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

Christopher R. Bollinger Jenny Minier*

May 2007

Preliminary and Incomplete: Please Do Not Circulate or Quote without Permission

Abstract

This paper considers the most effective use of multiple proxy measures for the same unobserved variable. It extends the results of Lubotsky and Wittenberg (2006) to examine the impact of proxy variables on correctly measured variables. We find that including all proxy variables in the regression minimizes the bias on all other coefficients in the regression. Unlike previous results, estimates of coefficients on other regressors do not require a scaling assumption. We derive a set of bounds based on results from Klepper and Leamer (1984) and Bollinger (2003) for parameters in the model. These results are compared to Extreme Bounds Analysis. We find through Monte Carlo results that our bounds perform better than extreme bounds in most circumstances. We also find that our results may overturn many of the results found through extreme bounds analysis. 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. We find that the coefficient estimate on initial income is "robust," as previous extreme bound analyses have concluded. However, in contrast to 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.

*Both authors are Department of Economics, University of Kentucky, Lexington, KY 40506. Email: crboll@uky.edu and jminier@uky.edu respectively. We thank Helle Bunzel, Josh Ederington, Brent Krieder, John Pepper, Justin Tobias, Ken Troske, Jim Ziliak, and participants in seminars at Iowa State University, the University of Oregon, and the Southern Economic Association meetings for helpful comments and discussion.

1 Introduction

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

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

where the usual regression assumptions hold. We assume that the structural model is well specified; that is, the researcher is interested in the specific model above: in particular, estimation of $\underline{\alpha}$. However, the researcher does not observe the variable Z_{2i}, but only observes a set of variables $\underline{X}_i$ (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 Whittenberg (2006) where y_i is consumption and Z_{2i} is permanent income. In our example in Section 4, following the Solow model, y_i is economic growth, while $\underline{Z}_{1i}$ 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, p. 976), whose examples include 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 score (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 Whittenberg, 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. Many other studies use outcome measures to proxy for the stock, rather than the flow, of human capital, generally based on the school attainment measures most recently updated in Barro and Lee (2001).

Our approach is related in a sense to the model specification literature, including extreme bounds analysis (EBA), but in practice, we will show, provides very different results. It appears that the main conceptual difference between our approach and EBA is that EBA does not have a specific structural model in mind, but is interested in how $E\left[y_i|\underline{Z}_{1i}, \underline{X}_i\right]$ varies across different choices of $\underline{X}_i$. Researchers employing EBA assume that regressions of y_i on $\underline{Z}_{1i}$ and various combinations of variables for $\underline{X}_i$ are consistently estimated by OLS. In the growth literature, EBA (and other model specification tests) have been used most frequently as tests of "robustness": that is, identifying the coefficient estimates that are always (EBA) or nearly always (the distributional approach of Sala-i-Martin 1997 and Bayesian Model Averaging) of the same sign as the combination of $\underline{X}_i$ variables changes.

Implicit in the EBA approach is the assumption that one of the specifications is the "true" specification, so that the high and low estimates of the coefficient must contain the true estimate. We argue that this may not always be the most accurate approach when a variable that belongs in the estimation is unobserved. To elaborate on this distinction with our example, the human-capital-augmented neoclassical growth model tells us that "human capital" belongs in the growth regression, and we have three proxy variables related to investment in human capital or its stock: literacy rates, primary school enrollment rates, and secondary school enrollment rates. Taking EBA literally, some combination of these three proxy variables is the correct specification; the problem is that the researcher does not know the correct specification, and so can only bound the (other) coefficient estimates with the high and low estimates from all possible combinations of these measures. 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.

These two approaches sound similar, yet 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, simulations, and an empirical example, we show that our generalized proxy bounds tend to outperform EBA, particularly in the most relevant cases.

The analytic section develops three results. First, we extend the result of Lubotsky and Whittenberg (2006) (hereafter L-W) and focus upon the bias in estimation of $\underline{\alpha}$. Those results are related to Bollinger (2003), which considers the case where the dimensions of $\underline{X}_i$ are the same as the dimensions of Z_{2i}. We show that the minimum bias estimates of the parameters $\underline{\alpha}$ and β can be achieved from the results of the regression which includes all proxy variables (L-W showed the minimum bias on β, but did not explicitly examine the impact of the bias on $\underline{\alpha}$). L-W show that the minimum bias estimate of β is only estimable if there is an element of $\underline{\rho}$ that is equal to one: that is, they assume that $\rho_1 = 1$. Our results for β require the same assumption. However, we show that the minimum bias estimates for $\underline{\alpha}$ do not require this assumption. We further show that, as in Bollinger (2003), minimum bias estimates of β/ρ_1 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 ($\underline{\alpha}$, β). The minimum bias estimates of $\underline{(\alpha, \beta)}$ form one bound, and a reverse regression (like that used in Klepper and Leamer, 1984, and Bollinger, 2003) provides the other bound. We compare these results to extreme bounds analysis. What is important in comparing these results to the extreme bounds literature is that regressions which include only a subset of these variables have at least as large a bias as the regression which includes all proxy variables. We show that the extreme bounds approach will not provide bounds for the parameters $\underline{\alpha}$ (or β) in the model above, while our results do provide such bounds.

Next, we demonstrate 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. 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 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 follow the results of L-W and establish that there is a linear combination of proxy variables which simultaneously minimizes the bias on all coefficients in the model. However, forming this linear combination requires knowledge of unknown variances. Second, we show that 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 bias 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 bias estimate of the ratio of β/ρ_1 (if, as in L-W, $\rho_1 = 1$, this achieves a minimum bias estimate of β), 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 of Bollinger (2003).

The relationship between the observed proxies and the variable of interest is

$$\underline{X}_i = \underline{\rho} Z_{2i} + \underline{\varepsilon_i}. \tag{2}$$

The vector $\underline{X}_i$ contains multiple measures of the variable Z_{2i}. We assume that $Cov(\underline{\varepsilon_i}, Z_{2i}) = Cov(\underline{\varepsilon_i}, \underline{Z}_{1i}) = Cov(\underline{\varepsilon_i}, u_i)$, but $V(\underline{\varepsilon_i}) = \Sigma$, which is unknown, and no restrictions are placed upon it except that it be positive definite. These assumptions are relatively benign: The relationship expressed in equation 2 and the assumption that $Cov(\underline{\varepsilon_i}, Z_{2i}) = 0$ is simply the linear projection of $\underline{X}_i$ on Z_i and exists provided that both $\underline{X}_i$ and Z_i have finite first and second moments. The assumptions that $Cov(\underline{\varepsilon_i}, \underline{Z}_{1i}) = Cov(\underline{\varepsilon_i}, u_i) = 0$ simply state that, except as measures of Z_{2i}, there is no additional information contained in these proxy variable. We assume that the researcher observes $(y_i, \underline{Z}_{1i}, \underline{X}_i)$.

Like L-W, we begin by considering the problem of choosing a linear combination of $\underline{X}_i$ to minimize the bias on the resulting coefficient. That is, L-W are interested in the regression of y_i on $X_i^\delta = \underline{\delta}^T \underline{X}_i$: the problem is to choose $\underline{\delta}$ to minimize the bias 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,

$$X_i^\delta = \underline{\delta}^T \underline{X}_i = \underline{\delta}^T \underline{\rho} Z_{2i} + \underline{\delta}^T \underline{\varepsilon_i}.$$

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 $\gamma = \underline{\delta}' \underline{\rho} = 1$, then classical errors-in-variables results reveal that measurement error bias from the regression of y_i on $\underline{Z}_{1i}$ and X_i^δ is minimized when $V\left(e_i^\delta\right)$ 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 the expressions for the bias on $\underline{\alpha}$ and show that the result of L-W also minimizes the bias on all regressors in the model.

Proposition 1 *Define $\gamma = \underline{\delta}^T\underline{\rho} > 0$ and $\theta = \beta/\gamma$. Let $(\underline{a}^\delta, t^\delta)$ be the coefficients from the regression of y_i on $\underline{Z}_{1i}, X_i^{\delta(\gamma)}$ for any δ and a given value of γ. Then, $\underline{\delta} = \gamma(\Sigma^{-1}\underline{\rho})(\underline{\rho}^T\Sigma^{-1}\underline{\rho})$ solves both $\min_\delta \left(t^\delta - \theta\right)^2$ and $\min_\delta \left(a^\delta - \alpha\right)$ for $\gamma = \underline{\delta}^T\underline{\rho}$ if Σ is positive definite. If Σ is not full rank, then a solution may be achieved provided there exists a δ such that $\Sigma\delta = 0$ (which must be) and $\gamma = \underline{\delta}^T\underline{\rho}$. In this case, identification of α and β/γ is achieved.*

Proof. By Lemma 1 (see appendix) we can write

$$\left(\underline{a}^\delta - \underline{\alpha}\right) = \left(V_1 - \underline{C}V_2^{-1}\underline{C}'\right)^{-1}\underline{C}\left(\frac{V_2 - \underline{C}'V_1^{-1}\underline{C}}{V_2}\right)\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)\right) + (\underline{\delta}'\Sigma\underline{\delta})}\right)\beta \tag{4}$$

and

$$\left(t^\delta - \theta\right) = -\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)\right) + (\underline{\delta}'\Sigma\underline{\delta})}\right)\theta. \tag{5}$$

The common term, $\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)\right) + (\underline{\delta}'\Sigma\underline{\delta})}\right)$, is a scalar. Clearly, if Σ is not full rank, then there exists a δ so that $\left(\underline{\delta}^t\Sigma\underline{\delta}\right) = 0$. Provided that $\delta'\rho = \gamma > 0$ can also be solved, identification of all parameters can be achieved. When Σ is positive definite, the common term $\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)\right) + (\underline{\delta}'\Sigma\underline{\delta})}\right)$ is positive (see Lemma 2), hence the bias on any coefficient (as measured by any norm) is minimized when this term is minimized. It is trivial to show that $\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)\right) + (\underline{\delta}'\Sigma\underline{\delta})}\right)$ is increasing in the term $\left(\underline{\delta}^t\Sigma\underline{\delta}\right)$. Hence, choosing the bias-minimizing $\underline{\delta}$ is equivalent to solving

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

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

L-W derived the bias expression for t^{δ} when $\delta'\rho = \gamma = 1$, although they do also discuss the more general case. The bias minimization, relative to $\underline{\alpha}$ 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 bias from classical measurement error is determined by the variance of the error term, which in this case is $\mathrm{Var}(e_i^{\delta}) = \underline{\delta}'\Sigma\underline{\delta}$. Hence the goal in combining $X's$ is to choose a linear combination which 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 bias is minimized by choosing $\underline{\delta}$ 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 important 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 $\underline{X}_i$. Another reason is that if there exists a δ so that $\delta'\Sigma\delta = 0$, then the choice $\gamma = 1$ achieves no measurement error bias. 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 $\rho_1 = 1$.

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

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

We leave the proof of corollary 1 to the appendix and focus here upon the proof of the second corollary.

Proof. As noted above,

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

Substitution of the optimal choice of δ (give a value of γ) from Proposition 1 results in

$$(\underline{a}^{\delta}-\underline{\alpha})=\left(V_1-\underline{C}V_2^{-1}\underline{C}'\right)^{-1}\underline{C}\left(\frac{V_2-\underline{C}'V_1^{-1}\underline{C}}{V_2}\right)\left(\frac{\gamma^2\left(\underline{\rho}'\Sigma^{-1}\underline{\rho}\right)^{-1}}{\gamma^2\left(V_2-\underline{C}'V_1^{-1}\underline{C}\right)+\gamma^2\left(\underline{\rho}'\Sigma^{-1}\underline{\rho}\right)^{-1}}\right)\beta \tag{6}$$

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

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

The intuition is simple: unless $\Sigma = 0$, measurement error exists and severs the relationship between $\underline{Z}_{1i}$ 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 $\underline{Z}_{1i}$ must be rescaled as well.

It also important here to note that since the term $\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2-\underline{C}'V_1^{-1}\underline{C}\right)\right)+(\underline{\delta}'\Sigma\underline{\delta})}\right)$ is positive, the direction of the bias on each coefficient is determined solely by the sign of $\underline{C}$ and β. Note that if $\underline{C}=0$, there is no bias for $\underline{a}^{\delta}$, and that the magnitude of the bias is increasing in $\underline{C}$. The magnitude of the bias for different linear combinations of $\underline{X}_i$ is determined solely by $\left(\frac{(\underline{\delta}'\Sigma\underline{\delta})}{\left(\gamma^2\left(V_2-\underline{C}'V_1^{-1}\underline{C}\right)\right)+(\underline{\delta}'\Sigma\underline{\delta})}\right)$. 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 $\underline{\varepsilon}$), $\underline{C}$ (the covariance between $\underline{Z}_{1i}$ 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 $\underline{Z}_{1i}$ and $\underline{X}_i$ provides slope coefficients on $\underline{Z}_{1i}$ equivalent to the coefficients from the regression of y_i on $\underline{Z}_{1i}$ and X^{δ} for the optimal choice of δ. Further, a linear combination of the coefficients on $\underline{X}_i$ can be combined to achieve the minimum bias 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* $(\underline{a}, \underline{b})$ *be coefficients from the population least squares regression of* y_i *on* $\underline{Z}_{1i}$ *and* $\underline{X}_i$. *If* Σ *is non-singular, then* $\underline{a} = \underline{a}^\delta$ *and* $\underline{\rho}'\underline{b} = t^\delta$ *for* $\delta = \Sigma^{-1}\rho/\left(\rho'\Sigma^{-1}\rho\right)$.

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

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

Lubotsky and Wittenberg (2005) show that X^δ can be formed empirically if ρ_1, the first element of $\underline{\rho}$, is known to be 1. We note that this is a desirable case, but generalize to the more realistic case where this assumption cannot be made. L-W note that

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

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

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

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

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

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

Proposition 3 *The sign of* $\underline{\rho^{*\prime}\underline{b}}$ *is the sign of* β*. Further,* $\left|\underline{\rho^{*\prime}\underline{b}}\right| \leq |\beta| \leq |1/d|$ *and* $\alpha_j \in [a_j, -g_j/d]$.

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 $\underline{\alpha}$ 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 4, the bounds achieved will not contain any of the coefficients obtained through extreme bounds analysis: extreme bounds analysis provides the set of all biased coefficients that can be obtained. There is one special case in which extreme bounds analysis provides a set of consistent estimates of $\underline{\alpha}$: when $\underline{C} = 0$. In this case, any regressions of y_i on $\underline{Z}_{1i}$ and any (or no) elements of X_i provide consistent estimates of $\underline{\alpha}$.

A potential objection to the above procedure is that including many proxy variables will increase the standard errors on all variables. Although analytic results are still in progress, intuitively we note that including all proxy variables is equivalent to simply including the optimal X^{δ} linear combination. Indeed, the estimates can all be obtained through a two stage procedure: first regress y_i on $\underline{Z}_{1i}$ and $\underline{X}_i$. Using $\underline{b}$ and the estimated covariances, construct $\widetilde{X}^{\delta}$. Then regress y_i on $\underline{Z}_{1i}$ and $\widetilde{X}^{\delta}$. We claim that, like a plug-in two-stage least squares estimator, the sampling variance of the coefficients in this second stage are not affected by the sampling variance of the coefficients used to construct $\widetilde{X}^{\delta}$. Unlike 2SLS, there is no need to correct the standard errors as there is no prediction error; inclusion of the estimated X^{δ} is asymptotically equivalent to inclusion of the true X^{δ}. We also note empirically that the standard errors on a are nearly identical in the first and second stages. The intuition is relatively straightforward: standard errors on a are affected by the correlation between $\underline{Z}_{1i}$ and $\underline{X}_i$ or X^{δ}. These correlations are identical and are bounded above by the correlation between $\underline{Z}_{1i}$ and Z_{2i}. In the extreme, as $\underline{\delta}'\Sigma\underline{\delta}$ tends to zero, the standard errors on $\underline{a}$ will tend toward the standard errors that would be obtained if Z_{2i} were available. It should be noted that the standard errors on elements of b may be quite large, but these are not the standard errors of interest. Rather, the standard error on either $\rho' b$ or the coefficient on X^δ should be the focus: these also tend toward the standard errors that would be obtained if Z_{2i} were available. Formal proofs of these conjectures will be given in future drafts.

3 Simulation Results

In order to illustrate and evaluate the above results, we provide a set of simulations. We use the following model

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

We let

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

For simplicity, we generate $(v_i, \underline{e}_i, u_i)$ 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 correlation). 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 $\alpha = 1$ provides a standard on how large the bias from the regression of y_i on Z_{1i} and $\underline{X}_i$ (or subsets of $\underline{X}_i$) is likely to be. It also demonstrates how the extreme bounds approach can either over- or understate the coefficient on Z_{1i}, depending on C. Similarly the case of $\alpha = 0$ demonstrates that failure to include all X_i 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: $0, \pm 0.25, \pm 0.5, \pm 0.75$ 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. Panel A provides a great deal of intuition. In the first row, we present the proportion of times that the slope coefficient from the regression of y_i on only Z_{1i} would reject the null hypothesis that $\alpha = 1$, the true value (in other panels the test changes with values of α). This demonstrates the impact from omitted variable bias of not including Z_{2i}. As noted in Bollinger (2003), the inclusion of proxy variables mitigates omitted variable bias, but the bias on α (as can be seen in equation 6) is similar to omitted variable bias. As can be seen, when $C = 0$, the test accepts the null 95% of the time (as one would expect). For other values of C, failure to include any measure of Z_{2i} leads to rejection of the true null hypothesis in every sample.

The second row of panel A presents the test that $\alpha = 1$ when all proxy variables are included. As can be seen, when $C = 0$, this test also has the correct nominal size: there is no penalty to including the proxy variables. As C moves toward either 1 or -1, the nominal size falls. However, even at $C = \pm 0.75$, we accept the true null over 30% of the time.

The third row of panel A presents the percentage of simulations where the bounds derived in proposition 3 contain the true value of α. When $C = 0$, these generalized proxy bounds seldom contain the true value: the reverse regression is not informative about the coefficient on Z_{1i}, since it is identified in the direct regression. However, as the covariance gets higher, the percentage of times that the proxy bounds contain the true coefficient also tends toward 1. At $C = \pm 0.25$, the proxy bounds capture the true coefficient over 70% of the time. The fourth row of the panel shows that it is only sampling variance which prevents the proxy bounds from capturing the true coefficient at least 95% of the time. In the case where $C = 0$, the inclusion of the standard errors captures the true coefficient 95% of the time, while at higher values of C, it is 99% of the time or higher.

The fifth row presents the percentage of times that extreme bounds analysis would capture the true coefficient α. In the row labelled "xbnd1," the bounds are formed as the maximum coefficient estimate plus two times the standard error, and the minimum estimate minus two times the standard error. Notice that when $C = 0$, this actually performs better than not including the proxy variables at all. In contrast to the proxy bounds, the extreme bounds perform less well as the correlation between Z_{1i} and Z_{2i} increases.

The sixth row presents an alternative version of extreme bounds (labelled "xbnd2," where the bounds are defined as the maximum and minimum estimates from the EBA estimation. As one would expect, since these are narrower bounds, these perform decidedly worse than the first extreme bounds case, and very poorly compared to the proxy bounds. When sampling variance is included in the extreme bounds 2 case, by adding and substracting 2 times the standard errors of the bounds, the performance mimics that of the first extreme bounds case (as one would expect).

In panels E-H, we double the amount of noise in the proxy variables. This actually improves the performance of the generalized proxy bounds in an absolute sense, in that the proportion of times that the proxy bounds include the true coefficient on α increases for all values of C. In sharp contrast, this decreases the performance of the extreme bounds analysis. For some perspective on this, we note that when $s = 1$, the correlation between X^δ (the optimal linear combination of the proxy variables) and Z_{2i} is 0.93 when $s = 1$ and falls to 0.77 when $s = 1.4$. Indeed, the correlation between X_{3i} and Z_{2i} is 0.9 and 0.81 for the two cases respectively. These are cases where the noise in the proxy variables is quite low relative to the signal. Clearly, in cases where the noise is much higher, extreme bounds will likely perform even more poorly.

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 unobservable variables such as "technology" or "human capital."

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 Classical Model Averaging of Doppelhofer, Miller, and Sala-i-Martin (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 the 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.

Our goals in this section are more modest than identifying "robustness" in growth regressions: we use this application to illustrate the difference between extreme bounds and proxy bounds with a structural estimation of the neoclassical growth model including human capital. In future drafts we anticipate adapting this approach to robustness tests of this kind.

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:

$$Y(t) = K(t)^{\alpha} H(t)^{\beta} (A(t)L(t))^{1-\alpha-\beta} \tag{7}$$

where Y is aggregate output, K is the (physical) capital stock, H is the stock of human capital, A represents (labor-augmenting) technological progress, and L is the labor force. Income is invested in physical and human capital at the constant fractions s_k and s_h, respectively.

Allowing lower-case letters to denote quantities in terms of effective units of labor (i.e., $y = Y/AL$, $k = K/AL$, and $h = H/AL$), the economy evolves following:

$$\dot{k}(t) = s_k y(t) - (n + g + \delta)k(t); \tag{8}$$
$$\dot{h}(t) = s_h y(t) - (n + g + \delta)h(t)$$

where δ denotes the depreciation rate (assumed identical for human and physical capital), n is the population growth rate, and g is the rate of exogenous technological progress (i.e., the rate at which A grows).

The equations given in (8) imply that there is a steady state characterized by:

$$k^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n + g + \delta}\right)^{1/(1-\alpha-\beta)} ; \qquad h^* = \left(\frac{s_k^{\alpha} s_h^{a-\alpha}}{n + g + \delta}\right)^{1/(1-\alpha-\beta)} \tag{9}$$

In the steady state, income per capita $y^*$is given by:

$$\ln(y^*) = \ln A(0) + gt - \frac{\alpha + \beta}{1 - \alpha - \beta}\ln(n + g + \delta) + \frac{\alpha}{1 - \alpha - \beta}\ln(s_k) + \frac{\beta}{1 - \alpha - \beta}\ln(s_h). \tag{10}$$

However, equation (10) only describes income in the steady state. As economies converge toward their steady-state values of income, the Solow model implies what has come to be known as conditional convergence, or convergence (of income) conditional on the determinants of the steady state.

The speed at which a country converges from its income at time t to its steady state y^* can be approximated by:

$$\frac{d\ln(y(t))}{dt} = \lambda[\ln(y^*) - ln(y(t))] \tag{11}$$

where $\lambda = (n + g + \delta)(1 - \alpha - \beta)$.

Equation (11) implies that:

$$\ln(y(t)) = (1 - e^{-\lambda t})\ln(y^*) + e^{-\lambda t}\ln(y(0)), \tag{12}$$

where $y(0)$ is initial income per effective worker. Subtracting $\ln(y(0))$ from both sides and substituting for y^* yields a regression that can be estimated to measure the actual rate of convergence:

$$\begin{aligned}\ln(y(t)) - \ln(y(0)) = (1 - e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\ln(s_k) + (1 - e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\ln(s_h) \\ - (1 - e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) - (1 - e^{-\lambda t})\ln(y(0)).\end{aligned} \tag{13}$$

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 (13) is the key equation for our application, and 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 in physical capital is the investment/GDP ratio, also from the Penn World Tables. The annual rate of population growth is also taken from the Penn World Tables, and we follow Mankiw et al. (1992) in setting $(g + \delta) = 0.5$. We believe that these variables are measured correctly; the key question is how to accurately measure "human capital." (Mankiw et al. (1992), among many others, discuss this issue at length.) 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 rate, primary school enrollment rates, and secondary school enrollment rates. Our sample is a cross-section of 88 countries at all levels of development.

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

$$y_i = \underline{\alpha}' \underline{Z}_{1i} + \beta Z_{2i} + u_i. \tag{14}$$

Z_{2i} is not measured directly, but a vector $\underline{X}_i$ of correctly measured variables exists such that $\underline{X}_i = \underline{\rho} Z_{2i} + \underline{\epsilon}_i$. In this application $\underline{X}_i$ contains primary school enrollment, secondary school enrollment, and literacy rates. $\underline{Z}_1$ 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 $\underline{\alpha}$. Theory defines the structural relationship. The problem is that we do not have a measure of human capital (Z_{2i}) but rather 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 key 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.[1] After estimating the regression with each possible combination of the three human capital proxy variables (yielding seven regressions), we compute the upper and lower bounds for each of the correctly measured ($\underline{Z}_1$) regressors (initial GDP, investment, and population growth) as the maximum and minimum values of $\beta \pm 2\sigma$, which is also the cutoff used by Levine and Renelt (1992). Table 2 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).

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 slightly lower than the estimate of 0.014 in the full 98-country sample in Mankiw et al. (1992).[2] These estimates of λ imply that a country moves halfway toward its steady state in between 70 and 87 years.

[1] Because Levine and Renelt included over 30 possible 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).

[2] The Mankiw et al. sample covers the period 1960-85, which is extended here to 1960-2000. Their measure of human capital is secondary school enrollment rates.

4.2 Generalized Proxy Bounds

A problem with extreme bounds analysis is that when the control variable is measured with error, the estimates on the correctly measured variables are not guaranteed to be within the extreme bounds; in fact, by Proposition 2 and Corollary 3, the extreme bounds will *not* include the true estimates.

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:

$$y_i = \underline{\beta}'_{Z1}\underline{Z}_{1i} + \underline{\beta}'_{X}\underline{X}_i + \epsilon_i \tag{15}$$

and retain the $\underline{\beta}_{Z1}$ and $\underline{\beta}_X$ coefficients. Following Propositions 2 and 3, the estimates of $\underline{\beta}_{Z1}$ identify one of the bounds for each of the $\underline{Z}_1$ 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 $\underline{\rho}$. There are four consistent estimates for ρ_2 and ρ_3: for example, ρ_2 could be estimated by $\frac{\mathrm{cov}(x_2,y)}{\mathrm{cov}(x_1,y)}$, $\frac{\mathrm{cov}(x_2,z_1)}{\mathrm{cov}(x_1,z_1)}$, $\frac{\mathrm{cov}(x_2,z_2)}{\mathrm{cov}(x_1,z_2)}$, or $\frac{\mathrm{cov}(x_2,z_3)}{\mathrm{cov}(x_1,z_3)}$. 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:

$$B_1 = \beta_{X1} + \rho_2 \cdot \beta_{X2} + \rho_3 \cdot \beta_{X3} \tag{16}$$

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

$$\hat{Z}_{2i} = \frac{\beta_{X1} + \beta_{X2}X_{2i} + \beta_{X3}X_{3i}}{B_1} \tag{17}$$

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

$$\hat{Z}_{2i} = \delta_y y_i + \underline{\delta}'_{Z1}\underline{Z}_{1i} + \nu_i \tag{18}$$

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

$$B_2 = 1/\delta_y \tag{19}$$

Finally, the second bound for the $\underline{Z}_1$ variables is given by:

$$-\delta_{Z1j}/\delta_y \tag{20}$$

The results for the generalized proxy bounds are in Table 3. Several comparisons to the extreme bounds analysis in Table 2 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, in that investment 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). Second, the coefficient estimates on the population growth term $(n + g + \delta)$ are statistically significant and negative at both bounds, and are larger in magnitude than the estimates from the extreme bounds analysis. Although the extreme bounds analysis in Table 2 also yielded coefficient estimates that would be considered robust, population growth was *not* a robust variable in Levine and Renelt (1992).

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 3 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.)

5 Conclusions

We have demonstrated that when a researcher has a series of proxy variables which are thought to be correlated with an unobserved regressor from a structural model, including all proxy variables in the regression will result in estimates which are the least biased of any combination of the proxy variables. Further, we provide an approach which derives bounds on all coefficients in the model. The extreme bounds literature has a similar goal: when a series of proxy variables is available, they examine coefficients on all possible combinations. We show that this may lead to conclusions about the robustness of results that are unwarranted.

Our results are preliminary at this stage. One important issue is how the inclusion of many proxy variables will affect the standard errors of estimates of structural coefficients. We plan on examining this issue more closely in the future. This question suggests that the focus on bias in our section 2 and in L-W is only part of the story; the focus should be on mean squared error. Our future research will address this issue.

A second issue is further reconciling our results with the model specification literature (including Bayesian model averaging techniques as well as EBA), particularly as it has been applied in the empirical growth literature to address the question of which variables are "robustly" correlated with growth.

Table 1: Simulation Results

Panel A: a=1; s=1					
	C=0	C=0.25	C=0.5	C=0.75	C=0.9
shorthattest	0.958	0.000	0.000	0.000	0.000
minbiastest	0.950	0.924	0.788	0.352	0.016
obtestt	0.250	0.700	0.886	0.988	1.000
obtestse	0.946	0.990	1.000	1.000	1.000
xbnd1testt	0.994	0.942	0.806	0.374	0.022
xbnd2testt	0.266	0.304	0.114	0.012	0.000
xbnd2testtse	0.992	0.938	0.790	0.360	0.022
Panel B: a=1; s=1					
		C=–0.25	C=–0.5	C=–0.75	C=–0.9
shorthattest		0.000	0.000	0.000	0.000
minbiastest		0.922	0.722	0.336	0.014
obtestt		0.722	0.882	0.992	1.000
obtestse		0.996	1.000	1.000	1.000
xbnd1testt		0.938	0.740	0.350	0.018
xbnd2testt		0.292	0.118	0.008	0.000
xbnd2testtse		0.932	0.724	0.338	0.014
Panel C: a=0; s=1					
	C=0	C=0.25	C=0.5	C=0.75	C=0.9
shorthattest	0.954	0.000	0.000	0.000	0.000
minbiastest	0.948	0.898	0.738	0.340	0.016
obtestt	0.276	0.720	0.900	0.994	1.000
obtestse	0.966	0.994	0.998	1.000	1.000
xbnd1testt	0.976	0.916	0.758	0.356	0.020
xbnd2testt	0.278	0.284	0.100	0.006	0.000
xbnd2testtse	0.974	0.908	0.744	0.342	0.018
Panel D: a=0; s=1					
		C=–0.25	C=–0.5	C=–0.75	C=–0.9
shorthattest		0.000	0.000	0.000	0.000
minbiastest		0.884	0.728	0.334	0.018
obtestt		0.714	0.928	0.992	1.000
obtestse		0.992	0.998	1.000	1.000
xbnd1testt		0.898	0.750	0.364	0.020
xbnd2testt		0.290	0.072	0.008	0.000
xbnd2testtse		0.896	0.734	0.336	0.020

Table continued on following page

Table 1, continued

Panel E: a=1; s=1.4					
	C=0	C=0.25	C=0.5	C=0.75	C=0.9
shorthattest	0.944	0.000	0.000	0.000	0.000
minbiastest	0.948	0.850	0.406	0.012	0.000
obtestt	0.286	0.868	0.994	1.000	1.000
obtestse	0.912	1.000	1.000	1.000	1.000
xbnd1testt	0.984	0.860	0.428	0.012	0.000
xbnd2testt	0.270	0.136	0.006	0.000	0.000
xbnd2testtse	0.980	0.852	0.416	0.012	0.000
Panel F: a=1; s=1.4					
		C=–0.25	C=–0.5	C=–0.75	C=–0.9
shorthattest		0.000	0.000	0.000	0.000
minbiastest		0.818	0.420	0.018	0.000
obtestt		0.824	0.986	1.000	1.000
obtestse		1.000	1.000	1.000	1.000
xbnd1testt		0.838	0.440	0.018	0.000
xbnd2testt		0.180	0.014	0.000	0.000
xbnd2testtse		0.830	0.422	0.018	0.000
Panel G: a=0; s=1.4					
	C=0	C=0.25	C=0.5	C=0.75	C=0.9
shorthattest	0.952	0.000	0.000	0.000	0.000
minbiastest	0.958	0.816	0.378	0.018	0.000
obtestt	0.302	0.842	0.988	1.000	1.000
obtestse	0.980	0.996	1.000	1.000	1.000
xbnd1testt	0.988	0.838	0.400	0.020	0.000
xbnd2testt	0.300	0.166	0.014	0.000	0.000
xbnd2testtse	0.982	0.828	0.382	0.018	0.000
Panel H: a=0; s=1.4					
		C=–0.25	C=–0.5	C=–0.75	C=–0.9
shorthattest		0.000	0.000	0.000	0.000
minbiastest		0.836	0.394	0.008	0.000
obtestt		0.856	0.990	1.000	1.000
obtestse		1.000	1.000	1.000	1.000
xbnd1testt		0.844	0.404	0.008	0.000
xbnd2testt		0.154	0.010	0.000	0.000
xbnd2testtse		0.838	0.394	0.008	0.000

Notes to Table:

Table 2: Extreme Bounds Analysis

		β (s.e)	p-value	Bound	HK measures
$\ln(GDP)$	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
$\ln(n+g+\delta)$	upper:	-2.36 (0.71)	0.001	-0.94	lit
	lower:	-2.98 (0.68)	0.000	-4.33	lit, pri, sec

Notes to Table: The table reports the coefficient estimates, standard errors, and p-values associated with each bound. The bounds are defined as $\beta \pm 2\sigma$.

Table 3: Generalized Proxy Bounds

		β (s.e)	p-value	Bound
$\ln(GDP)$	upper:	-0.47 (0.09)	0.000	-0.29
	lower:	-1.14 (0.20)	0.000	-1.54
$\ln(INV)$	upper:	0.29 (0.09)	0.002	0.47
	lower:	-0.31 (0.20)	0.062	-0.71
$\ln(n+g+\delta)$	upper:	-2.98 (0.64)	0.000	-1.70
	lower:	-3.69 (1.32)	0.003	-6.33

Notes to Table: The table reports the coefficient estimates, standard errors, and p-values associated with each bound. The bounds are defined as $\beta \pm 2\sigma$.

References

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

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

[3] Doppelhofer, Gernot, Ronald Miller and Xavier Sala-i-Martin (2004), "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach" *American Economic Review* 94(4): 813-35.

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

[5] Klepper, Steven (1988), "Bounding the Effects of Measurement Error in Regressions Involving Dichotomous Variables," *Journal of Econometrics* 37: 343-359.

[6] Klepper, Steven and Edward E. Leamer (1984), "Consistent sets of Estimates for Regressions with Errors in All Variables," *Econometrica* 52: 163-183.

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

[8] Lubotsky, Darrin and Martin Whittenberg (2006), "Interpretation of Regressions with Multiple Proxies," *Review of Economics and Statistics* 88(3): 549-562.

[9] 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.

[10] 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.

[11] Sala-i-Martin, Xavier X. (1997), "I Just Ran Two Million Regressions," *American Economic Review* 87(2): 178-83.

6 Appendix

Let

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

Let $\underline{\delta}$ be an arbitrary vector such that $\underline{\delta}'\underline{\rho} = \dot{\gamma} > 0$ for some given value γ. Let $\theta = \beta/\gamma$.

The next three Lemmas establish key results for Proposition 3.

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

Then

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

Rewriting yields

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

which is equivalent to

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

This yields

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

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

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

Rearranging gives

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

Thus

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

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

and

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

QED.

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

The term $\left(\gamma^2\left(V_2-\underline{C}'V_1^{-1}\underline{C}\right)\right)+(\underline{\delta}'\Sigma\underline{\delta})$ is positive provided that $\gamma \neq 0$ and Σ is positive semi-definite. The term $V_1V_2 - C^2$ is the determinant of the $\mathrm{V}(Z_{1i}, Z_{2i})$, and so is, by necessary assumption, positive. The term $(\underline{\delta}'\Sigma\underline{\delta})$ will be non-negative provided Σ is positive semi-definite and V_1 is also non-negative. The respective derivatives, with respect to the term $(\underline{\delta}'\Sigma\underline{\delta})$ are

$$\begin{aligned}&\frac{\gamma\left(\gamma^2V_1V_2-\gamma^2C^2+V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)\right)-\gamma\left(\underline{\delta}'\Sigma\underline{\delta}\right)V_1}{\left(\gamma^2V_1V_2-\gamma^2C^2+V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)\right)^2} \\ = &\frac{\gamma^3V_1V_2-\gamma^3C^2}{\left(\gamma^2V_1V_2-\gamma^2C^2+V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)\right)^2} > 0\end{aligned}$$

and

$$\frac{V_1\left(\gamma^2V_1V_2-\gamma^2C^2+V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)\right)-V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)V_1}{\left(\gamma^2V_1V_2-\gamma^2C^2+V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)\right)^2} = \frac{V_1\left(\gamma^2V_1V_2-\gamma^2C^2\right)}{\left(\gamma^2V_1V_2-\gamma^2C^2+V_1\left(\underline{\delta}'\Sigma\underline{\delta}\right)\right)^2} > 0.$$

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

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

The Lagrangian is

$$(\underline{\delta}'\Sigma\underline{\delta}) - \lambda\left(\underline{\delta}'\underline{\rho} - \gamma\right).$$

FOC are

$$\begin{aligned} 2\Sigma\underline{\delta} - \lambda\underline{\rho} &= 0 \\ \underline{\delta}'\underline{\rho} - \gamma &= 0. \end{aligned}$$

Solving:

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

Substitution yields

$$\begin{aligned} \underline{\delta} &= \gamma\Sigma^{-1}\underline{\rho}\left(\underline{\rho}'\Sigma^{-1}\underline{\rho}\right)^{-1} \\ \lambda &= 2\gamma\left(\underline{\rho}'\Sigma^{-1}\underline{\rho}\right)^{-1}. \end{aligned}$$

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

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

From Lemma 1 we have that

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

Alternatively,

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

$$= \frac{\beta}{\gamma}\left(\frac{\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)}{\left(\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)\right) + \left(\underline{\delta}'\Sigma\underline{\delta}\right)}\right).$$

Substitute the optimal choice of δ from proposition 1 which yields

$$t = \frac{\beta}{\gamma}\left(\frac{\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right)}{\gamma^2\left(V_2 - \underline{C}'V_1^{-1}\underline{C}\right) + \gamma^2\left(\underline{\rho}'\Sigma^{-1}\underline{\rho}\right)^{-1}}\right)$$

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

Hence, by choosing

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

we have $t = \beta$: no bias 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 $(A + XBX')^{-1} = A^{-1} - A^{-1}X\left(B^{-1} + X'A^{-1}X\right)^{-1}X'A^{-1}$.*

Proof. Proof of Proposition 2:

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

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

Rewriting yields

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

which is equivalent to

$$\begin{bmatrix} V_1 & \underline{C}\underline{\rho}' \\ \underline{\rho}\underline{C}' & I\underline{\rho}'\underline{\rho}V_2 \end{bmatrix}^{-1} \begin{bmatrix} V_1 & \underline{C}\underline{\rho}' \\ \underline{\rho}\underline{C}' & \left(\underline{\rho}\underline{\rho}'V_2 + \Sigma\right) \end{bmatrix} \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix}$$

$$= \begin{bmatrix} V_1 & \underline{C}\underline{\rho}' \\ \underline{\rho}\underline{C}' & I\underline{\rho}'\underline{\rho} \end{bmatrix}^{-1} \begin{bmatrix} V_1\underline{\alpha} + \underline{C}\beta \\ \underline{\rho}\underline{C}'\underline{\alpha} + \underline{\rho}V_2\beta \end{bmatrix},$$

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

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

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

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

Substitution and simplification yields

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

or

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

We can write

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

and

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

Turning first to the term $\underline{a}$ and applying the Sherwin-Morrison_Woodbury Matrix

Inversion Lemma:

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

Simplification yields

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

or

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

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

Turning now to $\underline{b}$, consider

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

Again using the Sherwin-Morrison_Woodbury Matrix Inversion Lemma,

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

$$= \left(\underline{\rho}'\Sigma^{-1}\underline{\rho} - \underline{\rho}'\Sigma^{-1}\underline{\rho}\left(\left(V_2 - \underline{C}' V_1^{-1} \underline{C}\right)^{-1} + \underline{\rho}'\Sigma^{-1}\underline{\rho}\right) \underline{\rho}'\Sigma^{-1}\underline{\rho}\right) \left(V_2 - \underline{C}' V_1^{-1} \underline{C}\right) \beta$$

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

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

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