JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ., **13**, 361–375 (1998)

ROBUSTNESS TESTS OF THE AUGMENTED SOLOW MODEL

JONATHAN R. W. TEMPLE*
Hertford College, Oxford OX1 3BW, and Institute of Economics and Statistics, Manor Road, Oxford OX1 3UL

SUMMARY

This paper demonstrates some techniques for testing the robustness of cross-section and panel data regressions, and applies them to the influential augmented Solow growth model. The paper focuses on robust estimation and analysis of sensitivity to measurement error. In particular, it is shown that estimated technology parameters and convergence rates are highly sensitive to measurement error. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

In this paper I discuss some techniques for assessing the robustness of cross-section and panel data growth regressions. I then show how they can be used to test the influential augmented Solow model introduced by Mankiw, Romer, and Weil (1992), henceforth MRW. The task is useful for two reasons. First, MRW's paper has been extremely influential in the growth literature. Second, a case can be made that empirical researchers should apply these robustness tests as a matter of course. In showing how MRW's results change when the tests are applied, the paper demonstrates the potential importance of this kind of sensitivity analysis in empirical work.

There are two main strands to the robustness tests. The first strand is concerned with parameter heterogeneity and outliers. A frequent concern with cross-country growth work is that the countries are unlikely to fall on a common surface. As so often in economics, the regression model can be no more than a crude approximation to the underlying data-generating process. It is likely that the model will fit some observations particularly badly, and it is possible that these observations will act as influential outliers. Measurement error may also give rise to unrepresentative observations. The paper shows how techniques from the robust statistics literature can be applied to assess sensitivity to outliers.

The paper next considers the robustness of well-known results to measurement errors in the data. Sometimes the likely presence of measurement error is used to dismiss empirical literature out of hand. This is an unfortunate outcome, but partly arises because too few researchers are aware of techniques that can be used to assess the sensitivity of results to measurement error, and thereby indicate whether or not empirical work is robust to flaws in the data.

The tests are applied to the augmented Solow model of MRW. The next section sets out the model and discusses some possible weaknesses. Section 3 demonstrates the use of robust estimation, while Section 4 analyses the sensitivity of MRW's results to measurement error. Finally, Section 5 presents conclusions.

* Correspondence to: Jonathan Temple, Hertford College, Oxford OX1 3BW, UK. E-mail: jon.temple@economics.ox.ac.uk
Contract grant sponsor: ESRC; Hertford College, Oxford

CCC 0883–7252/98/040361–15$17·50

Received 3 October 1996
Revised 2 September 1997

2. THE AUGMENTED SOLOW MODEL

The underlying idea of the augmented Solow model is that aggregate technology is usefully described by a common Cobb–Douglas production function in which human capital plays an important role. MRW marshalled some impressive evidence in favour of this version of the Solow–Swan growth model. Mankiw (1995) argues strongly, partly based on this evidence, that the augmented Solow model can account for the known stylized facts about growth. This is an ambitious claim, but it is certainly true that the MRW paper has been much cited and discussed, and subsequent researchers have often chosen to build on their work.[1] Several authors, including Solow himself, have cited the paper as support for diminishing returns to accumulable factors, rather than the *Ak* formulations that have been popular in the literature.

Part of the appeal of the model lies in its simplicity. Following MRW, start with a Cobb–Douglas production function, in standard notation:

$$\begin{aligned} Y &= K^{\alpha} H^{\beta} (AL)^{1-\alpha-\beta} \\ L &= L(0)e^{nt} \\ A &= A(0)e^{gt} \end{aligned} \tag{1}$$

Define the stock of capital per efficiency unit of labour, $k = K/AL$, and similarly for output $y = Y/AL$. The evolution of capital is given by

$$\begin{aligned} \dot{k} &= s_k y - (n + g + \delta)k \\ &= s_k k^{\alpha} h^{\beta} - (n + g + \delta)k \end{aligned} \tag{2}$$

where δ is the rate of depreciation. MRW assume that a similar equation holds for human capital. Solving these equations for the steady-state, substituting them into the production function, and taking logs, gives steady-state income per capita as

$$\begin{aligned} \ln\left[\frac{Y^*}{L}\right] = {} & \ln A(0) + gt - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) \\ & + \frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \frac{\beta}{1-\alpha-\beta}\ln(s_h) \end{aligned} \tag{3}$$

Note the implicit restriction that the last three coefficients should sum to zero.

Equations like (3) seek to explain the variation in levels of per capita income across countries. The second set of regressions in MRW seeks to explain the change in log income between 1960 and 1985. The equation to estimate is obtained by using equation (3) and approximating around the steady state. Substituting for y^* gives

$$\begin{aligned} \ln y(t) - \ln y(0) = {} & \theta\frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \theta\frac{\beta}{1-\alpha-\beta}\ln(s_h) \\ & - \theta\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) - \theta \ln y(0) \end{aligned} \tag{4}$$

[1] Among them are Atje and Jovanovic (1993), Caselli *et al.* (1996), Durlauf and Johnson (1995), Helliwell (1992), Holtz-Eakin (1993), Islam (1995), Knight *et al.* (1993), Knowles and Owen (1995) and Lee *et al.* (1997).

where $\theta = 1 - e^{-\lambda t}$ and λ is the rate of convergence:

$$\frac{\mathrm{d}\ln y(t)}{\mathrm{d}t} = \lambda[\ln y^{*} - \ln y(0)] \tag{5}$$

$$\lambda = (n + g + \delta)(1 - \alpha - \beta) \tag{6}$$

We can rewrite equation (4) in terms of per capita output, to obtain

$$\begin{aligned}\ln\frac{Y(t)}{L(t)} - \ln\frac{Y(0)}{L(0)} = {}& \theta\ln A(0) + gt + \theta\frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \theta\frac{\beta}{1-\alpha-\beta}\ln(s_h) \\ & - \theta\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) - \theta\ln\frac{Y(0)}{L(0)}\end{aligned} \tag{7}$$

Writing the growth equation in this way makes clear the role of initial efficiency, $A(0)$. Since this variable is unobserved it must be omitted. If initial efficiency is correlated with the regressors, the estimates will be biased, a point familiar from the work of Islam (1995) and Caselli *et al.* (1996). Those papers treat initial efficiency as a fixed effect, and use panel data methods to analyse the augmented Solow model after eliminating the fixed effects.

As Barro (1996) points out, panel data methods have their own problems. The use of time-series variation may introduce unwanted business cycle effects, while the transformations used to eliminate fixed effects also reduce precision and can exacerbate the effects of measurement errors. These problems are familiar from the production function literature. Griliches and Mairesse (1995) suggest that the best way forward may be to proxy for the unobserved fixed effects, which leaves more identifying variance in the regressors, and is informative in itself.

That is the approach adopted in this paper. I use dummy variables for sub-Saharan Africa, Latin America and the Caribbean, East Asia, and the set of industrialized countries. The justification for doing this is the finding of Koop *et al.* (1995) that most of the variation in technical efficiency is between regional country groupings rather than within them. It is natural to think of the industrialized countries sharing a high level of efficiency by 1960, and sub-Saharan Africa a low one.

A second problem with the work of MRW is their key assumption that rates of technical progress are the same across countries, implying that per capita incomes will grow at the same rate in the steady state. As Lee *et al.* (1997) demonstrate, this assumption is not borne out by the data, and again tends to bias the estimates of convergence rates.[2] In this paper, though, I treat the assumption of a common rate of technical progress as a useful simplification, in the spirit of Box (1979): 'all models are wrong, but some are useful.' What I find is that, even when making this assumption, some widely cited results are unreliable.

In particular I will claim that aspects of the cross-section evidence are inconsistent with the augmented Solow model. The model proposed by MRW has almost no explanatory power for the OECD once a couple of outliers are removed. More seriously, one of the most discussed stylized facts, conditional convergence at a rate of 2% a year, is dubious. This paper shows that estimated convergence rates are highly sensitive to measurement error in initial income and the conditioning variables. That this might be a problem was indicated by Baumol (1986), De Long (1988) and Romer (1990), but since then the problem has been largely ignored.

[2] See also Canning *et al.* (1995).

These results suggest that it is useful to revisit the cross-section evidence. Work in this tradition has come under attack by those who advocate panel data or time-series methods. However, Mankiw (1995) and Barro (1996) have cast some doubt on the usefulness of panel data techniques. The findings in this paper mean that believers in the augmented Solow model must confront some important difficulties with the cross-section evidence, as well as that from panel data and time-series studies.

Given parameter heterogeneity, the time series methods of Lee *et al.* (1997) often have much to recommend them, but there are some situations in which such techniques cannot be applied. For instance, several of the most interesting variables used in the growth literature are fixed over time. This suggests that alternative approaches to parameter heterogeneity, such as those advocated here, are potentially useful in the cross-country growth literature and elsewhere. The next section explains the use of robust estimation and its relation to parameter heterogeneity.

3. PARAMETER HETEROGENEITY, OUTLIERS, AND ROBUST ESTIMATION

3.1. Robust Estimation

The underlying idea of robust estimation is to characterize the most coherent part of the data set, and so restrict the influence of small groups of observations that are not representative of the rest of the sample. Used thoughtfully, the techniques provide a relatively objective way of dealing with outliers. Otherwise, the choice of sample can quickly become controversial, and will lead to debates that are difficult to resolve with traditional techniques.[3]

Outliers may arise through measurement error, omitted variables or parameter heterogeneity. The last is particularly relevant to the growth literature, given the frequency of the objection that very different countries are unlikely to fall on a common surface. Often the problem is ignored, but sometimes it is addressed using single-case diagnostics, as discussed in Belsley, Kuh, and Welsch (1980). Cook's distance measure, the Studentized residuals, and DFITS fall into this category. Although easy to implement and interpret, the disadvantages of these diagnostics are increasingly well known to statisticians. These diagnostics may fail to pick up a group of outliers or leverage points (the 'masking effect') or wrongly identify some observations as unrepresentative, due to the presence of others (the 'swamping effect').

Robust estimators provide a useful alternative. There is a large literature on robust statistics, but the techniques have rarely been applied in economics. This is strange, since the methods are specifically designed for occasions where the chosen model is known to be only a crude approximation to the data-generating process.

The particular estimator chosen here is least trimmed squares (LTS) due to Rousseeuw (1984). The LTS estimator works by minimizing the sum of squares over half the observations, choosing the half with the smallest residual sum of squares. A way of thinking about this informally is that the estimator seeks out that part of the data which most nearly fits a common surface, and then bases the parameter estimates on that surface.

One problem is that typically LTS estimates are confined to coefficients, and programs are unable to report standard errors and diagnostic statistics. Although standard errors can be obtained by bootstrapping, it is harder to overcome the lack of diagnostics. To remedy this, I use the LTS estimates to classify some of the observations as unrepresentative, and then omit

[3] See Auerbach *et al.* (1994) and De Long and Summers (1994) for an example of one such debate.

these from an otherwise straightforward OLS regression. This procedure, a simple version of reweighted least squares (RWLS), is recommended by Rousseeuw and Leroy (1987) among others. Observations with high residuals are classed as unrepresentative, since they are some distance away from a robustly fitted regression line.

This raises the obvious problem that the cut-off for a 'high' residual is inevitably arbitrary. The solution used here is to drop observations in such a way that the parameter estimates of interest are close to the LTS estimates. A second apparent difficulty is more subtle. If an observation is identified as an outlier within, say, the OECD, we might expect to find it classified as unrepresentative in larger samples. In practice this does not always happen. The fitted model changes depending on the sample, and as it changes, so do the observations identified as unrepresentative. What this suggests is that the underlying technology is sufficiently different across samples for the classification of outliers to change. This is a problem for the augmented Solow model, rather than the procedure for identifying outliers used here.

Finally, the issue of leverage points requires some discussion. A leverage point is an observation which takes extreme values for one or more of the regressors. One drawback of least trimmed squares is that it will identify influential outliers, but will miss 'good' leverage points, those that affect the estimated precision of a parameter rather than its point estimate. Some statisticians advocate procedures which bound the influence of any observation on the results. Although these methods are more likely to provide estimates of standard errors that are robust to outliers, they usually require the applied researcher to specify tuning constants. Hence the procedure used to arrive at the final results will sometimes be less transparent than the relatively simple use of least trimmed squares, an important practical consideration.

3.2. Outliers and the Augmented Solow Model

The RWLS technique can be applied to MRW's regressions, and that is the task of the remainder of this section. MRW begin by examining the explanatory power of the augmented Solow model for per capita incomes. They note the startling finding that it is possible to explain 80% of the international variation in per capita income using just three variables: the investment ratio, population growth, and a proxy for investment in human capital.

However, the details of their findings are clearly problematic. The regressions take 1985 per capita income as a proxy for its steady-state value. This can be seen as an errors-in-variables problem, in which the measurement error is non-classical, or more simply as omission of a relevant variable: initial income should be included to control for transitional dynamics. Thus, I mainly concentrate on the growth regressions, which include initial income. However, since the findings for 1985 per capita incomes have received some attention, I briefly consider their robustness.

All three samples used by MRW include the OECD countries. Grossman and Helpman (1994) argue that the MRW results depend mainly on differences in schooling and population growth between rich and poor countries. If MRW's model is a good one, it should be capable of explaining per capita income when the sample is restricted to developing countries and NICs, or to the OECD. Hence my method is to divide the countries into the OECD and the remainder, and then estimate the levels equations by least trimmed squares.

It is sometimes argued that since the OECD is a club that countries join when they become rich, sample selection bias may be a problem. However, OECD membership changed relatively little over the sample period (1960–85) and indeed did not change at all between 1973 and 1985.[4] At least for the period under consideration, dividing the countries in this way is likely to provide some useful information on parameter heterogeneity.

For the regressions explaining 1985 per capita incomes (not reported) the key point is the poor performance of the augmented Solow model for the OECD. Use of LTS immediately identifies Greece, Portugal, and Turkey as unrepresentative observations: fitting a line to the most coherent part of the data leaves these countries with large residuals. Their effect on the OLS estimates is a striking demonstration of sample sensitivity. When just Portugal and Turkey are removed from the OECD sample, the R^2 falls from 0·35 to 0·02. It appears that, when one concentrates on the most coherent part of the OECD, the augmented Solow model in this form has almost no explanatory power. Given the variation in investment rates and the schooling variable across the OECD, and the relative accuracy with which the data are measured, this failure of the model is surprising.[5]

Recently, Nonneman and Vanhoudt (1996) have argued that the poor performance of the augmented Solow model for the OECD can be remedied by the inclusion of a variable for R&D investment. However, there are two problems with this set of results. First, the coefficient estimate on the R&D share is not significant at any conventional level when heteroscedasticity is allowed for. Second, and perhaps more seriously, robust estimation immediately identifies Japan as an outlier. There is no precisely estimated relation between growth and the R&D share when Japan is excluded from the sample. Omitting other possible influential outliers, again identified by least trimmed squares, only reinforces the result that the R&D share is not significant.

We can investigate another way of splitting the sample, by initial per capita income. In an important paper Durlauf and Johnson (1995) split the MRW sample using 1960 income and literacy rates, and found evidence that technology parameters varied across the samples, suggesting that the assumption of a common technology is a poor one. However, in their samples, the results may be driven by just one or two observations. As a simple test for multiple regimes, I divided the sample into four quartiles, from 1 (the poorest 25% of countries in 1960) to 4 (the richest 25%). Outliers are dropped from each quartile after robust estimation; this leaves 18–20 observations in each sample.

Looking at the results in Table I it is clear that in some ways the augmented Solow model stands up rather well. It has high explanatory power within each quartile, with R^2's around 0·6, or 0·5 when outliers are included. The adding-up restriction on the coefficients is rejected in only one sample. However, it seems that the relation between per capita income and population growth (the term in $\ln(n + g + \delta)$) is not properly captured by the augmented Solow model, as one might expect, given the complexities of the demographic transition. An alternative explanation is that rates of technical progress (g) and capital depreciation (δ) vary across countries. Finally, note the variation in estimated technology parameters across the samples.

After seeking to explain the variation in per capita income across countries, Mankiw, Romer, and Weil next examine the issue of growth and convergence, using estimates of equation (7). Note that, if one is only interested in the question of whether countries are converging or not, this is probably better studied by examining the international income distribution directly (Quah,

[4] When the OECD was formed in 1961, it had 20 member countries. Japan joined in 1964, Finland in 1969, Australia in 1971, and New Zealand in 1973.

[5] The mean of the investment ratio for the OECD is 25·8, with a standard deviation of 5. For the schooling variable, the figures are 9·1 and 2·1 respectively. The schooling variable ranges from 4·8 in Switzerland to 11·9 in the USA and New Zealand, so there is some variation even when Portugal and Turkey are excluded.

Table I. Robust regression estimates, by RWLS, stratified sample. Dependent variable: log GDP per working-age person in 1985

Quartile observations	Poorest 20	Second 20	Third 18	Richest 20
Constant	7·64	11·9	8·20	8·39
	(1·70)	(1·98)	(1·02)	(1·03)
$\ln(I/GDP)$	0·18	0·04	0·44	0·32
	(0·08)	(0·24)	(0·20)	(0·25)
$\ln(n+g+\delta)$	−0·15	1·17	−0·53	−0·90
	(0·61)	(0·81)	(0·31)	(0·28)
$\ln(SCHOOL)$	0·20	0·40	0·16	0·38
	(0·06)	(0·12)	(0·16)	(0·17)
R^2	0·58	0·58	0·62	0·67
Restriction p-value	0·62	0·07	0·91	0·68
RESET p-value	0·97	0·82	0·80	0·14
Implied α	0·13	0·03	0·28	0·19
Implied β	0·14	0·28	0·10	0·22

Quartile	Unrepresentative observations dropped in RWLS
Poorest	Botswana, Cameroon, Egypt, Indonesia, Zaire
Second	Brazil, Ghana, Korea, Papua New Guinea, Tunisia
Third	Hong Kong, Jamaica, Japan, Mexico, Singapore, Spain
Richest	Argentina, Chile, Ireland, Uruguay

Notes: MacKinnon–White (1985) HCSEs in parentheses. The technology parameters α and β are calculated using the coefficients on $\ln(I/GDP)$ and $\ln(SCHOOL)$.

1993a,b, 1996).[6] However, conditional convergence is interesting for a different reason: one can compare the estimated speed of convergence with that predicted by the augmented Solow model.

First, though, I briefly consider the question of unconditional convergence: do poorer countries grow faster than rich ones, on average? MRW found that the answer was no, except for the OECD sample. My investigation of this question using robust regression estimates indicated that the conclusion is not sensitive to unrepresentative observations. When outliers are excluded, the conclusion for the OECD is strengthened.[7]

It is perhaps more important to assess the robustness of the widely cited results on conditional convergence, drawn from the growth regressions. As discussed earlier, those results are likely to be biased by the omission of initial technical efficiency, $A(0)$. To lessen this bias, I estimate the equation including regional dummies, as before. A test for normality rejects at the 1% level in the intermediate sample, making the use of robust techniques all the more important. In Table II, I present estimates of MRW's conditional convergence regression made using reweighted least squares. One of the main results is that now, when excluding outliers and including regional dummies, the schooling variable is no longer significant. Hence it seems likely that the correlation

[6] Among other problems, the MRW approach has to face the possibility of 'local' convergence to one of several equilibria, as pointed out by Durlauf and Johnson (1995).

[7] A more detailed study of unconditional convergence would make use of the techniques in Barro and Sala-i-Martin (1992) and Erickson (1993) for dealing with measurement error.

Table II. Tests for conditional convergence: RWLS estimation. Dependent variable: log difference GDP per working-age person, 1960–1985

Sample observations	Non-oil 92	Intermediate 69	OECD 21	Non-oil, non-OECD 71	Intermediate, non-OECD 50
Constant	3·82	3·50	1·34	4·37	4·18
	(0·79)	(0·75)	(1·19)	(1·14)	(1·29)
$\ln(Y60)$	−0·30	−0·30	−0·32	−0·23	−0·24
	(0·06)	(0·07)	(0·07)	(0·08)	(0·10)
$\ln(I/GDP)$	0·59	0·66	0·13	0·56	0·61
	(0·09)	(0·12)	(0·20)	(0·10)	(0·13)
$\ln(n+g+\delta)$	−0·04	−0·24	−0·94	0·45	0·29
	(0·24)	(0·24)	(0·29)	(0·35)	(0·43)
$\ln(SCHOOL)$	−0·01	0·00	0·13	−0·07	−0·04
	(0·06)	(0·10)	(0·17)	(0·06)	(0·10)
AFRICA	−0·38	−0·38		−0·49	−0·37
	(0·11)	(0·14)		(0·11)	(0·14)
LATINCA	−0·11	−0·17		−0·21	−0·19
	(0·10)	(0·11)		(0·10)	(0·11)
EASTASIA	0·39	0·29		0·45	0·42
	(0·19)	(0·18)		(0·19)	(0·19)
INDUST	0·20	0·04			
	(0·14)	(0·15)			
R^2	0·71	0·75	0·74	0·70	0·72
Restriction p-value	0·05	0·13	0·05	0·01	0·05
Implied λ	0·014	0·014	0·015	0·010	0·011

Sample	Unrepresentative observations dropped in RWLS
Non-oil	Chad, Chile, Hong Kong, Mauritania, Somalia, Zambia
Intermediate	Argentina, Cameroon, Chile, Hong Kong, India, Zambia
OECD	Japan
Non-oil, non-OECD	Cameroon, Chad, Papua New Guinea, Somalia, Zambia
Intermediate, non-OECD	Cameroon, Chile, Zambia

Note: MacKinnon–White (1985) HCSEs in parentheses.

between growth and MRW's schooling measure is partly driven by outliers and a high correlation between schooling and the unobserved variable, initial efficiency.

Extending the analysis to the non-OECD samples is interesting. Importantly, the finding of conditional convergence, and its estimated rate, do not depend on the inclusion of the OECD in the sample. In some respects the RWLS estimates are satisfactory, with little evidence of heteroscedasticity or non-normality, and insignificant RESET statistics.

Note, however, that in the two non-OECD samples, the population growth and schooling variables are wrongly signed and insignificant, and the add-up-to-zero coefficient restriction is easily rejected. The dummy variables indicate that there is systematic variation in growth across continents which is not captured by the model. Hence MRW's claim that 'the augmented Solow model provides an almost complete explanation of why some countries are rich and other countries are poor' seems too strong. The statement seems to be true only at a broad-brush

Table III. Tests for conditional convergence, stratified sample. Dependent variable: log difference GDP per working-age person, 1960–85

Quartile observations	Poorest 21	Second 22	Third 21	Richest 21
Constant	6·68	−0·04	0·23	2·35
	(2·73)	(3·92)	(2·15)	(2·37)
$\ln(Y60)$	−0·90	0·23	−0·14	−0·37
	(0·29)	(0·32)	(0·19)	(0·18)
$\ln(I/GDP)$	0·26	0·83	0·49	0·56
	(0·12)	(0·16)	(0·21)	(0·26)
$\ln(n+g+\delta)$	−0·17	−0·15	−1·06	−0·96
	(0·58)	(1·00)	(0·39)	(0·36)
$\ln(SCHOOL)$	0·11	0·02	0·13	0·14
	(0·13)	(0·11)	(0·18)	(0·13)
R^2	0·57	0·69	0·75	0·84
Restriction p-value	0·77	0·41	0·38	0·46
Implied λ	0·092	—	0·006	0·018

Quartile	Unrepresentative observations dropped in RWLS
Poorest	Botswana, Cameroon, Egypt, Indonesia
Second	Korea, Morocco, Zambia
Third	Hong Kong, Jamaica, Singapore
Richest	Argentina, Chile, Uruguay

Note: MacKinnon–White (1985) HCSEs in parentheses.

level: the evidence presented here suggests that the augmented Solow model is suspect as a means of understanding the growth of developing countries.

We can also look at conditional convergence within the quartile samples considered earlier. Again this can be seen as a response to the heterogeneity problem: initial efficiency should not vary much across countries with similar levels of initial income. The RWLS estimates are presented in Table III. The most interesting aspect of the results is the variation in rates of convergence across samples: very fast in the poorest group, hardly present at all in the second and third quartiles. For countries in the middle income range in 1960, it does not appear to be true that relatively poor countries grew faster than others, holding other things constant. This potentially supports the finding of Quah (1996) and Bianchi (1997) that the world income distribution is polarizing into two groups, with a thinning in the middle of the distribution.

It is clear that cross-section regressions for the whole sample, as carried out by MRW, seem to obscure complex dynamics within the income distribution. These might be best captured by a more direct method, as advocated by Quah, rather than cross-section regressions including initial per capita output. As it stands, it may be that the conditional convergence result for the whole sample is driven by countries at the extremes of the income distribution.

Also worthy of note is the way the coefficient on the investment ratio varies across samples. There is some evidence here for the finding of Durlauf and Johnson (1995), that the underlying technology differs across countries. The variation in the constant term across samples suggests that initial levels of efficiency, and/or rates of technical progress, vary systematically across country groupings. This contradicts the usual assumptions made in estimating the augmented Solow model, and hence reinforces the results of Lee *et al.* (1997).

4. SENSITIVITY TO MEASUREMENT ERROR

I now turn to measurement error issues. Empirical work, including the 'new growth evidence', has not properly addressed measurement error biases, with one or two exceptions. The approach implicit in MRW, and nearly all other studies, is that the problem can be ignored, since good results can apparently be obtained even in its presence. This ignores the fact that, with several badly measured variables, coefficients can be biased away from zero as well as towards it. Inference should be far more careful given the likelihood of severe measurement error.

There are reasons to be particularly suspicious of the data used in cross-section growth regressions. Data quality for observed variables is often poor and, for others, rather crude proxies are used. Some of the most serious difficulties are noted by Heston and Summers (1996). They point out the lack of good data on labour force participation rates, which 'vary enough to make GDP per capita a very unsatisfactory proxy for GDP per worker' (p. 23). In the context of the augmented Solow model, one might also point to the proxy for human capital investment, the schooling variable, as another which is likely to be badly mismeasured.

If we are willing to make the classical errors-in-variables assumptions, we can use multivariate reverse regression to obtain information about sensitivity to measurement error. The modern treatment is due to Klepper and Leamer (1984), and their diagnostics have been extended by Klepper (1988), Klepper and Stapleton (1987), and Klepper *et al.* (1993). These diagnostics exploit the fact that the covariance matrices of the true variables, of the disturbances in the true model, and of the measurement errors, must all be positive semidefinite. These conditions, sometimes together with further restrictions, allow bounds to be placed on the value of the true coefficients. The diagnostics typically abstract from sampling error, but since this is often small in relation to possible measurement error biases, this tends to be a minor omission in practice.

One problem in using this technique is that, given sufficient measurement error, it may be that exact collinearity of the true regressors cannot be ruled out. In this case the coefficients are arbitrary and it is impossible to say anything about even their signs. However, we can overcome this by using educated guesses about the likely extent of measurement error. To do this we need to introduce the concept of R^{*2}, the value the R^2 would take if all variables were correctly measured. By imposing a sufficiently low bound, R, on R^{*2} it is possible to obtain bounds on the set of point estimates. Say our R bound is 0·90. Then we are assuming that our model would not explain more than 90% of the variation in per capita income even if all the variables were correctly measured. If we choose a lower value of R, nearer the actual R^2, we are implicitly making a more optimistic assumption about the likely extent of measurement error in our particular data set.

The sensitivity of the results to measurement error can also be shown by using classical method-of-moments estimators (see, for instance, Carroll *et al.*, 1995). Unlike the Klepper–Leamer method, these always require estimates of the reliability r of the regressors, where

$$r = 1 - \frac{\text{Noise variance}}{\text{Total variance}}$$

In this paper, I use both reverse regression and method-of-moments estimators to demonstrate the sensitivity of MRW's results to measurement error. First, consider the application of reverse regression to the MRW levels regressions. This indicates rather wide bounds on the technology parameters. For instance, for the non-oil sample and an R bound of 0·9 (compared to an actual R^2 of about 0·8) then the capital share ranges between 0·1 and 0·6, and the human capital share between 0·1 and 0·5.[8] To try and narrow these bounds, I carried out the reverse regression imposing the restriction that the coefficients sum to zero. I also experimented with a lower R bound of 0·85. We can use the d-statistics described in Klepper (1988) to give some idea of the optimism involved. If the R^2 of the model with correctly measured variables is to lie below 0·85, this must mean that less than 15% of the variation in the variable $\ln(SCHOOL) - \ln(n + g + \delta)$ is due to measurement error, using the d-statistic for this variable. Given the imperfection of *SCHOOL* as a proxy for the human capital investment rate, the mismeasurement of population growth, and the imputation of common growth and depreciation rates, we are clearly making a rather optimistic assumption about the likely extent of measurement error.

Yet even with this degree of optimism, the technology parameters α and β can lie anywhere within (0·20, 0·49) and (0·15, 0·38) respectively. For the intermediate sample, which omits the countries with the worst data, the bounds are still wide at (0·18, 0·47) and (0·17, 0·41) respectively. Although the finding that returns to accumulable factors are diminishing does seem to be robust, it is clearly a mistake to draw detailed conclusions about technology parameters from cross-section regressions of this kind.

Using estimates of the reliability of the variables, sensitivity to measurement error can also be assessed using method-of-moments estimators. If, optimistically, each regressor is thought to be measured with a reliability of 0·95, then the estimates of the parameters (α, β) for the non-oil sample are little changed from MRW's estimates of (0·25, 0·38). However, if investment is measured with a reliability of 0·9 while the schooling and population growth variables have a lower reliability (0·8), then the estimates of the technology parameters are (0·13, 0·54). Given the many uncertainties in data measurement, these reliabilities do not seem unrealistic, but it is clear that the results can be highly sensitive to flaws in the data.

Turning now to the growth regressions, it may be that the convergence result is entirely due to measurement error, a point made by Baumol (1986) and Romer (1990). To see this, rearrange equation (4) to obtain:

$$\ln y(t) = \text{conditioning variables} + (1 - \theta) \ln y(0) \tag{8}$$

If initial income is measured with error, while the other variables are correctly measured, then the coefficient on initial income will be biased towards zero. In turn, this will lead to an over-estimate of the speed of convergence.

Two considerations introduce further complications. First, the measurement error in final income may be correlated with that in initial income. Second, if one allows for measurement error in at least some of the other variables, the estimated rate of convergence could be biased in either direction. This should also be clear from equation (6), which shows how the convergence rate is related to population growth and the technology parameters. Since we have already established that the technology parameters are imprecisely estimated, it should be clear that it will be difficult to obtain clear-cut results for rates of convergence.

We can address this question more directly using Klepper's measurement error diagnostics. Table IV reports extreme lower and upper bounds for the rate of convergence, given bounds on

[8] Note that, as in Klepper and Leamer, these bounds do not allow for sampling error, which would make them even wider.

R^{*2}. It is clear that convergence may be so slow in the non-oil and intermediate samples that it is barely taking place at all. This finding is reinforced when using a method-of-moments estimator to correct for measurement error. If initial income is measured with a reliability of 0·9, the estimated convergence rate is 0·04% a year. Even if the reliability is as high as 0·95, which seems unlikely, the estimated convergence rate is still just 0·75% a year.

Hence, the measurement error problem is almost certainly too great to draw firm conclusions about convergence from cross-section regressions. Convergence may not be occurring, or may be rather faster than the 2% figure which has been much quoted in the literature. We certainly cannot take conditional convergence as a stylized fact, or use the estimated speed to draw conclusions about the likely share of capital in the production function. In the presence of measurement error, it is far too unreliably estimated.

It is possible to object that the R bounds imposed on R^{*2} seem rather high, but these reflect the high R^2's obtained using the regression (8) and shown in Table IV. These R bounds are not particularly pessimistic. For instance, to accept them, we must believe that less than 10% of the variation in initial income is due to measurement error. And if the other variables are measured incorrectly, as they surely are, this condition will be rather more stringent.[9]

Table IV. Convergence rates allowing for measurement error

	Non-oil	Intermediate	OECD	Non-oil, non-OECD	Intermediate, non-OECD
Lower bound	0·003	0·005	0·015	0·009	0·007
Upper bound	0·067	0·063	0·036	0·029	0·057
R^2	0·91	0·90	0·88	0·84	0·82
R bound	0·95	0·95	0·95	0·87	0·89

5. CONCLUSIONS

It is often argued that tests of robustness should be used more routinely in applied research than they are at present. One reason for their absence is ignorance about the kind of tests available, and how they can be implemented. In this paper, I have discussed some useful robustness tests, and applied them to the influential work of Mankiw, Romer, and Weil (1992). Accompanying computer software to implement the tests is available from the JAE's data archive.

The first issues to be addressed by the paper are parameter heterogeneity and influential outliers. In cross-country growth work, the likely heterogeneity of parameters across countries leads some to advocate time-series studies of single countries. For many developing countries, time-series data is not available on some of the most interesting variables in the literature, and even when a time series is available, often it has been interpolated from two or three census years. This suggests that alternative techniques for handling parameter heterogeneity could sometimes be useful, when researchers are restricted to cross-section or panel data work. Robust estimators like least trimmed squares provide a useful way of identifying and characterizing the

[9] To obtain narrower bounds on the rate of convergence, I experimented with imposing the restriction that the coefficients on investment, schooling, and population growth should add to zero. This did little to narrow the bounds.

most coherent part of the sample, and clearly have much potential, particularly in the growth literature.

The application of robust estimators to MRW's results suggests that they stand up fairly well, with the important exception of the results for the OECD. The augmented Solow model has almost no explanatory power for this group of countries. The extension of Nonneman and Vanhoudt (1996) for the OECD is also flawed; use of a robust estimator suggests that their results are driven by the presence of just one country, Japan.

The second set of robustness tests considered here focus on measurement error. When several variables are measured with error, coefficient estimates can be biased away from zero, and researchers should take this into account. Method-of-moments adjustments for measurement error are not difficult to implement, and in any case are often available in computer software. Where researchers wish to avoid estimating the reliability of variables, this can sometimes be achieved by using the measurement error diagnostics introduced by Klepper and Leamer (1984) and Klepper (1988).

The importance of this kind of sensitivity analysis is shown by applying it to the results of MRW on convergence. Much has been made of the finding that countries and regions seem to converge to their steady states at a rate of 2% a year. Some have used this new 'stylized fact' to draw important conclusions about the roles of physical capital and education in growth. This paper has shown that the finding is unreliable even if fixed effects and parameter heterogeneity are negligible difficulties, and even when the assumptions made about measurement error are rather optimistic. This casts some doubt on the regional convergence literature, as well as that addressed at convergence of nations. This emphasis on the inherent uncertainty in measuring convergence rates echoes that in Lee *et al.* (1997). MRW's claim that 'countries converge at about the rate the augmented Solow model predicts' is not well founded.

Finally, it is worth noting that the growth literature is moving towards panel data studies, as in Islam (1995) and Caselli *et al.* (1996). Measurement error and influential outliers remain serious difficulties in panel data applications, but robustness tests are even rarer in this context than when cross-section methods are applied. When a GMM estimator is used, as in Caselli *et al.* (1996), this may overcome measurement error biases to some extent, but GMM is sometimes regarded as sensitive to outliers. At present, techniques to address this difficulty (such as bounded-influence GMM) are still in their infancy.

However, the techniques implemented here can easily be applied at a preliminary stage. When fixed effects models are used, estimation proceeds by ordinary least squares after a simple transformation of the data. Hence, for linear models, all the sensitivity testing used here (Klepper's measurement error diagnostics, method-of-moments estimators, and robust estimation) can be applied after the within-groups transformation. This can be used as part of an exploratory analysis, for instance to identify observations which are particularly likely to be influential in GMM estimates of a dynamic model.

COMPUTING APPENDIX

OLS and reweighted least squares were carried out using PcGive 8·0, which allows calculation of the jackknife HCSEs. Robust regressions and the Klepper diagnostics were obtained using programs written in the statistics language S-Plus. Note that the current implementation of the language only approximates the LTS estimator, although in general this approximation should be reasonably good. Finally, method-of-moments corrections for measurement error were made using the statistics package Stata.

PcGive 8.0, International Thomson Publishing, Berkshire House, 168–173 High Holborn, London WC1V 7AA.

S-Plus, Statsci Division, 1700 Westlake Avenue N., Suite 500, Seattle, WA 98109. E-mail: mktg@statsci.com.

Stata, Release 4, StataCorp, College Station, TX.

ACKNOWLEDGEMENTS

An earlier version of this paper circulated under the title 'Testing the augmented Solow model'. The paper has been greatly improved by comments from Christopher Bliss, Paul Johnson, Steven Klepper, John Muellbauer, Steve Nickell, Danny Quah, Steve Redding, and an anonymous referee. The remaining errors are mine. I am grateful to the ESRC and Hertford College, Oxford, for financial support.

REFERENCES

Atje, R. and B. Jovanovic (1993), 'Stock markets and development', *European Economic Review*, **37**, 632–40.

Auerbach, A. J., K. A. Hassett, and S. D. Oliner (1994), 'Reassessing the social returns to equipment investment', *Quarterly Journal of Economics*, **109**, 789–802.

Barro, R. J. (1996), 'Determinants of economic growth: a cross-country empirical study', NBER working paper No. 5698.

Barro, R. J. and X. Sala-i-Martin (1992), 'Convergence', *Journal of Political Economy*, **100**, 223–51.

Baumol, W. J. (1986), 'Productivity growth, convergence and welfare: what the long-run data show', *American Economic Review*, **76**, 1072–85.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980), *Regression Diagnostics*, Wiley, New York.

Bianchi, M. (1997), 'Testing for convergence: evidence from non-parametric multimodality tests', *Journal of Applied Econometrics*, **12**, 393–409.

Box, G. E. P. (1979), 'Robustness in the strategy of scientific model building', in R. L. Launer and G. N. Wilkinson (eds), *Robustness in Statistics*, Academic Press, New York.

Canning, D., P. Dunne, and M. Moore (1995), 'Testing the augmented Solow and endogenous growth models', manuscript, Queen's University of Belfast.

Carroll, R. J., D. Ruppert, and L. A. Stefanski (1995), *Measurement Error in Non-linear Models*, Chapman and Hall, London.

Caselli, F., G. Esquivel, and F. Lefort (1996), 'Reopening the convergence debate: a new look at cross-country growth empirics', *Journal of Economic Growth*, **1**, 363–90.

De Long, J. B. (1988), 'Productivity growth, convergence and welfare: comment', *American Economic Review*, **78**, 1138–54.

De Long, J. B. and L. H. Summers (1994), 'Equipment investment and economic growth: reply', *Quarterly Journal of Economics*, **109**, 803–07.

Durlauf, S. N. and P. A. Johnson (1995), 'Multiple regimes and cross-country growth behaviour', *Journal of Applied Econometrics*, **10**, 365–84.

Erickson, T. (1993), 'Restricting regression slopes in the errors-in-variables model by bounding the error correlation', *Econometrica*, **61**, 959–69.

Griliches, Z. and J. Mairesse (1995), 'Production functions: the search for identification', manuscript, NBER, August.

Grossman, G. M. and E. Helpman (1994), 'Endogenous innovation in the theory of growth', *Journal of Economic Perspectives*, **8**, 23–44.

Helliwell, J. F. (1992), 'International growth linkages: evidence from Asia and the OECD', NBER working paper No. 4245.

Heston, A. and R. Summers (1996), 'International price and quantity comparisons: potentials and pitfalls', *American Economic Review*, **86**, 20–4.
Holtz-Eakin, D. (1993), 'Solow and the States: capital accumulation, productivity and economic growth', *National Tax Journal*, **46**, 425–39.
Islam, N. (1995), 'Growth empirics: a panel data approach', *Quarterly Journal of Economics*, **110**, 1127–70.
Klepper, S. (1988), 'Regressor diagnostics for the classical errors-in-variables model', *Journal of Econometrics*, **37**, 225–50.
Klepper, S., M. S. Kamlet, and R. G. Frank (1993), 'Regressor diagnostics for the errors-in-variables model—an application to the health effects of pollution', *Journal of Environmental Economics and Management*, **24**, 190–211.
Klepper, S. and E. E. Leamer (1984), 'Consistent sets of estimates for regressions with errors in all variables', *Econometrica*, **52**, 163–83.
Klepper, S. and D. C. Stapleton (1987), 'Consistent sets of estimates for restricted regressions with errors in all variables', *International Economic Review*, **28**, 445–57.
Knight, M., N. Loayza, and D. Villanueva (1993), 'Testing the neoclassical theory of economic growth', *IMF Staff Papers*, **40**, 512–41.
Knowles, S. and P. D. Owen (1995), 'Health capital and cross-country variation in income per capita in the Mankiw–Romer–Weil model', *Economics Letters*, **48**, 99–106.
Koop, G., J. Osiewalski, and M. F. J. Steel (1995), 'Measuring the sources of output growth in a panel of countries', CORE discussion paper No. 9542.
Lee, K., M. H. Pesaran, and R. Smith (1997), 'Growth and convergence in a multi-country empirical stochastic Solow model', *Journal of Applied Econometrics*, **12**, 357–92.
MacKinnon, J. G. and H. White (1985), 'Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties', *Journal of Econometrics*, **29**, 305–25.
Mankiw, N. G. (1995), 'The growth of nations', *Brookings Papers on Economic Activity*, 275–310.
Mankiw, N. G., D. Romer, and D. N. Weil (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, **107**, 407–37.
Nonneman, W. and P. Vanhoudt (1996), 'A further augmentation of the Solow model and the empirics of economic growth for OECD countries', *Quarterly Journal of Economics*, **111**, 943–53.
Quah, D. T. (1993a), 'Galton's Fallacy and tests of the convergence hypothesis', *Scandinavian Journal of Economics*, **95**, 427–43.
Quah, D. T. (1993b), 'Empirical cross-section dynamics in economic growth', *European Economic Review*, **37**, 426–34.
Quah, D. T. (1996), 'Empirics for economic growth and convergence', *European Economic Review*, **40**, 1353–75.
Romer, P. M. (1990), 'Human capital and growth: theory and evidence', *Carnegie-Rochester Conference Series on Public Policy*, **32**, 251–86.
Rousseeuw, P. J. (1984), 'Least median of squares regression', *Journal of the American Statistical Association*, **79**, 871–80.
Rousseeuw, P. J. and A. M. Leroy (1987), *Robust Regression and Outlier Detection*, Wiley, New York.

