bound for the ratio $\frac{\beta}{\rho_1}$, and the slope coefficients on $\mathbf{Z}_{1i}$ form one bound (upper or lower depending upon sign of **C**) for the coefficients $\boldsymbol{\alpha}$. Bollinger (2003) also shows that the reverse regression $X_i^{\rho_1}$ on y_i and $\mathbf{Z}_{1i}$ provides the upper bound on $\frac{\beta}{\rho_1}$ and the other bound on $\boldsymbol{\alpha}$. Since X^δ can be formed from the results of the regression of y_i on $\mathbf{Z}_{1i}$ and $\mathbf{X}_i$, the reverse regression can also be estimated. Let d_2 be the coefficient on y_i from the reverse regression and let $\mathbf{d}_1$ be the vector of coefficients on $\mathbf{Z}_{1i}$ from the reverse regression. This allows us to state the following proposition:

Proposition 3 *The sign of $\boldsymbol{\rho}'\mathbf{b}$ is the sign of β. Further, $|\boldsymbol{\rho}'\mathbf{b}|\leq\left|\frac{\beta}{\rho_1}\right|\leq|1/d_2|$ and α_j is bounded by a_j and $-d_{1j}/d_2$.*

The proof follows from the results in propositions 1 and 2, the definition of X^{ρ_1} and Theorem 1 and corollary 1 in Bollinger (2003).

This result provides an approach to achieve bounds on $\boldsymbol{\alpha}$ and a rescaled measure of β. These bounds correspond to the tightest bounds achievable using any linear combination of the available proxy variables. It is interesting to note that by corollary 3 and proposition 3, the bounds achieved will not contain any of the coefficients obtained through extreme bounds analysis: extreme bounds analysis provides a set of inconsistent coefficients. The inconsistency is related to omitted variable inconsistency for failing to include Z_{2i}: the proxy variables only partially control for Z_{2i}. Thus the difference in estimates from different combinations represent different levels of omitted variable inconsistency. There is one special case in which extreme bounds analysis provides a set of consistent estimates of $\boldsymbol{\alpha}$: when $\mathbf{C}$=0. In this case, any regressions of y_i on $\mathbf{Z}_{1i}$ and any (or no) elements of X_i provide consistent estimates of $\boldsymbol{\alpha}$.

The estimated bounds can all be obtained through a two-stage procedure: first regress y_i on $\mathbf{Z}_{1i}$ and $\mathbf{X}_i$. Estimate elements of $\rho^\star$ using the estimated covariance ratios. We found that estimation was more stable when all covariance ratios were used and simply averaged to arrive at a minimum distance type estimator. One could potentially improve upon this by using a GMM approach. Using b and the estimated $\rho^\star$ terms, construct $X_i^{\rho_1}$. Then regress y_i on $\mathbf{Z}_{1i}$ and $X_i^{\rho_1}$ to obtain the direct regression bounds, and regress $X_i^{\rho_1}$ on y_i and $\mathbf{Z}_{1i}$ to obtain the terms to construct the reverse regression bounds.

Standard errors are analytically difficult, and typical asymptotic approximation results are unlikely to perform well in finite sample for three reasons. First, the procedure above involves "constructed regressors" (for a discussion, see Wooldridge 2010). Second, the reverse regression bounds are nonlinear functions of estimated parameters. Third, even if there is no heteroskedasticity in the structural equation, the estimated reduced form equations will be highly heteroskedastic. We recommend, and use in what follows, bootstrapped standard errors.

3 Simulation Results

In order to illustrate and evaluate the above results, we provide a set of simulations. In particular, this section highlights differences between the results of the Extreme Bounds procedure and our Generalized Proxy Bounds. We should emphasize that we have chosen to compare our results to Extreme Bounds because of its relative transparency, but other model specification procedures (such as those in Sala-i-Martin 1997, and Sala-i-Martin, Doppelhofer, and Miller 2004) would yield similar results to EBA.[4] We use the following model:

$$y_i = \alpha Z_{1i} + Z_{2i} + u_i.$$

We let

$$\begin{aligned} X_{1i} &= Z_{2i} + s\times(0.25\times v_i + \sqrt{(1-0.25^2)}e_{1i}) \\ X_{2i} &= 2Z_{2i} + s\times(0.25\times v_i + \sqrt{(1-0.25^2)}e_{2i}) \\ X_{3i} &= 3Z_{2i} + s\times(0.25\times v_i + \sqrt{(1-0.25^2)}e_{3i}). \end{aligned} \tag{8}$$

For simplicity, we generate (v_i, $\mathbf{e}_i$, u_i) as jointly standard normally distributed and mutually independent. We generate (Z_{1i}, Z_{2i}) as jointly standard normally distributed with a covariance C (which is also the

4 One difference between EBA and alternative model specification approaches is that since EBA considers the entire distribution (i.e., it does not disregard the tails of the distribution), it shares a bound with our Generalized Proxy Bounds (i.e., the least-biased estimate from the direct regression, when all of the proxy variables are included). The bounds from the other model specification procedures are further from the true parameter.

correlation). For simplicity, we impose the L-W assumption that ρ_1=1; this has no implications on the bounds for α, which are the focus of the simulation. The term s determines the total amount of measurement error in the proxy variables (and also in the optimal linear combination of the proxy variables). Two values of α are interesting: 1 and 0. Using α=1 provides a benchmark estimate of how large the inconsistency from the regression of y_i on Z_{1i} and $\mathbf{X}_i$ (or subsets of $\mathbf{X}_i$) is likely to be. It also demonstrates that the extreme bounds approach can either over- or understate the coefficient on Z_{1i}, depending on C. Similarly, the case of α=0 demonstrates that the extreme bounds procedure may lead one to conclude that Z_{1i} is an important explanatory variable when in fact it is not. We examine nine values of C: 0,±0.25,±0.5,±0.75 and ±0.9. We examine two values of s: 1 and $1.4\approx\sqrt{2}$ (these result in error variances of 1 and 2, respectively). The simulation results are based on 500 replicates of samples of size 1000.

Table 1 summarizes the results of our Monte Carlo simulations. Our results focus on inference about α. In the first row of Panel A, we present the proportion of times that the estimated generalized proxy bounds include the true value of 1. The second row allows for sampling variance of the bounds and presents the proportion of times where the bounds plus and minus 1.96 times the standard error contain the true coefficient (that is, $GPB_{lower}-1.96\times\text{s.e.}<1<GPB_{upper}+1.96\times\text{s.e.}$). The third and fourth rows provide comparable statistics for the extreme bounds approach. In row 3 we present the proportion of times that the extreme bounds contain the true coefficient, while in row 4 we allow for sampling variance and present the proportion of times that the true coefficient falls within the bounds ±1.96 times the standard error. The fifth and sixth rows present the average values of the lower and upper bounds from each of the two procedures.

When C is small, the inconsistency on estimates when Z_2 is omitted is quite small. Indeed, in the case where C=0 (the first column of Panel A), there is no inconsistency from using the proxy variables (nor would there be from simply omitting Z_2 completely). This shows up in that both the Generalized Proxy Bounds we

Table 1 Simulation Results varying *C*.

	***C*=0**	***C*=0.25**	***C*=0.5**	***C*=0.75**	***C*=0.9**
P($\alpha\in$ GPB)	0.252	0.714	0.916	0.993	1.000
P($\alpha\in$ GPB+se)	0.981	0.997	1.000	1.000	1.000
P($\alpha\in$ EBA)	0.265	0.291	0.087	0.007	0.000
P($\alpha\in$ EBA+se)	0.985	0.923	0.749	0.336	0.019
mean (EBA)	(0.987, 1.011)	(1.02, 1.13)	(1.05, 1.29)	(1.11, 1.52)	(1.26, 1.76)
mean (left GPB)	0.999	1.02	1.05	1.11	1.26
s.d. (actual, boot)	(0.032, 0.033)	(0.034, 0.034)	(0.036, 0.037)	(0.048, 0.047)	(0.066, 0.065)
mean (right GPB)	1.002	0.73	0.34	−0.72	−3.72
s.d. (actual, boot)	(0.046, 0.048)	(0.054, 0.052)	(0.070, 0.070)	(0.148, 0.148)	(0.513, 0.518)
		***C*=−0.25**	***C*=−0.5**	***C*=−0.75**	***C*=−0.9**
P($\alpha\in$ GPB)		0.729	0.915	0.991	1.000
P($\alpha\in$ GPB+se)		0.991	0.999	1.000	1.000
P($\alpha\in$ EBA)		0.279	0.086	0.009	0.000
P($\alpha\in$ EBA+se)		0.917	0.743	0.313	0.014
mean (EBA)		(0.87, 0.98)	(0.71, 0.95)	(0.48, 0.89)	(0.24, 0.74)
mean (left GPB)		0.98	0.95	0.89	0.74
s.d. (actual, boot)		(0.034, 0.034)	(0.036, 0.037)	(0.048, 0.047)	(0.066, 0.065)
mean (right GPB)		1.27	1.66	2.72	5.69
s.d. (actual, boot)		(0.054, 0.052)	(0.074, 0.070)	(0.150, 0.148)	(0.521, 0.512)

Notes: In each panel, the first row gives the proportion of times in our simulations that the true value of *a* (=1) falls within the Generalized Proxy (GPB) bounds. The second row gives this proportion when sampling variance is accounted for. The next two rows give these proportions for Extreme Bounds (EBA) as described in the text. The last rows give the means of the extreme bounds (EBA) and Generalized Proxy bounds (GPB), along with the actual and bootstrapped standard errors for each of the GP bounds.

propose and the extreme bounds perform quite well when the standard errors are included. Without standard errors, the GPB and the EBA bounds are really two consistent estimators of the same parameter α. In the limit, these two bounds will converge to the true parameter. In sample, they may or may not bound the actual parameter. This can be seen most starkly, in the last two rows where the range is extremely narrow. Including the standard errors provides a confidence interval slightly larger than 95%, because it is based upon multiple estimates. In the case where C=0, the identification of the parameters ρ rests solely on the covariance with y as used in Lubostky and Wittenberg (2006).[5] Caution should be used when it may be that C=0, in that estimation of ρ may be imprecise. However, we postulate that by the very nature of the problem, it is an unusual case.

Moving across the table, C increases in absolute value. The third column, for example, provides a case where the covariance between Z_1 and Z_2 is equal to 0.5. The first two rows of column 3 demonstrate that the Generalized Proxy (GP) Bounds work quite well. The bounds themselves capture the true coefficient in over 91% of the samples. If sampling variance is allowed for, the bounds contain the true coefficient 100% of the time. In contrast, as shown in the next two rows, the extreme bounds perform quite poorly, only capturing the true coefficient 8.7% of the time. Allowing for sampling error, this rises to nearly 75% of samples.

Examining the bounds in the last two rows of column 3, we can see that the two sets of bounds share a common value: 1.05 (on average). This is the average upper bound for the GP bounds, and the average lower bound for the extreme bounds. As noted above, this occurs when all three proxy variables for Z_2 are included in the regression. We note that on average, the extreme bounds lie above 1.05, with the largest being 1.29. This represents the fact (as discussed above) that the omitted variable inconsistency from failure to include Z_2 generally biases the estimate of α upward. It is also worth noting that the estimate using all three proxy variables (1.05) is not dramatically inconsistent and, as discussed above and in Lubotsky and Wittenberg (2006), represents the "least inconsistent" estimate of all combinations of proxy variables. The GP bounds that we propose are, on average, the interval [0.34,1.05]. As is often the case, the bound associated with the reverse regression (in this case, 0.34) is quite far away from the true coefficient. As discussed at length in Klepper (1988), Bollinger (1996), and Bollinger (2003), the reverse regression bound is consistent with the case where all error in the relationship between Y and the proxy variables is due to the error terms in the proxy variables.

Comparing across all five columns of panel A, we note two important aspects. First, as the covariance between Z_1 and Z_2 rises, the GP bounds proposed here perform increasingly well. As expected, the extreme bounds approach results in a bounded region which is increasingly far away from the true coefficient. While the extreme bounds may initially appear attractive, as they are quite narrow, it is clear that they fail to bound the true coefficient because they fail to address the inconsistency in the estimates. The GP bounds do become disturbingly large. Klepper (1988) and Bollinger (2003) propose approaches to tighten these bounds when additional information is available, which we do in the next section.

Panel B presents the mirror image case where the covariance is negative. Examining the last two rows, here again we see that the GP bounds and the extreme bounds share a common bound; however, in this case it is the lower bound for the GP bounds and the upper bound for the extreme bounds. Because the covariance between Z_1 and Z_2 is negative, the inconsistency on α is downward. As a result, regressions including any combination of the proxy variables result in a downward-inconsistent estimate of α, which is demonstrated by the values of the extreme bounds. The lower GP bound is again the regression with all proxy variables included and is necessarily closer to the true coefficient than any subset of proxy variables. Panel B represents a potentially worrisome case for extreme bounds: because the inconsistency is downward, the researcher might conclude that Z_2 is not significantly correlated with y when indeed it is.

Table 2 presents two additional sets of simulations, with C fixed at 0.5. In the first column of Table 2, we set α to zero. On average, the extreme bounds do not contain zero; they both lie above zero because of the upward inconsistency, potentially leading researchers to incorrectly conclude that α is significant or robustly correlated with y, when in fact it is zero. The GP bounds we propose do contain the true coefficient of zero.

5 We are grateful to a referee for pointing this out.

Table 2 Simulation Results varying α and s.

	α=0; s=1	α=1; s=1.4
P($\alpha \in$ GPB)	0.892	0.988
P($\alpha \in$ GPB+se)	1.000	1.000
P($\alpha \in$ EBA)	0.108	0.012
P($\alpha \in$ EBA+se)	0.746	0.416
mean (EBA)	(0.05, 0.29)	(1.08, 1.36)
mean (left GPB)	0.05	1.08
s.d. (actual, boot)	(0.039, 0.037)	(0.038, 0.038)
mean (right GPB)	−0.66	0.33
s.d. (actual, boot)	(0.067, 0.070)	(0.075, 0.075)

Notes: C=0.5 for all simulations. See notes to Table 1. s is from the definitions of the proxy variables (equation 8).

The second column of Table 2 examines a case where the error variance for the proxy variable is larger (a variance equal to 2 rather than 1, as in all other simulations). Note that this column is directly comparable to column 3 of panel A in Table 1. We note that the Generalized Proxy bounds actually perform even better, containing the true coefficient over 99% of cases. The extreme bounds do not perform as well, containing the true coefficient less than half of the time. Again, note that the upper GP bound and the lower bound for the extreme bounds represent the case where all proxy variables are included. The inconsistency is larger than in column 3 of Panel A in Table 1.

The bottom two rows on each panel present the average bootstrapped standard error and the actual standard deviation for the bounds on α The "left" bound is associated with the lower bound on the proxy variable and the case where measurement error is zero, while the right bound is associated with the upper bound on the proxy variable and is associated with the case where the residual error is zero and all error in the relationship is from the measurement error. Here we can evaluate the bootstrapped standard errors, and find that they are actually quite close to the observed variation in the estimates. We take this as suggestive that the bootstrap is an appropriate method in this situation.

The simulation demonstrates that the generalized proxy bounds are designed to contain the true coefficient, while the extreme bounds do not share this feature. The downside of the generalized proxy bounds is that they are quite wide. The width of this type of reverse regression bound has been a concern in this literature. As noted above, Bollinger (2003), Klepper (1988), and others have proposed approaches that can be used to tighten the bounds very effectively, which we do for our empirical example in the following section.

4 Application

To demonstrate the difference between extreme bounds and proxy bounds, in this section we use an illustrative example from the economic growth literature, where a common problem is that the researcher wants to estimate a structural relationship between growth and a variable (or variables) of interest, but the conditioning variables include variables such as "technology" or "human capital" that are not directly observed.

In our example, we use extreme bounds analysis, which has been used to gauge the "robustness" of variables included in economic growth regressions, most influentially by Levine and Renelt (1992). Variants of the extreme bounds approach include the Bayesian Averaging of Classical Estimates of Sala-i-Martin, Doppelhofer, and Miller (2004) and the related distributional approach of Sala-i-Martin (1997). With all of these approaches, the problem is framed as one of model specification: the "correct" specification of the model is one containing some subset of the control variables, and the purpose of the exercise is to bound the coefficient estimates on other variables included in the regression. In contrast, with our proxy bounds approach, the set of conditioning variables is correlated with some variable (human capital, technology, institutions) that belongs in the regression but is not directly observable.

In their empirical test of the Solow (1956) model, Mankiw, Romer, and Weil (1992) augment the original Solow model with a separate measure of human capital, as follows. Consider a production function given by:

$$Y(t)=K(t)^{\alpha}H(t)^{\beta}(A(t)L(t))^{1-\alpha-\beta} \tag{9}$$

where Y is aggregate output, K and H are stocks of physical and human capital respectively, A represents the level of technology (which grows at the exogenous rate g), and L is the labor force. Income is invested in physical and human capital at the constant fractions s_k and s_h, respectively.

With standard assumptions, it is straightforward to derive an expression that can be estimated to measure the actual rate of convergence to the steady-state level of income per effective worker ($y=Y/(AL)$):

$$\begin{aligned}&\ln(y(t))-\ln(y(0))\\ &=(1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\ln(s_k)+(1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\ln(s_h)\\ &-(1-e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta)-(1-e^{-\lambda t})\ln(y(0)).\end{aligned} \tag{10}$$

where $\lambda=(n+g+\delta)(1-\alpha-\beta)$, n is the population growth rate, g is the technology growth rate, and δ is the depreciation rate of both physical and human capital.

Thus, growth in GDP per capita is a function of investment in physical capital (s_k), investment in human capital (s_h), a term including population growth, technological progress, and depreciation ($n+g+\delta$), and initial income ($y(0)$). In addition, the coefficient estimate on the log of initial income can be used to infer the speed of convergence toward the steady state (λ).

Equation (10) is the model underlying our application and corresponds to the known structural equation (1) discussed above. As is typically done, we focus upon the linear specification here, specifically

$$\Delta y_t=a_0+a_1\ln GDP_0+a_2\ln INV_t+a_3\ln NGD_t+bS_h+u_t$$

This is a standard regression in the empirical growth literature. Our estimates of GDP per capita are adjusted for purchasing power parity, and come from the Penn World Tables database, frequently used in the empirical growth literature. Investment and annual population growth are also taken from Penn World Tables, and we follow Mankiw, Romer, and Weil (1992) in setting $(g+\delta)=0.5$. Although there may be other issues associated with general measurement error in these variables, there is widespread agreement that they correspond fairly directly to the theory. For our purposes, we treat them as measured correctly. A much more difficult question is how to accurately measure "human capital." Indeed, Mankiw, Romer, and Weil (1992), among many others, discuss this issue at length. It is likely that "human capital" cannot be measured directly; however, numerous variables correlated with human capital are available. We include three variables correlated with stocks and accumulation rates of human capital, all of which are commonly used in the empirical growth literature: literacy rates, primary school enrollment rates, and secondary school enrollment rates. Our sample is a cross-section of 88 countries at all levels of development. Hence, the question is not one of specification: equation (10) is well specified. Rather, the question is one of measurement.

Using the notation of Section 1, we are interested in estimating the structural relationship

$$y_i=\boldsymbol{\alpha}'\mathbf{Z}_{1i}+\beta Z_{2i}+u_i. \tag{11}$$

Z_{2i} is not measured directly, but a vector $\mathbf{X}_i$ of correctly measured variables exists such that $\mathbf{X}_i=\boldsymbol{\rho}Z_{2i}+\boldsymbol{\epsilon}_i$. In this application $\mathbf{X}_i$ contains primary school enrollment, secondary school enrollment, and literacy rates. $\mathbf{Z}_1$ is a vector including the variables measured without error (in our application, initial GDP per capita, physical capital investment, and the term including population growth).

The focus of this estimation is the relationship between growth and the three regressors: initial income, physical capital investment, and population growth; that is, we focus on the coefficients $\boldsymbol{\alpha}$. Theory defines the structural relationship, but we do not have a measure of human capital (Z_{2i}). Rather, we have a set of variables (primary and secondary school enrollment and literacy rates) that are correlated with human capital.

The model specification literature has a similar approach: the primary concern is the coefficient estimates on a set of key variables (frequently initial income for estimating the speed of convergence, but often also a particular variable of interest), but the claim is that the correct set of additional conditioning variables is unknown. More generally, the traditional approach in much empirical work is to include different sets of control variables, under the assumption that the correct coefficient estimates on the variables of interest fall somewhere in that range. This is also the idea behind more formal approaches to model selection, such as extreme bounds analysis and Bayesian model averaging. For example, in extreme bounds analysis, the researcher includes all possible combinations of a set of control variables, and identifies the "extreme bounds" as the minimum and maximum estimates, accounting for standard errors.

4.1 Extreme Bounds

We adapt extreme bounds analysis first suggested by Leamer and Leonard (1983) and employed in the growth literature by Levine and Renelt (1992) in the following way.[6] After estimating the regression with each possible combination of the three human capital proxy variables (yielding seven regressions with estimates of $\boldsymbol{\alpha}$ denoted as $\tilde{a}_k$ and their corresponding estimated standard errors denoted as SE_k), we compute the upper and lower bounds for each of the correctly measured ($\mathbf{Z}_1$) regressors (initial GDP, investment, and population growth) as the maximum and minimum values of $\tilde{a}_k \pm 2 * SE_i$, which is also the cutoff used by Levine and Renelt (1992). Table 3 presents the results for initial income, physical capital investment, and population growth; each would be considered "robust" by their definition (i.e., for each variable, the highest and lowest bounds are statistically significant at 95% or greater and of the same sign).[7]

In general, empirical growth researchers have been concerned primarily with identifying variables that are "robustly" correlated with growth (i.e., consistently positively or negatively correlated with growth, conditional on other variables), and extreme bounds analysis and other approaches to model specification have been employed primarily to identify these variables. However, the coefficient estimates on initial GDP also allow for inference about the speed of convergence to the steady state. In the extreme bounds analysis, the coefficient estimate range of −0.47 to −0.36 implies an estimate of λ of between 0.008 and 0.010, which is

Table 3 Extreme Bounds Analysis.

	$\hat{\beta}$ s.e.	*p*-Value	Bound	Proxies Included
ln(*GDP*)				
Upper	−0.36 (0.10)	0.001	−0.16	lit
Lower	−0.47 (0.10)	0.000	−0.66	lit, pri, sec
ln(*INV*)				
Upper	0.41 (0.09)	0.000	0.59	lit
Lower	0.29 (0.09)	0.002	0.11	lit, pri, sec
ln(*n*+*g*+δ)				
Upper	−2.36 (0.71)	0.001	−0.94	lit
Lower	−2.98 (0.68)	0.000	−4.33	lit, pri, sec

Notes: The table reports the coefficient estimates, standard errors, and *p*-values associated with each bound. The bounds are defined as $\hat{\beta} \pm 1.96\sigma$, where σ are estimated standard errors. The last column gives the proxies included in the specification that identified each bound.

6 Because Levine and Renelt included seven control variables, primarily policy variables (the analogue here is the three measures of human capital), they limited their control variables to exactly three in each regression. We allow for all possible measures of our three measures of human capital (yielding seven regressions).

7 Although Levine and Renelt also found that initial income and physical capital investment were robustly correlated with growth, they did not find this for population growth. Our data cover a longer time period, and our control variables differ from theirs.

slightly lower than the estimate of 0.014 in the full 98-country sample in Mankiw et al. (1992).[8] These estimates of λ imply that a country moves halfway toward its steady state in between 70 and 87 years.

4.2 Generalized Proxy Bounds

In this section, we present the estimation of the alternative generalized proxy bounds. We proceed as follows. First, we estimate the base regression by OLS, including all of the proxies for Z_2:

$$y_i = \mathbf{a}'\mathbf{Z}_{1i} + \mathbf{b}'\mathbf{X}_i + \epsilon_i \tag{12}$$

and retain the **a** and **b** coefficients. Following Propositions 2 and 3, the estimates **a** provide one of the bounds for each of the $\mathbf{Z}_1$ variables.

To find the other set of bounds, we must construct X^δ, for the reverse regression. The first step is in finding estimates for $\boldsymbol{\rho}$. There are four consistent estimates for ρ_2 and ρ_3: for example, ρ_2 could be estimated by $\frac{cov(x_2,y)}{cov(x_1,y)}$, $\frac{cov(x_2,z_1)}{cov(x_1,z_1)}$, $\frac{cov(x_2,z_2)}{cov(x_1,z_2)}$, or $\frac{cov(x_2,z_3)}{cov(x_1,z_3)}$. For both ρ_2 and ρ_3, we take the average of these four estimators, which is equivalent to a minimum-distance GMM estimator. We then construct the lower bound on the slope of the unobserved Z_2 variable:

$$B_{lb} = b_1 + \rho_2 \cdot b_2 + \rho_3 \cdot b_3 \tag{13}$$

The minimum-inconsistent weighted average estimate of the unobserved Z_2 variable is then:

$$\hat{Z}_{2i} = \frac{b_1 X_{1i} + b_2 X_{2i} + b_X X_{3i}}{B_{lb}} \tag{14}$$

Finally, we regress these estimates $\hat{Z}_{2i}$ on y_i and $\mathbf{Z}_{1i}$:

$$\hat{Z}_{2i} = d_2 * y_i + \mathbf{d}_1'\mathbf{Z}_{1i} + v_i \tag{15}$$

The upper bound on the slope for the Z_2 variable is given by:

$$B_{ub} = 1/d_2 \tag{16}$$

and the second bound for the $\mathbf{Z}_1$ variables is given by:

$$-d_{1j}/d_2 \tag{17}$$

The results for the generalized proxy bounds are in Table 4. Several comparisons to the extreme bounds analysis in Table 3 merit attention. First, the coefficient estimates on investment are no longer considered "robust," in that one bound is positive and one is negative. This is somewhat surprising, since investment is generally considered to be one of the variables most strongly correlated with growth.[9] Second, the coefficient estimates on the population growth term ($n+g+\theta$) are statistically significant and negative at both bounds, and are larger in magnitude than the estimates from the extreme bounds analysis.[10]

8 The Mankiw et al. sample covers the period 1960–1985, which is extended here to 1960–2000. Their measure of human capital is secondary school enrollment rates.

9 For example, it was one of only three variables out of over 30 tested identified as "robustly" correlated with growth in the extreme bounds analysis of Levine and Renelt (1992).

10 Although the extreme bounds analysis in Table 3 also yielded coefficient estimates that would be considered robust, population growth was *not* a robust variable in Levine and Renelt (1992).

Table 4 Generalized Proxy Bounds.

	$\hat{\beta}$ s.e.	*p*-Value	Bound
ln(*GDP*)			
Upper	−0.47 (0.10)	0.000	−0.27
Lower	−1.14 (0.26)	0.000	−1.65
ln(*INV*)			
Upper	0.29 (0.11)	0.012	0.51
Lower	−0.31 (0.28)	0.274	−0.86
ln(*n*+*g*+*δ*)			
Upper	−2.98 (0.65)	0.000	−1.75
Lower	−3.69 (1.59)	0.020	−6.81

Notes: The table reports the coefficient estimates, bootstrapped standard errors, and *p*-Values associated with each bound. The bounds are defined as $\hat{\beta}\pm 1.96\sigma$, where σ are bootstrapped standard errors.

Finally, the bounds on the coefficient estimates for initial GDP are interesting, because the point estimates can be used to infer the speed of convergence toward the steady state. The upper generalized proxy bound is identical to the lower extreme bound, suggesting that the coefficient estimate is more negative, and the speed of convergence faster. The speed of convergence (λ) implied by the generalized proxy bounds in Table 4 is between 0.010 and 0.019, which would imply that a country would move half of the distance toward its steady state in between 70 and 35 years. (The range for λ in the extreme bounds is 0.008 to 0.010.)

As noted in the previous section, we can tighten these bounds by bringing additional information or restrictions to bear on the problem, as outlined by Klepper (1988). Note that the upper bound on human capital and the associated bounds on other parameters represent the case where all error in the relationship stems from measurement error (ε) in the relationship between the proxies and the true human capital. This implies that if the true human capital measure were obtained, the R^2 of the resulting regression would be 1. Clearly, this is not likely. By choosing an upper bound on the R^2 from this ideal regression, we can tighten the bounds on the other parameters. However, choosing such an R^2 is controversial. The approach we adopt here is to ask what bound on R^2 would be necessary to arrive at a particular economic conclusion from the coefficients in the model. In this case, we focus upon the coefficient on investment. The fact that investment is no longer consistently positive is of concern. As shown in Bollinger (2003), we can solve for the R^2 where the lower bound on the investment coefficient is zero. Column 2 of Table 5 presents this bound and the implications for the other coefficients. The implied R^2 is 0.62, while the base R^2 from the direct regression is 0.56. This restriction also has implications for the correlation between true human capital and its proxies in that it increases this correlation compared to the reverse regression case. In the reverse regression, where all error is associated with the human capital measure, the implied correlation between human capital and the proxy is 0.697; by tightening the bounds, we raise the implied correlation to 0.762. Klepper (1988) shows that these two restrictions are one-to-one: we can achieve tighter bounds either by

Table 5 Tightening the GP Bounds.

	Original Lower	Restrict Inv	Restrict GDP	Upper
ln(*GDP*)	−1.14	−0.80	−0.99	−0.47
ln(*INV*)	−0.31	0.00	−0.17	0.29
ln(*n*+*g*+*δ*)	−3.69	−3.32	−3.53	−2.98
R^2	1.00	0.617	0.826	0.569
Corr(HC, proxy)	0.697	0.762	0.719	1.00

Notes: Columns 1 and 4 report the regression results from Table 4. Columns 2 and 3 report the results of, respectively, restricting the coefficient on investment to be positive, and the coefficient on GDP to be >−0.985, which yields a plausible magnitude for estimated convergence.

placing an upper bound on the R^2 or by placing a lower bound on the correlation between the proxy and the true variable. The implications here are actually quite plausible. In Column 3, we conduct a similar analysis restricting the coefficient on initial GDP to be −0.985, which yields an estimate of λ of approximately 0.017, closer to the estimate in Mankiw, Romer, and, Weil (1992). Again, the implications are reasonable. We note that this restriction is not sufficient to restrict the coefficient on log investment to be positive. While the restriction on investment leads to a coefficient on initial GDP of −0.8 an implied λ of 0.015. Notice that these restrictions affect only the lower bound (the original results are presented in Column 1), leaving the upper bound unchanged (which is presented in column 4).

5 Conclusions

We have demonstrated that when a researcher has a series of proxy variables that are thought to be correlated with an unobserved regressor from a structural model, including all proxy variables in the regression results in estimates that are the least inconsistent of any combination of the proxy variables. Furthermore, we provide an approach that derives bounds on other coefficients in the model. The model specification literature has a similar goal: when a series of proxy variables is available, they examine how coefficients change as different combinations of proxies are included. We show that this may lead to conclusions about the robustness of results that are unwarranted.

The problem considered here – estimation and inference when a regressor is unobserved but proxy variables are available – is quite general. Indeed, in many empirical applications the theoretical variables of interest are unobserved, and researchers have typically used so-called proxy variables in their place. This paper extends the work of Bollinger (2003) and Lubotsky and Wittenberg (2006), combining important results from both papers and establishing a general approach that provides both an indication of the direction of the inconsistency and its potential magnitude. As noted in Lubotsky and Wittenberg (2006), researchers have often been concerned about which and how many proxies to include. This paper extends those results to focus on inconsistency in other parameters of the model, and provides a direct comparison to extreme bounds analysis as a representative approach from the model specification literature.

Further work on this problem should be considered. It is straightforward to extend the results here to a case where there are multiple sets of proxies for multiple unobserved variables, but this requires the somewhat stringent assumptions that the proxies can be uniquely assigned to a specific unobserved variable and that the errors across the groups of proxies are uncorrelated. However, this is rather limiting. At this writing, extending this approach to a general multiple proxies for multiple unobserved variables has not been accomplished. The problem is that there is no unique scaling parameter like γ in the above discussion.

This paper demonstrates that including all proxy variables not only minimizes the inconsistency in estimating the coefficient on the unobserved regressor (as shown by Lubotsky and Wittenberg 2006), but also the inconsistency in estimating coefficients on any other regressors in the model. We further show the direction of the inconsistency on coefficients on measured variables by establishing bounds. If the strong assumption that the scale is known, as assumed by Lubotsky and Wittenberg (2006), is added, our results bound the coefficient on the unobserved variable as well, and in any event bound the rescaled coefficient. These results have important implications in the interpretation of nearly all empirical work.

Acknowledgment: We thank Helle Bunzel, Steven Durlauf, Josh Ederington, Per Hjerstrand, Brian Krauth, Brent Krieder, Mike McCracken, John Pepper, Shinichi Sakata, Justin Tobias, Ken Troske, Tom Wansbeek, Hendrik Wolff, Jim Ziliak, and participants in seminars at the Universities of California, Berkeley and Santa Cruz, University of Oregon, University of Washington, Iowa State University, IUPUI, the International Measurement Error Conference, Canadian Economics Association, and the Southern Economic Association meetings for helpful comments and discussion.

Previously published online March 13, 2014

Appendix: Proofs

Let

$$V\begin{pmatrix}\mathbf{Z}_{1i}\\ Z_{2i}\end{pmatrix}=\begin{bmatrix}V_1 & \mathbf{C}\\ \mathbf{C}' & V_2\end{bmatrix}.$$

The matrix V_1 is the $k\times k$ variance matrix for $\mathbf{Z}_{1i}$, $\mathbf{C}$ is the $k\times 1$ covariance, and V_2 is the scalar variance of Z_{2i}. Let $\boldsymbol{\delta}$ be an arbitrary $l\times 1$ vector such that $\boldsymbol{\delta}'\boldsymbol{\rho}=\gamma>0$ for some given value γ. Let $\theta=\beta/\gamma$.

The next three Lemmas establish key results for Proposition 3.

Lemma 1 *Expressions for* $(\boldsymbol{\alpha}-\mathbf{a})$ *and* $(\theta-t)$.

Then

$$\begin{bmatrix}\mathbf{a}\\ t\end{bmatrix}=\begin{bmatrix}V_1 & \gamma\mathbf{C}\\ \gamma\mathbf{C}' & \gamma^2V_2+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})\end{bmatrix}^{-1}\begin{bmatrix}V_1\alpha+\theta\gamma\mathbf{C}\\ \gamma\mathbf{C}'\alpha+\theta\gamma^2V_2\end{bmatrix}.$$

Rewriting yields

$$\begin{bmatrix}V_1 & \gamma\mathbf{C}\\ \gamma\mathbf{C}' & \gamma^2V_2+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ t\end{bmatrix}=\begin{bmatrix}V_1\alpha+\theta\gamma\mathbf{C}\\ \gamma\mathbf{C}'\alpha+\theta\gamma^2V_2\end{bmatrix},$$

which is equivalent to

$$\begin{bmatrix}V_1 & \gamma\mathbf{C}\\ \gamma\mathbf{C}' & \gamma^2V_2\end{bmatrix}^{-1}\begin{bmatrix}V_1 & \gamma\mathbf{C}\\ \gamma\mathbf{C}' & \gamma^2V_2+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ t\end{bmatrix}$$
$$=\begin{bmatrix}V_1 & \gamma\mathbf{C}\\ \gamma\mathbf{C}' & \gamma^2V_2\end{bmatrix}^{-1}\begin{bmatrix}V_1\alpha+\theta\gamma\mathbf{C}\\ \gamma\mathbf{C}'\alpha+\theta\gamma^2V_2\end{bmatrix}.$$

This yields

$$\begin{bmatrix}I & \gamma(V_1-\mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}(1-(\gamma^2V_2)^{-1}(\gamma^2V_2+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})))\\ 0 & (\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))^{-1}(\gamma^2V_2+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})-\gamma^2\mathbf{C}'V_1^{-1}\mathbf{C})\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ t\end{bmatrix}=\begin{bmatrix}\alpha\\ \theta\end{bmatrix}.$$

Noting that V_2, γ, and $(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$ are all scalars, this can be written as

$$\begin{bmatrix}I & -\gamma(V_1-\mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\dfrac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{\gamma^2V_2}\right)\\ \mathbf{0}' & \left(1+\dfrac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))}\right)\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ t\end{bmatrix}=\begin{bmatrix}\alpha\\ \theta\end{bmatrix}.$$

Rearranging gives

$$\begin{bmatrix}\mathbf{a}-\gamma(V_1-\mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\dfrac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{\gamma^2V_2}\right)t\\ \left(1+\dfrac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))}\right)t\end{bmatrix}=\begin{bmatrix}\alpha\\ \theta\end{bmatrix}.$$

Thus

$$(\mathbf{a}-\alpha)=\gamma(V_1-\mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\frac{V_2-\mathbf{C}'V_1^{-1}\mathbf{C}}{V_2}\right)\left(\frac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}\right)\theta$$
$$=(V_1-\mathbf{C}V_2^{-1}\mathbf{C})^{-1}\mathbf{C}\left(\frac{V_2-\mathbf{C}'V_1^{-1}\mathbf{C}}{V_2}\right)\left(\frac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}\right)\beta,$$

and

$$(t-\theta)=\left(\frac{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}\right)\theta-\theta$$
$$=-\left(\frac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}\right)\theta.$$

QED.

Lemma 2 *The term* $\left(\frac{(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}\right)$ *is positive and increasing in* $(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$.

The term $(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$ is positive provided that $\gamma\neq 0$ and Σ is positive semi-definite. The term $V_2-\mathbf{C}'V_1^{-1}\mathbf{C}$ is the determinant of the $V(Z_{1i}, Z_{2i})$, and so is, by necessary assumption, positive. The term $(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$ will be non-negative provided Σ is positive semi-definite. The derivative with respect to the term $(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$ is

$$\frac{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})-(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})}{((\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}))^2}$$
$$=\frac{\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{((\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}))^2}>0.$$

Hence the inconsistency in both are increasing in $(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$. QED.

Lemma 3 *The solution to* $\min_\delta(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})$ *s.t.* $\boldsymbol{\delta}\boldsymbol{\rho}=\gamma$ *is* $\boldsymbol{\delta}=\gamma\Sigma^{-1}\boldsymbol{\rho}(\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho})^{-1}$.

The Lagrangian is

$$(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta})-\lambda(\boldsymbol{\delta}'\boldsymbol{\rho}-\gamma).$$

FOC are

$$2\Sigma\boldsymbol{\delta}-\lambda\boldsymbol{\rho}=0$$
$$\boldsymbol{\delta}'\boldsymbol{\rho}-\gamma=0.$$

Solving:

$$\boldsymbol{\delta}=\frac{1}{2}\lambda\Sigma^{-1}\boldsymbol{\rho}$$
$$\frac{1}{2}\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho}\lambda=\gamma.$$

Substitution yields

$$\boldsymbol{\delta}=\gamma\Sigma^{-1}\boldsymbol{\rho}(\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho})^{-1}$$
$$\lambda=2\gamma(\boldsymbol{\rho}'\Sigma^{-1}\boldsymbol{\rho})^{-1}.$$

QED.

Proof. The proof of proposition 1 follows from the details in the text combined with the above lemmas. ■

Proof. Proof of Corollary 1. Substitution of the results from proposition 1 into the expressions in Lemma 1 yields

$$(\delta'\Sigma\delta)=\frac{\gamma^2\rho'\Sigma^{-1}\Sigma\Sigma^{-1}\rho}{(\rho'\Sigma^{-1}\rho)^2}=\frac{\gamma^2}{(\rho'\Sigma^{-1}\rho)}.$$

From Lemma 1 we have that

$$(t-\theta)=-\theta\left(\frac{(\delta'\Sigma\delta)}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\delta'\Sigma\delta)}\right)$$

Alternatively,

$$\begin{aligned}t&=\theta\left(1-\left(\frac{(\delta'\Sigma\delta)}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\delta'\Sigma\delta)}\right)\right)\\&=\frac{\beta}{\gamma}\left(\frac{\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{(\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C}))+(\delta'\Sigma\delta)}\right).\end{aligned}$$

Substitute the optimal choice of δ from proposition 1 which yields

$$\begin{aligned}t&=\frac{\beta}{\gamma}\left(\frac{\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{\gamma^2(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})+\gamma^2(\rho'\Sigma^{-1}\rho)^{-1}}\right)\\&=\frac{\beta}{\gamma}\left(\frac{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})+(\rho'\Sigma^{-1}\rho)^{-1}}\right).\end{aligned}$$

Hence, by choosing

$$\gamma=\frac{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})+(\rho'\Sigma^{-1}\rho)^{-1}}{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}=1+\frac{1}{(\rho'\Sigma^{-1}\rho)(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})},$$

we have $t=\beta$: no inconsistency in the coefficient on X^δ. QED ■

Lemma 4 *(Sherwin-Morrison_Woodbury Matrix Inversion Lemma): If A and B are non-singular matrices, and X is conformable, then* $(A+XBX')^{-1}=A^{-1}-A^{-1}X(B^{-1}+X'A^{-1}X)^{-1}X'A^{-1}$.

Proof. Proof of Proposition 2:

The linear regression of y_i on $\mathbf{Z}_{1i}$ and $\mathbf{X}_i$ yields slope coefficients consistent for

$$\begin{pmatrix}\mathbf{a}\\\mathbf{b}\end{pmatrix}=\begin{bmatrix}V_1 & \mathbf{C}\rho'\\ \rho\mathbf{C}' & (\rho\rho'V_2+\Sigma)\end{bmatrix}^{-1}\begin{bmatrix}V_1\alpha+\mathbf{C}\beta\\ \rho\mathbf{C}'\alpha+\rho V_2\beta\end{bmatrix}.$$

Rewriting yields

$$\begin{bmatrix}V_1 & \mathbf{C}\rho'\\ \rho\mathbf{C}' & (\rho\rho'V_2+\Sigma)\end{bmatrix}\begin{bmatrix}\mathbf{a}\\\mathbf{b}\end{bmatrix}=\begin{bmatrix}V_1\alpha+\mathbf{C}\beta\\ \rho\mathbf{C}'\alpha+\rho V_2\beta\end{bmatrix},$$

which is equivalent to

$$\begin{aligned}&\begin{bmatrix}V_1 & \mathbf{C}\rho'\\ \rho\mathbf{C}' & I\rho'\rho V_2\end{bmatrix}^{-1}\begin{bmatrix}V_1 & \mathbf{C}\rho'\\ \rho\mathbf{C}' & (\rho\rho'V_2+\Sigma)\end{bmatrix}\begin{bmatrix}\mathbf{a}\\\mathbf{b}\end{bmatrix}\\&=\begin{bmatrix}V_1 & \mathbf{C}\rho'\\ \rho\mathbf{C}' & I\rho'\rho V_2\end{bmatrix}^{-1}\begin{bmatrix}V_1\alpha+\mathbf{C}\beta\\ \rho\mathbf{C}'\alpha+\rho V_2\beta\end{bmatrix},\end{aligned}$$

where I is the identity matrix of appropriate dimensions. The inverse of the leading matrix (a partitioned matrix) can be written as

$$\begin{bmatrix} (V_1-\mathbf{C}\rho'(I\rho'\rho V_2)^{-1}\rho\mathbf{C}')^{-1} & -(V_1-\mathbf{C}\rho'(I\rho'\rho V_2)^{-1}\rho\mathbf{C}')^{-1}\mathbf{C}\rho'(I\rho'\rho V_2)^{-1} \\ -(I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1}\rho\mathbf{C}'V_1^{-1} & (I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1} \end{bmatrix}.$$

Since $\boldsymbol{\rho}'\boldsymbol{\rho}V_2$ is a scalar, this reduces to

$$\begin{bmatrix} (V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1} & -(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}\mathbf{C}\rho'(\rho'\rho V_2)^{-1} \\ -(I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1}\rho\mathbf{C}'V_1^{-1} & (I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1} \end{bmatrix}.$$

Substitution and simplification yields

$$\begin{bmatrix} I & -(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}(\mathbf{C}\rho'(\rho'\rho V_2)^{-1}\Sigma) \\ 0 & (I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1}(\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\rho'+\Sigma) \end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}$$
$$=\begin{bmatrix} \alpha \\ (I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1}\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta \end{bmatrix},$$

or

$$\begin{bmatrix} \mathbf{a}-(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}(\mathbf{C}\rho'\left(\rho'\rho V_2\right)^{-1}\Sigma)\mathbf{b} \\ (I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1}(\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\rho'+\Sigma)\mathbf{b} \end{bmatrix}$$
$$=\begin{bmatrix} \alpha \\ (I\rho'\rho V_2-\rho\mathbf{C}'V_1^{-1}\mathbf{C}\rho')^{-1}\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta \end{bmatrix}.$$

We can write

$$\mathbf{b}=(\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\rho'+\Sigma)^{-1}\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta,$$

and

$$\begin{aligned}\mathbf{a}=&\alpha\\ &+(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}(\mathbf{C}\rho'(\rho'\rho V_2)^{-1}\Sigma)\\ &\times(\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\rho'+\Sigma)^{-1}\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta.\end{aligned}$$

Turning first to the term **a** and applying the Sherwin-Morrison_Woodbury Matrix Inversion Lemma:

$$\begin{aligned}\mathbf{a}=&\alpha\\ &+(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}(\mathbf{C}\rho'(\rho'\rho V_2)^{-1}\Sigma)\\ &\times(\Sigma^{-1}-\Sigma^{-1}\rho((V_2-\mathbf{C}'V_1^{-1}\mathbf{C})^{-1}+\rho'\Sigma^{-1}\rho)^{-1}\rho'\Sigma^{-1})\\ &\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta.\end{aligned}$$

Simplification yields

$$\begin{aligned}\mathbf{a}=&\alpha\\ &+(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}\mathbf{C}\left(\frac{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{V_2}\right)\\ &\times\left(1-\frac{\rho'\Sigma^{-1}\rho}{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})^{-1}+\rho'\Sigma^{-1}\rho}\right)\beta\end{aligned}$$

or

$$=\alpha$$
$$+(V_1-\mathbf{C}V_2^{-1}\mathbf{C}')^{-1}\mathbf{C}\left(\frac{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{V_2}\right)$$
$$\times\left(\frac{(\rho'\Sigma^{-1}\rho)^{-1}}{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})+(\rho'\Sigma^{-1}\rho)^{-1}}\right)\beta,$$

which is the expression for **a** when the error-variance-minimizing choice of δ is used to construct X^δ (See Corollary 2).

Turning now to **b**, consider

$$\rho'\mathbf{b}=\rho'(\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\rho'+\Sigma)^{-1}\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta.$$

Again using the Sherwin-Morrison_Woodbury Matrix Inversion Lemma,

$$\rho'\mathbf{b}=\rho'(\Sigma^{-1}-\Sigma^{-1}\rho((V_2-\mathbf{C}'V_1^{-1}\mathbf{C})^{-1}$$
$$+\rho'\Sigma^{-1}\rho)^{-1}\rho'\Sigma^{-1})\rho(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta$$
$$=(\rho'\Sigma^{-1}\rho-\rho'\Sigma^{-1}\rho((V_2-\mathbf{C}'V_1^{-1}\mathbf{C})^{-1}$$
$$+\rho'\Sigma^{-1}\rho)\rho'\Sigma^{-1}\rho)(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta$$
$$=\left(\frac{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})^{-1}(\rho'\Sigma^{-1}\rho)+(\rho'\Sigma^{-1}\rho)^2-(\rho'\Sigma^{-1}\rho)^2}{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})^{-1}+\rho'\Sigma^{-1}\rho}\right)$$
$$(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})\beta$$
$$=\left(\frac{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})}{(V_2-\mathbf{C}'V_1^{-1}\mathbf{C})+(\rho'\Sigma^{-1}\rho)^{-1}}\right)\beta.$$

This is equal to the expression for **a** when the error variance minimizing choice of δ is used to construct X^δ in Corollary 1 if $\gamma=1$.

QED ■

References

Barro, Robert J. 1991. "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106 (2): 407–443.

Barro, Robert J., and Jong-Wha Lee. 2001. "International Data on Educational Attainment: Updates and implications." *Oxford Economic Papers* 53 (3): 541–563.

Bollinger, Christopher R. 1996. "Bounding Mean Regressions When A Binary Regressor is Mismeasured." *Journal of Econometrics* 73 (2): 387–399.

Bollinger, Christopher R. 2003. "Measurement Error in Human Capital and the Black-White Wage Differential." *Review of Economics and Statistics* 85 (3): 578–585.

Brock, William A., and Steven N. Durlauf. 2001. "Growth Empirics and Reality." *World Bank Economic Review* 15 (2): 229–272.

Brock, William A., Steven N. Durlauf, and Kenneth D. West. 2003. "Policy Evaluation in Uncertain Economic Environments." *Brookings Papers on Economic Activity* 2003 (1): 235–301.

Durlauf, Steven N., Andros Kourtellos, and Chih Ming Tan. 2008. "Are Any Growth Theories Robust?." *Economic Journal* 2008 (119): 329–346.

Goldberger, Arthur S., and Karl G. Jöreskog. 1975. "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable." *Journal of the American Statistical Associaton* 70 (351): 631–639.

Griliches, Zvi. 1974. "Errors in Variables and Other Unobservables." *Econometrica* 42 (6): 971–998.

Klepper, Steven. 1988. "Regressor Diagnostics for the Classical Errors-in-Variables Model." *Journal of Econometrics* 37: 225–243.

Klepper, Steven, and Edward E. Leamer. 1984. "Consistent sets of Estimates for Regressions with Errors in All Variables." *Econometrica* 52: 163–183.

Leamer, Edward E., and Herman B. Leonard. 1983. "Reporting the Fragility of Regression Estimates," *The Review of Economics and Statistics* 65 (2): 306–317.

Levine, Ross, and David Renelt. 1992. “A Sensitivity Analysis of Cross-Country Growth Regressions.” *American Economic Review* 82 (4): 942–963.

Lubotsky, Darrin, and Martin Wittenberg. 2006. “Interpretation of Regressions with Multiple Proxies.” *Review of Economics and Statistics* 88 (3): 549–562.

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. “A Contribution to the Empirics of Economic Growth.” *Quarterly Journal of Economics* 107 (2): 407–437.

Neal, Derek A., and William R. Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences.” *Journal of Political Economy* 104 (5): 869–895.

Sala-i-Martin, Xavier X. 1997. “I Just Ran Two Million Regressions.” *American Economic Review* 87 (2): 178–183.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald Miller. 2004. “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” *American Economic Review* 94 (4): 813–835.

Solow, Robert M. 1956. “A Contribution to the Theory of Economic Growth.” *Quarterly Journal of Economics* 70 (1): 65–94.

Wittenberg, Martin. 2007. “Testing For A Common Latent Variable In A Linear Regression.” Working Paper, Available at SSRN: http://ssrn.com/abstract=978395.

Wooldridge, Jeffrey. 2010. *Econometric Analsyis of Cross Section and Panel Data,* second edition. Cambridge: MIT Press.