centre for microdata methods and practice

Identification of a triangular two equation system without instruments

Arthur Lewbel
Susanne M. Schennach
Linqi Zhang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP41/20

Economic
and Social
Research Council

Identification of a Triangular Two Equation System Without Instruments

Arthur Lewbel, Susanne M. Schennach, and Linqi Zhang*
Boston College, Brown University, and Boston College

August 2020

Abstract

We show that a standard linear triangular two equation system can be point identified, without the use of instruments or any other side information. We find that the only case where the model is not point identified is when a latent variable that causes endogeneity is normally distributed. In this non-identified case, we derive the sharp identified set. We apply our results to Acemoglu and Johnson's (2007) model of life expectancy and GDP, obtaining point identification and comparable estimates to theirs, without using their (or any other) instrument.

1 Introduction

Consider a standard linear triangular structural model

$$Y = X'b_1 + \varepsilon_1 \tag{1}$$

$$W = \gamma Y + X'b_2 + \varepsilon_2 \tag{2}$$

for some endogenous variables Y and W, exogenous covariates X, and unobserved errors ε_1 and ε_2. For example, W could be a worker's wages or earnings and Y could be her level of schooling. Or, as in our later empirical application, W could be a country's GDP growth and Y a health measure like growth in life expectancy. The primary goal is identification of

****JEL codes**: C14, C30, **Keywords**: Returns to schooling, identification, triangular Systems, Kotlarski, deconvolution. Corresponding Author: Arthur Lewbel, Department of Economics - Maloney 315, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, https://sites.google.com/bc.edu/arthur-lewbel/ Susanne Schennach acknowledges support from NSF grant SES-1950969. Vincent Starck is gratefully acknowledged for valuable comments.

γ, the direct causal effect of Y on W, though we will also obtain identification of b_1, b_2, and the joint distribution of the errors.[1]

The main obstacle to identification and estimation of γ is that ε_1 and ε_2 may be correlated, because both depend on a common unobserved U (ability in the case of schooling and wages, technology in the case of GDP and health). That is, in its simplest form,

$$\varepsilon_1 = U + V \quad \text{and} \quad \varepsilon_2 = \beta U + R \tag{3}$$

where U, V, and R are unobserved, mutually independent (conditional on X) random variables and β is a constant. After projecting off covariates X, the V and R errors represent idiosyncratic shocks to Y and W, while U is what makes Y an endogenous regressor in the W equation.

Similar triangular structural models arise whenever we have one variable Y affecting another variable W, and a common unobservable that affects them both. For example, consider a two period dynamic model with autocorrelated errors. In this case W equals Y in a subsequent time period, and U represents the autocorrelation in the errors. Another example is production, where W could be a firm's value-added output per unit of capital, Y is the firm's labor per unit of capital, and U is unobserved entrepreneurship, which affects both productivity and the chosen level of inputs.

Such models are traditionally identified in econometrics by finding an instrument, i.e., a variable that correlates with Y but not ε_2, or equivalently, a variable that correlates with V but not U or R. However, such instruments can be difficult to find. For example, Card (1995, 2002) and others propose using measures of access to schooling, such as distance to or cost of colleges in one's area, as wage equation instruments, while others raise objections to the validity of these instruments, e.g. Carneiro and Heckman (2002). Other wage equation instruments may raise fewer questions of validity but can be weak, like Angrist and Krueger's (1991, 2001) quarter of birth instruments.

Similarly, Acemoglu and Johnson (2007) propose using changes in predicted mortality, constructed based on innovations in health care, as an instrument for life expectancy growth Y in their regression of GDP growth W on Y. However, such health innovations could be correlated with other technological advances that increase GDP, leading to instrument invalidity. Comparable questions can be raised regarding the instruments or identifying side information in other similar studies, such as Aghion, Howitt, and Murtin (2010), who find a positive γ, in contrast to Acemoglu and Johnson's (2007) negative γ. Ecevit (2013) summarizes results from eleven similar studies, finding estimates of γ that range from strongly negative to insignificant to strongly positive. This range of estimates raises serious questions regarding the validity of instruments or other side information that different authors use to identify γ.[2]

Rather than propose any new instrument, we address the more fundamental question of whether and when this model can be point identified and estimated *without* side information such as instruments whose validity can be hard to ascertain (noting that the alternative

[1]Throughout this paper we focus on the traditional homogeneous effects model where γ is a constant, rather than a heterogeneous treatment effects model.

[2]Of course, differences are also due to variation in data sets and in how Y and W are defined and constructed. As another way to explain these differing results, Cervellati and Sunde (2011) suggest that the true effect might be non-monotonic.

of a randomized experiment is not feasible for a macro question like this). If so, then we can estimate the model without relying on side information, and/or test the validity of side information like instruments via overidentification tests.

We provide conditions for point identification of the model

$$Y = U + V \tag{4}$$

$$W = \gamma Y + \beta U + R \tag{5}$$

with U, V, and R being unobserved, mutually independent random variables with unknown distributions. The same identification theorem can then be applied conditioning on covariates X, to show point identification of more general models, because the entire distributions of U, V, and R could depend nonparametrically on X. A special case of this general identification result is then identification of equations (1), (2) and (3). In this special case, variables V and R that depend nonparametrically on X in equations (4) and (5) are instead replaced with $X'\beta_1 + V$ and $X'\beta_2 + R$, where these new V and R do not depend on X.

Our main result is surprising: under minimal regularity assumptions, the coefficients γ and β, and the distributions of U, V, and R (and b_1 and b_2 in that model) are all point identified without instruments or other side information, unless either U or V is normally distributed (after appropriately conditioning on or projecting off covariates X).

In addition to proving this general identification result, we also: 1. Provide a few low order moments yielding simple GMM estimators of the model, 2. Show how infinitely many additional moments conditions can be systematically constructed to provide identification under weaker conditions, 3. Provide the sharp identified set for the coefficients γ and β in the case where either U or V is normal and hence point identification fails, 4. Investigate the behavior of these GMM estimators in some Monte Carlo exercises, and 5. Provide an empirical application where we establish that our identification and estimation strategy is viable even with a very small sample size. Specifically, we estimate the Acemoglu and Johnson (2007) model without using any instruments, and obtain estimates that are very similar to what they found with their instrument.

The identification of equations (4) and (5) without instruments has been previously considered by Rigobon (2003), Klein and Vella (2010), and Lewbel (2012), but these results neither nest nor are nested by ours because they *require* that the errors be heteroskedastic, and identification is obtained by imposing varying restrictions on the structure of that heteroskedasticity.[3]

A number of special cases of our results do appear in the literature, but all of them assume $\gamma = 0$, and so they omit the most important feature of the model in applications like ours. Kotlarski (1967) is the special case of our model where it is known that $\gamma = 0$ and $\beta = 1$, and in that case Kotlarski's Lemma shows that point identification of the distribution of all the latent variables holds even under normality. Similarly, Reiersøl (1950) uses a special case of our model where it is known that $\gamma = 0$ and Y plays the role of a measurement of U contaminated by an error V and establishes conditions under which β would be identified. As noted in Lewbel (2020), with $\gamma = 0$ and Reiersøl's identification of β, one could rewrite

[3] Rigobon(2003) and Klein and Vella (2010) impose different parametric restrictions on the error variances, while Lewbel (2012) imposes a nonparametric restriction.

Reiersøl's model as $Y = U + V$ and $W/\beta = U + R/\beta$, and then apply Kotlarski's lemma to the joint distribution of Y and W/β to identify the distributions of U, V, and R.[4]

Our results, showing necessary and sufficient conditions to identify the more general model of equations (4) and (5) with unknown nonzero γ, turns out to be a difficult extension. In particular, the methods of proof used by Reiersøl (1950) and Kotlarski (1967) do not extend to our problem. Moreover, due to the presence of γ, the condition our model requires for point identification (non-normality of both U and V) turns out to be much simpler than Reiersøl's (1950) condition, which depends on the concept of factors of a distribution (this comparison is discussed in detail later).

In section 2, we provide a few simple moments that will often suffice to point identify our model, and can be used to construct a correspondingly simple GMM estimator. In Section 3, we present our general identification results, including constructing more moments like those in Section 2, and showing that, with minimal regularity, the model is point identified as long as both U and V are not normal. In sections 4 and 5 we derive the sharp identified set when either U or V is normal, and derive some inequalities regarding our model relative to ordinary least squares. Section 6 provides a Monte Carlo analysis of our simple GMM estimators. In section 7 we provide an empirical application based on Acemoglu and Johnson (2007), in which we obtain estimates comparable to theirs, without using their (or any other) instrument. Section 8 concludes with some suggestions for further work.

2 Simple Identification and Estimation

We begin with a simple special case of our general results, by providing some moments that can easily be used to identify and estimate (by standard GMM) the models described in the introduction. These results are not as general as our main identification theorem, but are likely to suffice for many empirical applications.

We first consider identification and estimation of equations (4) and (5) without covariates X, and then we extend the results to equations (1) and (2).

Assumption 1 *We observe the joint distribution of two real valued, nondegenerate random variables Y and W.*

With data, we could assume independent, identically distributed observations of Y and W, and then identify their joint distribution to satisfy Assumption 1 using the Glivenko Cantelli theorem.

[4] A special case of non-normality is when the components U and V are asymmetric. Lewbel (1997) and Erickson and Whited (2002) exploit asymmetry to construct simple estimators for the Reiersøl (1950) model. Other papers propose estimators for models like equations (4) and (5) with $\gamma = 0$, by assuming that coefficients like β are point identified using higher moments, but without explicitly characterizing when that is possible. Examples include Bonhomme and Robin (2010), Fruehwirth, Navarro, and Takahashi (2016), and Navarro and Zhou (2017). Generalizations of Kotlarski's lemma to models with more components (but again still assuming $\gamma = 0$) include Székely and Rao (2000) and Li and Zheng (2020). A nonlinear extension of Reiersøl (1950) is Schennach and Hu (2013).

Assumption 2 *The unobserved real valued random variables U, V, and R are mean zero and mutually independent, with unknown distributions.*

Assumption 3 *R has finite variance, and U and V each have finite fourth moments.*

Assumption 4 *The unknown constants γ and β are real valued, finite, and $\beta > 0$.*

We can assume our data Y and W have been demeaned, rationalizing the assumption that the unobservables have mean zero. To see why we need a sign restriction on β, observe that we can rearrange equations (4) and (5) to get $W = (\gamma + \beta) Y - \beta V + R$, which, except for the sign of β, is observationally equivalent to the original model, switching the roles of V and U. Usually, the sign of β should be clear from the economics of the application, e.g., in a returns to schooling model, $\beta > 0$ is a natural assumption, since it says that unobserved ability that increases (decreases) education outcomes will increase (decrease) wages. If we instead believed β was negative, we could just replace Y with $-Y$ everywhere to make β positive (redefining γ, U, and V accordingly).

We also rule out $\beta = 0$, because if $\beta = 0$ then trivially we could not separately identify V and U. Moreover, having $\beta = 0$ would be nonsensical in the types of applications we consider, since it would mean that Y is exogenous, making identification and estimation of γ trivial.

Substituting equation (4) into equation (5) gives the reduced form expression for W

$$W = \gamma V + \alpha U + R \quad \text{with} \quad \alpha = \gamma + \beta \tag{6}$$

The following Lemma provides two moments that can often suffice to point identify γ and α, which then trivially also point identifies β.

Lemma 1 *Let Assumptions 1-4 and equations (4) and (5) (and therefore also equation 6) hold. Then*

$$E\left[(W - \gamma Y)(W - \alpha Y) Y\right] = 0 \tag{7}$$

$$cov\left[(W - \gamma Y)(W - \alpha Y), Y^2\right] = 2E\left(WY - \gamma Y^2\right) E\left(WY - \alpha Y^2\right) \tag{8}$$

Proofs are all in Appendix A. The proof of Lemma 1 works by substituting $W - \gamma Y = \beta U + R$ and $W - \alpha Y = -\beta V + R$ into equations (7) and (8), and then uses the mutual independence of U, V, and R to verify that these equations hold.

Lemma 1 provides two equations in the two unknowns α and γ. If we solve the first equation for α and substitute that into the second, we obtain a quadratic in γ. The sign restriction that $\beta > 0$ then determines which root is the correct one for γ.

We later provide the formal conditions under which these two equations suffice to point identify α and γ. The main condition, derived in Theorem 1 below, is equation (21). Equation (21) shows that the main cases in which equations (7) and (8) by themselves fail to provide point identification are when U and V have the exact same distribution, or when both are symmetrically distributed, or if either U or V is normally distributed. We later show that infinitely many additional equations in α, γ, Y and W can be constructed, based on higher moments of Y and W than those used in Lemma 1. These higher moments can help identify α and γ in applications where Lemma 1 does not suffice.

A simple estimator for α and β can be constructed by rewriting equations (7) and (8) as moment conditions, and applying standard method of moments or GMM. One can immediately check that these equations take the form

$$E\left(YW - \mu_{yw}\right) = 0, \qquad E(Y^2 - \mu_{yy}) = 0 \tag{9}$$

$$E\left[\left(W - \gamma Y\right)\left(W - \left(\gamma + \beta\right) Y\right) Y\right] = 0 \tag{10}$$

$$E\left[\left(W - \gamma Y\right)\left(W - \left(\gamma + \beta\right) Y\right)\left(Y^2 - \mu_{yy}\right) - 2\left(\mu_{yw} - \gamma \mu_{yy}\right)\left(W - \left(\gamma + \beta\right) Y\right) Y\right] = 0 \tag{11}$$

where $\mu_{yw} = E\left(YW\right)$ and $\mu_{yy} = E\left(Y^2\right)$. The parameters μ_{yw} and μ_{yy} are estimated along with γ and β by putting equations (9), (10), and (11) into any standard GMM estimation routine. One could replace β with e^b in these equations to impose the sign restriction that $\beta > 0$.

Lemma 1 uses up to fourth moments of the data. Based on results derived in the next section, in Appendix B we provide additional equations (using up to fifth moments) that can provide overidentification of γ and β, or point identification in some cases where Lemma 1 does not suffice.

Let σ_U^2, σ_V^2, and σ_R^2 denote the variances of the error components U, V, and R. It may be of economic interest to estimate these variances, to identify how much of the variance of the model errors is due to unobserved ability U versus the idiosyncratic components V and R. From the model we have $E\left(\left(W - \gamma Y\right) Y\right) = \beta \sigma_U^2$, $E\left(Y^2\right) = \sigma_U^2 + \sigma_V^2$, and $E\left(\left(W - \gamma Y\right)^2\right) = \beta^2 \sigma_U^2 + \sigma_R^2$, which implies

$$\sigma_U^2 = E\left(\left(W - \gamma Y\right) Y\right)/\beta, \qquad \sigma_V^2 = E\left(Y^2\right) - \sigma_U^2, \qquad \sigma_R^2 = E\left(\left(W - \gamma Y\right)^2\right) - \beta^2 \sigma_U^2 \tag{12}$$

Given estimates of β and γ, we can replace the expectations in equation (12) with sample averages to estimate these variances.

Alternatively, we can estimate these variances jointly with the model parameters by observing that

$$\mu_{yy} = \sigma_U^2 + \sigma_V^2, \qquad \mu_{yw} = \beta \sigma_U^2 + \gamma\left(\sigma_U^2 + \sigma_V^2\right). \tag{13}$$

So, in equations (9), (10), and (11) we can replace μ_{yy} and μ_{yw} with their expressions in equation (13), and apply GMM using those equations along with the additional equation

$$E\left(\left(W - \gamma Y\right)^2 - \beta^2 \sigma_U^2 - \sigma_R^2\right) = 0 \tag{14}$$

to simultaneously estimate β, γ, σ_U^2, σ_V^2, and σ_R^2. We can further replace σ_U^2 with $\sigma_U^2 = e^{\tau_U}$ and similarly for σ_V^2 and σ_R^2, to impose the constraint that variances are positive. See Appendix B for details on these moments.

Higher moments of U, V, and R can be estimated analogously. Alternatively, as discussed later, once we have have identified and estimated β and γ, we can apply Kotlarski's Lemma to recover the entire distributions of U, V, and R.

We can also easily extend this identification and associated estimation to allow for covariates. Suppose we have the model

$$Y = b_1' X + U + V \tag{15}$$

$$W = \gamma Y + b_2' X + \beta U + R \tag{16}$$

where X is exogenous and is therefore uncorrelated with U, V, and R. The reduced form for W is now

$$W = (\gamma b_1 + b_2)' X + (\gamma + \beta) U + \gamma V + r$$

So we can estimate the coefficient vectors b_1 and b_2 along with γ and β by replacing Y and W in equations (9), (10), and (11) with $Y - b_1' X$ and $W - (\gamma b_1 + b_2)' X$, respectively and estimate those moments along with the moments

$$E\left(\left(W - (\gamma b_1 + b_2)' X\right) X\right) = 0, \qquad E\left((Y - b_1' X) X\right) = 0 \tag{17}$$

The complete set of moments for estimating this model via GMM, which we use in our empirical application, is provided in Appendix B.

3 General Point Identification

We now provide a more general and systematic analysis of the identification of our model, using more information than the low order moments of Lemma 1. We provide four main results. First, we show that it is possible to construct infinitely many moments like those of Lemma 1, which can be used to construct simple GMM estimators, and we give the conditions under which these moments point identify the coefficients α and γ (equivalently, β and γ). Second, we apply Kotlarski's lemma to point identify the distributions of U, V, and R given point identification of α and γ. Third, we demonstrate that, using the entire joint distribution of Y and W (instead of just some moments) the only case where point identification is not possible is when U or V (or both) are normal. Finally, in the not point identified case, we fully characterize the sharp identified set.

We make extensive use of the characteristic function and its logarithm. Knowing the (log) characteristic function of a vector of random variables is equivalent to knowing the joint distribution of those variables (Theorem 3.1.1 in Lukacs (1970)).

Definition 1 *Given two random variables Y and W, let $\phi_{Y,W}(\zeta, \xi) \equiv E\left[e^{i\zeta Y + i\xi W}\right]$ denote their joint characteristic function. Similarly for a single random variable, let $\phi_Y(\zeta) \equiv E\left[e^{i\zeta Y}\right]$. Moreover, let $\Phi_{Y,W}(\zeta, \xi) \equiv \ln \phi_{Y,W}(\zeta, \xi)$ and $\Phi_Y(\zeta) \equiv \ln \phi_Y(\zeta)$ denote log characteristic functions (which are also called cumulant generating functions).*

Definition 2 *Given two random variables Y and W, define the cumulant of order k, ℓ (Lukacs (1970), p. 27) as*

$$\Phi_{Y,W}^{k,\ell} \equiv \left[\frac{\partial^{k+\ell} \Phi_{Y,W}(\zeta, \xi)}{i^{k+\ell} \partial \zeta^k \partial \xi^\ell}\right]_{\zeta=0, \xi=0}.$$

Similarly for a single random variable, define the cumulant of order k as

$$\Phi_Y^k \equiv \left[\frac{\partial^k \Phi_Y(\zeta)}{i^k \partial \zeta^k}\right]_{\zeta=0}.$$

All cumulants can be expressed in terms of standard moments, as obtained by an explicit differentiation of the log characteristic function and by exploiting the characteristic function moment theorem (e.g. $E\left[Y^k\right] = \left[\frac{\partial^k \phi(\xi)}{i^k \partial \xi^k}\right]_{\xi=0}$)[5]. Also note that the joint and marginal characteristic functions as well as the corresponding cumulants are directly related, e.g., $\phi_Y(\zeta) = \phi_{Y,W}(\zeta, 0)$, $\Phi_Y(\zeta) = \Phi_{Y,W}(\zeta, 0)$ and $\Phi_Y^k = \Phi_{Y,W}^{k,0}$.

With these tools in hand, we are ready to state a general identification result based on moment constraints. As in Lemma 1, we start by rewriting the model of equations (4) and (5) in the reduced form of equations (4) and (6), and focus on the parameters α and γ.

Theorem 1 *Let Assumptions 1, 2, and Equations (4) and (6) hold. Assume* $-\infty < \gamma < \alpha < \infty$ *and let*

$$M_p(\alpha, \gamma) \equiv \Phi_{Y,W}^{1+p,2} - \alpha^2 \Phi_Y^{3+p} - (\gamma + \alpha)\left(\Phi_{Y,W}^{2+p,1} - \alpha \Phi_Y^{3+p}\right). \tag{18}$$

Let $q, \tilde{q} \in \mathbb{N} \equiv \{0, 1, \ldots\}$ *with* $q < \tilde{q}$. *If* $E\left[|U|^{\tilde{q}}\right]$, $E\left[|V|^{\tilde{q}}\right]$ *and* $E\left[|R|^{\tilde{q}}\right]$ *exist and* $\Phi_Y^{3+\tilde{q}} \Phi_{Y,W}^{2+q,1} \neq \Phi_Y^{3+q} \Phi_{Y,W}^{2+\tilde{q},1}$ *(or, equivalently, if* $\Phi_U^{3+\tilde{q}} \Phi_V^{3+q} \neq \Phi_V^{3+\tilde{q}} \Phi_U^{3+q}$*), then the moment constraints*

$$M_q(\alpha, \gamma) = 0 \tag{19}$$
$$M_{\tilde{q}}(\alpha, \gamma) = 0 \tag{20}$$

point identify the parameters of the model as $(\alpha, \gamma) = (\alpha_+, \alpha_-)$, *where*

$$\alpha_\pm = \frac{F^{3012}}{2F^{3021}} \pm \sqrt{\left(\frac{F^{3012}}{2F^{3021}}\right)^2 + \frac{F^{1221}}{F^{3021}}}$$

and where $F^{abcd} \equiv \Phi_{Y,W}^{a+\tilde{q},b} \Phi_{Y,W}^{c+q,d} - \Phi_{Y,W}^{a+q,b} \Phi_{Y,W}^{c+\tilde{q},d}$.

The proof, provided in Appendix A, proceeds by a judicious choice of cumulants of (Y, W) that do not depend on cumulants of R, and by exploiting the fact that cumulants of (Y, W) of order k, ℓ that share the same value of $k + \ell$ involve the same cumulants of U and V with prefactors that only differ in how they depend on α and γ. These observations then lead to specific functions of cumulants that can be analytically solved for α and γ.

Note that if we had assumed $\beta < 0$ instead of $\beta > 0$, then the same Theorem would hold except that now α and γ would be point identified by $(\alpha, \gamma) = (\alpha_-, \alpha_+)$. We next formally show that Theorem 1 contains Lemma 1 as a special case.

Corollary 2 *The assumptions of Theorem 1 with* $q = 0$ *and* $\tilde{q} = 1$ *imply that the assumptions of Lemma 1 hold. Equations (19) and (20) in Theorem 1 with* $q = 0$ *and* $\tilde{q} = 1$ *are equivalent to equations (7) and (8) in Lemma 1.*

[5] For high-order cumulants, these otherwise tedious algebraic manipulations could be handled with symbolic algebra packages.

Equations (9), (10), and (11), used for GMM estimation of α and γ, were obtained by converting equations (7) and (8) into moments suitable for GMM. Equivalently, equations (9), (10), and (11) could have been directly derived from $M_0(\alpha,\gamma)=0$ and $M_1(\alpha,\gamma)=0$. This is done explicitly in the proof of Corollary 2.

As noted above, all cumulants can be expressed in terms of standard moments, specifically, cumulants equal sums of products of moments. To fit within a GMM framework, the cumulants in the expressions $M_p(\alpha,\gamma)=0$, after being converted to functions of moments, must be linearized. This is done by introducing nuisance parameters. To illustrate, the cumulant Φ_Y^4 appears in the equation $M_1(\alpha,\gamma)=0$. Now Φ_Y^4 equals $E[Y^4]-3[E(Y^2)]^2$, so, e.g., to convert the expression $\Phi_Y^4=c$ into a form suitable for GMM, we rewrite this expression as $E[Y^4-3Y^2\mu_{YY}-c]=0$ and $E[Y^2-\mu_{YY}]=0$, using the nuisance parameter μ_{YY} that was introduced in the previous section.

Theorem 1 shows that one can obtain any number of additional, potentially overidentifying, moments to use for GMM estimation, based on the fact $M_p(\alpha,\gamma)=0$ holds for any nonnegative integer p (as long as the associated moments of U, V, and R exist). We illustrate this in Appendix B, where, in addition to the moments based on Lemma 1, we provide the additional moments suitable for GMM estimation that are obtained from $p=2$. In our later Monte Carlo simulations and empirical application, we provide results using the exactly identifying set of GMM moments based on $p=0$ and 1, and also using the generally over identifying set of GMM moments based on $p=0,1$ and 2.

Theorem 1 provides explicit conditions under which any pair of cumulant functions $M_q(\alpha,\gamma)=0$ and $M_{\widetilde{q}}(\alpha,\gamma)=0$ suffice to identify the parameters α and γ. In particular, point identification based on the moments in Lemma 1, corresponding to $M_0(\alpha,\gamma)=0$ and $M_1(\alpha,\gamma)=0$, requires that $\Phi_U^4\Phi_V^3\neq\Phi_V^4\Phi_U^3$, or equivalently

$$\left(E\left(U^4\right)-3\left[E\left(U^2\right)\right]^2\right)E\left(V^3\right)\neq\left(E\left(V^4\right)-3\left[E\left(V^2\right)\right]^2\right)E\left(U^3\right) \tag{21}$$

which is violated, for instance, if either U or V is normal, or if both U and V are symmetric, or if both U and V have the exact same distribution. If we add the additional moments corresponding to $M_2(\alpha,\gamma)=0$, then point identification only requires that at least one of the inequalities $\Phi_U^4\Phi_V^3\neq\Phi_V^4\Phi_U^3$, $\Phi_U^5\Phi_V^3\neq\Phi_V^5\Phi_U^3$, or $\Phi_U^5\Phi_V^4\neq\Phi_V^5\Phi_U^4$, hold. For example, if the second of these holds then Theorem 1 applies with $q=0$ and $\widetilde{q}=2$. If more than one of these inequalities holds, then we are generally overidentified.

Once the parameters α and γ have been identified, the full distribution of all unobservables can be determined under the following Assumption.[6]

Assumption 5 *The characteristic functions of U,V and R are nonvanishing on the real line.*

Corollary 3 *If Assumptions 1, 2, 5 and Equations (4) and (6) hold, $E[|Y|]<\infty$ and if α,γ are point identified, then the distributions of U, V and R are point identified from the*

[6]This can be relaxed to nonvanishing everywhere, except at isolated points, under slightly stronger moment existence conditions; see Schennach (2000) and Evdokimov, K. and H. White (2012).

joint distribution of Y and W through

$$\begin{aligned}\Phi_V(\xi) &= \int_0^{\xi} \frac{E\left[iYe^{i\zeta\frac{W-\alpha Y}{\gamma-\alpha}}\right]}{E\left[e^{i\zeta\frac{W-\alpha Y}{\gamma-\alpha}}\right]} d\zeta \\ \Phi_U(\zeta) &= \Phi_Y(\zeta) - \Phi_V(\zeta) \\ \Phi_R(\xi) &= \Phi_W(\xi) - \Phi_U(\alpha\xi) - \Phi_V(\gamma\xi).\end{aligned} \tag{22}$$

A more explicit expression for the distributions of these unobserved variables can be obtained by an inverse Fourier transform. For instance, if V admits a density, it is given by

$$f_V(v) = (2\pi)^{-1} \int_{-\infty}^{\infty} \exp(\Phi_V(\xi))\, e^{-i\xi v} d\xi \tag{23}$$

and similarly for the other densities. More general distributions (e.g. discrete and/or singular) can be recovered as well, if equation (23) is interpreted in the appropriate measure theoretic sense.

Although Theorem 1 is quite general, it does require the condition $\Phi_U^{3+\tilde{q}}\Phi_V^{3+q} \neq \Phi_V^{3+\tilde{q}}\Phi_U^{3+q}$ to deliver identification, so it is natural to ask whether this is fundamentally necessary. It is in fact possible to formulate an estimation strategy that relaxes this condition. For instance, as discussed above, one could stack the moment conditions of the form (19) and (20) obtained with different values of $(q, \tilde{q})$. The resulting moment conditions would only fail to identify (α, γ) if the condition $\Phi_U^{3+\tilde{q}}\Phi_V^{3+q} \neq \Phi_V^{3+\tilde{q}}\Phi_U^{3+q}$ fails simultaneously for all the choices of q and $\tilde{q}$ considered.

An even more general strategy could be to start from the fundamental relationships between the log characteristic functions of the observables and unobservables ($\Phi_{Y,W}(\zeta,\xi) = \Phi_U(\zeta + \alpha\xi) + \Phi_V(\zeta + \gamma\xi) + \Phi_R(\xi)$) and cast identification as an optimization problem that minimizes deviations between the observed and predicted quantities:

$$\begin{aligned}&(\alpha, \gamma, \Phi_U, \Phi_V, \Phi_R) \\ =\ & \arg \min_{(\alpha,\gamma,\Phi_U,\Phi_V,\Phi_R)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\Phi_U(\zeta+\alpha\xi) + \Phi_V(\zeta+\gamma\xi) + \Phi_R(\xi) - \Phi_{Y,W}(\zeta,\xi)\right|^2 d\xi d\zeta,\end{aligned} \tag{24}$$

subject to zero mean constraints ($\Phi_U'(0) = 0, \Phi_V'(0) = 0, \Phi_R'(0) = 0$) and that (Φ_U, Φ_V, Φ_R) be valid log characteristic functions. This approach circumvents requiring existence of the moments $E\left[|U|^{\tilde{q}}\right]$, $E\left[|V|^{\tilde{q}}\right]$ and $E\left[|R|^{\tilde{q}}\right]$. However, the introduction of nuisance functions (Φ_U, Φ_V, Φ_R) would complicate estimation, as these would have to be parameterized by series or other expansions to construct a corresponding sieve estimator.

The question remains, do there exist situations where neither this nor any other estimator can consistently estimate the model, due to lack of point identification? The following theorem fully addresses this question, by showing that there exist cases that are not point identified. However, all such cases are when U or V (or both) are normal.

This differs from, and is simpler than, Reiersøl's (1950) well-known result in linear univariate errors-in-variables models, where the nonidentified cases arise when the model contains normal factors (see below). However, the required methods of proof differ significantly. For instance, the presence of two slope parameters α and γ (instead of one), and the presence of both latent variables U and V in both equations of the model, prevents us from using Reiersøl's proof method, which is based on the fact that two functions of different variables that are equal to each other must be constant. In our case, we have sums of many different functions of different variables on each side of an equality, and possible cancellation between terms that complicates the argument significantly. Also, Reiersøl's approach heavily relies on the concept of *factors* of a distribution, while we show that this abstract concept does not affect the analysis of point identification of our model. We will, however, return to this concept when we later construct identified sets in the not point identifed case.

Assumption 6 $E\left[|U|^3\right], E\left[|V|^3\right], E\left[|R|^3\right]$ *are finite.*

Theorem 4 *Let Assumptions 1, 2, 5, 6 and Equations (4) and (6) hold and assume that* $-\infty < \gamma < \alpha < \infty$. *If neither* U *nor* V *are normally distributed, then* α, γ *are uniquely determined by the joint distribution of* Y *and* W *by Equation (24).*

In the next section, we address what happens when either U or V (or both) are normally distributed.

4 Set Identification

In the case where Theorem 4 does not apply, so that the parameters are not point identified, the objective function of Equation (24) is maximized over a set rather than at a single point. In order to precisely characterize this *identified set*, we first need to introduce the notion of *factor*, which is used by Reiersøl (1950) and by Schennach and Hu (2013).

Definition 3 *If a random variable* Z *can be decomposed as* $Z = Z_1 + Z_2$ *where* Z_1 *and* Z_2 *are independent, then* Z_1 *and* Z_2 *are called* factors *of* Z. *(The term* factor *can also be used to refer to the distributions of these variables.)*

While for given characteristic functions $\phi_{Z_1}(\xi)$ and $\phi_{Z_2}(\xi)$, we automatically have that $\phi_Z(\xi) = \phi_{Z_1}(\xi)\phi_{Z_2}(\xi)$ by the convolution theorem, the notion of factor embodies the fact that, if one is instead given the two characteristic functions $\phi_Z(\xi)$ and $\phi_{Z_1}(\xi)$, it is not automatic that there exists a random variable Z_2 with characteristic function $\phi_{Z_2}(\xi) = \phi_Z(\xi)/\phi_{Z_1}(\xi)$. The inverse Fourier transform of $\phi_{Z_2}(\xi)$, may not actually yield a proper probability measure (it could assign negative weights to some sets, for instance).

Next we consider what it means for a random variable to have a normal factor.

Lemma 2 *Let* Z *be an observed zero mean random vector. Then* Z *admits a unique decomposition into two unobserved zero mean independent factors*

$$Z = Z_g + Z_n, \tag{25}$$

where Z_g *is Gaussian with variance* $\bar{\Lambda}$ *and* Z_n *has no Gaussian factors. Furthermore, the variance of* Z_g *is determined (from the observed distribution of* Z*) from the unique* $\bar{\Lambda}$ *such that*

$$\bar{\Lambda} - \Lambda \text{ is positive semidefinite} \iff \phi_Z(\xi)\exp(\xi'\Lambda\xi/2) \text{ is a characteristic function.}$$

(Note that either Z_g *or* Z_n *or both could be zero.)*

Intuitively, Lemma 2 indicates that the decomposition into a Gaussian and a non-Gaussian factor can, in principle, be found by attempting to deconvolve Z by a Gaussian of variance Λ and seeking the "largest" (in a positive definite sense) possible Λ that will still yield a proper distribution. In Fourier representation, this amounts to dividing $\phi_Z(\xi)$ by $\exp(-\xi'\Lambda\xi/2)$ and checking if the result is a valid characteristic function (e.g., by verifying if the inverse Fourier transform is a nonnegative measure). An alternative check for the validity of a given function $\phi(\xi)$ to be a valid characteristic function can be based on Bochner's Theorem (Theorem 4.2.2 in Lukacs (1970): ϕ is a characteristic function iff

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j^* \phi\left(\xi_i - \xi_j\right) \geq 0 \text{ for all } c_1,\ldots,c_n \in \mathbb{C} \text{ for all } \xi_1,\ldots,\xi_n \in \mathbb{R} \text{ for all integer } n \geq 1$$

(Bochner's Theorem also includes the conditions that $\phi(\xi)$ be continuous and $\phi(0) = 1$ but these are automatically satisfied in our context.)

Using Lemma 2, we can decompose the observed $Z = (Y, W)$ into Gaussian (g) and non-Gaussian (n) factors

$$(Y, W) = (Y_g, W_g) + (Y_n, W_n) \tag{26}$$

This decomposition can be accomplished without the knowledge of α or γ. The non-Gaussian or Gaussian nature of the two factors is important in our context, because it is associated with the features that can or cannot be point-identified. This type of decomposition is not a purely theoretical construct; it can be empirically implemented. Independent Component Analysis techniques, which are widely used in signal processing, (see Hyvärinen and Oja (2000) for a review) specifically rely on such decompositions into Gaussian and non-Gaussian components.

Define

$$B_s = \frac{E[W_s Y_s]}{E[Y_s^2]} \tag{27}$$

$$D_s = \frac{E[W_s^2]\,E[Y_s^2] - (E[W_s Y_s])^2}{(E[Y_s^2])^2} \geq 0 \tag{28}$$

where the subscript s is either set to "g", or to "n", or is removed. We can now state our set-identification theorem:

Theorem 5 *Let Assumptions 1, 2 and Equations (4) and (6) hold and assume that $E[Y^2]$, $E[W^2]$, $E[R^2] < \infty$ and that $-\infty < \gamma < \alpha < \infty$. Then, the following bounds (illustrated in Figure 1) are sharp:*

1. *If both U and V are Gaussian (and $E[Y^2] > 0$), then*

$$\alpha \geq B_g \tag{29}$$

$$B_g - \frac{D_g}{\alpha - B_g} \leq \gamma \leq B_g. \tag{30}$$

2. If V is Gaussian but U is not (and $E[Y_n^2], E[Y_g^2] > 0$), then

$$\alpha = B_n \tag{31}$$
$$B_g - \frac{D_g}{\alpha - B_g} \leq \gamma \leq B_g. \tag{32}$$

3. If U is Gaussian but V is not (and $E[Y_n^2], E[Y_g^2] > 0$), then

$$\gamma = B_n \tag{33}$$
$$B_g \leq \alpha \leq B_g + \frac{D_g}{B_g - \gamma}. \tag{34}$$

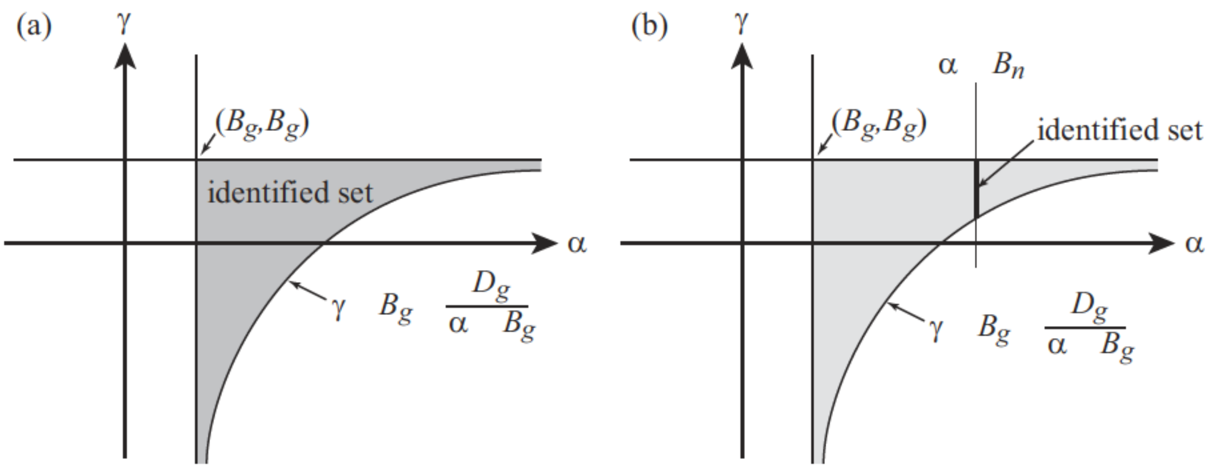

Figure 1: Identified set of Theorem 5 for (a) Case 1 and (b) Case 2 (Case 3, analogous to Case 2, is not shown).

For each of the possible values of (α, γ) in the set given by Theorem 5, there corresponds a unique implied distribution for U, for V, and for R, given by Corollary 3. To distinguish between the three cases in Theorem 5, we have that case 1 holds only if Y is normal, in case 2 $B_n > B$, and in case 3 $B_n < B$.

Although the quantities B_n, B_g, D_n, D_g are, in principle, observable quantities, they may be difficult to estimate. For this reason, we also provide below a coarser bound that is only based on the covariances matrix of the observed Y and W:

Corollary 6 *The following bounds on α, γ always hold:*

$$\alpha \geq B$$
$$B - \frac{D}{\alpha - B} \leq \gamma \leq B.$$

It is no accident that these bounds have the same form as Case 1 of Theorem 5: Both are solely based on covariance information, but in the Gaussian case, covariances exhaust all available information and yield sharp inequalities while, in general, that is not the case.

This looser bound is also related to the measurement error bounds in Frisch (1934). If one is willing to rely on this relaxed bound, then a simple GMM estimator for the resulting identified set could be obtained based on the moment conditions

$$E\left[\alpha^2\sigma_U^2 + \gamma^2\left(Y^2 - \sigma_U^2\right) + \sigma_R^2 - W^2\right] = 0 \quad (35)$$

$$E\left[\alpha\sigma_U^2 + \gamma\left(Y^2 - \sigma_U^2\right) - YW\right] = 0 \quad (36)$$

while optimizing over $\alpha, \gamma, \sigma_U^2, \sigma_R^2$, subject to the constraints $\gamma < \alpha$ (equivalent to $\beta > 0$), $\sigma_U^2 \geq 0$ and $\sigma_R^2 \geq 0$. These moment conditions are obtained from Equations (64) and (63) in the proof of Theorem 5, without extracting the Gaussian parts. The bounds of Corollary 6 are also obeyed in the case of point identified models, since they are obtained solely from positive variance considerations that must always be satisfied. This implies that, if one is unsure whether Y is normal or not, the moment conditions (35) and (36) could be stacked with the ones of Theorem 1 to yield an estimator that is robust to loss of point identification.[7]

5 Ordinary Least Squares

It is instructive to analyze in more detail how the parameters of our model relate to the slope coefficient of a naive OLS regression (in the population limit). The coefficient B given by Equation (27) is the slope coefficient of the least-square regression of W on Y (in the population limit). Regardless of whether the model is point identified or not, an implication of the model (i.e., of equations (4) and (5)) is that B always lies between γ and α. This can be immediately verified by observing that

$$B = \frac{E\left[YW\right]}{E\left[Y^2\right]} = \frac{E\left[\left(U+V\right)\left(\alpha U + \gamma V\right)\right]}{E\left[\left(U+V\right)^2\right]} = \frac{\alpha E\left[U^2\right] + \gamma E\left[V^2\right]}{E\left[U^2\right] + E\left[V^2\right]} = \alpha\lambda + \gamma\left(1-\lambda\right) \quad (37)$$

where $\lambda = E\left[U^2\right] / \left(E\left[U^2\right] + E\left[V^2\right]\right)$ and so lies between zero and one. So in particular, if $\beta > 0$ we get $\gamma \leq B \leq \alpha$.

This type of inequality has been noted before in the context of estimating returns to education (e.g. by Card (2001), in a more detailed model that allows for some individual heterogeneity). In particular, in the returns to schooling context, we would expect both β and γ to be positive (because unobserved ability U should affect schooling Y and wages W in the same direction, and increased schooling should increase wages). By the above analysis, this in turn means that we would expect $0 < \gamma \leq B$.

However, as noted by Card (2001), most returns to schooling empirical applications yield estimates of γ, using instrumental variables methods, that are greater than B, which contradicts this inequality and hence also contradicts the model. One possible explanation for this contradiction is that, in the returns to schooling context, Y may also contain significant measurement error. Standard attenutation bias under classical measurement error implies that the ordinary least squares coefficient B is biased towards zero relative to γ, which if $0 < \gamma$ would imply $B < \gamma$. If the model is correct for returns to education, but in addition Y is mismeasured, then B could be either larger or smaller than γ, depending on the relative magnitude of the measurement error.

[7] In this case the maximizing estimands could be sets rather than points, requiring nonstandard inference.

6 Monte Carlo

To assess the finite sample performance of our simple GMM estimators, we generate data from the model of equations (4) and (5) without covariates. All of our designs are chosen to satisfy equation (21), so the model is point identified just from the moments in Lemma 1.[8] The true values of the coefficients are $\gamma = \beta = 1$. It is widely recognized that estimators based on higher moments can behave poorly with small sample sizes, so to see if our estimators suffer from these issues, we work with relatively small sample sizes of $n = 100$ and $n = 400$.

We generate $5,000$ replications of four different designs. In design 1, U is log normal while V and R are each standard Gumbel. In design 2, U is log normal while V and R are uniform. We then reverse these, making U Gumbel and V and R log normal in design 3, and making U uniform with V and R log normal in design 4. For each design, we report results using two different estimators. The exactly identified estimator is GMM using moments corresponding to Lemma 1, given by equations (74), (75), and (76) (without covariates, so $\widetilde{Y} = Y$ and $\widetilde{W} = W$). The over-identified estimator is GMM using these same equations, plus equations (78) and (79).

Tables 1 to 4 report results from designs 1 to 4, respectively. Each Table has four panels, corresponding to the two different GMM estimators, each with the two different sample sizes. We report estimates of γ, β, the error component variances σ_U^2, σ_V^2, and σ_R^2, and, when over-identified, μ_{WW}. Reported summary statistics of each parameter estimate across the simulations are the mean (MEAN), the standard deviation (SD), the 25% quantile (LQ), the median (MED), the 75% quantile (UQ), the root mean squared error (RMSE), the mean absolute error (MAE), and the median absolute error (MDAE).

Some general tendencies stand out in these simulations. First, consider the trade off between the exactly identified vs over identified estimators. The latter uses more information, but that information takes the form of up to fifth order moments, which can be noisy and more sensitive to outliers. In general we find that the overidentified estimator performs better than the exactly identified estimator, particularly at the larger sample size.

The primary parameter of interest, γ, tends to be estimated reasonably precisely in all of the designs, with most RMSEs in the range of .3 to .7. In contrast, β is generally much less precisely estimated, often having much larger RMSEs (except in design 2). Estimates of the variances σ_U^2, σ_V^2, and σ_R^2, are mostly similar to each other, usually being less precise than γ but more than β. The estimate of μ_{WW} is noisier, since it only appears in the highest order moment equations of the over identified model. The designs where U was log normal (designs 1 and 2) generally had more accurate estimates than the other designs. We conclude that our estimator performs reasonably well even with rather small sample sizes.

7 GDP and Life Expectancy

There is a long literature studying the causal effect of health on economic growth. Examples include Acemoglu and Johnson (2007) (which we will hereafter refer to as AJ),

[8] In particular in all of our designs, U and V have different, non-normal distributions, and at least one is asymmetically distributed. U, V, and R are also mutually independent and centered at mean zero.

Well (2007), Lorentzen, McMillan, and Wacziarg (2008), Aghion, Howitt, and Murtin (2010), Cervellati and Sunde (2011), Ecevit (2013), Bloom, Canning, and Fink (2014), and Bloom, Canning, Kotschy, Prettner, and Schünemann (2019).

Based on a neo-classical growth model, AJ estimate a model in the form of equations (1) and (2), where Y is the change in the log of a country's life expectancy at birth between 1940 and 1980, W is the change in that country's log GDP in the same time span, and X is either just a constant, or a constant and a measure of the country's quality of institutions, or a constant and GDP per capita in 1930. The main goal is estimation of γ, the coefficient of Y in the W equation.

AJ observe that ordinary least squares estimation of the W equation is inconsistent, because the health measure Y is endogenous, with improvements and investments in a country's productive technology over time positively impacting both health outcomes and GDP. This technology change corresponds to our unobserved factor U (with $\beta > 0$) in equations (15) and (16), while V and R are the idiosyncratic shocks to health and economic outcomes, respectively.

To deal with the endogeneity caused by U, AJ construct an instrument, called predicted mortality, that combines each country's 1940 mortality rates from specific diseases with a set of global interventions that addressed those diseases. As noted in the introduction, one may question the validity of such constructed instruments.

In Table 5, we replicate selected results appearing in Table 3 and Table 9 of AJ.[9] Column 1 gives AJ's ordinary least squares (OLS) estimates, while columns 2, 3, and 4 replicate AJ's estimates using two stage least squares (2SLS) with the above listed combinations of covariates X, and using their predicted mortality instrument. AJ's OLS estimate of γ (corresponding to B in the previous section) is -0.813, while their 2SLS estimates of γ are considerably larger in magnitude, ranging from -1.316 to -1.643. As we noted earlier, having $\gamma < B$, as AJ find, is an implication of our model when $\beta > 0$. Note that the sample size is quite small in this application, with only 47 countries. Nevertheless, AJ's estimates of γ are statistically significant.

Now suppose we had not observed predicted mortality, or we are uncertain of its validity as an instrument. We can instead consider applying our GMM estimators. First, consider the distribution of Y. Heuristically, if Y is close to normal, then it may be that U or V is normal, which would prevent point identification of our model. Y has a skewness of 0.170 and a kurtosis of 1.791, which is reasonably far from normal in terms of the low order moments our GMM estimator is based on. The p-value of a Shapiro-Wilk test of normality of Y is .02, and even lower if one tests the residuals after regressing Y on either of the covariates in X. This suggests that the model could be point identified (at least ruling out both U and V normal), so we attempt to apply our GMM estimators.

In Table 6, we report three sets of estimates. First are the columns labeled 2SLS1, 2SLS2, and 2SLS3, which in Panel A are AJ's estimates from Table 5. Next are columns labeled GMM1, GMM2, and GMM3. These are GMM estimates of equations (15) and (16), which do not make use of the predicted mortality instrument in any way. Specifically, these are estimates based on the over-identifying set of moments given by equations (74) to (79) in

[9] Our data are provided by AJ. Life expectancy is from UN data sources and the League of Nation reports. Pre-war GDP data are from Maddison (2003), and post-war data are from the UN. See AJ for details.

Appendix B. The last three columns of Table 6 then give GMM estimates that use both our over-identifying set of moments and the additional moment given by AJ's instrument (as discussed at the end of Appendix B).[10]

Panel A in Table 6 reports the main parameter of interest γ, and also reports b_2, the other covariate coefficients in equation (16). Our main takeaway from Panel A of Table 6 is that our estimates of γ are quite comparable to AJ's. In GMM4 and GMM5, the estimates of γ are -1.676 and -1.309, virtually the same range as AJ's 2SLS estimates, and are statisically significant. GMM6 gives an estimate of a higher magnitude -3.096, but this estimate is statistically insignificant with a very large standard error, suggesting that our higher moment based estimator is imprecise for this particular combination of covariates and small sample size. The last three columns of Table 6, which combine both our moments and the AJ instrument, give estimates very close to those of AJ, with somewhat smaller standard errors, which is exactly what one would expect to see if both sets of moments are valid and if AJ's instrument is strong. In the bottom row of Table 6 we report Hansen's J-test; we do not reject validity of the joint set of overidentifying restrictions in any of the GMM estimates.

Panels B and C of Table 6 provide the other estimated parameters of the model. Panel C gives the estimated b_1 coefficients from equation (15), while Panel B gives the estimates of β and the estimated variances of our error components. β appears to be difficult to precisely estimate, with large standard errors[11]. In the specifications where γ is statistically significant, the variance of U (the source of endogeneity in the model) is much smaller than the variances of the idiosyncratic components V and R, but very precisely estimated with small standard errors.

Later tables have the same format as Table 6, providing additional results. In Table 7, we re-estimate the model using the exactly identified set of moments from Lemma 1. As expected with fewer moments, these estimates are less efficient, and turn out to be quite a bit noisier than those of Table 6. GMM4, without covariates, is still reasonably comparable to AJ with γ of -1.799, while now both GMM5 and GMM6 are insignificant and more variable. The estimates combining these moments with AJ's instrument behave as before.

In Table 8, we reestimate the models using a larger sample of 56 countries. This expanded sample contains additional countries for which the data for life expectancy and GDP per capita in 1940 and 1980 were available. AJ excluded these additional countries from their analysis due to concerns about data quality. The estimates that do not include covariates, 2SLS1, GMM1 and GMM4, remain comparable to each other and to the earlier estimates, but both the AJ and GMM model estimates that include covariates are more erratic in this sample, presumably due to these data quality issues. All the estimates of γ in all the models have a negative sign.

We conclude that, in all specifications where the standard errors were small enough to yield statistically significant results, our estimates based on higher moments, without side information, are very close to those obtained by AJ that required an instrument.

[10]These GMM models are estimated in Stata, using the vce(robust) option to compute standard errors.

[11]In contrast α is, like γ, much more precisely estimated, but apparently the difference $\beta \equiv \alpha - \gamma$ is harder to pin down.

8 Conclusions

We have shown that a standard linear triangular structural model is generally point identified, without an instrument or other side information that is generally used to identify such models. We illustrate the result with Monte Carlo simulations and in an empirical application. Our application shows that, without using an instrument, GMM estimation of moments based on the model yields estimates very close to those that were obtained by previous authors using an instrument.

What makes point identification possible is the assumed error structure, which takes the standard form of a scalar common component U in each equation, plus additional scalar idiosyncratic components V and R. Possible goals for future work could include deriving alternative estimators for the model (perhaps based directly on characteristic functions rather than moments), extending the model to more equations, allowing the common component U to affect outcomes nonlinearly, and extending the model to also allow for measurement error in Y. Based on Card (2001), this last extension would likely be needed for returns to education applications.

A Appendix A: Proofs

Proof of Lemma 1. Define Q and P as $Q = W - \gamma Y = \beta U + R$ and $P = W - (\gamma + \beta) Y = -\beta V + R$. Then Lemma 1 claims

$$E(QPY) = 0 \quad \text{and} \quad cov\left(QP, Y^2\right) - 2E(QY)E(PY) = 0 \tag{38}$$

First verifying that $E(QPY) = 0$, we have

$$\begin{aligned} E(QPY) &= E\left[(\beta U + R)(-\beta V + R)(U + V)\right] \\ &= E\left(R^2U + R^2V + RU^2\beta - RV^2\beta - U^2V\beta^2 - UV^2\beta^2\right) = 0 \end{aligned}$$

This expectation is zero by U, V, and R being mutually independent with mean zero.

For the second equation, we have

$$\begin{aligned} cov\left(QP, Y^2\right) &= cov\left[(\beta U + R)(-\beta V + R), (U + V)^2\right] \\ &= cov\left(R^2 + RU\beta - RV\beta - UV\beta^2, U^2 + 2UV + V^2\right) \\ &= cov\left(-UV\beta^2, U^2 + 2UV + V^2\right) \\ &= -\beta^2 E\left[UV\left(U^2 + 2UV + V^2\right)\right] = -2\beta^2 E\left(U^2\right)E\left(V^2\right) \end{aligned}$$

and

$$\begin{aligned} 2E(QY)E(PY) &= 2E\left[(\beta U + R)(U + V)\right]E\left[(-\beta V + R)(U + V)\right] \\ &= 2E\left(\beta U^2\right)E\left(-\beta V^2\right) = -2\beta^2 E\left(U^2\right)E\left(V^2\right) \end{aligned}$$

■

Proof of Theorem 1. To show identification, we first compute the joint characteristic function of the two observed variables in terms of the unobserved variables:

$$\begin{aligned}\phi_{Y,W}(\zeta,\xi) &= E\left[e^{i\zeta(U+V)}e^{i\xi(\alpha U+\gamma V+R)}\right]\\ &= E\left[e^{i(\zeta+\alpha\xi)U}\right]E\left[e^{i(\zeta+\gamma\xi)V}\right]E\left[e^{i\xi R}\right]\\ &= \phi_U(\zeta+\alpha\xi)\,\phi_V(\zeta+\gamma\xi)\,\phi_R(\xi)\end{aligned}$$

where we have used mutual independence of U, V, R to factor the expectations. In terms of cumulant generating functions, we therefore have:

$$\Phi_{Y,W}(\zeta,\xi) = \Phi_U(\zeta+\alpha\xi)+\Phi_V(\zeta+\gamma\xi)+\Phi_R(\xi) \tag{39}$$

For $\xi = 0$, this specializes to:

$$\Phi_Y(\zeta) = \Phi_U(\zeta)+\Phi_V(\zeta). \tag{40}$$

Next, for any any $p \in \mathbb{N}$ and $\ell \in \{1,2\}$, we immediately get, from (39), the cumulant relationship

$$\Phi_{Y,W}^{3+p-\ell,\ell} = \alpha^\ell\Phi_U^{3+p}+\gamma^\ell\Phi_V^{3+p}.$$

Replacing Φ_U^{3+p} by its value implied from (40), we have:

$$\Phi_{Y,W}^{3+p-\ell,\ell} = \alpha^\ell\Phi_Y^{3+p}+\left(\gamma^\ell-\alpha^\ell\right)\Phi_V^{3+p}. \tag{41}$$

Now, (41) implies, for $\ell = 1,2$, the following system of equations:

$$\Phi_{Y,W}^{2+p,1}-\alpha\Phi_Y^{3+p} = (\gamma-\alpha)\,\Phi_V^{3+p} \tag{42}$$

$$\Phi_{Y,W}^{1+p,2}-\alpha^2\Phi_Y^{3+p} = \left(\gamma^2-\alpha^2\right)\Phi_V^{3+p}. \tag{43}$$

Factoring $(\gamma^2-\alpha^2)\,\Phi_V^{3+p}$ as $(\gamma+\alpha)(\gamma-\alpha)\,\Phi_V^{3+p}$ in Equation (53) and replacing $(\gamma-\alpha)\,\Phi_V^{3+p}$ by its value from Equation (42) yields:

$$\Phi_{Y,W}^{1+p,2}-\alpha^2\Phi_Y^{3+p} = (\gamma+\alpha)\left(\Phi_{Y,W}^{2+p,1}-\alpha\Phi_Y^{3+p}\right),$$

which is identical to the condition $M_p(\alpha,\gamma) = 0$ for $M_p(\alpha,\gamma)$ defined in statement of the Theorem. Now, considering two different values q and $\tilde{q}$ of p, we obtain:

$$\Phi_{Y,W}^{1+q,2}-\alpha^2\Phi_Y^{3+q} = (\gamma+\alpha)\left(\Phi_{Y,W}^{2+q,1}-\alpha\Phi_Y^{3+q}\right) \tag{44}$$

$$\Phi_{Y,W}^{1+\tilde{q},2}-\alpha^2\Phi_Y^{3+\tilde{q}} = (\gamma+\alpha)\left(\Phi_{Y,W}^{2+\tilde{q},1}-\alpha\Phi_Y^{3+\tilde{q}}\right) \tag{45}$$

Next, multiplying (44) by $\left(\Phi_{Y,W}^{2+\tilde{q},1}-\alpha\Phi_Y^{3+\tilde{q}}\right)$ yields:

$$\left(\Phi_{Y,W}^{1+q,2}-\alpha^2\Phi_Y^{3+q}\right)\left(\Phi_{Y,W}^{2+\tilde{q},1}-\alpha\Phi_Y^{3+\tilde{q}}\right) = (\gamma+\alpha)\left(\Phi_{Y,W}^{2+\tilde{q},1}-\alpha\Phi_Y^{3+\tilde{q}}\right)\left(\Phi_{Y,W}^{2+q,1}-\alpha\Phi_Y^{3+q}\right). \tag{46}$$

(We can verify that $\Phi_{Y,W}^{2+\tilde{q},1}-\alpha\Phi_Y^{3+\tilde{q}} \neq 0$ as follows:

$$\begin{aligned}\Phi_{Y,W}^{2+\tilde{q},1}-\alpha\Phi_Y^{3+\tilde{q}} &= \left(\Phi_{U,\alpha U}^{2+\tilde{q},1}+\Phi_{V,\gamma V}^{2+\tilde{q},1}+\Phi_{0,R}^{2+\tilde{q},1}\right)-\alpha\left(\Phi_U^{3+\tilde{q}}+\Phi_V^{3+\tilde{q}}\right)\\ &= \left(\alpha\Phi_U^{3+\tilde{q}}+\gamma\Phi_V^{3+\tilde{q}}\right)-\alpha\left(\Phi_U^{3+\tilde{q}}+\Phi_V^{3+\tilde{q}}\right) = (\gamma-\alpha)\,\Phi_V^{3+\tilde{q}}\end{aligned}$$

where $\gamma \neq \alpha$ by assumption. If $\Phi_V^{3+\tilde{q}}$ were zero, we could simply exchange the role of tilded and non-tilded variables, which would require instead that $\Phi_V^{3+q} \neq 0$. Under the assumption that $\Phi_U^{3+q}\Phi_V^{3+\tilde{q}} \neq \Phi_V^{3+q}\Phi_U^{3+\tilde{q}}$, it is not possible that both $\Phi_V^{3+\tilde{q}} = 0$ and $\Phi_V^{3+q} = 0$, so one of these two approaches has to yield a nonzero multiplier.)

In Equation (46), $(\gamma + \alpha)\left(\Phi_{Y,W}^{2+q,1} - \alpha\Phi_Y^{3+q}\right)$ can be replaced by its value from (45) to yield a single equation in α:

$$\left(\Phi_{Y,W}^{1+q,2} - \alpha^2\Phi_Y^{3+q}\right)\left(\Phi_{Y,W}^{2+\tilde{q},1} - \alpha\Phi_Y^{3+\tilde{q}}\right) = \left(\Phi_{Y,W}^{1+\tilde{q},2} - \alpha^2\Phi_Y^{3+\tilde{q}}\right)\left(\Phi_{Y,W}^{2+q,1} - \alpha\Phi_Y^{3+q}\right)$$

Expanding the products, simplifying and collecting by powers of α yields:

$$-\left(\Phi_Y^{3+\tilde{q}}\Phi_{Y,W}^{2+q,1} - \Phi_Y^{3+q}\Phi_{Y,W}^{2+\tilde{q},1}\right)\alpha^2 + \left(\Phi_Y^{3+\tilde{q}}\Phi_{Y,W}^{1+q,2} - \Phi_Y^{3+q}\Phi_{Y,W}^{1+\tilde{q},2}\right)\alpha + \left(\Phi_{Y,W}^{1+\tilde{q},2}\Phi_{Y,W}^{2+q,1} - \Phi_{Y,W}^{1+q,2}\Phi_{Y,W}^{2+\tilde{q},1}\right) = 0,$$

or, in the notation of the theorem (noting that $\Phi_Y^{3+q} = \Phi_{Y,W}^{3+q,0}$):

$$-F^{3021}\alpha^2 + F^{3012}\alpha + F^{1221} = 0. \tag{47}$$

The roots of (47) are thus:

$$\alpha_{\pm} = \frac{F^{3012}}{2F^{3021}} \pm \sqrt{\left(\frac{F^{3012}}{2F^{3021}}\right)^2 + \frac{F^{1221}}{F^{3021}}}. \tag{48}$$

Since the original problem is completely symmetric upon permutation of the role of (α, U) and (γ, V), if we had gone through the same steps after eliminated $\Phi_V(\zeta)$ instead of $\Phi_U(\zeta)$, we would have obtained the same Equation (47) with α replaced by γ. This implies that the two roots of (47) simply correspond to the values of α and γ. Since we have assumed $\gamma < \alpha$, it follows that we must make the assignment $(\alpha, \gamma) = (\alpha_+, \alpha_-)$. The condition $\Phi_Y^{3+\tilde{q}}\Phi_{Y,W}^{2+q,1} - \Phi_Y^{3+q}\Phi_{Y,W}^{2+\tilde{q},1} \neq 0$ of the Theorem ensures that $F^{3021} \neq 0$, so that Equation (48) is well-defined. This condition can also be phrased in terms of the log characteristic functions of U and V:

$$\begin{aligned}
& \Phi_Y^{3+\tilde{q}}\Phi_{Y,W}^{2+q,1} - \Phi_Y^{3+q}\Phi_{Y,W}^{2+\tilde{q},1} \\
= & \left(\Phi_U^{3+\tilde{q}} + \Phi_V^{3+\tilde{q}}\right)\left(\Phi_{U,\alpha U}^{2+q,1} + \Phi_{V,\gamma V}^{2+q,1} + \Phi_{0,R}^{2+q,1}\right) - \left(\Phi_U^{3+q} + \Phi_V^{3+q}\right)\left(\Phi_{U,\alpha U}^{2+\tilde{q},1} + \Phi_{V,\gamma V}^{2+\tilde{q},1} + \Phi_{0,R}^{2+\tilde{q},1}\right) \\
= & \left(\Phi_U^{3+\tilde{q}} + \Phi_V^{3+\tilde{q}}\right)\left(\alpha\Phi_U^{3+q} + \gamma\Phi_V^{3+q}\right) - \left(\Phi_U^{3+q} + \Phi_V^{3+q}\right)\left(\alpha\Phi_U^{3+\tilde{q}} + \gamma\Phi_V^{3+\tilde{q}}\right) \\
= & (\alpha - \gamma)\left(\Phi_U^{3+q}\Phi_V^{3+\tilde{q}} - \Phi_V^{3+q}\Phi_U^{3+\tilde{q}}\right)
\end{aligned}$$

Since we have assumed that $\alpha > \gamma$, the only condition needed is that $\Phi_U^{3+q}\Phi_V^{3+\tilde{q}} \neq \Phi_V^{3+q}\Phi_U^{3+\tilde{q}}$. ■

Proof of corollary 2. The equivalence is shown by starting from Equations (19) and (20) and using the definitions of cumulants in terms of moments for zero-mean random variables:

$$
\begin{aligned}
\Phi^{1,2}_{Y,W} &= E\left[YW^2\right] \\
\Phi^{2,1}_{Y,W} &= E\left[Y^2W\right] \\
\Phi^{3}_{Y} &= E\left[Y^3\right] \\
\Phi^{2,2}_{Y,W} &= E\left[W^2Y^2\right] - E\left[W^2\right]E\left[Y^2\right] - 2E\left[WY\right]E\left[WY\right] \\
\Phi^{3,1}_{Y,W} &= E\left[WY^3\right] - 3E\left[WY\right]E\left[Y^2\right] \\
\Phi^{4}_{Y} &= E\left[Y^4\right] - 3E\left[Y^2\right]E\left[Y^2\right].
\end{aligned}
$$

Some cumulants involve products of moments ($\Phi^{2,2}_{Y,W}, \Phi^{3,1}_{Y,W}, \Phi^{4}_{Y}$) and to linearize them to obtain proper GMM-type moment conditions, we need to introduce nuisance parameters. For the above expression, it is sufficient to define:

$$
\mu_{yy} = E\left[Y^2\right] \qquad (49)
$$

$$
\mu_{yw} = E\left[YW\right] \qquad (50)
$$

so that we can write, for instance, $\Phi^4_Y = E\left[Y^4\right] - 3\mu_{yy}E\left[Y^2\right] = E\left[Y^4 - 3\mu_{yy}Y^2\right]$. Equations (49) and (50) yield the moment conditions of Equation (9).

Next, to establish Equation (10), we start from Equation (19):

$$
\begin{aligned}
& \Phi^{1,2}_{Y,W} - \alpha^2\Phi^3_Y - (\gamma+\alpha)\left(\Phi^{2,1}_{Y,W} - \alpha\Phi^3_Y\right) \\
=\ & E\left[YW^2\right] - \alpha^2E\left[Y^3\right] - (\gamma+\alpha)\left(E\left[Y^2W\right] - \alpha E\left[Y^3\right]\right) \\
=\ & E\left[YW^2 - \alpha^2Y^3 - (\gamma+\alpha)Y^2W + \alpha(\gamma+\alpha)Y^3\right] \\
=\ & E\left[YW^2 - (\gamma+\alpha)Y^2W + \alpha\gamma Y^3\right] \\
=\ & E\left[\left(W^2 - (\gamma+\alpha)YW + \alpha\gamma Y^2\right)Y\right] \\
=\ & E\left[(W-\gamma Y)(W-\alpha Y)Y\right],
\end{aligned}
$$

which is Equation (10).

To establish Equation (11), we start from Equation (20):

$$
\begin{aligned}
& \Phi^{2,2}_{Y,W} - \alpha^2\Phi^4_Y - (\gamma+\alpha)\left(\Phi^{3,1}_{Y,W} - \alpha\Phi^4_Y\right) \\
=\ & \Phi^{2,2}_{Y,W} - \gamma\left(\Phi^{3,1}_{Y,W} - \alpha\Phi^4_Y\right) - \alpha\Phi^{3,1}_{Y,W} \\
=\ & E\left[W^2Y^2\right] - E\left[W^2\right]E\left[Y^2\right] - 2E\left[WY\right]E\left[WY\right] \\
& -\gamma\left(\left(E\left[WY^3\right] - 3E\left[WY\right]E\left[Y^2\right]\right) - \alpha\left(E\left[Y^4\right] - 3E\left[Y^2\right]E\left[Y^2\right]\right)\right) \\
& -\alpha\left(E\left[WY^3\right] - 3E\left[WY\right]E\left[Y^2\right]\right) \\
=\ & E\left[W^2Y^2\right] - \mu_{yy}E\left[W^2\right] - 2\mu_{yw}E\left[WY\right] - \gamma\left(E\left[WY^3\right] - 3\mu_{yy}E\left[WY\right]\right) \\
& -\gamma\alpha\left(E\left[Y^4\right] - 3E\left[Y^2\right]\mu_{yy}\right) - \alpha\left(E\left[WY^3\right] - 3E\left[WY\right]\mu_{yy}\right) \\
=\ & E\left[W^2Y^2 - \mu_{yy}W^2 - 2\mu_{yw}WY - \gamma WY^3 + 3\gamma\mu_{yy}WY + \alpha\gamma Y^4 - 3\alpha\gamma\mu_{yy}Y^2 - \alpha WY^3 + 3\alpha\mu_{yy}WY\right] \\
=\ & E[\left(W^2 - \alpha WY - \gamma WY + \alpha\gamma Y^2\right)Y^2 - \left(W^2 - \alpha WY - \gamma WY + \alpha\gamma Y^2\right)\mu_{yy} \\
& + \left(2\alpha WY + 2\gamma WY - 2\alpha\gamma Y^2\right)\mu_{yy} - 2\mu_{yw}WY] \\
=\ & E\left[(W-\gamma Y)(W-\alpha Y)\left(Y^2-\mu_{yy}\right) + 2\alpha WY\mu_{yy} + 2\gamma WY\mu_{yy} - 2\alpha\gamma Y^2\mu_{yy} - 2\mu_{yw}WY\right] \\
=\ & E\left[(W-\gamma Y)(W-\alpha Y)\left(Y^2-\mu_{yy}\right) - 2\left(\mu_{yw} - \gamma\mu_{yy}\right)(W-\alpha Y)Y\right],
\end{aligned}
$$

which is Equation (11). ■

Proof of Corollary 3. We first express the model of Equations (4) and (6) as:

$$\begin{aligned} Y &= U+V \quad (51)\\ W &= \alpha U+\gamma V+R \quad (52)\end{aligned}$$

Equations (51) and (52) can be re-written as:

$$\begin{aligned} Y &= V+U \\ \frac{W-\alpha Y}{\gamma-\alpha} &= V+\frac{1}{(\gamma-\alpha)}R \end{aligned}$$

from which one can see that Y and $\frac{W-\alpha Y}{\gamma-\alpha}$ provide two repeated error-contaminated measurements of V which satisfy the assumption of the Kotlarski identity (Equation (22) given in the statement of Theorem, and the distribution of V is thus known. (The condition $E[|Y|]<\infty$ ensures the numerator of (22) is well defined while Assumption 5 ensures that the denominator is nonvanishing.)

Next, from Equation (40), we have $\Phi_U(\zeta)=\Phi_Y(\zeta)-\Phi_V(\zeta)$ and the distribution of U is identified. Finally, from Equation (39), we have $\Phi_R(\xi)=\Phi_{Y,W}(0,\xi)-\Phi_U(\alpha\xi)-\Phi_V(\gamma\xi)=\Phi_W(\xi)-\Phi_U(\alpha\xi)-\Phi_V(\gamma\xi)$ and the distribution of R is thus identified. In these steps, we have used the fact that Assumption 5 implies that the log characteristic functions of Y,W,U,V,R all exist everywhere on the real line. ■

Proof of Theorem 4. Lack of point identification means that there exists an observationally equivalent alternative model with variables $Y,W,\tilde{V},\tilde{U},\tilde{R}$ (note that Y,W are the same since they are observable) and parameters $\tilde{\alpha},\tilde{\gamma}$. We first establish that $\alpha,\gamma,\tilde{\alpha},\tilde{\gamma}$ must all be different. We note that, by assumption, we have both $\gamma<\alpha$ and $\tilde{\gamma}<\tilde{\alpha}$, so $\alpha\neq\gamma$ and $\tilde{\alpha}\neq\tilde{\gamma}$. We can also show that if $\alpha=\tilde{\alpha}$, then $\gamma=\tilde{\gamma}$ and the alternative model would, in fact, be identical. This follows from the fact that, if we knew α, we could write the model as:

$$\begin{aligned} Y &= V+U \\ W-\alpha Y &= (\gamma-\alpha)V+R \end{aligned}$$

which is just a standard errors-in-variables model in which the slope $(\gamma-\alpha)$ and the latent distributions are identified under non-normality of V (Reiersøl (1950)). We can also permute the role of (α,U) and (γ,V) and show that the knowledge of γ implies the knowledge of α in the same fashion, under non-normality of U. Hence, the mapping between α and γ is one-to-one. Similarly, $\alpha=\tilde{\gamma}$ implies $\gamma=\tilde{\alpha}$, but this would violate the condition $\tilde{\gamma}<\tilde{\alpha}$. These considerations let us assume throughout that $\alpha,\gamma,\tilde{\alpha},\tilde{\gamma}$ are all different.

We now proceed by showing that if α,γ were not point identified, then V and U would be normal, leading to a contradiction of the assumptions of the theorem. Starting from (39), we calculate $\partial/\partial\zeta$:

$$\Phi^{10}_{Y,W}(\zeta,\xi)=\Phi^{1}_{U}(\zeta+\alpha\xi)+\Phi^{1}_{V}(\zeta+\gamma\xi). \quad (53)$$

Note that knowledge of $\Phi^1_V(\zeta)$ is sufficient to recover $\Phi_V(\zeta)$ since it is known that $\Phi_V(0)=\ln E[1]=0$. We have a similar expression for the alternative model:

$$\Phi^{10}_{Y,W}(\zeta,\xi)=\Phi^{1}_{\tilde{U}}(\zeta+\tilde{\alpha}\xi)+\Phi^{1}_{\tilde{V}}(\zeta+\tilde{\gamma}\xi). \quad (54)$$

Equating (53) and (54), we have:

$$\Phi_U^1(\zeta+\alpha\xi)+\Phi_V^1(\zeta+\gamma\xi)-\Phi_{\tilde{U}}^1(\zeta+\tilde{\alpha}\xi)-\Phi_{\tilde{V}}^1(\zeta+\tilde{\gamma}\xi)=0. \tag{55}$$

Equation (55) has the general structure:

$$A_1(a_1\cdot x)+A_2(a_2\cdot x)+A_3(a_3\cdot x)+A_4(a_4\cdot x)=0$$

where $x=(\zeta,\xi)$ and

$$\begin{aligned} A_1(\chi) &= \Phi_U^1(\chi),\ a_1=(1,\alpha)\\ A_2(\chi) &= \Phi_V^1(\chi),\ a_2=(1,\gamma)\\ A_3(\chi) &= -\Phi_{\tilde{U}}^1(\chi),\ a_3=(1,\tilde{\alpha})\\ A_4(\chi) &= -\Phi_{\tilde{V}}^1(\chi),\ a_4=(1,\tilde{\gamma}) \end{aligned}$$

in which the $A_k(\chi)$ are twice differentiable by Assumption 6.

Lemma 3 below shows that if no two vectors in $\{(1,\alpha),(1,\gamma),(1,\tilde{\alpha}),(1,\tilde{\gamma})\}$ are colinear (which is the case here since $\alpha,\gamma,\tilde{\alpha},\tilde{\gamma}$ are all different) then it must be that the functions $A_1(\cdot),A_2(\cdot),A_3(\cdot),A_4(\cdot)$ are polynomials. This implies that $\Phi_V^1(\cdot)$ and $\Phi_U^1(\cdot)$ must be polynomials and so must $\Phi_V(\cdot)$ and $\Phi_U(\cdot)$. But by Theorem 7.3.5 in Lukacs (1970), the only possibility is that $\Phi_V(\cdot)$ and $\Phi_U(\cdot)$ are quadratic and V and U are thus normal. Since U and V were assumed not normal, this is not possible. ■

Lemma 3 *For $k=1,\dots 4$, let $a_k\in\mathbb{R}^2\backslash\{(0,0)\}$ be given and let A_k be unknown twice differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. If no pair of vectors in $\{a_k\}_{k=1}^4$ are colinear and, for all $x\in\mathbb{R}^2$,*

$$A_1(a_1\cdot x)+A_2(a_2\cdot x)+A_3(a_3\cdot x)+A_4(a_4\cdot x)=0, \tag{56}$$

then the A_k are polynomials.

Proof of Lemma 3. Computing all distinct second derivatives (denoted by $''$) of Equation (56) with respect to elements of x yields the following system of equations, in matrix form:

$$Mb=0$$

where

$$\begin{aligned} M &= \begin{bmatrix} a_{11}^2 & a_{12}^2 & a_{13}^2 & a_{14}^2\\ a_{11}a_{21} & a_{12}a_{22} & a_{13}a_{23} & a_{14}a_{24}\\ a_{21}^2 & a_{22}^2 & a_{23}^2 & a_{24}^2 \end{bmatrix}\\ b &= \begin{bmatrix} A_1''(a_1\cdot x)\\ A_2''(a_2\cdot x)\\ A_3''(a_3\cdot x)\\ A_4''(a_4\cdot x) \end{bmatrix}. \end{aligned}$$

Note that rows of M are linearly independent under the assumptions that the a_k are pairwise noncolinear. (Indeed, the proportions $a_{1k}^2 : a_{1k}a_{2k}$ are different for two different k and so are the proportions $a_{1k}a_{2k} : a_{2k}^2$ and $a_{1k}^2 : a_{2k}^2$. Also, there exists no $f, g \in \mathbb{R}$ such that $fa_{1k}^2 + ga_{2k}^2 = a_{1k}a_{2k}$ for all k since this expression reduces to $f\frac{a_{1k}}{a_{2k}} + g\frac{1}{a_{1k}/a_{2k}} = 1$ where the ratios[12] $\frac{a_{1k}}{a_{2k}}$ differ for two different k.)

Since there are 3 constraints and 4 elements in b, and since M does not depend on x, the solution vector b is constrained to be of the form

$$b = B(x)\,c$$

where c is a constant vector satisfying $Mc = 0$ and $B(x)$ is a scalar-valued function defined on $\mathbb{R}^2$. Note that c must have at least 2 non zero elements because if it had only one, then a column of M would have to be zero. This would imply $a_{k1} = a_{k2} = 0$ for some k but then a_k would be colinear with all other a_j, in contradiction to the assumptions.

Let c_i and c_j $(i \neq j)$ denote two nonzero elements of c. Having $A_i''(a_i \cdot x) = B(x)\,c_i$ forces $B(x)$ to be of the form $\tilde{B}_i(a_i \cdot x)$ while having $A_j''(a_j \cdot x) = B(x)\,c_j$ forces $B(x)$ to be of the form $\tilde{B}_j(a_j \cdot x)$. These constraints are only compatible if $B(x)$ is in fact constant, as a_i and a_j are not colinear. But then, all $A_k''(a_k \cdot x)$ are constant, which implies that all $A_k(a_k \cdot x)$ are (second order) polynomials.

Note that this result can also be proven using Corollary 5 in Khatri and Rao (1972), under weaker differentiability conditions, but the proof of that result is distributed over multiple earlier papers and it is helpful to provide here a simple self-contained proof. ■

Proof of Lemma 2. We first need to establish that the decomposition (25) is unique. We will argue by contradiction by assuming that there exists another decomposition $Z = \tilde{Z}_g + \tilde{Z}_n$ with $\tilde{Z}_g$ being Gaussian with variance $\tilde{\Lambda}$ and $\tilde{Z}_n$ having no Gaussian factor. Define

$$\Delta^{\pm} = \pm\sum_i \mathbf{1}(\pm\lambda_i > 0)\,\lambda_i \nu_i \nu_i'$$

where λ_i and ν_i denote the i-th eigenvalue and corresponding eigenvector of the matrix $\left(\bar{\Lambda} - \tilde{\Lambda}\right)$, respectively. Then, observe that:

$$Z_g = N\left(0, \bar{\Lambda}\right) = N\left(0, \bar{\Lambda} - \Delta^+\right) + N\left(0, \Delta^+\right)$$

and that

$$\tilde{Z}_g = N\left(0, \tilde{\Lambda}\right) = N\left(0, \tilde{\Lambda} - \Delta^-\right) + N\left(0, \Delta^-\right),$$

where equalities hold in distributions and the Gaussians are mutually independent.

Next, we can show that $\bar{\Lambda} - \Delta^+ = \tilde{\Lambda} - \Delta^-$ since

$$\begin{aligned}
\bar{\Lambda} - \Delta^+ &= \tilde{\Lambda} + \left(\bar{\Lambda} - \tilde{\Lambda}\right) - \Delta^+ \\
&= \tilde{\Lambda} + \sum_i \lambda_i \nu_i \nu_i' - \sum_i \mathbf{1}(\lambda_i > 0)\,\lambda_i \nu_i \nu_i' \\
&= \tilde{\Lambda} + \sum_i \left(1 - \mathbf{1}(\lambda_i > 0)\right)\lambda_i \nu_i \nu_i' \\
&= \tilde{\Lambda} + \sum_i \mathbf{1}(\lambda_i < 0)\,\lambda_i \nu_i \nu_i' = \tilde{\Lambda} - \Delta^-.
\end{aligned}$$

[12]We conventionally assign the value "∞" to a fraction of the form $a_{1k}/0$. The indeterminate case $0/0$ cannot occur because then two (a_{1k}, a_{2k}) would be colinear.

The random variable Z can then be decomposed into independent factors in two ways:

$$\begin{aligned} Z &= N\left(0, \bar{\Lambda} - \Delta^+\right) + N\left(0, \Delta^+\right) + Z_n \\ Z &= N\left(0, \tilde{\Lambda} - \Delta^-\right) + N\left(0, \Delta^-\right) + \tilde{Z}_n \end{aligned}$$

Since $\bar{\Lambda} - \Delta^+ = \tilde{\Lambda} - \Delta^-$, we have that $N(0, \Delta^+) + Z_n$ and $N(0, \Delta^-) + \tilde{Z}_n$ must have the same distribution. Now, Δ^+ and Δ^- are degenerate covariance matrices whose null spaces are orthogonal. This then implies that Z_n and $N(0, \Delta^+ + \Delta^-) + \tilde{Z}_n$ must have the same distribution. But then, Z_n would have a normal factor, in contradiction to our assumptions. The only possibility is that $\Delta^+ + \Delta^- = 0$, which implies that the two decomposition are, in fact, the same.

Having shown uniqueness of the decomposition into Gaussian and non-Gaussian factors, we then observe that we can write the following decomposition into independent factors:

$$Z = N(0, \Lambda) + N\left(0, \bar{\Lambda} - \Lambda\right) + Z_n$$

if and only if $\bar{\Lambda} - \Lambda$ is positive definite. Since $\phi_Z(\xi) \exp(\xi' \Lambda \xi / 2)$ is the characteristic function of $N\left(0, \bar{\Lambda} - \Lambda\right) + Z_n$ (i.e. Z deconvolved by $N(0, \Lambda)$), the result follows. ■

Proof of Theorem 5. Thanks to Lemma 2, we have the unique factor decomposition $(Y, W) = (Y_g, W_g) + (Y_n, W_n)$ and the model defined by Equations (51) and (52) can be uniquely decomposed into a sum of two models of the form:

$$Y_s = U_s + V_s \tag{57}$$

$$W_s = \alpha U_s + \gamma V_s + R_s, \tag{58}$$

one Gaussian (with subscript s set to "g") and one non-Gaussian (with subscript s set to "n").

Consider Case 1: Since U and V are Gaussian, the non-Gaussian model reduces to:

$$\begin{aligned} Y_n &= 0 \\ W_n &= R_n \end{aligned}$$

and provides no information regarding α and γ. For the Gaussian model, the covariance matrix of Y_g and W_g exhausts all available information:

$$E\left[Y_g^2\right] = \sigma_{U_g}^2 + \sigma_{V_g}^2 \tag{59}$$

$$E\left[W_g^2\right] = \alpha^2 \sigma_{U_g}^2 + \gamma^2 \sigma_{V_g}^2 + \sigma_{R_g}^2 \tag{60}$$

$$E\left[Y_g W_g\right] = \alpha \sigma_{U_g}^2 + \gamma \sigma_{V_g}^2 \tag{61}$$

where the possible value of the parameters $\alpha, \gamma, \sigma_{U_g}^2 \equiv E\left[U_g^2\right]$, $\sigma_{V_g}^2 \equiv E\left[V_g^2\right]$ and $\sigma_{R_g}^2 \equiv E\left[R_g^2\right]$ have to be determined, under the constraints that $\sigma_{U_g}^2 \geq 0$, $\sigma_{V_g}^2 \geq 0$, $\sigma_{R_g}^2 \geq 0$ and $\alpha > \gamma$. From (59) we have

$$\sigma_{V_g}^2 = E\left[Y_g^2\right] - \sigma_{U_g}^2 \tag{62}$$

and (60) and (61) become:

$$E\left[W_g^2\right] = \alpha^2\sigma_{U_g}^2 + \gamma^2\left(E\left[Y_g^2\right] - \sigma_{U_g}^2\right) + \sigma_{R_g}^2 \tag{63}$$

$$E\left[Y_gW_g\right] = \alpha\sigma_{U_g}^2 + \gamma\left(E\left[Y_g^2\right] - \sigma_{U_g}^2\right) \tag{64}$$

From (64) we then have:

$$\sigma_{U_g}^2 = \frac{E\left[Y_gW_g\right] - \gamma E\left[Y_g^2\right]}{\alpha - \gamma} \geq 0 \tag{65}$$

Since $\alpha > \gamma$, we have

$$\gamma \leq \frac{E\left[Y_gW_g\right]}{E\left[Y_g^2\right]} \equiv B_g.$$

This incorporates the first constraint $\sigma_{U_g}^2 \geq 0$ and shows the first inequality (29) defining the identified set.

From (63), with $\sigma_{U_g}^2$ from (65), we have

$$E\left[W_g^2\right] = \alpha^2\frac{E\left[Y_gW_g\right] - \gamma E\left[Y_g^2\right]}{\alpha - \gamma} + \gamma^2\left(E\left[Y_g^2\right] - \frac{E\left[Y_gW_g\right] - \gamma E\left[Y_g^2\right]}{\alpha - \gamma}\right) + \sigma_{R_g}^2$$

which can be re-arranged as:

$$(\alpha - \gamma)E\left[W_g^2\right] + \alpha\gamma(\alpha - \gamma)E\left[Y_g^2\right] - (\alpha + \gamma)(\alpha - \gamma)E\left[Y_gW_g\right] = \sigma_{R_g}^2(\alpha - \gamma)$$

where we know that $\sigma_{R_g}^2 \geq 0$. Upon division by $(\alpha - \gamma) > 0$ we have:

$$E\left[W_g^2\right] + \alpha\gamma E\left[Y_g^2\right] - (\alpha + \gamma)E\left[Y_gW_g\right] \geq 0$$

and re-arranging:

$$\gamma\left(\alpha E\left[Y_g^2\right] - E\left[Y_gW_g\right]\right) \geq \alpha E\left[Y_gW_g\right] - E\left[W_g^2\right] \tag{66}$$

Now, we observe that

$$\begin{aligned}
\left(\alpha E\left[Y_g^2\right] - E\left[Y_gW_g\right]\right) &= E\left[Y_g\left(\alpha Y_g - W_g\right)\right] \qquad (67)\\
&= E\left[(U_g + V_g)(\alpha U_g + \alpha V_g - \alpha U_g - \gamma V_g - R_g)\right]\\
&= E\left[(U_g + V_g)((\alpha - \gamma)V_g - R_g)\right] = (\alpha - \gamma)E\left[(U_g + V_g)V_g\right]\\
&= (\alpha - \gamma)E\left[V_g^2\right] > 0.
\end{aligned}$$

Therefore (66) can be divided by $\alpha E\left[Y_g^2\right] - E\left[Y_gW_g\right]$ while preserving the inequality:

$$\begin{aligned}
\gamma &\geq \frac{\alpha E\left[Y_gW_g\right] - E\left[W_g^2\right]}{\alpha E\left[Y_g^2\right] - E\left[Y_gW_g\right]} = \frac{\alpha E\left[Y_gW_g\right]/E\left[Y_g^2\right] - E\left[W_g^2\right]/E\left[Y_g^2\right]}{\alpha - E\left[Y_gW_g\right]/E\left[Y_g^2\right]}\\
&= \frac{\alpha B_g - E\left[W_g^2\right]/E\left[Y_g^2\right]}{\alpha - B_g} = \frac{(\alpha - B_g)B_g + B_g^2 - E\left[W_g^2\right]/E\left[Y_g^2\right]}{\alpha - B_g}\\
&= B_g + \frac{B_g^2 - E\left[W_g^2\right]/E\left[Y_g^2\right]}{\alpha - B_g}
\end{aligned}$$

or

$$\gamma \geq B_g - \frac{D_g}{\alpha - B_g} \tag{68}$$

where $D_g \geq 0$ by Cauchy-Schwartz. This incorporates the second constraint $\sigma^2_{R_g} \geq 0$ and shows that (68) provides the second inequality (30) defining the identified set.

The last constraint $\sigma^2_{V_g} \geq 0$ turns out to then be automatically satisfied. Indeed, combining Equation (62) and (65), we have

$$\begin{aligned} \sigma^2_{V_g} &= E\left[Y_g^2\right] - \sigma^2_{U_g} = E\left[Y_g^2\right] - \frac{E\left[Y_g W_g\right] - \gamma E\left[Y_g^2\right]}{\alpha - \gamma} \\ &= \frac{\alpha E\left[Y_g^2\right] - E\left[Y_g W_g\right]}{\alpha - \gamma} > 0 \end{aligned}$$

since $\alpha > \gamma$ by assumption and $\alpha E\left[Y_g^2\right] - E\left[Y_g W_g\right] > 0$ was already shown in (67).

We now turn to Case 2: V is Gaussian but U is not. The non-Gaussian model thus reduces to:

$$Y_n = U_n \tag{69}$$

$$W_n = \alpha U_n + R_n, \tag{70}$$

where, under our assumptions, U_n is necessarily nondegenerate. Equations (69) and (70) just define a standard regression model with correctly measured regressors, thus implying that α is point-identified:

$$\alpha = \frac{E\left[Y_n W_n\right]}{E\left[Y_n^2\right]}. \tag{71}$$

Next, the Gaussian model reduces to:

$$Y_g = U_g + V_g \tag{72}$$

$$W_g = \alpha U_g + \gamma V_g + R_g \tag{73}$$

where we note that even though U is not Gaussian, it could still have a nondegenerate Gaussian factor U_g.We also note that $Y_g = Y - U_n$ and $W_g = W - \alpha U_n - R_n$, so that the left-hand sides of (72)-(73) do not depend on γ. Therefore, as one considers different possible value of γ, one does not need to take into account possible changes in the left-hand side. The left-hand side does depend on α but α has been determined already. Equations (72) and (73) thus have the form assumed in Case 1, except that α has a known value. This implies that the identified set for γ is:

$$B_g - \frac{D_g}{\alpha - B_g} \leq \gamma \leq B_g$$

for α given by (71).

Finally, Case 3 is analogous to case 2, with the roles of (α, U) and (γ, V) permuted. ■

Proof of corollary 6. The proofs follows the proof of Case 1 of Theorem 5 with the subscript g removed, so we work with the observed distributions directly rather than their

Gaussian factor. The proofs only relies on the fact that variances must be positive, so the inequalities are equally valid for non-Gaussian random variables. However, unlike the Gaussian case, these bounds are not sharp because covariances matrices are not a sufficient statistic for the whole distribution in general. ■

B Appendix B: Moments for GMM Estimation

To facilitate the application of our estimator, here we write out the moments required for simple estimation of our model with or without covariates, based on Lemma 1 and Theorem 1. Assume X is a J vector of covariates X_1,...,X_J. The model with covariates X is

$$W = \gamma Y + b_2' X + \varepsilon_1, \qquad Y = b_1' X + \varepsilon_2$$

where the mean zero errors ε_1 and ε_2 are

$$\varepsilon_1 = \beta U + R, \qquad \varepsilon_2 = U + V.$$

Here b_1 and b_2 are vectors of coefficients b_{11},...,b_{1J}, and b_{21},...,b_{2J}.Typically we would have $X_1 = 1$, so b_{11} and b_{21} are the constant terms in the regressions.

Define $\widetilde{Y}$, $\widetilde{W}$, Q, and P by

$$\widetilde{Y} = Y - b_1' X, \qquad \widetilde{W} = W - (\gamma b_1 + b_2)' X,$$

$$Q = W - \gamma Y - b_2' X, \qquad P = W - (\gamma + \beta) Y + (\beta b_1 - b_2)' X$$

The parameters we wish to estimate are γ, β, σ_U^2, σ_V^2, σ_R^2, b_{11},...,b_{1J}, and b_{21},...,b_{2J}. Substituting the above expressions for $\widetilde{Y}$, $\widetilde{W}$, Q, and P into the following equations gives the moments for GMM estimation of these parameters.

$$E\left(\widetilde{Y}\widetilde{W} - \beta\sigma_U^2 - \gamma\left(\sigma_U^2 + \sigma_V^2\right)\right) = 0, \qquad E\left(\widetilde{Y}^2 - \sigma_U^2 - \sigma_V^2\right) = 0, \tag{74}$$

$$E\left(Q^2 - \beta^2\sigma_U^2 - \sigma_R^2\right) = 0, \qquad E\left(QP\widetilde{Y}\right) = 0, \tag{75}$$

$$E\left[QP\left(\widetilde{Y}^2 - \sigma_U^2 - \sigma_V^2\right) - 2\beta\sigma_U^2 P\widetilde{Y}\right] = 0 \tag{76}$$

$$E\left(QX_j\right) = 0 \quad \text{and} \quad E\left(\widetilde{Y}X_j\right) = 0 \quad \text{for } j = 1, ..., J \tag{77}$$

In addition to these moments, we also have the inequality constraints that β, σ_U^2, σ_V^2, σ_R^2 are all positive. These inequalities can be imposed by replacing these parameters in the above expressions with $\beta = e^b$, $\sigma_U^2 = e^{\tau_U}$, $\sigma_V^2 = e^{\tau_V}$, and $\sigma_R^2 = e^{\tau_R}$, and instead estimating the parameters b, τ_U, τ_V, and τ_R.

For the model without covariates, one can replace b_1 and b_2 with zero in the above expressions, and drop equation (77). Note that in this case Y and W should be demeaned.

Theorem 1 showed that $0 = \Phi_{Y,W}^{1+p,2} - \alpha^2\Phi_Y^{3+p} - (\gamma + \alpha)\left(\Phi_{Y,W}^{2+p,1} - \alpha\Phi_Y^{3+p}\right)$ holds for non-negative integers p, and the moments of Lemma 1 are equivalent to this equation for $p = 0$

and $p = 1$. Straightforward but tedious algebra shows that, with $p = 2$, we get the additional moments

$$E(\widetilde{W}^2 - \mu_{ww}) = 0 \tag{78}$$

$$0 = E[-3\widetilde{Y}\widetilde{W}^2\mu_{yy} - 6\mu_{yw}\widetilde{Y}^2\widetilde{W} - \mu_{ww}\widetilde{Y}^3 + \widetilde{Y}^3\widetilde{W}^2 - \alpha^2\left(\widetilde{Y}^5 - 10\widetilde{Y}^3\mu_{yy}\right) \tag{79}$$

$$-(\gamma+\alpha)\left(-6\mu_{yy}\widetilde{Y}^2\widetilde{W} - 4\mu_{yw}\widetilde{Y}^3 + \widetilde{Y}^4\widetilde{W} - \alpha\left(\widetilde{Y}^5 - 10\widetilde{Y}^3\mu_{yy}\right)\right)]$$

where again we could replace $\mu_{ww} = e^{\tau_W}$ to impose the sign constraint that $\mu_{ww} > 0$.[13]

Estimation of standard GMM using, as moments, equations (74), (75), (76), and (77) yields the exactly identified models named GMM4, GMM5, and GMM6 in our empirical application. Estimation using equations (74), (75), (76), (77), (78), and (79) gives the overidentified models labeled GMM1, GMM2, and GMM3 in our application.

If one in addition has an external instrument Z, then instead of $Y = b_1'X + \varepsilon_2$ we would have $Y = b_1'X + \delta Z + \varepsilon_2$. In this case all of the above equations still hold if we redefine $\widetilde{Y}$ as $\widetilde{Y} = Y - b_1'X - \delta Z$, and we could then add the additional moment

$$E\left(\widetilde{Y}Z\right) = 0 \tag{80}$$

Estimation using equations (74) to (80) with this redefinition of $\widetilde{Y}$ gives the models GMM1+AJ, GMM2+AJ, and GMM3+AJ in our application, and the same without (78) and (79) gives GMM4+AJ, GMM5+AJ, and GMM6+AJ.

References

Acemoglu, D., and S. Johnson, (2007), "Disease and development: The effect of life expectancy on economic growth," Journal of political Economy, 115(6), 925-985.

Aghion, P., P. Howitt, P., and F. Murtin, (2010), "The relationship between health and growth: When Lucas meets Nelson-Phelps, " National Bureau of Economic Research.

Angrist, J. and A. Krueger, (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" Quarterly Journal of Economics, 106, 979-1014.

Angrist, J. and A. Krueger, (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," Journal of Economic Perspectives, 15, 69-85.

Bloom, D. E., D. Canning, and G. Fink (2014), "Disease and development revisited ", Journal of Political Economy, 122(6), 1355-1366.

[13] Just as Lemma 1 required $\Phi_U^4\Phi_V^3 \neq \Phi_V^4\Phi_U^3$ for identification (see equation 21), to have equations (78), and (79) provide useful overidentifying information requires that either $\Phi_U^5\Phi_V^3 \neq \Phi_V^5\Phi_U^3$ or $\Phi_U^5\Phi_V^4 \neq \Phi_V^5\Phi_U^4$. See Theorem 1 for details.

Bloom, D. E., D. Canning, R. Kotschy, K. Prettner, and J. J. Schünemann (2019), "Health and Economic Growth: Reconciling the Micro and Macro Evidence," NBER Working Paper No. 26003.

Bonhomme, S. and J. - M. Robin (2010), "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," The Review of Economic Studies, 77, 491–533.

Card, D (1995) "Using Geographic Variation in College Proximity to Estimate the Return to Schooling", In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp , University of Toronto Press, Toronto.

Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", Econometrica, 69, 1127–1160.

Carneiro, P. and J. J. Heckman, (2002), "The Evidence on Credit Constraints in Post-Secondary Schooling," The Economic Journal, 112(482), 705–734.

Cervellati, M., and U. Sunde, (2011), "Life expectancy and economic growth: The role of the demo- graphic transition ", Journal of economic growth, 16(2), 99-133.

Fruehwirth, J. C., S. Navarro, and Y. Takahashi (2016), "How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects," Journal of Labor Economics 34:4, 979-1021

Erickson, T. and T. M. Whited, (2002), "Two-step GMM estimation of the errors-in-variables model using high-order moments," Econometric Theory, 18(3), 776-799.

Evdokimov, K. and H. White (2012), "Some Extensions of a Lemma of Kotlarski," Econometric Theory, 28(4), 925–932.

Frisch, R. (1934), "Statistical confluence analysis by means of complete regression systems," Vol. 5, Universitetets Økonomiske Instituut.

Hyvärinen, A. and E. Oja (2000) "Independent component analysis: algorithms and applications", Neural Networks 13, 411–430.

Khatri, C. and Rao, C. R. (1972), "Functional equations and characterization of probability laws through linear functions of random variables", Journal of Multivariate Analysis 2, 162–173.

Klein, R., and F. Vella, (2010), "Estimating a class of triangular simultaneous equations models without exclusion restrictions," Journal of Econometrics 154(42), 154–164.

Kotlarski, I. I. (1967), "On characterizing the gamma and normal distribution," Pacific Journal of Mathematics, 20, 69–76.

Lewbel, A. (1997), ”Constructing Instruments for Regressions With Measurement Error When No Additional Data are Available, With an Application to Patents and R&D,” Econometrica, 65(5), 1201-1213.

Lewbel, A. (2012), ”Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models,” Journal of Business and Economic Statistics, 30, 67–80.

Lewbel, A. (2020), ”Kotlarski With a Factor Loading,” Unpublished Manuscript, Boston College.

Li, S. and X. Zheng, (2020), ”A Generalization of Lemma 1 in Kotlarski (1967),” Statistics and Probability Letters,165, article 108814.

Li, T. and Q. Vuong (1998), “Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators,” Journal of Multivariate Analysis, 65, 139–165.

Lorentzen, P., J. McMillan, and R. Wacziarg, (2008), ”Death and development,” Journal of Economic Growth, 13(2), 81–124.

Lukacs, E. (1970), “Characteristic Functions”, Second edition, Griffin, London.

Maddison, A. (2003), “Development centre studies the world economy historical statistics: Historical statistics,” OECD Publishing.

Navarro, S. and J. Zhou, (2017) ”Identifying agent’s information sets: An application to a lifecycle model of schooling, consumption and labor supply,” Review of Economic Dynamics, 25, 58-92.

Reiersøl, O. (1950), “Identifiability of a linear relation between variables which are subject to error,” Econometrica, 18, 375-389.

Rigobon, R. (2003), ”Identification Through Heteroskedasticity,” .Review of Economics and Statistics 85(4), 777–792.

Schennach, S. M. (2000), “Estimation of nonlinear models with measurement error”, Working Paper, University of Chicago.

Schennach, S. M. and Y. Hu (2013), ”Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models Without Side Information,” Journal of the American Statistical Association, 108, 177-186.

GJ Székely, G. J. and C. R. Rao (2000), ”Identifiability of distributions of independent random variables by linear combinations and moments” Sankhyā: The Indian Journal of Statistics, Series A, 62(2), 193-202.

Well, D. N. (2007), ”Accounting for the Effect Of Health on Economic Growth,” The Quarterly Journal of Economics, 122(3), 1265–1306.

Table 1: Design 1

Over Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.96	0.76	0.8	1.15	1.36	0.76	0.44	0.32
β	1	1.18	1.14	0.66	1.01	1.43	1.15	0.57	0.37
σ_U^2	1.72	1.16	1.28	0.09	0.8	1.86	1.4	1.16	1.2
σ_V^2	1.64	1.8	1.03	0.98	1.81	2.5	1.04	0.85	0.76
σ_R^2	1.64	1.72	1.84	1.25	1.69	2.08	1.84	0.57	0.43
μ_{WW}	10.17	8.74	3.78	6.41	7.83	9.93	4.04	3.13	2.91
Hansen-Sargan J stat		1.42	2.04	0.22	0.72	1.81			
p-val		0.23		0.64	0.4	0.18			

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.99	0.48	0.88	1.09	1.27	0.48	0.3	0.22
β	1	1.06	0.44	0.84	1	1.18	0.44	0.26	0.17
σ_U^2	1.72	1.33	0.98	0.55	1.21	1.96	1.06	0.88	0.8
σ_V^2	1.64	1.8	0.72	1.31	1.79	2.31	0.74	0.59	0.52
σ_R^2	1.64	1.73	0.46	1.45	1.73	2.2	0.47	0.35	0.28
μ_{WW}	10.17	9.32	2.39	7.74	8.83	10.34	2.54	2	1.83
Hansen-Sargan J stat		1.57	1.96	0.31	0.86	2.12			
p-val		0.21		0.58	0.35	0.15			

Exactly Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	1.05	0.49	0.89	1.44	1.33	0.5	0.33	0.27
β	1	1.92	7.73	0.78	1.08	1.48	7.79	1.19	0.33
σ_U^2	1.72	1.32	2.74	0.08	0.96	1.94	2.77	1.24	1.14
σ_V^2	1.64	2.03	1.04	1.3	1.98	2.66	1.11	0.86	0.72
σ_R^2	1.64	1.71	0.67	1.31	1.72	2.15	0.68	0.52	0.41

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	1	0.39	0.87	1.07	1.23	0.39	0.26	0.19
β	1	1.21	2.09	0.89	1.03	1.21	2.1	0.36	0.16
σ_U^2	1.72	1.53	1.23	0.84	1.45	2.08	1.24	0.81	0.65
σ_V^2	1.64	1.83	0.7	1.34	1.81	2.3	0.73	0.58	0.49
σ_R^2	1.64	1.71	0.43	1.45	1.7	1.98	0.44	0.33	0.26

Notes: Design 1: $\ln(U) \sim N(-0.5, 1), V \sim \text{Gumbel}(0, 1), R \sim \text{Gumbel}(0, 1)$. All resulting variables are standardized to have zero means. The four panels are GMM estimates based on over-identifying set of moments, and exactly identified set of moments with sample sizes $n = 100$ and $n = 400$. The reported summary statistics are the mean (MEAN), the standard deviation (SD), the 25% quantile (LQ), the median (MED), the 75% quantile (UQ), the root mean squared errors (RMSE), the mean absolute errors (MAE), and the median absolute errors (MDAE).

Table 2: Design 2

Over Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.96	1.40	0.80	1.05	1.36	1.40	0.42	0.29
β	1	1.07	1.50	0.59	0.96	1.33	1.51	0.53	0.37
σ_U^2	1.72	1.34	1.24	0.45	1.17	1.90	1.30	1.00	0.87
σ_V^2	1.33	1.42	0.80	0.92	1.30	1.82	0.80	0.59	0.44
σ_R^2	1.33	2.00	20.66	0.98	1.36	1.69	20.67	1.11	0.35
μ_{WW}	9.54	8.65	4.25	6.03	7.52	9.81	4.34	3.19	2.80
Hansen-Sargan J stat		2.19	3.44	0.26	0.91	2.59			
p-val		0.14		0.61	0.34	0.11			

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	1.04	0.37	0.89	1.02	1.27	0.37	0.27	0.17
β	1	0.96	0.37	0.72	0.98	1.15	0.38	0.28	0.21
σ_U^2	1.72	1.33	0.90	0.81	1.33	1.83	0.98	0.76	0.59
σ_V^2	1.33	1.51	0.62	1.15	1.35	1.74	0.64	0.43	0.25
σ_R^2	1.33	1.36	0.44	1.12	1.34	1.66	0.44	0.34	0.26
μ_{WW}	9.54	8.88	2.48	7.29	8.39	9.85	2.57	1.95	1.71
Hansen-Sargan J stat		4.10	7.55	0.57	1.53	3.71			
p-val		0.04		0.45	0.22	0.05			

Exactly Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	1.09	0.54	0.87	1.18	1.43	0.55	0.39	0.34
β	1	1.02	1.31	0.53	0.87	1.23	1.31	0.51	0.39
σ_U^2	1.72	1.11	1.45	0.00	0.84	1.67	1.57	1.15	1.11
σ_V^2	1.33	1.90	1.22	1.09	1.65	2.45	1.35	0.90	0.61
σ_R^2	1.33	1.49	0.63	1.13	1.55	1.94	0.64	0.52	0.46

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	1.14	0.49	0.91	1.23	1.50	0.51	0.41	0.40
β	1	0.86	0.49	0.49	0.79	1.09	0.51	0.41	0.38
σ_U^2	1.72	1.06	1.17	0.00	1.00	1.72	1.34	1.01	0.90
σ_V^2	1.33	1.99	1.03	1.19	1.76	2.72	1.22	0.90	0.65
σ_R^2	1.33	1.54	0.60	1.21	1.64	2.01	0.63	0.53	0.52

Notes: Design 2: $\ln(U) \sim N(-0.5, 1)$, $V \sim U(-2, 2)$, $R \sim U(-2, 2)$.

Table 3: Design 3

Over Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.94	0.66	0.79	1.01	1.27	0.66	0.37	0.24
β	1	1.21	1.46	0.56	0.97	1.53	1.48	0.70	0.47
σ_U^2	1.64	1.77	1.23	0.84	1.74	2.50	1.24	0.97	0.84
σ_V^2	1.72	1.32	1.30	0.19	1.08	2.09	1.36	1.11	1.09
σ_R^2	1.72	1.55	4.92	0.74	1.24	1.87	4.92	0.98	0.72
μ_{WW}	10	9.01	2.20	7.53	8.70	10.17	2.41	1.96	1.78
Hansen-Sargan J stat		1.74	2.71	0.23	0.82	2.19			
p-val		0.19		0.63	0.37	0.14			

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.93	0.40	0.86	0.96	1.06	0.41	0.19	0.11
β	1	1.06	1.04	0.72	0.93	1.20	1.04	0.40	0.25
σ_U^2	1.64	1.95	0.89	1.40	1.96	2.53	0.94	0.73	0.64
σ_V^2	1.72	1.24	1.02	0.44	1.10	1.86	1.13	0.93	0.87
σ_R^2	1.72	1.62	1.96	1.07	1.49	1.95	1.97	0.65	0.48
μ_{WW}	10	9.39	1.26	8.56	9.28	10.08	1.40	1.12	0.97
Hansen-Sargan J stat		2.30	6.35	0.29	0.96	2.46			
p-val		0.13		0.59	0.33	0.12			

Exactly Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.99	0.55	0.86	1.04	1.27	0.55	0.32	0.22
β	1	2.23	8.86	0.79	1.17	1.86	8.95	1.51	0.45
σ_U^2	1.64	1.58	1.24	0.56	1.46	2.34	1.24	0.99	0.89
σ_V^2	1.72	1.77	1.48	0.71	1.65	2.47	1.48	1.06	0.88
σ_R^2	1.72	1.40	1.46	0.57	1.14	1.86	1.49	1.02	0.84

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.96	0.33	0.89	0.99	1.09	0.33	0.18	0.10
β	1	1.68	6.85	0.83	1.06	1.38	6.88	0.91	0.26
σ_U^2	1.64	1.76	0.95	1.14	1.73	2.31	0.95	0.73	0.60
σ_V^2	1.72	1.60	1.07	0.86	1.57	2.24	1.08	0.83	0.69
σ_R^2	1.72	1.60	1.04	1.01	1.45	2.01	1.04	0.75	0.57

Notes: Design 3: $U \sim \text{Gumbel}(0,1), \ln(V) \sim N(-0.5,1), \ln(R) \sim N(-0.5,1)$.

Table 4: Design 4

Over Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.95	0.56	0.83	1.01	1.18	0.56	0.28	0.18
β	1	1.23	2.99	0.32	0.97	1.54	3.00	0.86	0.62
σ_U^2	1.33	1.67	1.41	0.86	1.36	2.15	1.45	0.92	0.59
σ_V^2	1.72	1.26	1.25	0.19	1.07	1.88	1.33	1.05	0.95
σ_R^2	1.72	1.42	4.82	0.57	1.17	1.77	4.83	1.02	0.76
μ_{WW}	8.76	8.06	1.86	6.93	7.79	8.89	1.99	1.53	1.32
Hansen-Sargan J stat		2.45	4.63	0.22	0.85	2.52			
p-val		0.12		0.64	0.36	0.11			

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.99	0.23	0.92	0.99	1.06	0.23	0.12	0.07
β	1	1.09	2.12	0.55	1.00	1.42	2.12	0.55	0.44
σ_U^2	1.33	1.63	1.10	0.98	1.32	1.99	1.14	0.73	0.44
σ_V^2	1.72	1.43	1.02	0.73	1.43	2.02	1.06	0.82	0.65
σ_R^2	1.72	1.39	1.31	0.79	1.33	1.87	1.36	0.74	0.60
μ_{WW}	8.76	8.43	1.10	7.74	8.30	8.94	1.15	0.88	0.74
Hansen-Sargan J stat		3.86	11.48	0.26	0.87	2.49			
p-val		0.05		0.61	0.35	0.12			

Exactly Identified Moments

N = 100

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.91	0.63	0.86	1.00	1.13	0.64	0.27	0.13
β	1	1.99	4.54	0.67	1.33	2.04	4.65	1.35	0.66
σ_U^2	1.33	1.47	1.30	0.67	1.21	1.93	1.30	0.87	0.64
σ_V^2	1.72	1.56	1.48	0.48	1.36	2.25	1.49	1.09	0.93
σ_R^2	1.72	1.26	3.13	0.00	0.93	1.74	3.16	1.23	1.05

N = 400

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
γ	1	0.99	0.16	0.93	1.00	1.06	0.16	0.09	0.06
β	1	1.70	2.62	0.48	1.27	1.94	2.71	1.13	0.67
σ_U^2	1.33	1.62	1.21	0.80	1.22	2.46	1.25	0.92	0.65
σ_V^2	1.72	1.44	1.16	0.15	1.47	2.19	1.19	0.95	0.83
σ_R^2	1.72	1.18	1.17	0.00	1.12	2.01	1.29	1.03	0.98

Notes: Design 4: $U \sim U(-2,2), \ln(V) \sim N(-0.5,1), \ln(R) \sim N(-0.5,1)$.

Table 5: Replication of Acemoglu and Johnson (2007): Dependent Variable: Growth in GDP per Capita, 1940-1980

	(1) OLS	(2) 2SLS1	(3) 2SLS2	(4) 2SLS3
Growth in life expectancy	-0.813***	-1.316***	-1.643***	-1.589*
	(0.258)	(0.350)	(0.521)	(0.876)
Quality of institutions			-0.0490	
			(0.0418)	
Log GDP per capita 1930				-0.0730
				(0.198)
Constant	1.163***	1.336***	1.681***	1.990
	(0.0907)	(0.119)	(0.367)	(1.807)
Observations	47	47	47	47
R-squared	0.135	0.083	0.065	0.029

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Over identified moments: Base sample 1940 and 1980

	(1) 2SLS1	(2) 2SLS2	(3) 2SLS3	(4) GMM1	(5) GMM2	(6) GMM3	(7) GMM1+AJ	(8) GMM2+AJ	(9) GMM3+AJ
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.316***	-1.643***	-1.589*	-1.676**	-1.309***	-3.096	-1.365***	-1.624***	-1.592*
	(0.350)	(0.521)	(0.876)	(0.744)	(0.496)	(2.217)	(0.323)	(0.500)	(0.863)
Institutions		-0.0490			-0.0326			-0.0468	
		(0.0418)			(0.0426)			(0.0388)	
Initial GDP			-0.0730			-0.340			-0.0617
			(0.198)			(0.409)			(0.191)
Constant	1.336***	1.681***	1.990		1.488***	4.545		1.661***	1.899
	(0.119)	(0.367)	(1.807)		(0.376)	(3.874)		(0.341)	(1.744)
Panel B. β and variances									
β				2.562	4.909	2.687	2.165	2.005	1.202
				(1.712)	(11.35)	(2.907)	(1.147)	(3.403)	(1.597)
σ_U^2				0.00941***	0.00159**	0.0142***	0.00826***	0.00707**	0.0136***
				(0.0138)	(0.00416)	(0.00626)	(0.00644)	(0.0153)	(0.0160)
σ_V^2				0.0202***	0.0178***	2.05e-06	0.0219***	0.0125***	1.86e-05
				(0.0140)	(0.00492)	(0.00535)	(0.00658)	(0.0137)	(0.0157)
σ_R^2				0.0665***	0.0918**	0.123***	0.0832***	0.108***	0.124***
				(0.0327)	(0.0955)	(0.0424)	(0.0178)	(0.0515)	(0.0318)
μ_{ww}				0.146***	0.143***	0.125***	0.129***	0.142***	0.125***
				(0.0258)	(0.0275)	(0.0207)	(0.0228)	(0.0274)	(0.0207)
Panel C. Dependent Variable: Growth in life expectancy 1940-1980									
Institutions		-0.0310***			-0.0496***			-0.0491***	
		(0.00755)			(0.00999)			(0.00959)	
Initial GDP			-0.117***			-0.183***			-0.183***
			(0.0309)			(0.0205)			(0.0202)
Constant		0.324***	1.122***		0.579***	1.751***		0.577***	1.744***
		(0.0595)	(0.267)		(0.0523)	(0.160)		(0.0502)	(0.157)
Observations		47	47	47	47	47	47	47	47
Hansen J				0.389	0.00192	0.136	2.419	0.0158	0.228
p-val				0.533	0.965	0.712	0.298	0.992	0.892

Notes: In all models, the endogenous regressor is the changes in log life expectancy between 1940 and 1980. 2SLS1 is the two-stage least squares regression of growth in GDP per capita on growth in life expectency, using predicted mortality as instrument. 2SLS2 includes a measure of quality of institutions as exogenous covariate. 2SLS3 adds log GDP per capita in 1930. GMM1-GMM3 are the same models as 2SLS1-2SLS3 estimated by GMM estimators based on over identified moments. GMM1-GMM3+AJ combine our over identified moments and the AJ moment, i.e. $E(IV\varepsilon) = 0$.

Table 7: Exactly identified moments: Base sample 1940 and 1980

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2SLS1	2SLS2	2SLS3	GMM4	GMM5	GMM6	GMM4+AJ	GMM5+AJ	GMM6+AJ
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.316***	-1.643***	-1.589*	-3.046	-1.376	-0.378	-1.364***	-1.627***	-1.598*
	(0.350)	(0.521)	(0.876)	(2.784)	(2.230)	(0.458)	(0.347)	(0.519)	(0.871)
Institutions		-0.0490			-0.0358			-0.0471	
		(0.0418)			(0.120)			(0.0415)	
Initial GDP			-0.0730			0.150			-0.0704
			(0.198)			(0.123)			(0.194)
Constant	1.336***	1.681***	1.990		1.527	-0.139		1.663***	1.967
	(0.119)	(0.367)	(1.807)		(1.335)	(1.098)		(0.364)	(1.775)
Panel B. β and variances									
β				2.791	13.47	2.090	2.147	0.757	1.243
				(1.968)	(484.6)	.	(1.204)	(5.985)	(1.231)
σ_U^2				0.0248***	0.000699	0	0.00836***	0.0187	0.0137***
				(0.0146)	(0.0221)	(0)	(0.00741)	(0.147)	(0.0126)
σ_V^2				0.00587**	0.0188***	0.0140***	0.0217***	0.000721	0.000243
				(0.0140)	(0.0228)	(0.00439)	(0.00739)	(0.148)	(0.0103)
σ_R^2				0.0838***	0.00466	0.122***	0.0846***	0.126***	0.122***
				(0.0227)	(5.082)	(0.0213)	(0.0194)	(0.0804)	(0.0319)
Panel C. Dependent Variable: Growth in life expectancy 1940-1980									
Institutions		-0.0310***			-0.0496***			-0.0494***	
		(0.007555)			(0.00997)			(0.00994)	
Initial GDP			-0.118***			-0.184***			-0.184***
			(0.0309)			(0.0216)			(0.0219)
Constant		0.324***	1.122***		0.579***	1.758***		0.577***	1.758***
		(0.0595)	(0.267)		(0.0523)	(0.168)		(0.0521)	(0.170)
Observations		47	47	47	47	47	47	47	47
Hansen J							2.257	0.0129	0.03
p-val							0.133	0.994	0.985

Notes: In all models, the endogenous regressor is the changes in log life expectancy between 1940 and 1980. 2SLS1 is the two-stage least squares regression of growth in GDP per capita on growth in life expectency, using predicted mortality as instrument. 2SLS2 includes a measure of quality of institutions as exogenous covariate. 2SLS3 adds log GDP per capita in 1930. GMM4-GMM6 are the same models as 2SLS1-2SLS3 estimated by GMM estimators based on exactly identified moments. GMM4-GMM6+AJ combine our exactly identified moments and the AJ moment, i.e. $E(IV\varepsilon) = 0$.

Table 8: Expanded sample 1940 and 1980

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2SLS1	2SLS2	2SLS3	GMM1	GMM2	GMM3	GMM4	GMM5	GMM6
Panel A. Dependent Variable: Growth in GDP per Capita 1940-1980									
Life expectancy	-1.629***	-1.897***	-2.405**	-1.799***	-1.114	-10.47*	-1.934*	-1.275	-9.016*
	(0.491)	(0.619)	(1.0114)	(0.633)	(6.147)	(5.633)	(1.019)	(8.586)	(5.011)
Institutions		-0.0582			-0.0368			-0.0382***	
		(0.0409)			(0.219)			(0.00856)	
Initial GDP			-0.265			-1.497*			-0.165***
			(0.220)			(0.807)			(0.0216)
Constant	1.474***	1.822***	3.744*		1.458	15.77**		0.500***	1.587***
	(0.182)	(0.392)	(2.00615)		(2.954)	(7.970)		(0.0431)	(0.168)
Panel B. β and variances									
β				4.008	0.00438	10.30***	3.427	35.58	8.340***
				(4.455)	(6.286)	(6.182)	(4.645)	(3,873)	(5.403)
σ_U^2				0.00573***	0.0210***	0.0148***	0.00807**	0.000134	0.0148***
				(0.0102)	(0.00373)	(0.00390)	(0.0191)	(0.00929)	(0.00383)
σ_V^2				0.0214***	8.73e-08	0.00143***	0.0196***	0.0218***	0.000972***
				(0.0113)	(0)	(0.00102)	(0.0193)	(0.00926)	(0.00114)
σ_R^2				0.0961***	0.135	0.00528	0.103***	5.95e-05	0.142***
				(0.0523)	(0.182)	(0.262)	(0.0492)	(25.19)	(0.0879)
μ_{ww}				0.193***	0.197***	0.184***			
				(0.0549)	(0.0647)	(0.0734)			
Panel C. Dependent Variable: Growth in life expectancy 1940-1980									
Institutions		-0.0254***			-0.0382***			-0.0382***	
		(0.00613)			(0.00860)			(0.00856)	
Initial GDP			-0.0943***			-0.160***			-0.165***
			(0.0216)			(0.0216)			(0.0216)
Constant		0.278***	1.504***		0.499***	1.550***		0.500***	1.587***
		(0.0458)	(0.182)		(0.0435)	(0.169)		(0.0431)	(0.168)
Observations	56	56	56	56	56	56	56	56	56
Hansen J				0.0152	0.0181	0.188			
p-val				0.902	0.893	0.665			

Notes: The expanded sample contains 56 countries, which have data for life expectancy and GDP per capita in 1940 and 1980. In all models, the endogenous regressor is the changes in log life expectancy between 1940 and 1980. GMM1 is the baseline model. GMM2 includes a measure of quality of institutions as exogenous covariate. GMM3 adds log GDP per capita in 1930. GMM1-GMM3 are estimated by GMM estimators based on over identified moments. GMM4-GMM6 are the same models as in GMM1-GMM3, but are estimated by exactly identified moments instead.