

A Monte Carlo study of growth regressions

William R. Hauk Jr. · Romain Wacziarg

Published online: 1 May 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Using Monte Carlo simulations, this paper evaluates the bias properties of estimators commonly used to estimate growth regressions derived from the Solow model. We explicitly allow for measurement error, country-specific fixed effects and regressor endogeneity. An OLS estimator applied to a single cross-section of variables averaged over time (the between estimator) performs best in terms of the extent of bias on each of the estimated coefficients. Fixed-effects and the Arellano–Bond GMM estimator overstate the speed of convergence under a wide variety of assumptions, while the between estimator understates it. Finally, fixed effects and Arellano–Bond bias towards zero the slope estimates on the human and physical capital accumulation variables, while the between estimator and the Blundell–Bond system GMM estimator bias these coefficients upwards.

Keywords Growth regressions · Measurement error · System-GMM

JEL Classification O47 · O57 · C15 · C23

1 Introduction

In the last decade, spurred by the early work of [Baumol \(1986\)](#) and [Barro \(1991\)](#), growth regressions have become an industry. There is no good alternative for addressing the

W. R. Hauk Jr.
Department of Economics, Moore School of Business, University of South Carolina,
1705 College St., Columbia, SC 29208, USA

R. Wacziarg (✉)
UCLA Anderson School of Management, 110 Westwood Plaza,
Los Angeles, CA 90095, USA

R. Wacziarg
NBER, Cambridge, MA, USA

R. Wacziarg
CEPR, London, UK

fundamental question of what accounts for vast observed differences in per capita income across countries. Yet the detractors of growth regressions have stressed their numerous drawbacks. These include an often excessive distance between measured variables and the theoretical concepts they are meant to capture; poor grounding of estimated functional forms in economic theory, and in particular the prevalence of reduced form relationships from which structural parameters cannot be inferred; unjustified claims of causality in explanations of growth; a small number of available observations; and the prevalence of prior-driven data-mining. These are but a few in a growing list, resulting in numerous methodological debates on the proper way to run growth regressions. Many of these debates are as yet unresolved, so research evaluating the effectiveness of current methodologies and suggesting improvements to cross-country growth empirics appears necessary.

This paper is such a study. We use simulation methods to evaluate the bias properties of several estimators commonly used in the empirical growth literature.

The main contribution of our approach is to consider explicitly the impact of measurement error on estimates of the determinants of growth. Measurement error is likely to be a central problem in cross-country growth empirics, yet it has received little attention in the literature.¹ In the absence of measurement error and other sources of endogeneity, a fixed-effects estimator unambiguously dominates estimators that use any between-country variation, when omitted variables such as the initial level of technology are correlated with included right-hand side variables. In the presence of measurement error, however, fixed-effects estimators will tend to exacerbate measurement error bias when the right-hand side variables are more time persistent than the errors in measurement. The issue then is whether the gains from reducing omitted variables are offset by an increase in measurement error bias under fixed-effects. We lack guidance from econometric theory to evaluate the resulting net bias in the multivariate context of growth regressions, making Monte Carlo simulations necessary. Our results evaluating alternative estimators in the presence of a multiplicity of data problems can guide researchers towards the least biased estimators for growth regressions.

Another contribution of this paper is to help resolve a long-running methodological debate in growth empirics: whether the appropriate way to control for time invariant cross-country heterogeneity is to use a fixed-effects estimator, identifying parameters only using within-country variation, or to retain at least some between-country variation in the data and directly include as right-hand side variables available proxies for technological differences.² Advancing this methodological debate can have important implications for our substantive understanding of the growth process, because results obtained using fixed-effects differ markedly from those obtained by attempting to include additional correlates of growth. First, fixed-effects lead to estimates of the speed of conditional convergence that are much higher than the

¹ One exception occurs in [De Long's \(1998\)](#) comment on [Baumol \(1986\)](#), where he stressed the potential for measurement error on income to bias upwards estimates of the speed of convergence. Another notable exception is [Barro \(1997, Chap. 1, p. 36\)](#), who briefly discusses the probable upward bias in fixed-effects estimates of the speed of convergence due to measurement error. [Griliches and Hausman \(1986\)](#) make a related point, though not in the specific context of growth regressions, and [Wansbeek \(2001\)](#) builds on their paper to develop a systematic way to consistently estimate models with mismeasured data using GMM. See also [Temple \(1998\)](#) for a discussion of the likely impact of measurement error (among other data problems) on estimates of the rate of convergence in the augmented Solow model, and [Durlauf et al. \(2005\)](#) for a general discussion of data problems that plague growth empirics.

² The first approach is associated with the work of [Knight et al. \(1993\)](#); [Islam \(1995\)](#); [Caselli et al. \(1996\)](#) and [Bond et al. \(2001\)](#) among others. Since the mid-1990s, the use of dynamic panel data estimators in growth empirics has become prevalent. The second approach can be likened to a “kitchen-sink” method, in which the unaccounted variation in economic growth is attributed to additional right-hand side variables, to capture institutions, policies and economic structures. See [Wacziarg \(2002\)](#) for a discussion of these approaches.

conventional 2% obtained in cross sectional studies, suggesting that any income gains can be obtained primarily from improvements in steady-state determinants and not the forces of convergence. Second, the fixed-effects estimator tends to reduce the magnitudes and statistical significance of the estimated coefficients on right-hand variables compared to cross-sectional alternatives, so that given improvements in steady-state determinants yield smaller steady-state income gains.³ Third, country-specific fixed effects account for a large share of the cross-country variation in income levels, suggesting that technological differences rather than differences in factor quantities account for a large share of differences in economic performance.⁴

To evaluate methodologies for growth empirics, we start with the [Solow \(1956\)](#) growth model, in its human capital-augmented version proposed by [Mankiw et al. \(1992\)](#). The Solow model is arguably the only solid theoretical foundation for the specific functional forms generally estimated by practitioners, which involves regressing growth rates on the log of initial income and a set of steady-state income determinants. A growth specification can be directly derived from this model, and reasonable values for the exogenous parameters of the model postulated. Thus, the Solow model appears ideally suited as a starting point for simulations. Modeling explicitly the dynamic nature of the Solow growth specification, we generate simulated data with moments resembling those of the empirical data. We then perform Monte Carlo simulations to evaluate the performance of commonly used estimators under a wide set of assumptions on the data generating process, including measurement error, reverse causality in the regressors, country fixed effects, etc. A recent line of research, starting with [Caselli et al. \(1996\)](#) focuses on GMM estimators that exploit the dynamic nature of the Solow model and use lagged variables as instruments, to address a variety of sources of endogeneity bias. We provide a detailed analysis of the Arellano–Bond GMM estimator and the more recent “system” GMM estimator introduced by [Arellano and Bover \(1995\)](#) and [Blundell and Bond \(1998\)](#).

Our results suggest that using a least-squares estimator applied to a single cross-section of variables averaged over time (the between estimator) best estimates the speed of conditional income convergence.⁵ The fixed-effects estimator, as well as the Arellano–Bond estimator, greatly overstate the speed of convergence under a variety of assumptions concerning the type and extent of measurement error. The random effects estimator also tends to overstate

³ For example, human capital variables that are highly significant in cross-sectional estimates become insignificantly different from zero in panel fixed-effects applications and sometimes reverse signs. See [Islam \(1995\)](#) and [Benhabib and Spiegel \(1994\)](#).

⁴ Partly as a result of this last finding, as well as skepticism about cross-country growth regressions, some researchers turned to a cross-country income accounting approach, aimed at accounting for variation in the level of income rather than in its growth rate—i.e. excluding lagged income from the right hand side of a regression explaining income levels. Salient examples include [Klenow and Rodríguez-Clare \(1997\)](#); [Hall and Jones \(1999\)](#); [Frankel and Romer \(1999\)](#); [Acemoglu et al. \(2001\)](#). This literature relies on calibrations or on income level regressions to identify the sources of cross-country income variation. See [Caselli \(2004\)](#) for an excellent survey. The specifications used in this new literature do not lend themselves well to Monte Carlo simulations: level regressions are less easily traceable to a single theoretical model than their growth counterparts, which in most cases can be traced back to some version of the neoclassical growth model. This relative lack of theoretical foundation makes simulations difficult, as it precludes setting structural parameters equal to theoretically-founded values. An exception are income levels regressions based on the augmented Solow model, as in [Mankiw et al. \(1992\)](#) and more recently in [Bernanke and Gürkaynak \(2001\)](#). The validity of these specifications, however, relies on the unrealistic assumption that countries are located at their steady-states. Moreover, growth regressions remain vastly more prevalent in the literature on the determinants of economic development. Thus, we focus on growth regressions in this paper, leaving the evaluation of cross-country income accounting to future research.

⁵ While it has been noted elsewhere that OLS may outperform other estimators in the presence of multiple sources of bias, we take this insight seriously and confirm it in the context of growth regressions.

the speed of convergence, though much less drastically than fixed-effects. The instrumental variables procedure of the Blundell–Bond GMM estimator goes some way towards correcting the deficiencies of the fixed effects and Arellano–Bond estimators. Finally, fixed-effects seriously biases toward zero the slope estimates on the determinants of the steady-state level of income (the accumulation and depreciation variables of the Solow model), in particular on the human capital accumulation rate. In contrast, random effects and between tend to overstate them (bias them away from zero).

This paper is structured as follows: Sect. 2 briefly discusses theoretical considerations related to the methodology of growth regressions. Section 3 presents our basic simulation methodology and results, contrasting OLS, fixed-effects, random effects, the Arellano–Bond and Blundell–Bond GMM estimators. Section 4 discusses extensions of our simulations to regressor-specific measurement error, autocorrelated measurement error, alternative true models, alternative regression specifications and additional estimators. Section 5 concludes by presenting new estimates of the speed of income convergence and of the effect of steady-state determinants using real data, and discusses them in light of our simulation results.

2 Theoretical framework

2.1 Growth regressions and the Solow model

The theoretical basis for our data-generating process is the Solow model. We choose it, not because of any priors that it is a particularly compelling model of growth, but because it is tractable and constitutes arguably the only strict theoretical basis for the specific functional forms often estimated in the vast cross-country growth literature. As we show below, this model is also well-suited for generating simulated data. Mankiw et al. (1992, henceforth MRW) and Islam (1995) showed that the Solow growth model can be written to allow its estimation through a simple application of linear regression techniques.

The derivation of the data-generating process is well-known and can be found in the working paper version of this paper (Hauk and Wacziarg 2004), as well as in Islam (1995) and MRW (1992). The result is the following growth equation:

$$\begin{aligned} \log y(t_2) = & (1 - e^{-\lambda\tau}) \frac{\alpha}{1 - \alpha - \beta} \log s_k + (1 - e^{-\lambda\tau}) \frac{\beta}{1 - \alpha - \beta} \log s_h \\ & - (1 - e^{-\lambda\tau}) \frac{\alpha + \beta}{1 - \alpha - \beta} \log (n + g + \delta) \\ & + e^{-\lambda\tau} \log y(t_1) + (1 - e^{-\lambda\tau}) \log A(0) + g(t_2 - e^{-\lambda\tau} t_1) \end{aligned} \quad (1)$$

where $y(t)$ is a country's per-capita income level at time t , s_k and s_h are the country's savings rates in physical and human capital, α and β are the shares of the country's income attributed to physical and human capital, n and g are the growth rates of the country's population and technology level, δ is the rate at which these variables depreciate, $\lambda = (n + g + \delta)(1 - \alpha - \beta)$ is the rate at which the economy converges to its steady-state level of income, and $\tau = t_2 - t_1$ is the time between two observations on the income variable. The $\log y(t_1)$ term on the right-hand side implies that this specification is a growth and not a levels regression. Adding an error term v_{it} with mean zero conditional on all the right-hand side variables, capturing inherent randomness in $\log y_{it}$, we can rewrite Eq. 1 as a fixed-effects panel data regression of the form:

$$\begin{aligned} \log y_{it} = & \gamma_0 + \gamma_1 \log s_{k,it-\tau} + \gamma_2 \log s_{h,it-\tau} + \gamma_3 \log (n + g + \delta)_{it-\tau} \\ & + \gamma_4 \log y_{it-\tau} + \mu_i + \eta_t + v_{it} \end{aligned} \tag{2}$$

where t denotes the end of a time period of duration τ and $t - \tau$ denotes the beginning of that period.⁶ The reduced form parameters and error terms are defined as:

$$\begin{aligned} \gamma_1 &= (1 - e^{-\lambda\tau}) \frac{\alpha}{1-\alpha-\beta} \\ \gamma_2 &= (1 - e^{-\lambda\tau}) \frac{\beta}{1-\alpha-\beta} \\ \gamma_3 &= - (1 - e^{-\lambda\tau}) \frac{\alpha+\beta}{1-\alpha-\beta} \\ \gamma_4 &= e^{-\lambda\tau} \\ \gamma_0 + \mu_i &= (1 - e^{-\lambda\tau}) \log A_i(0) \quad (\text{an intercept plus a country effect}) \\ \eta_t &= g(t - e^{-\lambda\tau}(t - \tau)) \quad (\text{a time specific effect}) \\ v_{it} & \quad (\text{a zero-mean error term, possibly correlated with the regressors}) \end{aligned}$$

Equation 2 is the functional form used as the data-generating process for the remainder of this paper. In what follows we will sometimes find it useful to rewrite Eq. 2 as:

$$\log y_{it} = \gamma' x_{it} + \mu_i + \eta_t + v_{it} \tag{3}$$

where we define $x'_{it} = [1, \log s_{k,it-\tau}, \log s_{h,it-\tau}, \log (n + g + \delta)_{it-\tau}, \log y_{it-\tau}]$ and $\gamma' = [\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4]$, with the dimension of these vectors denoted $Q = 5$.

2.2 Country-level heterogeneity

The log $A(0)$ term constitutes a stumbling block for growth regressions. This term captures the initial level of technology, often approximated using variables such as resource endowments, climate, institutions, government type, and so on. These variables vary widely across countries, so that we index $A(0)$ by i . Hence, we define $\gamma_0 + \mu_i \equiv (1 - e^{-\lambda\tau}) \log A_i(0)$, where γ_0 is a constant capturing the average level of initial technology across countries and μ_i is a zero-mean country-specific effect. There have been three basic ways of dealing with country-level heterogeneity (i.e. the $\mu_i = (1 - e^{-\lambda\tau}) \log A_i(0) - \gamma_0$ term) to estimate growth regressions. These methods are associated with the contributions of MRW (1992), Islam (1995), Caselli et al. (1996, henceforth CEL) and Bond et al. (2001, henceforth BHT), respectively.

2.2.1 MRW (1992) and Islam (1995)

MRW assumed the μ_i term had mean zero conditional on other right-hand side variables. As a result, they ran simple OLS regressions of growth on the log of initial income and time-averaged steady-state determinants (i.e. a single cross-section), including an intercept in the regression to account for γ_0 . This approach causes the estimated coefficients to be biased if the orthogonality assumption is violated, a major limitation.

Assume that we want to estimate the parameters of the panel data model of Eq. 3:

$$\log y_{it} = \gamma' x_{it} + \mu_i + \varepsilon_{it} \tag{4}$$

⁶ In our actual empirical application of Eq. 2 the determinants of the steady-state level of income $\log s_k$, $\log s_h$, and $\log(n + g + \delta)$ are entered as averages over the period $t - \tau$ to t , rather than their beginning of period values. This is consistent with the common practice of growth regressions, as in MRW and Islam, where introducing right-hand side variables as period averages is thought to limit the extent of classical measurement error. Note that theory gives us no guidance on this choice, as it considers these regressors to be constant.

where μ_i is *not* assumed independent from x_{it} and $\varepsilon_{it} = \eta_t + v_{it}$ is a well-behaved random-noise term.

Define X as the $NT \times Q$ matrix that stacks x_{it} over time periods $t = 1 \dots T$ and countries $i = 1 \dots N$, and μ as the $NT \times 1$ vector that similarly stacks μ_i . Let Σ_{xx} be the $Q \times Q$ covariance matrix of columns of X , and $\Sigma_{\mu x}$ the $Q \times 1$ vector of the covariances of μ with the columns of X . If a pooled OLS (POLS) regression is run on the stacked data, standard omitted variables bias will result:⁷

$$\text{plim } \hat{\gamma}^{POLS} = \gamma + \Sigma_{xx}^{-1} \Sigma_{\mu x} \tag{5}$$

Equation 5 implies that slope estimates will be biased if the country-specific effects are correlated with the regressors. In our particular application, the Solow model states that the omitted term captures some positive multiple of the initial level of technology $\log A_i(0)$. The observed data in X are initial income, rates of human and physical capital accumulation and population growth. While the Solow model strictly speaking is silent about the correlation between the $\log A_i(0)$ term and the right-hand side variables, there is a strong presumption that these four variables will be potentially highly correlated with $\log A_i(0)$. Hence, estimated coefficients could be seriously biased when the correlation between $\log A_i(0)$ and the steady-state determinants is ignored.

This is Islam’s (1995) original point, advocating the use of fixed-effects estimation instead of OLS on country means. Islam averaged annual data from the available sample of countries across time, into 5-year periods. μ_i is a time-invariant effect if λ is treated as a constant and τ does not vary with time (i.e. the panel involves equally spaced periods). Hence, it can be represented as a country fixed-effect in a panel regression. The term $g(t - e^{-\lambda\tau}(t - \tau))$ is a time effect. Using a fixed-effects estimator, Islam found the estimated rate of convergence was much higher than had been estimated by MRW, and the effect of some right-hand side variables smaller (particularly human capital).

2.2.2 CEL (1996)

Going one step further, CEL pointed out the *necessary* correlation between the country-specific effect μ_i and the log of initial income resulting from the dynamic nature of the specification. We can rewrite Eq. 3 as:

$$\log y_{it} = \gamma_0 + \gamma'_s w_{it} + \gamma_4 \log y_{it-\tau} + \mu_i + \eta_t + v_{it} \tag{6}$$

where $\gamma'_s = [\gamma_1, \gamma_2, \gamma_3]$ and $w_{it} = [\log s_{k,it-\tau}, \log s_{h,it-\tau}, \log(n + g + \delta)_{it-\tau}]$. Lagging Eq. 6 by one period, it is evident that $\log y_{it-\tau}$ contains μ_i . Thus, $\log y_{it-\tau}$ *must* be correlated with the error term unless μ_i is appropriately accounted for.

CEL transformed all variables used in the regressions into deviations from period means (thereby removing the need for a time-specific intercept η_t) and then eliminated the country-specific effects μ_i by taking first-differences. Their transformed regression is:

$$\begin{aligned} \widetilde{\log y_{i,t}} - \widetilde{\log y_{i,t-\tau}} &= \gamma'_s (\widetilde{w_{i,t}} - \widetilde{w_{i,t-\tau}}) \\ &+ \gamma_4 (\widetilde{\log y_{i,t-\tau}} - \widetilde{\log y_{i,t-2\tau}}) + (\widetilde{v_{i,t}} - \widetilde{v_{i,t-\tau}}) \end{aligned} \tag{7}$$

⁷ This bias is also known as heterogeneity bias. We will use the terms heterogeneity bias and omitted variables bias interchangeably.

where “ $\tilde{\cdot}$ ” denotes deviations of variables from period means. In this specification, while μ_i and η_t have been differenced away, the term $\widetilde{\log y_{i,t-\tau}}$ is clearly not independent from $\tilde{v}_{i,t-\tau}$.⁸ Hence, some sort of instrumental variables approach is required. CEL proposed a GMM estimator similar to the [Arellano and Bond \(1991, henceforth AB\)](#), estimator to deal with the problems of heterogeneity bias and endogeneity of the differenced lagged income term in Eq. 7.⁹ Their estimator results in a consistent estimator for the unknown parameters under the moment condition $E[\tilde{v}_{i,t}\tilde{v}_{i,t-\tau}] = 0$. They instrument for the differenced independent variables using all predetermined independent variables (in levels). Specifically, the panel consists of four time-periods, and the variables $(\widetilde{\log y_{i,1}} - \widetilde{\log y_{i,0}})$ and $(\tilde{w}_{i,1} - \tilde{w}_{i,0})$ for period 1 are instrumented using $y_{i,0}$ and $w_{i,0}$. Then, $(\widetilde{\log y_{i,2}} - \widetilde{\log y_{i,1}})$ and $(\tilde{w}_{i,2} - \tilde{w}_{i,1})$ are instrumented using $y_{i,0}, y_{i,1}, w_{i,0}$ and $w_{i,1}$ and so on. The exclusion of the current period $\tilde{w}_{i,t}$ term from the list of instruments is meant to deal with the possible endogeneity of the variables in w_{it} , a valid procedure if all of the instrumental variables are assumed to be predetermined.¹⁰ Consistent estimates are obtained even in the presence of measurement error on the right-hand side variables, as long as the instruments are not correlated with the errors in measurement, for example if these are white noise. An additional advantage of this procedure is to address bias arising from endogenous regressors, under the assumption that the lagged regressors in levels, used as instruments, are predetermined.

Estimators in the class of the AB estimator require losing at least two periods of data in order to implement the IV procedure, which could bias estimates when T is small. Another recently identified drawback of AB is the problem of weak instruments arising in small samples (small N), biasing GMM estimates towards their fixed-effects counterparts. This is likely to be a problem especially when the convergence parameter γ_4 is large and when the variance of the country-specific effect μ_i is large relative to the variance of $\log y_{it}$. There is now a sizable literature on weak instruments. For example, [Stock et al. \(2002\)](#) show that in a two-stage least squares (2SLS) context, if the instruments in the first stage do not help *at all* in predicting the endogenous regressors, 2SLS reduces exactly to OLS.¹¹ [Staiger and Stock \(1997\)](#) provide a rule of thumb for determining whether instruments are weak in the linear IV case with one endogenous regressor: if the first stage F -test for the joint significance of the instruments is smaller than 10, then the instruments are declared to be weak. For the case of multiple endogenous regressors, [Stock and Yogo \(2003\)](#) propose using the [Cragg and Donald \(1993\)](#) test statistic for underidentification, but using appropriately corrected critical values in order to use the statistic for a test of the null hypothesis of weak instruments.

We apply this test in Sect. 3 in order to assess whether the AB estimates of the Solow model are likely to be subject to the weak instruments problem, finding that they are.¹² We

⁸ We will refer to this source of bias as endogeneity bias, to differentiate it from heterogeneity bias.

⁹ Estimators in this class have been widely used in the empirical growth literature. See for instance [Forbes \(2000\)](#), [Benhabib and Spiegel \(1994\)](#), [Easterly et al. \(1997\)](#) among many others.

¹⁰ The Solow model, however, treats the w_{it} variables as exogenous, so the endogeneity of w_{it} should not be a problem within the strict confines of the model. However, we allow for endogeneity of all RHS variables throughout our simulations.

¹¹ [Stock et al. \(2002\)](#) define the concentration parameter μ^2 as a measure of goodness of fit of the first stage regression, or equivalently a measure of the strength of the instruments. They state that: “When $\mu^2 = 0$ (...), the instruments are not just weak, but irrelevant. In this case, the mean of the 2SLS estimator is the probability limit of the ordinary least squares (OLS) estimator, $\text{plim}(\hat{\beta}^{OLS})$. (...) When the instruments are relevant but weak, the 2SLS estimator is biased toward $\text{plim}(\hat{\beta}^{OLS})$.” (p. 519).

¹² To evaluate the extent of small sample bias in dynamic panel estimators used to estimate growth regressions, [Islam \(2000\)](#) conducted a Monte Carlo study. He concluded that panel-IV estimators such as the

will also examine the properties of the AB estimator when N is increased beyond the number of countries usually available in growth regressions, to assess the extent of small sample bias.

2.2.3 BHT (2001)

To address the problem of weak instruments, [Arellano and Bover \(1995\)](#) and [Blundell and Bond \(1998\)](#) developed a “system GMM” estimator (henceforth BB), and [Levine et al. \(2000\)](#) and [BHT \(2001\)](#) applied it to the study of economic growth. In addition to instrumenting for differenced variables using lagged levels (see Eq. 7), BHT use lagged differences to instrument for levels variables. Both the equation in first differences (with levels instruments) and the equation in levels (with first-differenced instruments) are included in the system, to exploit the additional moment conditions valid for the latter equation. This procedure leads to consistent estimates under a stationarity assumption that the covariance between $\log y_{i,t-\tau}$ and μ_i is constant across values of τ (this assumption is required in addition to all the assumptions required for the AB estimator). Indeed, if $E(\mu_i y_{i,t-s}) = E(\mu_i y_{i,t-r})$ for all s, r , then the moment condition $E(\mu_i (y_{i,t-s} - y_{i,t-r})) = 0$ can be used in a GMM procedure that employs $y_{i,t-s} - y_{i,t-r}$ as an instrument for $y_{i,t-\tau}$ (where $s > r \geq \tau$).

BHT argue that the BB estimator addresses the weak instruments problem of the AB estimator, and that their IV procedure eliminates classical measurement error bias. Their estimates of the convergence parameter in the Solow model are lower than estimates produced using the AB estimator, consistent with the weak instruments argument. However, the additional moment conditions must be valid. BHT argue that this assumption is reasonable in the long-run growth context, if $\log y_{i,t-\tau}$ is represented as a deviation from period means as in the AB estimator. However, the Solow model involves not only long-run growth, but short-run convergence towards steady-states. μ_i , a function of the level of initial technology, is a determinant of the steady-state income level, and as such should be expected, within countries, to correlate with economic growth along the transition. Hence, it is unclear that the moment conditions needed for the BB estimator will hold. In the context of our simulations, this assumption can be evaluated directly—we do so below. If these moment conditions are invalid, then the BB estimator will be inconsistent in large samples. However, because of its desirable properties in addressing the weak instruments problem, BB may still be a good estimator in practice for small samples.¹³

2.3 Measurement error and heterogeneity bias

In this subsection, we examine what happens once we allow for measurement error in the independent variables, in addition to heterogeneity bias. We assume $E(\log y_{it} | x_{it}) = \gamma' x_{it} + \mu_i$ where μ_i is unobserved and not necessarily orthogonal to the variables in x_{it} .¹⁴ Thus, the true

Footnote 12 continued

AB estimator suffer from serious small sample/weak instruments problem, and that the fixed-effects (least-squares dummy variables) and the minimum distance estimators had the best small sample performance. He did not, however, consider the issue of measurement error. As we argue below, measurement error is a first order problem, greatly exacerbating small sample bias in estimators of the AB class.

¹³ In a further attempt to address the small sample bias problem, [Arellano \(2003\)](#) suggests a method to find the optimal set of GMM instruments using linear forecasts of all available lags as instruments, and applies it to the growth context. He reports an estimate of the speed of convergence lying in between that reported by MRW and CEL (about 4%). To our knowledge this estimator has not been applied more broadly in the growth context.

¹⁴ For simplicity in the rest of this section we continue to ignore the well-known problem arising from using a fixed-effects estimator in the presence of a lagged dependent variable in x_{it} —i.e. the problem identified in

model is $\log y_{it} = \gamma'x_{it} + \mu_i + \varepsilon_{it}$, where $E(\varepsilon_{it}|x_{it}) = 0$, $V(\varepsilon_{it}) = \sigma_\varepsilon^2$ and $\varepsilon_{it} = \eta_t + v_{it}$. Define $\bar{x} \equiv E(x_{it})$, $\bar{\mu} \equiv E(\mu_i)$ and $\sigma_\mu^2 \equiv V(\mu_i)$.

We assume that x_{it} cannot be observed, but that we can observe $x_{it}^* = x_{it} + d_{it}$, where $E[d_{it}] = 0$. Define:

$$\text{var}[d_{it}|x_{it}] = D = \text{diag} \left\{ \sigma_{d_0}^2, \sigma_{d_1}^2, \dots, \sigma_{d_{Q-1}}^2 \right\} \tag{8}$$

In our application, we will set $\sigma_{d_0}^2 = 0$, i.e. we do not shock the intercept column in X .

Estimating the model using pooled OLS will involve two separate problems: 1) Omitted variables bias, due to μ_i being potentially correlated with the right hand side variables in x_{it} , 2) Measurement error bias, due to x_{it} being imperfectly observed. In this case, it is difficult to make statements about the sign and magnitude of the bias. This can be seen by deriving the probability limit of the pooled OLS estimator in the presence of both measurement error and heterogeneity bias.

The unconditional mean and covariance matrix of $(\log y_{it}, x_{it}^*, \mu_i)$ are:

$$E(\log y_{it}, x_{it}^*, \mu_i) = (\bar{x}'\gamma + \bar{\mu}, \bar{x}, \bar{\mu}) \tag{9}$$

and:

$$V(\log y_{it}, x_{it}^*, \mu_i) = \begin{bmatrix} \sigma_\varepsilon^2 + \gamma' \Sigma_{xx} \gamma + 2\gamma' \Sigma_{\mu x} + \sigma_\mu^2 & \gamma' \Sigma_{xx} + \Sigma'_{\mu x} & \gamma' \Sigma_{\mu x} + \sigma_\mu^2 \\ \Sigma_{xx} \gamma + \Sigma_{\mu x} & \Sigma_{xx} + D & \Sigma_{\mu x} \\ \gamma' \Sigma_{\mu x} + \sigma_\mu^2 & \Sigma'_{\mu x} & \sigma_\mu^2 \end{bmatrix} \tag{10}$$

Suppose that we estimate γ using pooled OLS with x_{it}^* as our observed regressor. Then the limiting value of the pooled OLS estimator for γ is:

$$\text{plim } \hat{\gamma}^{POLS} = (\Sigma_{xx} + D)^{-1} \Sigma_{xx} \gamma + (\Sigma_{xx} + D)^{-1} \Sigma_{\mu x} \tag{11}$$

If $D = 0$, pooled OLS would not involve measurement error bias, and we would recover Eq. 5. On the other hand, if $\Sigma_{\mu x} = 0$, we would have $\text{plim } \hat{\gamma}^{POLS} = (\Sigma_{xx} + D)^{-1} \Sigma_{xx} \gamma$, showing that pooled OLS regression would not involve any heterogeneity bias. Even in this simpler case, we cannot say anything about the sign and magnitude of the bias unless we can make assumptions about the correlation structure among the various independent variables and the covariance matrix of measurement error, i.e. if we can place restrictions on $(\Sigma_{xx} + D)$ and Σ_{xx} , based on some knowledge of the nature of measurement error. This is in general a tall order.¹⁵

If neither of these two issues were a problem ($D = 0$ and $\Sigma_{\mu x} = 0$), pooled OLS would be a consistent estimator for γ . In a context where both problems coexist, there may be a trade-off between reducing the extent of bias due to measurement error and reducing the bias attributable to heterogeneity. The common way to deal with heterogeneity is to estimate γ using the fixed-effects (FE) estimator $\hat{\gamma}^{FE}$. Appendix 1 derives the limiting value of $\hat{\gamma}^{FE}$ in the presence of measurement error, showing that it gets rid entirely of heterogeneity bias. It also derives the limiting value of the between (BE) estimator $\hat{\gamma}^{BE}$, obtained by computing country means of the data over time and running OLS regressions on these country means.

Footnote 14 continued

the discussion following Eq. 7. We will take this issue into account in our simulations, which explicitly model the dynamic nature of the empirical Solow model.

¹⁵ For a discussion of this point, see Klepper and Leamer (1984).

Finally, it derives the plim of the random effects (RE) estimator $\hat{\gamma}^{RE}$, a matrix-weighted average of BE and FE. These limiting values are:

$$\text{plim } \hat{\gamma}^{POLS} = (\Sigma_{xx} + D)^{-1} \Sigma_{xx} \gamma + (\Sigma_{xx} + D)^{-1} \Sigma_{\mu x} \tag{12}$$

$$\text{plim } \hat{\gamma}^{BE} = \left(\Sigma_{xx}^B + \frac{1}{T} D \right)^{-1} \Sigma_{xx}^B \gamma + \left(\Sigma_{xx}^B + \frac{1}{T} D \right)^{-1} \Sigma_{\mu x} \tag{13}$$

$$\text{plim } \hat{\gamma}^{FE} = \left(\Sigma_{xx}^W + \frac{T-1}{T} D \right)^{-1} \Sigma_{xx}^W \gamma \tag{14}$$

$$\begin{aligned} \text{plim } \hat{\gamma}^{RE} &= \left[\left(\Sigma_{xx}^W + \frac{T-1}{T} D \right) + \theta \left(\Sigma_{xx}^B + \frac{1}{T} D \right) \right]^{-1} \\ &\times \left[\left(\Sigma_{xx}^W + \frac{T-1}{T} D \right) \gamma^{FE} + \theta \left(\Sigma_{xx}^B + \frac{1}{T} D \right) \gamma^{BE} \right] \end{aligned} \tag{15}$$

where Σ_{xx}^B denotes the between-country variation in X , Σ_{xx}^W denotes the within-country variation (as defined in Appendix 1), and $\Sigma_{xx} = \Sigma_{xx}^W + \Sigma_{xx}^B$.

It is difficult to make general statements about what happens to bias from measurement error under FE and BE estimation in the multivariate case. Measurement error bias on specific slope coefficients under FE may or may not be exacerbated. This is the main justification for a Monte Carlo approach to evaluating the properties of these estimators.¹⁶ However, a few statements can be made to compare the properties of pooled OLS, BE and FE estimators using Eqs. 12, 13 and 15:

- 1) As noted above, FE gets rid entirely of the heterogeneity bias while there is in general both measurement error and heterogeneity bias when using the BE and pooled OLS estimators.
- 2) The BE estimator tends to reduce the extent of measurement error bias compared to the other estimators by averaging the imperfectly measured variables over time, reducing the variance of the measurement error relative to the true signal. Moreover, the greater is T , the smaller the bias from measurement error.
- 3) Both pooled OLS and BE will involve smaller heterogeneity bias, the greater the extent of measurement error. In other words, there is an interaction between these two sources of bias. All other things equal, measurement error reduces the correlation between the regressors and the country effects, alleviating the omitted variables problem. For the same reason, BE will on average involve larger heterogeneity bias compared to pooled OLS.
- 4) Comparing FE and BE, if Σ_{xx}^W relative to $\frac{T-1}{T} D$ is “smaller” in a matrix sense than Σ_{xx}^B is relative to $\frac{1}{T} D$, then the bias arising from measurement error will tend to be smaller under BE compared to FE. This is likely to hold if the within variation is small compared to the between variation (most of the variation in the panel arises from the cross-section rather than the time dimension—which is the case in growth applications), or if T is large.

To conclude, despite these general lessons, we can say little about the net biases to individual parameter estimates as they would result from each estimation method. Given the multivariate nature of growth regressions, only simulations can determine which estimator dominates in terms of bias under alternative assumptions about the data-generating process,

¹⁶ In Sect. 2.4, and in Appendix 2, we discuss conditions under which measurement error is exacerbated under FE in the case of univariate regression. We argue these conditions are likely to hold in our particular application, as in most. Only a Monte Carlo simulation can provide definitive answers for the multivariate case.

i.e. the covariance structure of the true data Σ_{xx} , the covariances between the true variables and the country-specific effects $\Sigma_{\mu x}$ and the covariance matrix of the measurement error D .

2.4 Autocorrelated measurement error: univariate example

Appendix 2 analyzes in detail a simple case illustrating the trade-offs identified above in a case where net biases can be signed: the case of $Q = 2$ (a single regressor x_{it} plus an intercept). This example is also useful to illustrate what happens when measurement error is autocorrelated.

The example shows clearly that, under FE estimation, eliminating heterogeneity bias may come at the cost of exacerbating measurement error bias. The greater the time persistence in x_{it} , the greater the extent to which measurement error bias is exacerbated, as the variance of the true signal gets differenced away relative to the variance of the error in measurement. In the context of growth regressions, where right-hand side variables are highly time persistent, this point is particularly central. We cannot say analytically whether this increase in measurement error bias is worth the elimination of heterogeneity bias unless we know the moments of the true underlying data and of measurement error.

As argued in Sect. 2.2, a GMM procedure such as AB could in principle deal with both sources of bias if measurement error is white noise, since measurement error in the instruments is uncorrelated with measurement error in the regressors. However, considering measurement error weakens the first stage relationship between predetermined regressors and the instruments, potentially making the weak instruments problem worse. Moreover, the validity of this procedure relies on the assumption of non-autocorrelated errors in measurement.

Appendix 2 shows that when measurement error is autocorrelated, where we define $\rho_d = \text{corr}(d_{it}, d_{it-\tau})$, FE exacerbates measurement error bias compared to pooled OLS whenever $\rho_d < \rho_x$, where ρ_x is the autocorrelation coefficient in x_{it} . In this case, instrumenting for differenced x_{it} using its lagged levels values (as in the AB procedure) no longer eliminates measurement error bias. In words, quite apart from the weak instruments problem, as long as $\rho_d > 0$, we cannot produce a consistent estimator of the desired parameters using the AB estimator. We consider the case of autocorrelated measurement error in Sect. 4.

2.5 Endogeneity

Another concern we address is the issue of reverse causality between the regressors and the residual term v_{it} . This possibility is precluded by the strict version of the exogenous growth model on which we based our simulation. However, the endogenous nature of the steady-state determinants is empirically plausible. For example, a positive shock to growth might encourage investors to save more, thus raising $\log s_{k,it-\tau}$ —the so-called accelerator model of investment. Similarly, a positive shock to growth might cause higher values of $\log s_{h,it-\tau}$. There is also some evidence suggesting that, as incomes improve, people tend to have fewer children, thereby lowering $\log(n + g + \delta)_{it-\tau}$ while v_{it} rises.

These considerations would introduce another source of bias in estimators that do not instrument for steady-state determinants. All least-squares estimators rely on some form of the moment condition $\text{plim}[X'v] = 0$ for consistency. If this condition does not hold due to regressor endogeneity, then the plim of our estimates of γ will be:

$$\text{plim } \hat{\gamma} = \gamma + (\Sigma_{xx})^{-1} \Sigma_{xv} \tag{16}$$

where Σ_{xv} is the $k \times 1$ covariance matrix between the regressors and v_{it} .

Endogeneity was an explicit concern of the CEL and BHT studies cited above. In principle, both the AB and BB estimators should address the problem of endogeneity by exploiting the panel nature of growth data: both use lagged values of the regressors to instrument for current first-differences, and the BB system GMM estimator exploits a second set of moment conditions to use lagged first differences as an instrument for current levels. Both approaches produce valid instruments under the assumption that current-period shocks in v_{it} do not affect past values of the regressors and that past values of the regressors do not affect current values of $\log y_{it}$ directly. However, these instruments may be weak in small samples.

It is unclear how important regressor endogeneity is in practice. For example, assume that $Q = 2$. Then Eq. 16 reduces to:

$$\text{plim } \hat{\gamma} = \gamma + \frac{\sigma_{xv}}{\sigma_x^2}$$

where σ_{xv} is the covariance between x and v and σ_x^2 is the variance of x . It is apparent that the bias in $\hat{\gamma}$ is a function of the size of σ_{xv} relative to σ_x^2 . However, we cannot make σ_{xv} arbitrarily large. By definition, $\sigma_{xv} = \rho_{xv}\sigma_x\sigma_v$, where ρ_{xv} is the correlation coefficient between x and v , which also by definition must take on a value between -1 and 1 . Therefore,

$$\text{plim } \hat{\gamma} = \gamma + \rho_{xv} \frac{\sigma_v}{\sigma_x}$$

It necessarily follows that:

$$-\frac{\sigma_v}{\sigma_x} < \text{plim } \hat{\gamma} - \gamma < \frac{\sigma_v}{\sigma_x}$$

In words, the maximum possible bias due to endogeneity is bounded above and below by the ratio $\frac{\sigma_v}{\sigma_x}$. The smaller $|\rho_{xv}|$, the smaller this range. This general principle applies, in matrix form, for a multivariate regression. In the data-generating process for all our simulations, we explicitly allow for correlations between v_{it} and the regressors.

2.6 Summary

Seven factors can cause biased estimates of γ in panel growth regressions. The first is an omitted-variables bias resulting from the possible correlation between country-specific effects and the regressors, affecting the consistency of pooled OLS, BE and RE estimates. The second is the endogeneity problem specific to dynamic panels, identified after Eq. 7, making FE and RE estimates inconsistent. The third problem is the more general issue of the endogeneity of the steady-state determinants due to reverse causation. The fourth is classical measurement error on the independent variables, which affects the consistency of pooled OLS, BE, RE and FE estimator, though the bias tends to be exacerbated under FE and partly averaged away under BE. The fifth is possible autocorrelation in measurement errors, which results in bias for the AB estimator. The sixth is the weak instruments problem that can cause bias in the AB estimator in small samples. Finally, the seventh problem is the possible invalidity of the specific moment condition needed for the BB estimator to be consistent. A summary of these potential sources of bias can be found in Table 1.

Each of the estimators under consideration involves a trade-off: pooled OLS suffers from heterogeneity bias but limits the incidence of measurement error bias relative to FE; the BE estimator reduces measurement error through time averaging of the regressors, but does not deal with heterogeneity bias; FE addresses the problem of heterogeneity bias, but tends to exacerbate the problem of measurement error. The AB will not work well in small samples

Table 1 Potential sources of bias across estimators

Source of bias	Estimator				
	FE	BE	RE	AB	BB
Omitted country-specific effect	No bias	Bias	Bias	No bias	No bias
Endogeneity of lagged income term	Bias	No bias	Bias	No bias	No bias
Reverse causality in regressors	Bias	Bias	Bias	No bias in large samples	No bias in large samples
Regressor measurement error	Bias	Bias	Bias	No bias in large samples	No bias in large samples
Autocorrelated regressor measurement error	Bias	Bias	Bias	Bias	Bias
Weak instruments	N/A	N/A	N/A	Bias	Bias
GMM moment condition violations	N/A	N/A	N/A	No bias	Bias

when instruments are weak or when measurement error is autocorrelated, whereas the BB estimator will be inconsistent if the necessary moment conditions do not hold. To conclude, each of the estimators discussed above have their strengths and weaknesses, most of which cannot be evaluated analytically. Hence, to evaluate the bias properties of these estimators in the presence of a wide array of data problems, we turn to Monte Carlo simulation.

3 Monte Carlo simulations

3.1 Simulation methodology

Since it is impossible to derive analytical results about the extent and sign of omitted variables, measurement error biases and small sample biases in a multivariate context, we use Monte Carlo simulations to evaluate the bias properties of FE, BE, RE, AB and BB estimators. We consider additional estimators in Sect. 4.¹⁷ In our simulations, the data-generating process for the true data (the data not measured with error) is exactly Eq. 2, i.e. the augmented Solow model.

3.1.1 Simulated data

Underlying data. We define a period by five years ($\tau = 5$). Our underlying data spans 40 years, from 1960 to 2000, and our 8 five year periods are defined as 1960–1965, 1965–1970, . . . , 1995–2000. In Eq. 1, the variables $\log s_{k,it-\tau}$, $\log s_{h,it-\tau}$, $\log(n + g + \delta)_{i,t-\tau}$ and $\log y_{it-\tau}$ are simulated with moments resembling those of the corresponding observed variables. To obtain these moments, we captured $\log s_k$ using the log of investment rates as a share of real GDP from the Penn World Tables, version 6.1 (Heston et al. 2002, henceforth, PWT6.1). $\log s_h$ is the log of the secondary school gross enrollment ratio from Barro and Lee (2000) and n is the rate of population growth calculated from the PWT6.1 population

¹⁷ Results for pooled OLS estimates are available upon request. The pooled OLS estimator is rarely used in cross-sectional growth regressions since it is less efficient than random effects.

series.¹⁸ In calculating $\log(n + g + \delta)$, we assumed, conventionally, that $g + \delta = 0.07$. Finally, $\log y_{it-\tau}$ is the log of per capita income in purchasing power parity from PWT6.1, measured at the beginning of the first time period (1960).¹⁹

We averaged the variables over relevant time periods and arrayed them in a $N \times (T(Q - 1) + 1)$ matrix. Since our underlying data was available in all periods for 69 countries, and we are seeking a balanced panel, we set $N = 69$.

One assumption implied by our analysis is that the observed data surely incorporate measurement error. If we do not take this assumption into account, we will overstate the variances of the underlying true variables, and understate their cross-correlations. While we do not know with certainty how great the level of measurement error is in reality, we can adjust for our assumed levels of measurement in the simulations. It can easily be shown that, if two variables x_1 and x_2 are measured with white noise error d_1 and d_2 such that the observed values are $x_1^* = x_1 + d_1$ and $x_2^* = x_2 + d_2$, then:

$$\text{corr}(x_1, x_2) = \text{corr}(x_1^*, x_2^*) \times \sqrt{\left(1 + \frac{\text{var}(d_1)}{\text{var}(x_1)}\right) \left(1 + \frac{\text{var}(d_2)}{\text{var}(x_2)}\right)} \quad (17)$$

In our application, we set $\text{var}(d_1)/\text{var}(x_1) = \text{var}(d_2)/\text{var}(x_2) \equiv F$, where F is the assumed “error-to-truth” ratio. Therefore, for any choice of F , we adjust the values of variances and covariances of the regressors from the observed data using the assumed level of measurement error, to obtain the moments of the “true” data.

Simulating the fixed effects term. One difficulty we face is simulating the country fixed effects $\mu_i = (1 - e^{-\lambda\tau}) \log A_i(0) - \gamma_0$.²⁰ Theory provides no guidance as to the values of $A_i(0)$ for different countries, since it is taken as exogenous. The problem is important because the covariance structure linking the country-specific effects to the observed regressors determines how much heterogeneity bias will be present. To obtain simulated fixed effects and their corresponding covariance structure with the right-hand side variables, we used our observed panel data set to run a fixed-effects regression on the specification in Eq. 2. We computed the fitted country-specific effects from this regression. We then used this series and treated it as an additional variable, as if it were observed, to generate the moments of the simulated data.

Obviously, given that the underlying data must incorporate measurement error, this procedure will lead to biased estimates of the country-specific effects. If this is the case, the moments of μ_i and especially its covariance structure with the other right-hand side variables will also be flawed. It is therefore critical that we also present results with alternative assumptions about the covariance structure between the simulated fixed effects and the simulated regressors in x_{it} . We do so in Sect. 3.2.3, comparing simulation results obtained when varying the extent of the assumed correlations between μ_i and x_{it} .

¹⁸ Our measure of s_h differs slightly from that in MRW, who used the gross enrollment ratio in secondary education multiplied by the fraction of the working age population aged 15–19. We use the gross enrollment ratio because, as noted in Islam (1995), it is more widely available for a broad panel of countries (this led Footnote 18 continued

Islam to measure human capital as a stock rather than a flow rate, a route we do not pursue in order to keep closer to the strict Solow model). At any rate, the correlation between our gross enrollment ratio and MRW’s schooling variable for the 1960–1985 cross-sectional average is 95.4%, so the difference should be immaterial in practice.

¹⁹ Since the model is dynamic, subsequent values of the initial income term $\log y_{it-\tau}$ will be generated by iterating on income using the Solow specification, starting from a value drawn for the first period.

²⁰ The time fixed effect $g(t - e^{-\lambda\tau}(t - \tau))$, which is identical for all countries at each date, was generated for each period t simply by setting the parameters g and λ to their assumed values, and $\tau = 5$.

Table 2 Correlation structure among regressors and fixed-effects

	$\log s_{k,it-\tau}$	$\log s_{h,it-\tau}$	$\log(n + g + \delta)_{it-\tau}$	$\log y_{it-\tau}$	μ_i
<i>Panel A: Pooled data</i>					
$\log s_{k,it-\tau}$	1.0000				
$\log s_{h,it-\tau}$	0.6046	1.0000			
$\log(n + g + \delta)_{it-\tau}$	-0.3800	-0.5763	1.0000		
$\log y_{it-\tau}$	0.6220	0.8086	-0.6640	1.0000	
μ_i	0.6248	0.8031	-0.5957	0.9273	1.0000
<i>Panel B: Between variation</i>					
$\log s_{k,it-\tau}$	1.0000				
$\log s_{h,it-\tau}$	0.7160	1.0000			
$\log(n + g + \delta)_{it-\tau}$	-0.5004	-0.6594	1.0000		
$\log y_{it-\tau}$	0.7107	0.8691	-0.7154	1.0000	
μ_i	0.7096	0.9070	-0.6629	0.9622	1.0000
<i>Panel C: Within variation</i>					
$\log s_{k,it-\tau}$	1.0000				
$\log s_{h,it-\tau}$	0.2104	1.0000			
$\log(n + g + \delta)_{it-\tau}$	0.0763	-0.2531	1.0000		
$\log y_{it-\tau}$	0.1497	0.5400	-0.3799	1.0000	
μ_i	0.0000	0.0000	0.0000	0.0000	1.0000

Moments of the underlying data. Table 2, panel A presents the matrix of correlations among the $Q + 1$ variables of interest in the pooled data used to generate simulated datasets. For example, once stacked over time and countries, $\log s_{k,it}$ bears a correlation of 0.60 with $\log s_{h,it}$. The estimated country-specific effect bears high correlations with the right hand side variables, suggesting a big scope for heterogeneity bias when using estimators that do not correct for it. For instance, the correlation between initial income $\log y_{i,t-\tau}$ and μ_i is 0.93.²¹ Panel B isolates the between correlations among variables, by taking time means (\bar{x}_i, μ_i) and computing their correlation matrix. The between correlations are quite close to the pooled data correlations, suggesting that cross-sectional variation dominates in our data. For example, the between correlation between $\log s_{k,it}$ and $\log s_{h,it}$ is 0.72. Finally, Panel C displays the within correlations, obtained by computing $x_{it} - \bar{x}_i$. These correlations are always much lower than either the pooled or between correlations, again suggesting that the cross-country variation dominates in the pooled data. For example, the within correlation between $\log s_{k,it}$ and $\log s_{h,it}$ is 0.21.

Draws of simulated data. We are now in the presence of N observations for $T(Q - 1) + 2$ variables.²² We computed the $(T(Q - 1) + 2) \times 1$ vector of means for these variables, denoted $\hat{m}_{x,\mu}$ and their variance covariance matrix, denoted $\hat{\Omega}_{x,\mu}$. Stacking the data in this way allows us to provide a realistic simulation of the relative weights of between and within variations—by specifying explicitly the autocorrelation structure of the right-hand side variables in addition to their cross-correlations. Depending on the size of the assumed measurement error (as discussed above) the rows and columns of matrix corresponding to a mismeasured variable x are then multiplied by a factor of $1/(\sqrt{1 + F_x})$, where F_x is the assumed extent of measurement error on regressor x . For each run of our simulation, we then drew N observations

²¹ For the sake of illustration, in Table 1 we used every 5-year time-interval observation between 1960 and 1995 for the *real* data on $\log y_{it-\tau}$. In contrast, in our simulations, we are generating $\log y_{it-\tau}$ from the model, for all but the first period—due to the dynamic nature of Eq. 18 below. The dynamically simulated data on $\log y_{it-\tau}$ and their observed counterparts are very highly correlated (correlations are available upon request).

²² That is, N observations per period for $\log s_{k,it-\tau}$, $\log s_{h,it-\tau}$, $\log(n + g + \delta)_{it-\tau}$, N observations for $\log y_{it-\tau}$ in 1960 and N observations on the time invariant country effects.

for the $T(Q - 1) + 2$ variables from a multivariate normal distribution with mean $\hat{m}_{x,\mu}$ and variance $\hat{\Omega}_{x,\mu}$.

The next part of the data generation procedure is to simulate the residuals v_{it} . A major concern of many researchers who have studied growth regressions is the potential endogeneity of the regressors. As discussed above, we might expect investment and educational attainment to be positively correlated with shocks to the economy. Population growth might also be affected by shocks to economic growth. Consequently, while generating the residuals for the simulations, we explicitly allow the residual term to be correlated with $\log s_{k,it-\tau}$, $\log s_{h,it-\tau}$ and $\log(n + g + \delta)_{i,t-\tau}$ and treat these correlations as adjustable parameters in our simulation. In the baseline simulations, the correlation between $\log s_{k,it}$ and v_{it} is set to 0.25, as is the correlation between $\log s_{h,it}$ and v_{it} , while the correlation between $\log(n + g + \delta)_{i,t-\tau}$ and v_{it} is assumed to be -0.25 .

We opted to let the variance of the residual differ across time periods, and the residuals covary across time periods. To do this, we generated the fitted residuals from the fixed-effects regression using observed data for each period, and arrayed them in a $N \times T$ matrix. We computed their $T \times T$ covariance matrix $\hat{\Omega}_v$. Finally, we generated N sets of T normally distributed residuals with mean zero and covariance matrix $\hat{\Omega}_v$. An interesting aspect of this exercise is that the variance of the fixed-effects estimated residual term \hat{v}_{it} was found to be a small fraction of the variance of $\log y_{it}$.

Using all the parameters and simulated data, we computed the simulated dependent variable $\log y_{i,1965}$ for period 2, using Eq. 2 and the simulated data on $\log y_{i,1960}$. We used this generated value of $\log y_{i,1965}$ to similarly generate $\log y_{i,1970}$, and so on iteratively until we obtained $\log y_{i,2000}$. Formally, log income in period t for country i was simulated as:

$$\log y_{it} = \gamma_s \sum_{j=0}^{t-1} \gamma_4^j w_{i,t-j-1} + \gamma_4^t \log y_{i0} + \mu_i \sum_{j=0}^{t-1} \gamma_4^j + \sum_{j=0}^{t-1} \gamma_4^j v_{it-j} \tag{18}$$

3.1.2 Parameter values

There is no guarantee that the generated income data resembles in any way the underlying real world data. Equation 18 shows that simulated income is a function of past values of the steady-state determinants in w_{it} , the log of income at the beginning of the first period $\log y_{i0}$, the fixed effects μ_i and a weighted sum of the current and past residuals v_{it} , as well as the model’s reduced form parameters in γ . As t increases, the moments of the generated values of income might diverge more and more from those observed in the true income data.

To address this issue we calibrated the model’s parameters α and β so that the generated income variables in a typical draw of the data have moments resembling those of the observed variables. We found that we did not need to diverge greatly from commonly assumed values of α and β to obtain a good calibration: in a typical draw of the data, setting $\alpha = \beta = 0.27$ delivers moments of generated income variables that look similar to those seen in the PWT6.1 data.²³ These variables are conventionally both set to 1/3 in the context of the Solow model (as discussed for instance in Barro and Sala-i-Martin 1995).

²³ Details of our calibration exercise, including a comparison of the moments of the generated data with those of the observed data, are available upon request.

The other parameters of the structural model, g , δ and n are set to their conventional values as in Barro and Sala-i-Martin (1995):

$$g = 0.02; n = 0.01; \delta = 0.05$$

These parameters imply a speed of convergence $\lambda = (n + g + \delta)(1 - \alpha - \beta) = 3.68\%$, slightly higher than the value of 2.67% implied by conventional values of α and β .²⁴ With these assumed structural parameters and $\tau = 5$, the implied reduced form parameters are as follows:²⁵

$$\gamma_1 \approx 0.099; \gamma_2 \approx 0.099; \gamma_3 \approx -0.197; \gamma_4 \approx 0.832$$

3.1.3 Measurement error

The dataset generated above is free from measurement error. If we were to run fixed-effects regressions of $\log y_{it}$ on $\log s_{k,it-\tau}$, $\log s_{h,it-\tau}$, $\log(n + g + \delta)_{i,t-\tau}$ and $\log y_{i,t-\tau}$ using repeated draws of the simulated data, the only source of bias in the fixed-effects regression would be the endogeneity problem that stems from the dynamic nature of the model.²⁶ Similarly, if we were to run between regressions on simulated data, the estimated coefficients would not be tainted by measurement error bias.

To evaluate the merits of various estimators used to estimate growth regressions in the presence of measurement error, we shocked our simulated variables by adding white-noise.²⁷ This can be done in several ways. In our baseline simulations, we simply added a normally distributed shock with variance equal to some fraction $F < 1$ of the variance of the underlying variable (F is the error-to-truth ratio). This was applied to simulated variables period-by-period. Formally, consider first the determinants of the steady-state level of income, $\log s_{k,it-\tau}$, $\log s_{h,it-\tau}$ and $\log(n + g + \delta)_{i,t-\tau}$. For independent variable x_q in period t , we computed $\log x_{q,it}^*$ as:

$$\log x_{q,it}^* = \log x_{q,it} + d_{q,it} \tag{19}$$

with $d_{q,it} \sim N\left(-F\hat{\sigma}_{qt}^2/2, F\hat{\sigma}_{qt}^2\right)$ for all i and $q = 2 \dots 4$, where $\hat{\sigma}_{qt}^2$ is the sample variance of $\log x_{q,it}$ in period t .²⁸ We proceeded in exactly the same way for the income variable $\log y_{it}$:

²⁴ We have rerun our simulations assuming $\alpha = \beta = 1/3$ and all the results were qualitatively unchanged. These results are available upon request.

²⁵ Note that in empirical applications of the Solow growth model, a small contradiction exists between the theoretically derived estimating equation and the linear specification actually estimated: insofar as the rate of population growth n enters the equation as a variable (in the term $\log(n + g + \delta)$), then terms such as $(1 - e^{-\lambda t})$, where λ depends in part on n , should not be treated as constant. However, doing so is an acceptable approximation since variation in n is likely to have a small impact on variation in $e^{-(1-\alpha-\beta)(n+g+\delta)\tau}$, while it will have a larger impact on variation in $\log(n + g + \delta)$. At any rate, we follow common practice in treating $e^{-\lambda\tau}$ as a constant and $\log(n + g + \delta)$ as a variable.

²⁶ Additionally, if we simulated the income data without the error term v_{it} , fixed effects would lead us to recover *exactly* the reduced form theoretical parameters γ of the model. We did this to check our simulation program for errors, and the corresponding results are available upon request.

²⁷ Obviously, we did not shock the fixed-effects μ_i nor the intercept.

²⁸ d_{it} is distributed with a non-zero mean because if $\log x_{it}^* = \log x_{it} + d_{it}$ then $x_{it}^* = x_{it}e^{d_{it}}$. That is, $e^{d_{it}}$ is log-normally distributed and multiplicative. Because a log-normal distribution has a mean of $e^{\mu_d + \sigma_d^2/2}$

$$\log y_{it}^* = \log y_{it} + d_{y,it} \quad (20)$$

where $d_{y,it} \sim N\left(-F\hat{\sigma}_{y_{it}}^2/2, F\hat{\sigma}_{y_{it}}^2\right)$ for all i , where $\hat{\sigma}_{y_{it}}^2$ is the sample variance of $\log y_{it}$ in period t .

In the specifications above, the variance of the measurement error can vary period-by-period insofar as the variance of the underlying true data varies. In other words, the $T(Q-1) \times T(Q-1)$ variance covariance matrix of the errors-in-variables, $\hat{\Omega}_d$, is diagonal, with the diagonal elements allowed to differ across regressors and time.

The parameter F is set to four values: 0%, 5%, 10%, and 15%.²⁹ There are two limitations to this approach. First, we are assuming that the error-to-truth ratio is common to all variables. We do not have strong priors on the relative levels of measurement error for different variables, and this assumption has the advantage of simplicity. We relax it in Sect. 4 by varying the extent of measurement error on income, for reasons explained below. Second, it is difficult to know what value of F is appropriate. Hence, it is essential to vary F to assess the robustness of our results. However, we do have some priors on reasonable ranges for F : note that all variables are entered in logs, so even a value of $F = 5\%$ may imply rather large shocks, especially on the underlying income variable y_{it} . Our additive term $d_{y,it}$ translates into a multiplicative term $e^{d_{y,it}}$ applied to y_{it} (income in 1996 PPP dollars).

We can display draws of the mismeasured variables pooled across time periods and compute the average absolute value of the shocks, a summary measure of the extent of measurement error. Table 3 displays these values as well as the pooled sample averages of the underlying unshocked variables, for comparison. To construct Table 3, we drew simulated data for 2,000 countries in the 8-period panel, i.e. 16,000 observations.

Consider first measurement error on income y_{it} . For $F = 5\%$, the average absolute value of measurement error was \$230, for $F = 10\%$ it was \$323 and for $F = 15\%$ it was \$396. These are to be compared to the pooled sample mean of simulated income, which is \$4,997. The magnitudes we obtain on the other variables seem more moderate, due to the fact that their values are between 0 and 1. Consider for instance the rate of physical capital accumulation s_k : for $F = 5\%$, the average absolute value of measurement error was 0.30 percentage points, for $F = 10\%$ it was 0.41 percentage points and for $F = 15\%$ it was 0.50 percentage points. The pooled sample mean of s_k was roughly 17%. Similar relative orders of magnitudes hold for s_h and $(n + \delta + g)$, as shown in Table 3. While it is hard to know what the appropriate level of measurement error would be, the range of values displayed in Table 3 does not seem unreasonable.

Footnote 28 continued

(where μ_d and σ_d^2 are the mean and variance of d_{it} , respectively) we adjust μ_d so that $E\left[e^{d_{it}}\right] = 1$ and $E\left[x_{q,it}^*\right] = x_{q,it}$. This adjustment is necessary in order to avoid systematically understating or overstating each mismeasured regressor.

²⁹ As an additional check, we also ran our simulations with extreme values of F : 25% and 50%. The qualitative properties of our results were unchanged, but the extent of bias quickly became unreasonably large. Note also that Eq. 17 can help place an upper bound on the extent of white noise measurement error, since we must have $-1 \leq \text{corr}(x_1, x_2) \leq 1$. Looking at the matrix of correlations between the underlying observables, it appears that values of F in the 0–10% range preclude most adjusted correlations from being greater than 1 in absolute value, and are thus reasonable. Values above 10% quickly lead to implied correlations of many true variables that are greater than 1, and are thus excessive a priori. The systematic analysis of the interesting limitations imposed on the nature and extent of classical measurement error by this type of inequality restriction (along with the required positive definiteness of the covariance matrix of the true data) is a topic we leave for future research.

Table 3 Magnitude of measurement error on the underlying data (based on various values of F)

	%	Income (\$)	Investment rate (% GDP)	Secondary enrollment rate (%)	Population growth (%)
Average value of unshocked variable		4997.15	17.17	51.40	1.59
Average absolute value of shock with F	1	102.821	0.13	0.44	0.01
	2.5	158.571	0.21	0.68	0.02
	5	230.383	0.30	1.00	0.03
	10	323.118	0.41	1.38	0.04
	15	396.056	0.50	1.63	0.04

Averages computed from simulated data for 2,000 countries in the 8-period panel, i.e. 16,000 pooled observations

3.1.4 Regressions on simulated data

Having generated simulated data and shocked it with classical measurement error, we can now evaluate the bias properties of alternative estimators in the presence of correlated country-specific effects, endogenous regressors and measurement error. We estimated Eq. 2 on draws of simulated data using five estimators: fixed-effects (FE), between (BE), random effects (RE), Arellano–Bond (AB) and Blundell–Bond (BB). We stored the estimated slope coefficients from each run, and repeated this procedure 1000 times. We then computed the means of the resulting estimates, and compared those to the known true parameters. The difference between the mean estimates and the corresponding true parameters gives a measure of bias for each estimates of the slope parameters in γ . The average absolute value of these biases across parameters is used as summary measures of bias across the slope elements of γ . Although our discussion of the results focuses on bias, the standard errors of the estimates from simulated data are also available to examine the efficiency properties of the estimators.³⁰

3.2 Baseline simulation results

3.2.1 10% Measurement error

In our baseline case, we set the error-to-truth ratio F equal to 10% for all the right-hand side variables in the model and the correlation between the regressors and v_{it} at 25%. Table 4, column 3 presents the resulting estimates based on averages over 1000 runs. In terms of the average absolute value of bias on the slope parameters, our results reveal that the BE estimator dominates by a wide margin: average absolute bias is 34%, while the same number for the other four estimators exceeds 100% in all cases. As suggested by econometric theory, estimators that use the within variation exacerbate measurement error bias, and BE averages it out.

Turning to individual coefficient estimates, BE tends to bias the estimate of the convergence parameter γ_4 upward by 20%—the average simulated coefficient is 0.995 versus a true

³⁰ A discussion of the efficiency properties of the various estimators we considered can be found in the working paper version of this study, [Hauk and Wacziarg \(2004\)](#). On a related topic, [Windmeijer \(2005\)](#) discusses the small sample efficiency properties of two-step GMM estimators that are efficient asymptotically, and develops a correction.

Table 4 Varying the error–truth ratio: average estimated coefficients and bias (1000 runs, residual correlation = 25%)

Variable	True Coeffs	(1)		(2)		(3)		(4)	
		$F = 0\%$		$F = 5\%$		$F = 10\%$		$F = 15\%$	
		Mean	Bias (%)	Avg Coeff	Bias (%)	Avg Coeff	Bias (%)	Avg Coeff	Bias (%)
<i>Fixed effects</i>									
$\log s_{k,it-\tau}$	0.099	0.094	-5	-0.033	-133	0.005	-95	-0.015	-115
$\log s_{h,it-\tau}$	0.099	0.077	-22	-0.068	-169	-0.114	-216	-0.105	-207
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.279	41	-0.731	270	-1.136	476	-1.065	440
$\log y_{it-\tau}$	0.832	0.787	-5	-0.026	-103	0.189	-77	0.121	-85
Avg. Abs. Bias			19		169		216		212
Implied λ	3.68%	0.048	30	N/A	N/A	0.333	804	0.422	1047
<i>Between</i>									
$\log s_{k,it-\tau}$	0.099	0.081	-18	0.073	-26	0.077	-22	0.075	-23
$\log s_{h,it-\tau}$	0.099	0.116	18	0.086	-12	0.105	6	0.101	2
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.025	-87	-0.036	-82	-0.026	-87	-0.019	-91
$\log y_{it-\tau}$	0.832	0.982	18	1.019	23	0.995	20	1.000	20
Avg. Abs. Bias			35		36		34		34
Implied λ	3.68%	0.004	-90	-0.004	-110	0.001	-97	0.000	-100
<i>Random effects</i>									
$\log s_{k,it-\tau}$	0.099	0.110	11	0.337	241	0.210	113	0.235	138
$\log s_{h,it-\tau}$	0.099	0.173	75	0.412	317	0.326	231	0.359	264
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.175	-11	-2.636	1236	-1.401	610	-1.705	764
$\log y_{it-\tau}$	0.832	0.924	11	0.453	-46	0.701	-16	0.642	-23
Avg. Abs. Bias			27		460		242		297
Implied λ	3.68%	0.016	-57	0.158	330	0.071	93	0.089	141
<i>Arellano–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.095	-3	-0.097	-198	-0.069	-170	-0.092	-193
$\log s_{h,it-\tau}$	0.099	0.070	-29	-0.135	-237	-0.261	-365	-0.230	-333
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.239	21	-0.127	-36	-0.649	229	-0.485	146
$\log y_{it-\tau}$	0.832	0.797	-4	-0.041	-105	0.099	-88	0.053	-94
Avg. Abs. Bias			15		144		213		192
Implied λ	3.68%	0.045	24	N/A	N/A	0.462	1154	0.586	1493
<i>Arellano–Bover/Blundell–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.104	6	0.339	244	0.232	135	0.263	167
$\log s_{h,it-\tau}$	0.099	0.158	60	0.361	266	0.326	230	0.356	261
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.249	26	-3.382	1614	-2.320	1076	-2.685	1261
$\log y_{it-\tau}$	0.832	0.931	12	0.388	-53	0.644	-23	0.578	-30
Avg. Abs. Bias			26		544		366		430
Implied λ	3.68%	0.014	-61	0.189	414	0.088	139	0.110	198

coefficient of 0.832 (the implied speeds of convergence, i.e. λ parameter, are respectively 0.1% and 3.68%).³¹ In contrast, both the FE and AB severely bias this coefficient downwards, with average biases of -77% and -88% respectively, implying very high speeds of convergence (respectively 33.3% and 46.2%). In terms of the pattern of coefficients, this broadly replicates the finding of the literature—where the FE or AB estimates of the convergence speed are an order of magnitude higher than the between estimate. CEL, for example, report a speed of convergence of 10% per year based on the AB estimator—5 times larger than the

³¹ To calculate λ from the estimate of γ_4 , recall that $\gamma_4 = e^{-\lambda\tau}$. Therefore, it follows that $\lambda = -\log(\gamma_4)/\tau$ where $\tau = 5$.

2% cross-sectional estimate in MRW.³² On the other hand, the BB estimator significantly better estimates the convergence coefficient than the AB estimator, though it still exhibits a downwards bias. The γ_4 parameter is estimated as 0.644 by BB, which implies a λ of 8.8%. Our results suggest that the finding of fast convergence in the literature employing fixed-effects estimators may be traceable to the incidence of exacerbated measurement error bias.

Turning to the other slope parameters of the Solow model, interesting results also emerge. While all the estimators involve some bias, the direction and magnitudes of the biases differs sharply. BE tends to bias the human capital parameter slightly away from zero: the coefficient on $\log s_{h,it}$ is biased upward by 6%. The corresponding bias when using FE is a downward bias of 95%. Our simulation accounts for differences between estimators found in the literature—where FE typically lead to a coefficient estimate on the human capital accumulation variable that is closer to zero than BE. For example, Islam (1995) shows that the estimated BE coefficient on $\log s_{h,it}$ is roughly 0.182, and equals -0.071 when using FE. Our corresponding BE estimate is 0.105, and our FE estimate is -0.114 . Similar comparisons would hold when we turn to AB rather than FE—the accumulation parameters are both severely biased towards zero, and the depreciation parameter γ_3 is biased away from zero. The BB estimator in this instance, behaves similarly to the RE estimator in that all of the slope parameters of the Solow model are significantly biased upwards. Again, our results suggest that the finding of smaller effects of the accumulation variables in the fixed-effects literature compared to the cross-sectional literature may be largely attributable to measurement error bias.

AB estimates are very close to the FE estimates, suggesting that the weak instruments problem may be prevalent here. To evaluate this more formally, we implemented the test of the null hypothesis of weak instruments suggested by Stock and Yogo (2003), using the real world data that serves as a basis for our simulations. This test is based on computing the Cragg and Donald (1993) statistic, a generalization of the first-stage F -test for the case of multiple endogenous regressors.³³ The statistic is then compared to the critical values in Stock and Yogo (2003). The critical values depend on parameter b , the maximum amount of squared bias that the researcher is willing to accept relative to squared OLS bias (in our case, FE bias). For instance, a value $b = 0.1$ indicates that the maximal allowable bias of the IV estimates is 10% of the maximum OLS bias. In our case, the value of the Cragg-Donald statistic was 1.513, which is smaller than all critical values whatever the value of b presented in Stock and Yogo (2003) (these range from 5% to 30%). Thus, we fail to reject the null of weak instruments at the 5% significance level even when we are willing to accept a high level of AB bias relative to FE bias.³⁴

³² The precise extent to which γ_4 is biased downwards when using FE and AB in our simulations obviously depends on the postulated extent of measurement error and the postulated correlations between μ_i and the elements of x_{it} . If the error-to-truth ratio is brought down to 5%, the estimated speed of convergence is brought down to 22.05% for FE and to 32.50% for AB. These values remain higher than those reported in the literature. In Sect. 4.1 we discuss how to obtain more reasonable values of FE and AB estimates of the convergence speed by reducing the extent of measurement error on the income term $\log y_{it}$ —Sect. 3.2.3 suggested that the average absolute value of the shock to y_{it} might be too high, on the order of \$2,000 for the pooled sample in the baseline simulation with $F = 10\%$. This is to be compared for a pooled sample mean of simulated income equal to roughly \$5,000.

³³ In our application we allow all the right hand side variables to be predetermined—so that all four regressors are instrumented for. In doing this we follow the practice in CEL. See the discussion after Eq. 7. Formally, the Cragg-Donald statistic is the smallest eigenvalue of the matrix analog of the F -statistic from the first stage regression. See Stock and Yogo (2003).

³⁴ We also implemented the Staiger and Stock (1997) rule of thumb based on the first stage F -statistics. Strictly speaking, the rule of thumb is only valid for the case of one endogenous regressor. But the values of our first stage F -statistics were sufficiently below 10 to reinforce our confidence that the weak instruments problem is important here. The F -statistics for the first-stage regression of lagged first-differenced initial

As long as the first stage relationship between the instruments (in levels) and the regressors (in first differences) is not exactly zero, the weak instruments bias should vanish in large samples. To further examine whether weak instruments are responsible for the AB bias we observe in our simulations even without measurement error, we reran our simulations setting $N = 1000$ instead of $N = 69$. The results (available upon request) decisively indicate that the weak instruments problem is almost entirely responsible for the AB bias. The bias almost goes away when the cross sectional sample size is raised to 1000 (the average absolute bias is 3% instead of 192%). As soon as measurement error is introduced, however, large AB biases reappear even when $N = 1000$. The reason is that the introduction of measurement error weakens the first stage relationship between the differenced regressors and the levels instruments, considerably slowing down convergence to the true parameters as the sample size is raised. In Sect. 4, we further examine the sensitivity of our results to varying the sample size N .

This weak instruments problem regarding the AB estimator has been noted by BHT, who propose the BB system GMM estimator as a corrective for empirical growth models. In principle, the BB estimator should solve the weak instruments problem, but, as noted above, it also requires additional moment conditions to hold to ensure its consistency. Consistent with the findings of BHT, we find that the estimate of the convergence coefficient using the BB estimator is in between the estimate from the AB estimator and the BE estimator—in this case with a downwards bias of 23%. This makes it clearly superior to the AB estimator, but it is still somewhat farther away from the true parameter than the BE estimator. On the other estimated coefficients, there is a strong upwards bias (for example, an upwards bias of 135% on the $\log s_{k,it}$ parameter). It seems likely, as a result, that the additional moments conditions necessary for the consistency of the BB estimator may not hold.

Finally, Table 4 also reveals that RE performs more poorly than several other estimators when the summary measure of bias is the average absolute value of the bias (here 242% compared to 34% for the between estimator). This is important because this estimator is frequently used by growth regression practitioners who wish to retain the panel dimension but are unwilling to discard all the between variation in the data. However, RE does quite well in estimating the convergence parameter, displaying a bias of only -16% . The other slope parameters are all biased upwards. For instance the coefficient on $\log s_{k,it}$ is biased upward by 113%. One possible reason is that, contrary to between, RE does not average measurement error over time, nor does it address the problem of heterogeneity bias. The interaction of these two biases, which is hard if not impossible to characterize analytically, turns out to result in large net biases in this particular application.

3.2.2 Varying the extent of measurement error

Increasing the error-to-truth ratio to 15% or reducing it to 5% does not generally change the conclusions reached above (Table 4, columns 2 and 4). As expected, the average absolute bias tends to (weakly) increase with the error-to-truth ratio for most estimators, though this is not necessarily true for individual parameter estimates. An interesting feature of our simulations is how increases in the extent of measurement error across columns of Table 4 seem to little

Footnote 34 continued

income on its instruments was 4.73, and the corresponding values for savings on physical capital, savings on human capital and the depreciation term were, respectively, 4.03, 2.71 and 2.22.

affect the BE estimate of the coefficient on lagged income per capita—in fact the upward bias on γ_4 remains equal to 18–23% whatever the value of F .

Interesting lessons can also be learnt when measurement error is shut down entirely. While unrealistic, this exercise allows us to isolate the overall incidence of other sources of bias, besides measurement error. Table 4, column (1) presents simulation results when $F = 0$. As predicted, BE still tends to create an upward bias on the lagged income coefficient, an upward bias on the effect of human capital, and a downward bias on the depreciation term. Clearly, the strong positive correlation between the country-specific effect and the lagged income term, built into our simulations, accounts for the upward bias on $\log y_{it-\tau}$ when country fixed effects are not included in the regression. The results suggest that most of the bias in the BE estimates seen when F is set to a value different from zero is attributable to heterogeneity and endogeneity bias, as the biases on individual coefficients change little as F is increased. In other words, BE does a good job at averaging away measurement error.

When $F = 0$, FE estimates are also biased. This is because FE does not address the endogeneity problem inherent to dynamic panels and reverse causality between growth and the regressors. Our simulations allow us to isolate these problems. The biases are relatively small, especially on the main parameter of interest γ_4 (biased downwards by -5%). This bias quickly gets swamped by measurement error bias when F is increased. The AB and BB estimators, which are supposed to get rid of endogeneity bias, do display biases of similar magnitudes as FE. Moreover the biases on the various slope parameters are similar in signs and relative magnitudes to the FE biases. As discussed above, this is due to the weak instruments problem in AB, which tends to bias AB estimates towards FE. In the BB estimator, a violation of the necessary moment conditions may cause the observed bias.

To summarize, as predicted by theory, when measurement error is not present, BE is tainted by heterogeneity bias, the other estimators perform better and the AB estimator in particular performs best. Even in the presence of a small amount of measurement error ($F = 5\%$), large biases appear when using FE, and the BE estimator asserts itself as the dominant estimator in terms of average absolute bias. Moreover, in that case we are able to broadly replicate the pattern of estimates found in the literature across estimators.

3.2.3 Varying the extent of endogeneity bias

Our next step is to examine how the results changes when we vary the extent to which the regressors are correlated with the regression residual term v_{it} . We assumed a correlation of 25% between contemporaneous values of $\log s_{k,it}$ and $\log s_{h,it}$ and v_{it} , and a correlation of -25% between $\log(n + g + \delta)_{i,t-\tau}$ and v_{it} . We maintained the practice of keeping the absolute value of the correlations equal for all three steady-state determinants, but varied the correlation to 0%, 40% and 55%.³⁵

The results appear in Table 5. For comparison purposes, the baseline simulation where the correlation is 25% appears in column 2. A result that immediately jumps out is that the BE

³⁵ This methodology in which all of the correlations are varied simultaneously prevents us from choosing arbitrarily high levels of endogeneity. In particular, the residual terms have to be chosen so that the variance-covariance matrix between the simulated residuals and all of the other simulated variables is positive semi-definite. When we make the residuals highly correlated with all of the RHS variables, this criterion may not be satisfied. In practice, we found that it was difficult to avoid singularity problems as the correlation level approached 60%.

Table 5 Varying the residual correlation with regressors: average estimated coefficients and bias (1000 runs, $F = 10\%$)

Variable	True Coeffs	(1)		(2)		(3)		(4)	
		Mean	Bias (%)	Avg Coeff	Bias (%)	Avg Coeff	Bias (%)	Avg Coeff	Bias (%)
Error correlation:		0%		25%		40%		55%	
<i>Fixed effects</i>									
$\log s_{k,it-\tau}$	0.099	-0.005	-105	0.005	-95	-0.003	-103	-0.001	-101
$\log s_{h,it-\tau}$	0.099	-0.115	-217	-0.114	-216	-0.115	-217	-0.109	-210
$\log(n+g+\delta)_{it-\tau}$	-0.197	-1.120	468	-1.136	476	-1.159	487	-1.100	458
$\log y_{it-\tau}$	0.832	0.191	-77	0.189	-77	0.192	-77	0.193	-77
Avg. Abs. Bias			217		216		221		211
Implied λ	3.68%	0.331	800	0.333	804	0.330	796	0.329	794
<i>Between</i>									
$\log s_{k,it-\tau}$	0.099	0.077	-22	0.077	-22	0.077	-22	0.077	-22
$\log s_{h,it-\tau}$	0.099	0.105	7	0.105	6	0.105	6	0.106	7
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.021	-89	-0.026	-87	-0.020	-90	-0.022	-89
$\log y_{it-\tau}$	0.832	0.995	20	0.995	20	0.995	20	0.994	19
Avg. Abs. Bias			34		34		34		34
Implied λ	3.68%	0.001	-97	0.001	-97	0.001	-97	0.001	-97
<i>Random effects</i>									
$\log s_{k,it-\tau}$	0.099	0.205	108	0.210	113	0.207	110	0.204	107
$\log s_{h,it-\tau}$	0.099	0.328	233	0.326	231	0.326	230	0.328	233
$\log(n+g+\delta)_{it-\tau}$	-0.197	-1.383	601	-1.401	610	-1.387	603	-1.365	592
$\log y_{it-\tau}$	0.832	0.702	-16	0.701	-16	0.703	-16	0.703	-16
Avg. Abs. Bias			239		242		240		237
Implied λ	3.68%	0.071	92	0.071	93	0.071	92	0.071	92
<i>Arellano–Bond</i>									
$\log s_{k,it-\tau}$	0.099	-0.078	-179	-0.069	-170	-0.078	-179	-0.074	-175
$\log s_{h,it-\tau}$	0.099	-0.266	-370	-0.261	-365	-0.264	-368	-0.257	-361
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.644	226	-0.649	229	-0.660	235	-0.627	218
$\log y_{it-\tau}$	0.832	0.098	-88	0.099	-88	0.099	-88	0.102	-88
Avg. Abs. Bias			216		213		217		210
Implied λ	3.68%	0.465	1164	0.462	1154	0.463	1159	0.456	1138
<i>Arellano–Bover/Blundell–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.226	129	0.232	135	0.230	133	0.224	127
$\log s_{h,it-\tau}$	0.099	0.331	236	0.326	230	0.329	234	0.332	237
$\log(n+g+\delta)_{it-\tau}$	-0.197	-2.270	1051	-2.320	1076	-2.299	1065	-2.246	1039
$\log y_{it-\tau}$	0.832	0.646	-22	0.644	-23	0.644	-23	0.645	-23
Avg. Abs. Bias			359		366		364		356
Implied λ	3.68%	0.087	138	0.088	139	0.088	139	0.088	139

estimator is substantively unaffected by the extent of endogeneity. The average absolute bias for all four levels of endogeneity is 34%, which is also the lowest level of bias across all five estimators for all four levels of endogeneity. Perhaps surprisingly, the relative performance of the AB and BB estimators is not affected either by endogeneity bias. For all four levels of endogeneity, the average absolute bias ranges between 210% and 217% for AB and 356% and 366% for BB. These results indicate that the presence of higher levels of endogeneity than found in our baseline simulation does not overturn our basic result that BE outperforms the other estimators. As a result, even with relatively high correlations between the regressors and v_{it} , the actual impact of endogeneity on the estimated coefficients is relatively small.

Why is this the case? A priori, the relative performance AB and BB, both GMM estimators that use lagged variables as instruments, should improve as the endogeneity problem becomes greater. However, results from our initial regressions to determine the moments of the simulated variables indicate that variation in v_{it} is small relative to variation in the independent variables (including the country fixed effects).³⁶ Thus, in terms of the discussion in Sect. 2.5, there is limited scope for endogeneity bias when the baseline estimates of the simulations rely of the fixed effects estimator. In Sect. 4.5, we examine the sensitivity of our results to using alternative estimators (or “true models”) in order to generate simulated data.

3.2.4 Varying the extent of heterogeneity bias

We now examine how our results change when we vary the extent of heterogeneity bias, holding measurement error fixed at some baseline level. As described in Sect. 3.1.2, the assumed correlations between the right hand side variables and the country fixed effect μ_i used to draw simulated data were based on estimated values of μ_i from an FE regression. In the presence of measurement error, these estimated μ_i s will be inconsistent estimates of the true country fixed-effects, so their sample correlations with the regressors will themselves be flawed. Hence, it is critical to examine what happens when we change these assumed correlations.

Table 6 displays simulation results when varying the correlations between the country-specific effects and the regressors, while setting $F = 10\%$ and the correlation between the RHS variables and the residuals at 25%. Table 1 showed that the correlations used for our baseline simulations were high. For example, in the pooled sample the correlation between our estimated μ_i and $\log y_{it-\tau}$ was 0.93, and the correlation with $\log s_{h,it-\tau}$ was 0.80. Here we simply multiply all these assumed correlations, variable-by-variable and period-by-period, by a single constant $C < 1$, prior to generating the simulated data.³⁷ We allowed C to take on the values 0%, 50% and 75%. For the sake of comparison, column (5) of Table 6 also reports the results obtained when $C = 100\%$ (i.e. column (3) of Table 4).

Table 6 demonstrates that the biases obtained under FE when varying C do change significantly, suggesting that at least some of the FE bias comes from endogeneity and not simply measurement error. Notably, the bias on the estimate of γ_4 declines (from -55% when $C = 0\%$ to -77% when $C = 100\%$). As expected from econometric theory, the AB estimate of γ_4 exhibits a greater degree of stability across values of C .

We now turn to the BE estimator in the extreme case where $C=0$. Again, this is an unrealistic assumption, but it allows us to evaluate the incidence of measurement error in isolation from heterogeneity bias. We observe that the average absolute value bias is increased (to 162% compared to 34% when $C=100\%$), but that the pattern of signs and relative magnitudes for the bias is roughly in line with the results in column (3) of Table 4. The convergence parameter γ_4 now exhibits a larger bias (35% rather than 20%). The same pattern holds for all the other slope parameters: as C rises, the extent of bias is progressively reduced, illustrating nicely a central message of this paper: as the incidence of heterogeneity bias rises, it

³⁶ In intuitive terms, when the variance of v_{it} is small relative to the variance of the regressors, even a large assumed correlation between these variables will translate into a small covariance and therefore a small bias term.

³⁷ That is, we modified the relevant entries of the data covariance matrix $\hat{\Omega}_{x,\mu}$ used to generate the simulated series.

Table 6 Varying heterogeneity bias (alternative correlations between μ_i and the regressors): $F = 10\%$, residual correlation = 25%, 1000 runs

Variable	True Coeffs	(1)		(2)		(3)		(4)	
		$C = 0\%$		$C = 50\%$		$C = 75\%$		$C = 100\%$	
		Mean	Bias (%)	Mean	Bias (%)	Mean	Bias (%)	Mean	Bias (%)
<i>Fixed effects</i>									
$\log s_{k,it-\tau}$	0.099	0.069	-30	0.032	-67	0.020	-80	0.005	-95
$\log s_{h,it-\tau}$	0.099	0.049	-50	-0.022	-123	-0.063	-164	-0.114	-216
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.204	3	-0.572	190	-0.860	336	-1.136	476
$\log y_{it-\tau}$	0.832	0.374	-55	0.283	-66	0.238	-71	0.189	-77
Avg. Abs. Bias			35		112		163		216
Implied λ	3.68%	0.196	434	0.253	587	0.287	680	0.333	804
<i>Between</i>									
$\log s_{k,it-\tau}$	0.099	0.007	-93	0.025	-75	0.039	-60	0.077	-22
$\log s_{h,it-\tau}$	0.099	-0.072	-173	-0.030	-130	0.015	-85	0.105	6
$\log(n+g+\delta)_{it-\tau}$	-0.197	0.491	-349	0.362	-284	0.248	-226	-0.026	-87
$\log y_{it-\tau}$	0.832	1.125	35	1.099	32	1.066	28	0.995	20
Avg. Abs. Bias			162		130		100		34
Implied λ	3.68%	-0.024	-164	-0.019	-151	-0.013	-135	0.001	-97
<i>Random effects</i>									
$\log s_{k,it-\tau}$	0.099	0.088	-11	0.130	32	0.164	67	0.210	113
$\log s_{h,it-\tau}$	0.099	0.074	-25	0.172	74	0.241	144	0.326	231
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.161	-18	-0.589	198	-0.924	369	-1.401	610
$\log y_{it-\tau}$	0.832	0.863	4	0.809	-3	0.764	-8	0.701	-16
Avg. Abs. Bias			15		77		147		242
Implied λ	3.68%	0.029	-20	0.042	15	0.054	46	0.071	93
<i>Arellano–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.015	-85	-0.055	-156	-0.068	-169	-0.069	-170
$\log s_{h,it-\tau}$	0.099	0.034	-65	-0.134	-235	-0.198	-301	-0.261	-365
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.490	148	-0.615	212	-0.664	237	-0.649	229
$\log y_{it-\tau}$	0.832	0.059	-93	0.004	-99	0.045	-95	0.099	-88
Avg. Abs. Bias			98		176		200		213
Implied λ	3.68%	0.565	1436	1.084	2845	0.621	1587	0.462	1154
<i>Arellano–Bover/Blundell–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.110	12	0.168	70	0.205	108	0.232	135
$\log s_{h,it-\tau}$	0.099	0.091	-7	0.217	120	0.288	192	0.326	230
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.290	47	-1.114	465	-1.726	775	-2.320	1076
$\log y_{it-\tau}$	0.832	0.827	-1	0.742	-11	0.687	-17	0.644	-23
Avg. Abs. Bias			17		166		273		366
Implied λ	3.68%	0.038	3	0.060	62	0.075	104	0.088	139

increasingly mitigates the problem of measurement error for the BE estimator. In this case, the two sources of bias tend to cancel each other out.

4 Extensions

This section considers various extensions to our basic simulation method. We consider what happens when we change assumptions on the extent and nature of measurement error, when we change the estimator used to generate the fixed effects and the residual term, when we depart from the strict confines of the Solow model. We also examine the properties of two additional estimators frequently used in the empirical literature on economic growth.

4.1 Varying measurement error on income

While our baseline simulation results replicate the broad differences in past findings on convergence and the determinants of steady-state income level across estimators, the estimated speed of convergence under the FE, AB and BB estimators was too high relative to the BE estimate—we obtained an FE estimate of $\lambda = 33.3\%$ and a BE estimate of $\lambda = 0.1\%$, while the literature finds values in the neighborhood of 10% and 2%, respectively. Moreover, in Sect. 3.1.3 we showed that an error-to-truth ratio of $F = 10\%$ implies an average absolute value error in measurement of roughly \$323, while the mean of simulated income was \$4,997. While it is difficult to know what the appropriate extent of measurement error is, this is possibly too big.³⁸ In this subsection we examine whether reducing the extent of measurement error on log income can help match convergence speeds estimated in the literature: we reduce the extent of measurement error on the income term (F_y), while maintaining $F = 10\%$ on the other variable. We consider values of F_y equal to 0%, 1%, 2.5%, and 5%.

We might expect reduced measurement error on $\log y_{it}$ to reduce attenuation bias on its coefficient and thereby improve the performance of the FE, AB and BB estimators (though this of course is not always true in the presence of measurement error on the other variables). Our intuition turns out to be borne out: Table 7 demonstrates that with a value of $F_y = 1\%$, we obtain FE and AB estimates of λ that are much closer to those found in the past literature: 9.3% and 10.9%, respectively. On the BB estimator, the convergence coefficient is quite close to the true value, slightly underestimating the rate of convergence for small levels of measurement error slightly overestimating it when $F_y = 5\%$. This performance is similar to our baseline simulations in Table 4. The BE estimate of γ_4 is unchanged compared to our baseline simulations, with an upward bias on the convergence coefficient, confirming our suggestion that most of this bias is attributable to the omission of the country-specific effect μ_i .

In general, the average absolute value bias becomes much lower for the AB, BB and FE estimators, largely because the bias on the income term is now reduced. In fact, when we set $F_y = 0$, these estimators get convergence almost right, suggesting that measurement error in this variable is important to replicate the pattern of γ_4 estimates found in the literature.

To summarize, when we allow for a smaller error-to-truth ratio on income, we are able to obtain FE, AB and BB estimated speeds of convergence that are much closer to those obtained when using real data. The extent of bias on the other parameters is not affected much by measurement error on income. We still get BE estimates of γ_4 that are too large (and therefore BE estimates of λ that are too small) relative to the existing cross-country literature.

4.2 Autocorrelated measurement error

So far, we have assumed classical measurement error, i.e. the error in measurement was purely white noise. However, errors in measurement could be autocorrelated. For instance, if a country has over-reported the amount of savings in physical capital in one period, it may be more

³⁸ In contrast, the extent of measurement error on the other variables, implied by setting $F = 10\%$, seemed more realistic. Some have argued that per capita income may be better measured than savings rates on human capital s_h , physical capital s_k and the depreciation variable $(n + g + \delta)$. In principle population growth n will be well-measured, but recall that we had to make an assumption of constancy across time and countries for δ and g , which surely introduces error. Similarly, s_h in the Solow model should be measured by dollars saved per unit of time for the purpose of financing education, but we followed the literature in proxying for this using gross enrollment rates in secondary education. However, it is well-known that different methods of computing price indices and PPP exchange rates can deliver vastly different estimates of PPP income.

Table 7 Varying measurement error on initial income, 1000 runs ($F = 10\%$ on the other variables, residual correlation = 25%)

Variable	True Coeffs	$F_y = 0\%$		$F_y = 1\%$		$F_y = 2.5\%$		$F_y = 5\%$	
		Mean	Bias (%)	Mean	Bias (%)	Mean	Bias (%)	Mean	Bias (%)
<i>Fixed effects</i>									
$\log s_{k,it-\tau}$	0.099	0.067	-32	0.051	-48	0.034	-66	0.017	-83
$\log s_{h,it-\tau}$	0.099	0.030	-70	-0.005	-105	-0.036	-136	-0.075	-176
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.146	-26	-0.392	99	-0.616	212	-0.872	342
$\log y_{it-\tau}$	0.832	0.771	-7	0.628	-24	0.487	-41	0.337	-60
Avg. Abs. Bias			34		69		114		165
Implied λ	3.68%	0.052	42	0.093	152	0.144	291	0.218	492
<i>Between</i>									
$\log s_{k,it-\tau}$	0.099	0.081	-18	0.080	-19	0.080	-19	0.079	-20
$\log s_{h,it-\tau}$	0.099	0.107	9	0.107	9	0.108	9	0.106	8
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.025	-87	-0.032	-84	-0.030	-85	-0.022	-89
$\log y_{it-\tau}$	0.832	0.989	19	0.990	19	0.990	19	0.992	19
Avg. Abs. Bias			33		33		33		34
Implied λ	3.68%	0.002	-94	0.002	-94	0.002	-94	0.002	-96
<i>Random effects</i>									
$\log s_{k,it-\tau}$	0.099	0.106	7	0.117	19	0.135	37	0.163	65
$\log s_{h,it-\tau}$	0.099	0.128	30	0.155	57	0.191	94	0.244	147
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.151	-23	-0.311	58	-0.531	169	-0.864	338
$\log y_{it-\tau}$	0.832	0.952	14	0.923	11	0.877	5	0.809	-3
Avg. Abs. Bias			19		36		76		138
Implied λ	3.68%	0.010	-73	0.016	-57	0.026	-29	0.042	15
<i>Arellano-Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.065	-34	0.034	-65	-0.010	-110	-0.044	-144
$\log s_{h,it-\tau}$	0.099	0.016	-84	-0.052	-153	-0.128	-229	-0.201	-304
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.129	-35	-0.335	70	-0.462	134	-0.569	188
$\log y_{it-\tau}$	0.832	0.768	-8	0.581	-30	0.390	-53	0.222	-73
Avg. Abs. Bias			40		79		132		177
Implied λ	3.68%	0.053	43	0.109	195	0.189	412	0.301	717
<i>Arellano-Bover/Blundell-Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.018	2	0.112	14	0.133	35	0.169	71
$\log s_{h,it-\tau}$	0.099	0.014	13	0.132	34	0.171	74	0.232	135
$\log(n + g + \delta)_{it-\tau}$	-0.197	0.142	11	-0.503	155	-0.869	340	-1.439	629
$\log y_{it-\tau}$	0.832	0.010	16	0.923	11	0.865	4	0.777	-7
Avg. Abs. Bias			10		53		113		211
Implied λ	3.68%	0.008	-79	0.016	-57	0.029	-21	0.050	37

likely to do so in subsequent periods. Hence, measurement error can be expected to persist. Moreover, as discussed in Sect. 2, persistent measurement error invalidates the IV procedure of the AB and BB estimator when it comes to addressing measurement error bias. In this subsection, we run simulations where autocorrelation is built into measurement error. Specifically, the $T(Q - 1) \times T(Q - 1)$ variance-covariance matrix of the errors-in-variables, $\hat{\Omega}_d$, is now block diagonal, with the diagonal elements identical to what they were before, and the off-diagonal elements of each $T \times T$ block equal to the postulate covariance between errors across periods. Errors are then drawn from a multivariate normal distribution with covariance matrix $\hat{\Omega}_d$, as before. We denoting the autocorrelation coefficient of the errors as ρ_d .

In this case, we do have some theoretical priors as to how persistence in measurement error might affect the results. If measurement error is highly persistent over time, we would expect FE to perform better relative to the case of $\rho_d = 0$, for two reasons: first, with persistence in $d_{q,it}$, some of the measurement error will be differenced away when the data is differenced from its country means. In other words, the greater the autocorrelation in measurement error, the larger the between component of measurement error and the smaller the within component. Since FE will difference away the between variation, we expect greater autocorrelation in the measurement error term, all else equal, to improve the performance of FE relative to BE. In the limit, when the autocorrelation coefficient ρ_d is 1, we would expect FE to get rid of all of the measurement error, as it will get entirely differenced away.³⁹

We ran simulations when the measurement error autocorrelation term was $\rho_d = 0.5$, $\rho_d = 0.75$ and $\rho_d = 0.90$, setting $F = 10\%$. The results in Table 8 show that our theory-driven priors are confirmed by the simulations. Average absolute value bias is at its lowest when $\rho_d = 90\%$ in FE, AB and BB estimators. When the autocorrelation term is only 50%, the convergence coefficient γ_4 exhibits a 20% bias using the BE estimator and a -48% bias using the FE estimator (in the baseline case of $\rho_d = 0$, the corresponding numbers were 20% and -77%, respectively). When the autocorrelation term is increased all the way to 90%, which is probably much too high, the BE coefficient has a bias of 21% and the FE bias of -24%. Therefore, while we confirm our intuition that the FE estimator improves relative to the BE estimator when we increase persistence in measurement error, BE tends to do better or as well as alternative estimators even as ρ_d is raised to implausibly high levels. Even high persistence in measurement error does not invalidate the overall conclusions reached in the baseline case.

4.3 Increasing sample size

As we indicated above in Sect. 3, we suspect that the primary deficiency with the AB estimator is a weak instruments problem. Indeed, when we ran the simulation with our sample size set to $N = 1000$, the bias in the AB estimator was substantially reduced. Noting that the AB estimator works well with such a large sample, though, will be of little practical importance in the context of economic growth, where sample sizes rarely exceed 150 countries. However, our baseline simulations use a balanced panel of 69 countries, which may be smaller than a growth regression that uses an unbalanced panel. Consequently, in this section, we repeat our baseline simulations with $F = 0\%$, 5%, 10% and 15%, but with a sample size of $N = 150$.

Our results from this exercise are found in Table 9. Results on the FE, BE and RE estimators are largely unaffected by sample size. This result is not surprising—for estimators that do not use instrumental variables, an increase in sample size will perhaps improve the efficiency of the estimators but will not affect biases. More interesting is the fact that increasing the sample size does not have a noticeable effect on the AB and BB GMM estimators. For the simulation where $F = 10$, the AB estimator had a 213% average absolute bias when the smaller sample size was used and a 207% average absolute bias when the larger sample was used. Similar results exist for the other levels of F . Therefore, while it may be the instruments used in the AB estimator are weak in small samples, the sample size has to be increased beyond a realistic amount before instruments are substantially strengthened. Also, while the BB estimator purportedly offers stronger instruments than AB, its performance does not improve either when the sample size is increased. With the smaller sample,

³⁹ We checked that this is the case, and the results are available upon request.

Table 8 Autocorrelated measurement error ($F = 10\%$, residual correlation = 25%, 1000 runs)

Variable	True Coeffs	$\rho_d = 50\%$		$\rho_d = 75\%$		$\rho_d = 90\%$	
		Mean	Bias (%)	Mean	Bias (%)	Mean	Bias (%)
<i>Fixed effects</i>							
$\log s_{k,it-\tau}$	0.099	0.029	-70	-0.001	-101	0.065	-34
$\log s_{h,it-\tau}$	0.099	-0.060	-160	-0.109	-210	-0.008	-108
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.771	291	-1.100	458	-0.542	175
$\log y_{it-\tau}$	0.832	0.432	-48	0.193	-77	0.629	-24
Avg. Abs. Bias			142		211		85
Implied λ	3.68%	0.168	356	0.329	794	0.093	153
<i>Between</i>							
$\log s_{k,it-\tau}$	0.099	0.076	-23	0.077	-22	0.071	-28
$\log s_{h,it-\tau}$	0.099	0.094	-4	0.106	7	0.079	-19
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.024	-88	-0.022	-89	-0.042	-79
$\log y_{it-\tau}$	0.832	1.001	20	0.994	19	1.010	21
Avg. Abs. Bias			34		34		37
Implied λ	3.68%	0.000	-101	0.001	-97	-0.002	-104
<i>Random effects</i>							
$\log s_{k,it-\tau}$	0.099	0.159	61	0.204	107	0.116	18
$\log s_{h,it-\tau}$	0.099	0.235	139	0.328	233	0.154	57
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.803	307	-1.365	592	-0.302	53
$\log y_{it-\tau}$	0.832	0.821	-1	0.703	-16	0.925	11
Avg. Abs. Bias			127		237		35
Implied λ	3.68%	0.040	7	0.071	92	0.016	-57
<i>Arellano-Bond</i>							
$\log s_{k,it-\tau}$	0.099	-0.020	-120	-0.074	-175	0.032	-68
$\log s_{h,it-\tau}$	0.099	-0.180	-282	-0.257	-361	-0.121	-223
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.315	60	-0.627	218	-0.191	-3
$\log y_{it-\tau}$	0.832	0.430	-48	0.102	-88	0.600	-28
Avg. Abs. Bias			128		210		80
Implied λ	3.68%	0.169	359	0.456	1138	0.102	178
<i>Arellano-Bover/Blundell-Bond</i>							
$\log s_{k,it-\tau}$	0.099	0.160	62	0.224	127	0.114	15
$\log s_{h,it-\tau}$	0.099	0.212	115	0.332	237	0.156	59
$\log(n + g + \delta)_{it-\tau}$	-0.197	-1.288	553	-2.246	1039	-0.507	157
$\log y_{it-\tau}$	0.832	0.798	-4	0.645	-23	0.907	9
Avg. Abs. Bias			183		356		60
Implied λ	3.68%	0.045	23	0.088	139	0.020	-41

its average absolute bias was 366%. When the sample size is increased, however, the average absolute bias actually to 393%. The problem with the BB estimator is not primarily one of strong vs. weak instruments, but rather one of invalid moment conditions needed for consistency.

4.4 Alternative underlying models

Throughout this study, the assumed underlying “true” model used to generate data for the simulations is a fixed-effects model with the Solow model structural parameters assumed to be $\alpha = \beta = 0.27$. As we discussed in Sect. 2, the simulations use the moments of the residuals from a FE regression on the real-world data in order to generate the simulated residuals. These residuals have two components: μ_i , a country-specific effect that is highly correlated with our regressors (see Table 1), and v_{it} , a white-noise residual term that, by construction, has a

Table 9 Varying the error–truth ratio with more observations: average estimated coefficients and bias (1000 runs, residual correlation = 25%, $N = 150$)

Variable	True Coeffs	(1)		(2)		(3)		(4)	
		$F = 0\%$		$F = 5\%$		$F = 10\%$		$F = 15\%$	
Error-to-truth ratio:		Mean	Bias (%)	Avg Coeff	Bias (%)	Avg Coeff	Bias (%)	Avg Coeff	Bias (%)
<i>Fixed effects</i>									
$\log s_{k,it-\tau}$	0.099	0.088	-11	0.025	-75	0.003	-97	-0.013	-114
$\log s_{h,it-\tau}$	0.099	0.056	-43	-0.101	-202	-0.112	-214	-0.105	-207
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.376	91	-1.026	420	-1.133	474	-1.114	465
$\log y_{it-\tau}$	0.832	0.747	-10	0.341	-59	0.202	-76	0.131	-84
Avg. Abs. Bias			39		198		215		217
Implied λ	3.68%	0.058	59	0.215	484	0.320	769	0.407	1006
<i>Between</i>									
$\log s_{k,it-\tau}$	0.099	0.082	-17	0.080	-19	0.079	-20	0.078	-21
$\log s_{h,it-\tau}$	0.099	0.118	20	0.113	14	0.108	10	0.104	6
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.042	-79	-0.033	-83	-0.031	-84	-0.032	-84
$\log y_{it-\tau}$	0.832	0.979	18	0.986	18	0.991	19	0.996	20
Avg. Abs. Bias			33		34		33		33
Implied λ	3.68%	0.004	-89	0.003	-92	0.002	-95	0.001	-98
<i>Random effects</i>									
$\log s_{k,it-\tau}$	0.099	0.111	12	0.166	68	0.208	111	0.237	140
$\log s_{h,it-\tau}$	0.099	0.177	80	0.277	181	0.330	234	0.360	265
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.215	9	-0.919	366	-1.394	606	-1.735	780
$\log y_{it-\tau}$	0.832	0.919	10	0.784	-6	0.698	-16	0.640	-23
Avg. Abs. Bias			28		155		242		302
Implied λ	3.68%	0.017	-54	0.049	32	0.072	95	0.089	143
<i>Arellano–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.090	-8	-0.048	-148	-0.084	-185	-0.107	-208
$\log s_{h,it-\tau}$	0.099	0.048	-51	-0.310	-414	-0.282	-386	-0.248	-352
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.247	25	-0.620	214	-0.545	176	-0.473	140
$\log y_{it-\tau}$	0.832	0.779	-6	0.261	-69	0.154	-82	0.106	-87
Avg. Abs. Bias			23		211		207		197
Implied λ	3.68%	0.050	36	0.269	631	0.375	918	0.449	1119
<i>Arellano–Bover/Blundell–Bond</i>									
$\log s_{k,it-\tau}$	0.099	0.108	9	0.168	70	0.224	127	0.256	159
$\log s_{h,it-\tau}$	0.099	0.160	62	0.253	157	0.301	205	0.317	222
$\log(n + g + \delta)_{it-\tau}$	-0.197	-0.331	68	-1.696	760	-2.599	1217	-3.108	1476
$\log y_{it-\tau}$	0.832	0.923	11	0.753	-10	0.643	-23	0.576	-31
Avg. Abs. Bias			38		249		393		472
Implied λ	3.68%	0.016	-57	0.057	54	0.088	140	0.110	200

correlation coefficient of 0.25 with contemporaneous regressors.⁴⁰ This procedure presents two potential weaknesses. The first is that, while the assumption that $\alpha = \beta = 0.27$ closely follows the values of $\alpha = \beta = \frac{1}{3}$ typically assumed in the Solow Model, we cannot be sure

⁴⁰ As before, we continue to include time effects in the residuals and correspondingly include time dummies in the regressions where applicable, but this does not play an important role in our analysis or results.

that these coefficients are in reality the true growth parameters.⁴¹ Also, as we note throughout the paper, using an FE regression to estimate these parameters is unlikely to reproduce the true coefficients. As a result, residuals generated from such a regression may not have the same moments as true Solow residuals. In this section, we relax these assumptions. Doing so, we can test the robustness of our results to differing assumptions about the true coefficients and associated residuals μ_i and v_{it} .

We use both the estimated values of γ obtained from FE, BE, RE, AB and BB regressions on the observed data as baseline “true” parameters, and from there generate moments for the simulated residuals. We maintain the practice of breaking the observed residuals into a country-specific effect μ_i and a white-noise term v_{it} and use the moments of both in our data-generating process.⁴² As before, we impose a correlation coefficient of 0.25 between v_{it} and its contemporaneous regressors and allow μ_i to have whatever correlations are observed using estimates from the original data. We also use a baseline measurement error value of $F = 10\%$ for all simulations. When switching between underlying models in this section, three sets of parameters change in the data-generating process: the values of γ used as true coefficients, the variances of μ_i and v_{it} , and the correlations between μ_i and the regressors.

Table 10 reports the results. A general lesson from this table is that when assuming a given underlying model, that model tends to perform relatively better compared to the theory-driven true model used in the simulations of Sect. 3. For instance, when the true slope parameters are assumed to be AB estimates on the real data, the AB estimator performs better than we found in Sect. 3. However, the relative rankings of estimators is largely preserved. So, for instance, BE still outperforms AB even when the true underlying coefficients are assumed to be the AB estimates.

In columns (1)–(3) we see the results of simulations using FE, AB and BB as the underlying models. In these simulations—all three of which use coefficients based on within-country variation and have residual terms that are highly correlated with the regressors—the BE estimator still performs well. It is clearly superior to all of the other estimators in terms of both average absolute bias and estimating the convergence coefficient when the AB and BB coefficients are assumed the true model. Under the true FE model, the FE estimator outperforms it slightly in terms of average absolute bias and the RE estimator comes somewhat closer to correctly estimating the convergence coefficient. By neither measure, however, is the difference very large. In columns (4) and (5)—the BE and RE true models, respectively—the performance of the BE estimator is even more striking. Because these models by construction generate values of μ_i that are not correlated with the regressors, the effect of country-level heterogeneity is largely taken out of the data-generating process. As a result, the BE estimator performs exceptionally well both in absolute and relative terms. Under the BE “true” model, the BE estimator nearly perfectly estimates the convergence coefficient and has an average absolute bias of only 4%. The performance is similarly strong for the FE model, where BE again comes quite close to the true convergence coefficient and has an average absolute bias of 14%. No other estimator even comes close to replicating this performance in both cases. We conclude our basic results are not very sensitive to the selection of our underlying true model.

⁴¹ In a past version of this paper, we also allowed α to vary across countries using data on capital shares from Caselli and Feyrer (2006). This did not affect our results.

⁴² μ_i is simply the average of the residual term across time periods within each country. v_{it} is the deviation from this average in each time period.

Table 10 Varying underlying model: average estimated coefficients and bias (1000 runs, $F = 10\%$, residual correlation = 25%)

Model used: Variable	(1) Fixed effects			(2) Arellano-Bond			(3) Blundell-Bond			(4) Between			(5) Random effects		
	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)
<i>Fixed effects</i>															
$\log s_{k,it-\tau}$	0.062	0.070	13	0.072	0.072	1	0.085	0.086	2	0.058	0.073	26	0.065	0.078	19
$\log s_{h,it-\tau}$	-0.028	-0.043	55	-0.060	-0.057	-4	0.025	-0.025	-201	0.041	-0.019	-146	0.040	-0.016	-140
$\log(n + g + \delta)_{it-\tau}$	-0.421	-0.604	43	-0.085	-0.443	422	-0.617	-0.754	22	-0.214	-0.483	126	-0.385	-0.582	51
$\log y_{it-\tau}$	0.796	0.057	-93	0.795	0.029	-96	0.944	0.085	-91	0.960	0.102	-89	0.940	0.098	-90
Avg. Abs. Bias			51			131			79			97			75
Implied λ	0.046	0.573	1155	0.046	0.706	1439	0.012	0.492	4170	0.008	0.456	5484	0.012	0.465	3657
<i>Between</i>															
$\log s_{k,it-\tau}$	0.062	0.057	-9	0.072	0.059	-18	0.085	0.061	-29	0.058	0.060	3	0.065	0.062	-4
$\log s_{h,it-\tau}$	-0.028	0.038	-236	-0.060	0.029	-148	0.025	0.042	68	0.041	0.040	-2	0.040	0.043	8
$\log(n + g + \delta)_{it-\tau}$	-0.421	-0.171	-59	-0.085	-0.182	115	-0.617	-0.211	-66	-0.214	-0.191	-11	-0.385	-0.221	-42
$\log y_{it-\tau}$	0.796	0.965	21	0.795	0.936	18	0.944	0.955	1	0.960	0.959	0	0.940	0.951	1
Avg. Abs. Bias			81			75			41			4			14
Implied λ	0.046	0.007	-84	0.046	0.013	-71	0.012	0.009	-20	0.008	0.008	2	0.012	0.010	-20
<i>Random effects</i>															
$\log s_{k,it-\tau}$	0.062	0.153	147	0.072	0.152	111	0.085	0.161	89	0.058	0.154	166	0.065	0.156	140
$\log s_{h,it-\tau}$	-0.028	0.171	-711	-0.060	0.142	-337	0.025	0.180	618	0.041	0.175	326	0.040	0.175	337
$\log(n + g + \delta)_{it-\tau}$	-0.421	-1.184	181	-0.085	-1.216	1331	-0.617	-1.290	109	-0.214	-1.123	425	-0.385	-1.189	209
$\log y_{it-\tau}$	0.796	0.682	-14	0.795	0.612	-23	0.944	0.658	-30	0.960	0.679	-29	0.940	0.672	-28
Avg. Abs. Bias			264			450			212			237			179
Implied λ	0.046	0.076	68	0.046	0.098	114	0.012	0.084	626	0.008	0.077	847	0.012	0.079	542

Table 10 continued

Model used:	(1) Fixed effects			(2) Arellano-Bond			(3) Blundell-Bond			(4) Between			(5) Random effects		
	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)	True Coefs	Mean	Bias (%)
<i>Arellano-Bond</i>															
$\log s_{k,it-\tau}$	0.062	0.009	-86	0.072	0.019	-74	0.085	0.007	-92	0.058	0.006	-90	0.065	0.014	-79
$\log s_{l,it-\tau}$	-0.028	-0.103	269	-0.060	-0.118	97	0.025	-0.104	-516	0.041	-0.090	-320	0.040	-0.086	-314
$\log(n + g + \delta)_{it-\tau}$	-0.421	0.009	-102	-0.085	0.206	-342	-0.617	0.037	-106	-0.214	0.217	-202	-0.385	0.083	-122
$\log y_{it-\tau}$	0.796	0.045	-94	0.795	0.042	-95	0.944	0.091	-90	0.960	0.085	-91	0.940	0.097	-90
Avg. Abs. Bias			138			152			201			176			151
<i>Blundell-Bond</i>															
Implied λ	0.046	0.622	1264	0.046	0.634	1282	0.012	0.480	4061	0.008	0.494	5946	0.012	0.466	3665
$\log s_{k,it-\tau}$	0.062	0.168	170	0.072	0.177	146	0.085	0.177	108	0.058	0.170	193	0.065	0.171	162
$\log s_{l,it-\tau}$	-0.028	0.194	-793	-0.060	0.177	-395	0.025	0.209	734	0.041	0.200	388	0.040	0.199	398
$\log(n + g + \delta)_{it-\tau}$	-0.421	-1.712	307	-0.085	-1.578	1757	-0.617	-1.916	210	-0.214	-1.562	630	-0.385	-1.619	320
$\log y_{it-\tau}$	0.796	0.614	-23	0.795	0.539	-32	0.944	0.576	-39	0.960	0.612	-36	0.940	0.607	-35
Avg. Abs. Bias			323			582			273			312			229
Implied λ	0.046	0.098	114	0.046	0.124	170	0.012	0.110	857	0.008	0.098	1101	0.012	0.100	707

4.5 Alternative specification

The estimation problems discussed in Sect. 2 are not specific to the Solow growth model, and are likely to exist in any dynamic panel data model with regressors that are both time-persistent and measured with error. To test the generalizability of our results to other growth specifications, and hence the broader applicability of our results to the empirical literature on the determinants of growth, we start from specification from Barro and Sala-i-Martin (2003), from their Table 12.3, column 2. This regression does not stem from or test a specific growth theory, but rather regresses economic growth on a number of potential growth determinants.⁴³ We collected data as close as possible to that used in the Barro and Sala-i-Martin regression, and replicated as closely as we could their regression results, using Seemingly Unrelated Regression (as they did).⁴⁴ We treated the resulting parameters as the true parameters, since we lack any theoretical guidance on the true parameters. We then repeated the Monte Carlo simulations described above, only using the moments-matrix of the newly collected data and the estimated true parameters as the basis for the simulated data. The country fixed effects were generated simply by averaging each country's error terms across time—the deviation from this average gave us the moments of the residual. The simulation of income and adding measurement error was done exactly as previously.

This approach has several drawbacks, which should be recognized up-front. In particular, theory provides no guidance as to the true values of the parameters, forcing us to rely on estimated parameters as the true ones. However, as discussed above, these parameters cannot be the true ones in any real or postulated sense, since their estimation is tainted by bias. Second, we choose a widely cited, but specific empirical model, and there is no guarantee that the results will hold up to alternative specifications. Hence, the results of this exercise should be taken only as suggestive.

Our results are shown in Table 11. We focus on the case of $F = 10\%$. The BE estimator outperforms the other four in terms of average absolute bias. Moreover, BE comes very close to getting the convergence parameter right, while FE and AB, as before, seriously overstate the speed of convergence and tend to understate the impact of steady-state determinants.⁴⁵ As before, BB and RE perform somewhat in-between: the convergence speed is overstated, and coefficients on the time persistent regressors tend to be biased away from zero. Despite all the caveats that apply to this exercise, we see much of the same qualitative pattern of biases in this specification as we do in the original simulations.

4.6 Other estimators

4.6.1 Flexible random effects (SUR)

In this subsection, we evaluate the bias properties of the Seemingly Unrelated Regressions (SUR) estimator, commonly used in the empirical growth literature (see, for example, Barro and Sala-i-Martin 1995, Chap. 12). This estimator is computationally close to the RE esti-

⁴³ This specification regresses log income at the end of the period on log income at the beginning of the period and the following variables: upper level schooling, inverse life expectancy, log fertility, the government consumption ratio, a rule of law index, a democracy index, the democracy index squared, trade openness, terms of trade change, the investment ratio and the inflation rate.

⁴⁴ We replicated this exercise using the fixed effects estimator as an alternative estimator to generate the true coefficients and the fixed effects. Our results, which are available upon request, were qualitatively unchanged.

⁴⁵ The extent to which they overstate it is probably excessive, for the reasons discussed in Sect. 4.1—i.e. the extent of measurement error on the income term is probably excessive.

Table 11 Barro-style regressions: 1000 runs

Variable	True Coefs	$F = 5\%$		$F = 10\%$		$F = 15\%$	
		Mean	Bias (%)	True Coefs	Mean	Bias (%)	
<i>Fixed effects</i>							
Log Lagged Income	0.825	-0.036	-104	-0.155	-119	-0.209	-125
Upper Level Schooling	0.018	-0.007	-141	-0.007	-138	-0.001	-107
Inverse Life Expectancy	-0.292	0.026	-109	0.047	-116	0.037	-113
Log Fertility	-0.091	-0.191	110	-0.161	77	-0.140	54
Govt. Consump. Ratio	-0.046	-0.079	71	-0.079	70	-0.077	66
Rule of Law Index	0.021	0.013	-36	0.010	-52	0.006	-72
Democracy Index	0.519	0.077	-85	0.041	-92	0.003	-99
Democracy-squared	-0.528	0.023	-104	0.040	-108	0.055	-110
Trade Openness	0.080	0.432	439	0.440	448	0.419	422
Terms of Trade Ratio	0.055	-0.021	-138	-0.013	-124	-0.001	-102
Investment Ratio	0.010	0.011	8	0.010	-4	0.008	-17
Inflation Rate	-0.022	-0.030	38	-0.030	38	-0.029	37
Avg. Abs. Bias			115		116		110
<i>Between</i>							
Log Lagged Income	0.825	0.861	4	0.854	4	0.847	3
Upper Level Schooling	0.018	0.023	28	0.024	30	0.023	27
Inverse Life Expectancy	-0.292	-0.410	40	-0.379	30	-0.376	29
Log Fertility	-0.091	0.048	-152	0.039	-143	0.022	-125
Govt. Consump. Ratio	-0.046	-0.045	-2	-0.041	-11	-0.046	-1
Rule of Law Index	0.021	0.003	-87	0.004	-81	0.005	-75
Democracy Index	0.519	-0.050	-110	-0.001	-100	0.008	-98
Democracy-squared	-0.528	-0.014	-97	-0.031	-94	-0.031	-94
Trade Openness	0.080	0.039	-52	0.038	-52	0.039	-51
Terms of Trade Ratio	0.055	-0.107	-293	-0.061	-210	-0.044	-180
Investment Ratio	0.010	0.011	6	0.011	8	0.010	3
Inflation Rate	-0.022	-0.031	45	-0.030	42	-0.029	34
Avg. Abs. Bias			76		67		60
<i>Random effects</i>							
Log Lagged Income	0.825	0.619	-25	0.520	-37	0.457	-45
Upper Level Schooling	0.018	0.042	129	0.052	184	0.059	224
Inverse Life Expectancy	-0.292	-0.606	107	-0.638	118	-0.661	126
Log Fertility	-0.091	-0.139	53	-0.194	114	-0.227	149
Govt. Consump. Ratio	-0.046	-0.072	56	-0.072	55	-0.078	68
Rule of Law Index	0.021	0.021	0	0.031	49	0.038	83
Democracy Index	0.519	0.009	-98	0.062	-88	0.091	-82
Democracy-squared	-0.528	0.026	-105	0.042	-108	0.050	-109
Trade Openness	0.080	0.078	-3	0.082	2	0.076	-6
Terms of Trade Ratio	0.055	0.010	-81	0.014	-74	0.011	-81
Investment Ratio	0.010	0.010	2	0.011	7	0.011	9

Table 11 continued

Error-to-truth ratio:	True Coefs	<i>F</i> = 5%		<i>F</i> = 10%		<i>F</i> = 15%	
		Mean	Bias (%)	True Coefs	Mean	Bias (%)	
Variable							
Inflation Rate	-0.022	-0.023	5	-0.022	0	-0.020	-5
Avg. Abs. Bias			56		70		82
<i>Arellano–Bond</i>							
Log Lagged Income	0.825	-0.092	-111	-0.187	-123	-0.218	-126
Upper Level Schooling	0.018	0.028	54	0.005	-70	0.006	-68
Inverse Life Expectancy	-0.292	-0.193	-34	-0.056	-81	-0.007	-98
Log Fertility	-0.091	-0.321	254	-0.234	157	-0.163	80
Govt. Consump. Ratio	-0.046	-0.028	-41	-0.001	-98	0.011	-124
Rule of Law Index	0.021	0.015	-29	0.017	-18	0.013	-35
Democracy Index	0.519	0.056	-89	-0.050	-110	-0.126	-124
Democracy-squared	-0.528	0.020	-104	0.089	-117	0.118	-122
Trade Openness	0.080	0.349	335	0.318	296	0.322	301
Terms of Trade Ratio	0.055	-0.048	-186	-0.038	-168	-0.017	-131
Investment Ratio	0.010	0.010	-5	0.003	-70	0.000	-105
Inflation Rate	-0.022	-0.033	53	-0.035	61	-0.035	60
Avg. Abs. Bias			108		114		115
<i>Arellano–Bover/Blundell–Bond</i>							
Log Lagged Income	0.825	0.569	-31	0.458	-44	0.392	-52
Upper Level Schooling	0.018	0.041	123	0.053	193	0.065	256
Inverse Life Expectancy	-0.292	-0.752	157	-0.762	160	-0.780	167
Log Fertility	-0.091	-0.237	161	-0.299	229	-0.306	237
Govt. Consump. Ratio	-0.046	-0.072	55	-0.065	41	-0.067	44
Rule of Law Index	0.021	0.002	-91	0.019	-6	0.030	43
Democracy Index	0.519	-0.112	-122	-0.057	-111	0.009	-98
Democracy-squared	-0.528	0.079	-115	0.110	-121	0.111	-121
Trade Openness	0.080	0.158	97	0.185	130	0.183	128
Terms of Trade Ratio	0.055	-0.051	-192	-0.048	-186	-0.047	-185
Investment Ratio	0.010	0.015	50	0.015	50	0.015	51
Inflation Rate	-0.022	-0.022	3	-0.021	-2	-0.020	-5
Avg. Abs. Bias			100		106		116

mator in that it also weighs the between and within variations in the data. However, in contrast to RE, the SUR estimator does not assume the within-country correlation in the error term to be the same across subsequent time periods, but instead allows it to vary. For example, the covariance between ε_{i1} and ε_{i2} is allowed to differ from the covariance between ε_{i2} and ε_{i3} . Thus, we can refer to the SUR estimator as a flexible RE estimator, as the residual covariance matrix is less restrictive. This is expected to lead to efficiency gains. Moreover, the weighing of the between and within variations will now differ from the RE weighing scheme, and be a complicated function of the variance of the fixed-effects, the variance of the error term $\mu_i + v_{it}$ and its autocorrelation structure. Thus, the bias properties of SUR may differ from those of random effects if the country-specific effects are correlated with the regressors, since it will weigh the within and between variations differently.

Results in Table 12 show that the estimates are not very different from RE (for the sake of comparison Table 12 also includes the RE results already presented in Table 4), but SUR does better overall than RE across values of *F*. For example, the convergence coefficient γ_4 displays a bias of -3% with SUR and -16% with RE when *F* = 10%. In fact, SUR overall is the best estimator when it comes to estimating the speed of convergence λ . In terms of

Table 12 Simulation results for alternative estimators: 1000 runs, residual correlation = 25%

Error-to-truth ratio:	True Coefs	$F = 5\%$		$F = 10\%$		$F = 15\%$	
		Mean	Bias (%)	Mean	Bias (%)	Mean	Bias (%)
<i>Random effects</i>							
$\log s_{k,it-\tau}$	0.099	0.337	241	0.210	113	0.235	138
$\log s_{h,it-\tau}$	0.099	0.412	317	0.326	231	0.359	264
$\log(n+g+\delta)_{it-\tau}$	-0.197	-2.636	1236	-1.401	610	-1.705	764
$\log y_{it-\tau}$	0.832	0.453	-46	0.701	-16	0.642	-23
Avg. Abs. Bias			460		242		297
Implied λ	3.68%	0.158	330	0.071	93	0.089	141
<i>SUR (flexible random effects)</i>							
$\log s_{k,it-\tau}$	0.099	0.134	36	0.165	67	0.183	86
$\log s_{h,it-\tau}$	0.099	0.212	115	0.250	153	0.275	179
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.520	163	-0.887	350	-1.140	478
$\log y_{it-\tau}$	0.832	0.867	4	0.803	-3	0.754	-9
Avg. Abs. Bias			80		143		188
Implied λ	3.68%	0.029	-22	0.044	19	0.056	53
<i>Between</i>							
$\log s_{k,it-\tau}$	0.099	0.073	-26	0.077	-22	0.075	-23
$\log s_{h,it-\tau}$	0.099	0.086	-12	0.105	6	0.101	2
$\log(n+g+\delta)_{it-\tau}$	-0.197	-0.036	-82	-0.026	-87	-0.019	-91
$\log y_{it-\tau}$	0.832	1.019	23	0.995	20	1.000	20
Avg. Abs. Bias			36		34		34
Implied λ	3.68%	-0.004	-110	0.001	-97	0.000	-100
<i>Mankiw–Romer–Weil (modified between)^a</i>							
$\log s_{k,it-\tau}$	0.452	0.606	34	0.612	35	0.617	36
$\log s_{h,it-\tau}$	0.452	0.939	108	0.922	104	0.914	102
$\log(n+g+\delta)_{it-\tau}$	-0.905	-0.784	-13	-0.775	-14	-0.697	-23
$\log y_{it-\tau}$	0.229	0.727	217	0.773	237	0.808	252
Avg. Abs. Bias			93		98		103
Implied λ	3.68%	0.008	-78	0.006	-83	0.005	-86

^a MRW true coefficients adjusted with $\tau = 40$ instead of $\tau = 5$ to ensure comparability of the point estimates with the other estimators

average absolute bias, SUR does better than RE whatever the level of measurement error. Both estimators do particularly well in estimating the speed of convergence, but both tend to greatly bias away from zero the estimates on steady-state determinants.

4.6.2 The Mankiw, Romer and Weil estimator

The BE estimator does not strictly correspond to the cross-sectional estimator often used in the cross-country growth literature. Indeed, it involves the time averaging of all variables, including the income term on the left-hand side and lagged income on the right-hand side.

In contrast, cross-sectional estimators in the class of MRW's OLS estimator are based on the following regression:

$$\log y_{i,2000} = \gamma_0 + \gamma_1 \overline{\log s_{k,i}} + \gamma_2 \overline{\log s_{h,i}} + \gamma_3 \overline{\log (n + g + \delta)_i} + \gamma_4 \log y_{i,1960} + \vartheta_{it} \quad (21)$$

where the bar denotes averages computed over the whole period. Contrary to the BE estimator, income enters as end and beginning of period values (our total period spans 1960–2000), so measurement error in the initial income term does not get averaged away.⁴⁶

Table 12 displays the simulation results. For comparability, we also reproduce the BE results from Table 4. The MRW and BE biases are very similar in terms of magnitudes and signs. However, BE does better than MRW in terms of average absolute bias across values of F . These results confirm that simple OLS cross-sectional estimators are best at limiting net overall bias resulting from heterogeneity and measurement error.

5 Conclusion

In this paper, we performed Monte Carlo simulations to evaluate the econometric methods commonly used to estimate growth regressions. Our results suggest that, in the presence of measurement error, the fixed-effects estimator and the Arellano–Bond GMM estimator lead to overestimating the speed of convergence and to underestimating the impact of several common determinants of the steady-state level of income, such as human capital. The Blundell–Bond system GMM estimator corrects for some of these deficiencies, but suffers from a violation of some of its moment conditions, leading to some bias. Simple OLS on variables averaged over time provides a closer estimate of the speed of convergence, but overestimates the magnitude of the effect of steady-state determinants. These findings were shown to be robust to several changes in the design of the simulations.

Until now, differences in speeds of convergence across estimators were interpreted as implying that heterogeneity bias was prevalent in cross-sectional growth regressions, since fixed-effects methods led to a speed of convergence roughly 5 times higher than that estimated using a between estimator. This paper shows instead that the difference in estimated convergence speeds is likely attributable to the different incidence of measurement error bias. The estimated speed from traditional cross-sectional regressions is in all likelihood closer to the correct speed of convergence. The role of physical and human capital accumulation is also greater and the role of technological differences smaller than fixed-effects estimates would lead to believe. We opened this paper by contrasting two sets of results: First, we described cross-sectional estimates, which initiated a neoclassical revival in growth empirics. Second, we surveyed panel estimates emphasizing the role of technology as the main source of growth differences. Our simulations lead us to fall on the side of the first set of results.

Of course, simulations can only characterize the properties of the estimators. They cannot inform us as to the actual speed of convergence or the impact of steady-state determinants, since we simulated our data by assuming values for these parameters implied by a strict application of the Solow model. This strict application led to postulated parameters that may or may not hold true in actual data. Another weakness of our approach is that, under the assumptions we made concerning the data generating process, none of the estimators that we considered (i.e. none of the estimators currently used in the literature) allow us to exactly identify the parameters of interest. What our simulations do suggest is that if we

⁴⁶ In this application τ is set to 40, requiring a correction to ensure the comparability of the estimated reduced form coefficients with BE estimates.

Table 13 Growth regressions on actual data—dependent variable: log per capita income—1960 to 1999^a

	(1)	(2)	(2')	(3)	(4)	(5)	(6)	(7)
	BE	MRW	MRW adj. ^b	RE	SUR	FE	AB	BB
$\log s_{k,it-\tau}$	0.058 (0.016)**	0.382 (0.104)**	0.062	0.065 (0.012)**	0.066 (0.011)**	0.062 (0.017)**	0.072 (0.024)**	0.085 (0.016)**
$\log s_{h,it-\tau}$	0.041 (0.014)**	0.405 (0.085)**	0.066	0.040 (0.011)**	0.038 (0.010)**	-0.028 (0.020)	-0.060 (0.033)*	0.025 (0.016)
$\log(n + g + \delta)_{it-\tau}$	-0.214 (0.111)*	-2.173 (0.734)**	-0.355	-0.385 (0.091)**	-0.284 (0.078)**	-0.421 (0.153)**	-0.085 (0.217)	-0.617 (0.157)**
$\log y_{it-\tau}$	0.960 (0.014)**	0.523 (0.078)**	0.922	0.940 (0.011)**	0.948 (0.010)**	0.796 (0.026)**	0.795 (0.042)**	0.944 (0.018)**
# of observations	—	—	—	552	552	552	414	483
Number of countries	69	69	69	69	69	69	69	69
Implied λ	0.82	1.62	1.62	1.24	1.07	4.56	4.59	1.15

Standard errors in parentheses

* Significant at 10%; ** Significant at 1%

All regressions include time effects (except columns 1 and 2). Regressions include a constant term (not reported), where applicable.

^a Income at the end of each 5-year period in all but columns 2 and 2', where the dependent variable is income in 1999.

^b MRW estimates adjusted with $\tau = 5$ instead of $\tau = 40$ to ensure comparability of the point estimates with the other columns

wish to control for the $\log A_i(0)$ term found in the Solow model, a panel data estimator that uses only within-country variation is not the way to go. A much better strategy is to use a cross-sectional estimator and add relevant covariates reflecting $\log A_i(0)$.

Nonetheless, our results about relative biases can be used to establish which set of estimates is likely to be closest to the truth when using real data. Table 13 displays the results from estimating our basic specification using PWT version 6.1 data and updated series for the secondary school enrollment rate from Barro and Lee. We are able to replicate the basic findings of the past literature in this data: the speed of convergence is roughly 5 times larger under FE and AB (respectively 4.6% and 5.3%) compared to BE (0.8%). Our simulation results suggest the latter number is likely to be much closer to the truth. All the other estimators, that do not isolate the within variation in the data (namely MRW, SUR and RE) lead to estimated speeds of convergence that lie between 0.8% and 1.6%, while FE and AB lead to estimates in the neighborhood of 5%.⁴⁷ The slope parameters on the determinants of the steady-state level of income are reduced in magnitude when using FE or the AB estimator. They are similar across estimators using at least some between variation in the data (BE, SUR, RE and MRW). For example, the impact of the log of the enrollment rate

⁴⁷ The speed of convergence we report based on the application of the BE estimator, less than 1%, falls short of the number typically reported in the cross-sectional literature. Barro and Sala-i-Martin (1995) cite a number closer to 2%, based on the previous version of PWT. This difference is not attributable to our use of new and extended data. Past cross-sectional estimates rely on an OLS specification closer to the MRW estimator described above, where the current and lagged income terms do not get averaged over time. Implementing this estimator, we obtain a convergence speed of 1.62%, closer to existing estimates. Given that BE and MRW somewhat understate the speed of convergence, a number in the neighborhood of 2% for λ does not seem unreasonable.

is equal to 0.04 using the BE estimator, and is significant at the 1% level. As in Islam, this estimate switches signs (to -0.03) when using the FE estimator, and is statistically insignificant.

Running Monte Carlo simulations required that we take the Solow model seriously. However, we allowed for a very rich data generating process as the basis for the simulations. In particular, we explicitly modeled the dynamic nature of the Solow model, we allowed for measurement error, cross-country heterogeneity, as well as endogenous regressors in the baseline simulations. In extensions, we allowed for complex specifications of measurement error, alternative underlying models for generating the country fixed effects and the residuals, alternative parameter values in the growth specification, and an alternative growth specification altogether. Our basic findings were remarkably robust to allowing for these complexities, suggesting that the findings of this paper may have implications beyond the narrow confines of the Solow model. This paper illustrates an econometric second best property: by addressing one source of bias (stemming from omitted variables or reverse causality), the application of a certain class of estimators makes another source worse (measurement error).

Acknowledgements We thank Francesco Caselli, Nazrul Islam, Norman Loayza, David McKenzie and seminar participants at Stanford University, UC Davis, the University of Houston, the IMF and Duke University for useful comments. The data and programs used in this paper are available upon request, and all errors are ours.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix 1: Limiting values of the between, fixed-effects and random effects estimators

This appendix derives the limiting values of the BE, FE and RE estimators in the multivariate case. As in the text, assume that the true model is:

$$\log y_{it} = \gamma'x_{it} + \mu_i + \varepsilon_{it} \tag{22}$$

where all these variables are as defined in Sect. 2. As in Sect. 2.3, we also abstract from the dynamic nature of the model (i.e. assume $\text{plim } X'\varepsilon = 0$).

Assume also that x_{it} is imperfectly measured. Instead of observing x_{it} , we can only see $x_{it}^* = x_{it} + d_{it}$, where $E[d_{it}|x_{it}] = 0$ for all observations and $\text{var}[d_{it}|x_{it}] = D = \text{diag}\{\sigma_{d_1}^2, \sigma_{d_2}^2, \dots, \sigma_{d_k}^2\}$. The unconditional expectations and unconditional variance for three variables of interest appear in the main text as Eqs. 9 and 10.

In order to analyze the properties of the BE and FE estimators further, it is useful to break down the variation on each variable into within-country variation and between-country variation. Define the between-country variance for x_{it} as:

$$\Sigma_{xx}^B \equiv E \left[\left(\frac{1}{T} \sum_{t=1}^T x_{it} \right) \left(\frac{1}{T} \sum_{t=1}^T x'_{it} \right) \right] - E \left(\frac{1}{T} \sum_{t=1}^T x_{it} \right) E \left(\frac{1}{T} \sum_{t=1}^T x'_{it} \right) \tag{23}$$

and the within-country variance to be:

$$\begin{aligned} \Sigma_{xx}^W \equiv & E \left[\left(x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \right) \left(x'_{it} - \frac{1}{T} \sum_{t=1}^T x'_{it} \right) \right] \\ & - E \left[x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \right] E \left[x'_{it} - \frac{1}{T} \sum_{t=1}^T x'_{it} \right] \end{aligned} \tag{24}$$

It is well-known that:

$$\Sigma_{xx} = \Sigma_{xx}^W + \Sigma_{xx}^B \tag{25}$$

It is also easy to show that, for the covariance between x_{it} and μ , $\Sigma_{\mu x} = \Sigma_{\mu x}^B$. Finally, Σ_{xx}^B , the between covariance matrix of the imperfectly observed data x^* , is defined as:

$$\begin{aligned} \Sigma_{xx}^B \equiv & E \left[\left(\frac{1}{T} \sum_{t=1}^T x_{it} + d_{it} \right) \left(\frac{1}{T} \sum_{t=1}^T x'_{it} + d'_{it} \right) - E [x_{it} + d_{it}] E [x'_{it} + d'_{it}] \right] \\ & = \Sigma_{xx}^B + \frac{1}{T} D \end{aligned} \tag{26}$$

It is also easy to show that:

$$\Sigma_{xx}^W = \Sigma_{xx}^W + \frac{T-1}{T} D \tag{27}$$

We are now ready to derive the plims of the BE and FE estimators in the presence of measurement error and in the multivariate case.

First consider the BE estimator (OLS on country means across time). Using standard OLS results, we can derive:

$$\text{plim } \hat{\gamma}^{BE} = \left(\Sigma_{xx}^B + \frac{1}{T} D \right)^{-1} \Sigma_{xx}^B \gamma + \left(\Sigma_{xx}^B + \frac{1}{T} D \right)^{-1} \Sigma_{\mu x} \tag{28}$$

Now consider FE. To eliminate the heterogeneity bias arising through the correlation between the time invariant country-specific effects and the regressors, the most obvious solution is to use the FE estimator. By the Frisch-Waugh theorem, we can show that:

$$\hat{\gamma}^{FE} = (X^{*'} M_C X^*)^{-1} X^{*'} M_C y \tag{29}$$

where

$$M_C = I - C (C' C)^{-1} C' \tag{30}$$

and C is an $(NT \times N)$ matrix that stacks dummy variables for the different countries (with subvectors of T ones along the diagonals, zero elsewhere). Then:

$$\text{plim } \hat{\gamma}^{FE} = \left(\Sigma_{xx}^W + \frac{T-1}{T} D \right)^{-1} \Sigma_{xx}^W \gamma \tag{31}$$

Finally, as is well-known, RE is simply a matrix-weighted average of BE and FE estimates:

$$\hat{\gamma}^{RE} = \left(\hat{\Sigma}_{xx}^W + \hat{\theta} \hat{\Sigma}_{xx}^B \right)^{-1} \left(\hat{\Sigma}_{xx}^W \hat{\gamma}^{FE} + \hat{\theta} \hat{\Sigma}_{xx}^B \hat{\gamma}^{BE} \right) \tag{32}$$

where $\hat{\Sigma}_{xx^*}^W$ and $\hat{\Sigma}_{xx^*}^B$ are the sample estimates of $\Sigma_{xx^*}^W$ and $\Sigma_{xx^*}^B$, respectively, and $\hat{\theta}$ is an estimate of θ where:

$$\theta = \frac{\sigma_\varepsilon^2}{T\sigma_\mu^2 + \sigma_\varepsilon^2} \tag{33}$$

i.e. θ is the weights given to the BE estimator. Then:

$$\begin{aligned} \text{plim } \hat{\gamma}^{RE} &= \left[\left(\Sigma_{xx}^W + \frac{T-1}{T}D \right) + \theta \left(\Sigma_{xx}^B + \frac{1}{T}D \right) \right]^{-1} \\ &\times \left[\left(\Sigma_{xx}^W + \frac{T-1}{T}D \right) \gamma^{FE} + \theta \left(\Sigma_{xx}^B + \frac{1}{T}D \right) \gamma^{BE} \right] \end{aligned} \tag{34}$$

Note that when the variance of the error term ε_{it} is zero, RE reduces to FE.

Appendix 2: A simple univariate example

To illustrate the effects at play in the presence of both heterogeneity bias and measurement error, we consider the case where x_{it} is unidimensional, and contrast estimation by pooled OLS and FE. Consider the following relationship with a single observed regressor and an intercept term:

$$\log y_{it} = \gamma_0 + \gamma_1 x_{it} + \mu_i + \varepsilon_{it} \tag{35}$$

Suppose that the observed variable x_{it}^* incorporates measurement error:

$$x_{it}^* = x_{it} + d_{it} \tag{36}$$

where d_{it} is independent of the true x_{it} . The variance of the measurement error and of x_{it} are, respectively, σ_d^2 and σ_x^2 .

By estimating (35) using pooled OLS, we get both an omitted variables bias due to the fact that μ_i is potentially correlated with x_{it} , and a measurement error bias due to the correlation between ε_{it} and x_{it}^* . The limiting value of the pooled OLS estimate of γ_1 is as follows:

$$\text{plim } \hat{\gamma}_1^{POLS} = \frac{\gamma_1}{1 + \frac{\sigma_d^2}{\sigma_x^2}} + \frac{\text{cov}(x_{it}, \mu_i)}{\sigma_x^2 + \sigma_d^2} \tag{37}$$

In Eq. 37, the two sources of bias appear clearly. The variance of measurement error contributes to lessen the extent of heterogeneity bias, as it appears in the denominator of the expression on the right hand side of (37).

Consider now FE estimation, still in the univariate case. To simplify things and without loss of generality, assume that we difference away the time invariant individual effects by taking first differences, rather than by taking differences from country means of the data. The limiting value of the FE estimate of γ_1 is then:

$$\text{plim } \hat{\gamma}_1^{FE} = \frac{\gamma_1}{1 + \frac{\sigma_{\Delta d}^2}{\sigma_{\Delta x}^2}} \tag{38}$$

where $\sigma_{\Delta d}^2$ is the variance of the first differenced measurement error, and $\sigma_{\Delta x}^2$ is the same for the “true” regressor x_{it} .

We have derived formal expressions for our estimate of interest in two cases. The second method, FE, allows us to remove the heterogeneity bias but will exacerbate measurement

error bias. To see why, note that the error-to-truth ratio in the denominator of Eq. 38 will *always* have increased compared to that under pooled OLS:

$$\sigma_{\Delta d}^2 = \text{var}d_{it} + \text{var}d_{it-\tau} - 2\text{cov}(d_{it}, d_{it-\tau}) = 2\sigma_d^2 \quad (39)$$

$$\sigma_{\Delta x}^2 = \text{var}x_{it} + \text{var}x_{it-\tau} - 2\text{cov}(x_{it}, x_{it-\tau}) = 2\sigma_x^2(1 - \rho_x) \quad (40)$$

where $\rho_x = \text{corr}(x_{it}, x_{it-\tau})$ is the autocorrelation of x_{it} . Thus,

$$\frac{\sigma_{\Delta d}^2}{\sigma_{\Delta x}^2} = \frac{\sigma_d^2}{\sigma_x^2(1 - \rho_x)} > \frac{\sigma_d^2}{\sigma_x^2} \quad (41)$$

In words, $\sigma_{\Delta x}^2$ will be smaller relative to σ_x^2 the greater the time persistence in x_{it} (i.e. the higher is ρ_x).

We have assumed until now that there was no time persistence in measurement error (i.e. we had white noise errors-in-variables). This assumption is problematic in the context of data used for growth regressions, where errors in measurement from one period are likely to carry over to the next. In the case of autocorrelated measurement error, where we define $\rho_d = \text{corr}(d_{it}, d_{it-\tau})$, the error-to-truth ratio under FE is:

$$\frac{\sigma_{\Delta d}^2}{\sigma_{\Delta x}^2} = \frac{\sigma_d^2(1 - \rho_d)}{\sigma_x^2(1 - \rho_x)} \quad (42)$$

It is then easy to see that FE will exacerbate measurement error bias compared to pooled OLS whenever $\rho_d < \rho_x$.

References

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91, 1369–1401.
- Arellano, M. (2003). Modelling optimal instrumental variables for dynamic panel data. In *Invited lecture at european meeting of the econometric society in Venice*, CEMFI Working Paper no. 310.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment questions. *Review of Economic Studies*, 58(2), 277–297.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variables estimation of error-components models. *Journal of Econometrics*, 68, 29–51.
- Barro, R. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106(2), 407–443.
- Barro, R. (1997). *Determinants of economic growth*. Cambridge: MIT Press.
- Barro, R. J., & Lee, J.-W. (2000). *International data on educational attainment: Updates and implications*. Working Paper no. 42. Center for International Development at Harvard University.
- Barro, R., & Sala-i-Martin, X. (1995). *Economic growth*. New York: McGraw Hill.
- Barro, R., & Sala-i-Martin, X. (2003). *Economic growth* (2nd ed.). New York: McGraw Hill.
- Baumol, W. (1986). Productivity growth, convergence and welfare: What the long run data show. *American Economic Review*, 76, 1072–1085.
- Benhabib, J., & Spiegel, M. (1994). The role of human capital in economic development: Evidence from aggregate cross-country data. *Journal of Monetary Economics*, 34, 143–173.
- Bernanke, B. S., & Gürkaynak, R. S. (2001). Is growth endogenous? Taking Mankiw, Romer, and Weil seriously. *NBER Macroeconomics Annual*.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87, 115–143.
- Bond, S., Hoeffler, A., & Temple, J. (2001). *GMM estimation of empirical growth models*. CEPR Discussion Paper #3048.
- Caselli, F. (2004). The missing input: Accounting for cross-country income differences. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth*. North Holland (forthcoming).

- Caselli, F., Esquivel, G., & Lefort, F. (1996). Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth*, 1(3), 363–389.
- Caselli, F., & Feyrer, J. (2006). *The marginal product of capital*, working paper. London School of Economics.
- Cragg, J. G., & Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9, 222–240.
- DeLong, J. B. (1998). Productivity growth, convergence, and welfare: Comment. *American Economic Review*, 78, 1138–1154.
- Durlauf, S., Johnson, P., & Temple, J. (2005). Growth econometrics. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1, part A, chapter 8, pp. 555–677). Amsterdam: North Holland.
- Easterly, W., Loayza, N., & Montiel, P. (1997). Has Latin America's post-reform growth been disappointing? *Journal of International Economics*, 43, 287–311.
- Forbes, K. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review*, 90, 869–887.
- Frankel, J. A., & Romer, D. (1999). Does trade cause growth? *American Economic Review*, 89(3), 379–399.
- Griliches, Z., & Hausman, J. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1), 93–118.
- Hall, R., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics*, 114(1), 83–116.
- Hauk, W. R., & Wacziarg, R. (2004). A Monte Carlo study of growth regressions. *NBER Technical Working Paper #T0296*
- Heston, A., Summers, R., & Aten, B. (2002). *Penn world table version 6.1*. Center for International Comparisons at the University of Pennsylvania (CICUP).
- Islam, N. (1995). Growth empirics: A panel data approach. *Quarterly Journal of Economics*, 110(4), 1127–1170.
- Islam, N. (2000). Small sample performance of dynamic panel estimators in estimating the growth convergence equation: A Monte Carlo study. *Advances in Econometrics*, 15, 317–339.
- Klenow, P. J., & Rodríguez-Clare, A. (1997). The neoclassical revival in growth economics: Has it gone too far? In B. Bernanke & J. Rotemberg (Eds.), *NBER macroeconomics annual 1997* (pp. 73–102). Cambridge, MA: MIT Press.
- Klepper, S., & Leamer, E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, 52(1), 163–184.
- Knight, M., Loayza, N., & Villanueva, D. (1993). Testing the Neoclassical theory of economic growth: A panel data approach. *International Monetary Fund Staff Papers*, 40(3), 512–541.
- Levine, R., Loayza, N., & Beck, T. (2000). Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, 46, 31–77.
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2), 407–437.
- Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1), 65–94.
- Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 557–586.
- Stock, J. H., Wright, J., & Yogo, M. (2002). A survey of weak instruments and weak identification in GMM. *Journal of Business and Economic Statistics*, 20(4), 518–529.
- Stock, J. H., & Yogo, M. (2003). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Festschrift in honor of Thomas Rothenberg*. Cambridge, UK: Cambridge University Press (forthcoming).
- Temple, J. (1998). Robustness tests of the augmented solow model. *Journal of Applied Econometrics*, 13(4), 361–375.
- Wacziarg, R. (2002). Review of easterly's the elusive quest for growth. *Journal of Economic Literature*, 40(3), 907–918.
- Wansbeek, T. (2001). GMM estimation in panel data models with measurement error. *Journal of Econometrics*, 104, 259–268.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126, 25–51.