# WHEN IS IT JUSTIFIABLE TO IGNORE EXPLANATORY VARIABLE ENDOGENEITY IN A REGRESSION MODEL?

RICHARD A. ASHLEY AND CHRISTOPHER F. PARMETER

ABSTRACT. The point of empirical work is commonly to test a very small number of crucial null hypotheses in a linear multiple regression setting. Endogeneity in one or more model explanatory variables is well known to invalidate such testing using OLS estimation. But attempting to identify credibly valid (and usefully strong) instruments for such variables is an enterprise which is arguably fraught and invariably subject to (often justified) criticism. As a modeling step prior to such an attempt at instrument identification, we propose a sensitivity analysis which quantifies the minimum degree of correlation between these possibly-endogenous explanatory variables and the model errors which is sufficient to overturn the rejection (or non-rejection) of a particular null hypothesis at, for example, the 5% level. An application to a classic model in the empirical growth literature illustrates the practical utility of the technique.

**JEL Classification: C2, C15**.

## 1. INTRODUCTION

No issue in econometrics has evoked as much literature (and angst) over the years as that of the likely endogeneity of explanatory variables in our regression models. The profession's main response – instrumental variables (IV) regression – has ameliorated this concern in some settings, albeit at the cost of a decrease in estimation precision. In large part, however, the use of IV regression has simply shifted the focus of attention to the exogeneity of the instruments, spawning a search for 'clever' instruments whose exogeneity can be argued – e.g., see Angrist and Krueger (1991) and Acemoglu, Johnson and Robinson (2001). For evidence that instrument validity is a continuing concern, see Angrist and Pischke (2010), Keane (2010), Leamer (2010), Murray (2006), Sims (2010) and Stock (2010). Even more recently, Bazzi and Clemens (2013) have strongly criticized the way IV is applied in growth regressions. In the present paper we suggest a different approach.

Applied economic analysis almost always culminates in the rejection of (or, occasionally, in the failure to reject) a very small number of crucial null hypotheses at some nominal level of significance, usually 5%. For any particular one of these hypothesis tests, this translates into a rejection $p$-value of less than or equal to 0.05. The endogeneity issue then becomes: is the reported rejection of the null hypothesis actually just an artifact of unaccounted for (or improperly accounted for) endogeneity in the supposedly exogenous explanatory variables?

But suppose that it was possible to determine that any likely degree of such endogeneity was insufficient to overturn our small set of key inferences? We could then base our analysis on OLS regression without having to expend resources (or credibility) on finding plausible instruments or on worrying about their validity.

Ashley and Parmeter (2015) provides an empirically implementable algorithm for performing exactly this kind of sensitivity analysis with respect to the validity (exogeneity) of instruments used in linear GMM regression modeling. Where the algorithm finds that a key hypothesis test rejection is overturned by very small amounts of correlation between the instruments and the (unobserved) model errors, this inference is deemed 'fragile.' In contrast, where it is found that quite substantial levels of instrument-error correlation – e.g., in excess of 0.50 in magnitude – are necessary in order to overturn this hypothesis test rejection, then this inference is deemed 'robust.'[1] Clearly, inference with respect to some null hypotheses may be fragile whereas others are robust, even within the same regression model.

Here we observe that OLS regression is equivalent to letting regressors act as instruments for themselves and apply the Ashley/Parmeter algorithm to the underlying model estimated via OLS. As an illustrative example, in the next section we analyze the impact of explanatory variable endogeneity on the inferential conclusions obtained in Mankiw, Romer and Weil (1992), a foundational paper in the economic growth literature an area that is routinely criticized for endogeneity.

## 2. A Sensitivity Analysis for Exogeneity

### 2.1. **Estimation/Inference in the Presence of Unaddressed Endogeneity.**

Consider the standard linear model, with the 'structural equation'

$$Y_1 = Y_2\alpha + W_1\beta + \varepsilon, \tag{1}$$

where $Y_2$ is an $n \times m$ matrix of (potentially) endogenous variables, $W_1$ is an $n \times k$ matrix of variables whose exogeneity is not in question, $\alpha$ and $\beta$ are $m \times 1$ and $k \times 1$ vectors of coefficients, respectively, and $\varepsilon$ is the structural error. For the present purpose we do not assume the presence of additional (instrumental) variables to correct for the endogeneity of $Y_2$.[2]

Accordingly, the moment conditions assumed here are:

$$\begin{aligned} E\left[Y'_{2i}\varepsilon_i\right] &= 0 \\ E\left[W'_{1i}\varepsilon_i\right] &= 0 \end{aligned} \tag{2}$$

The first of the two conditions in Equation 2 incorporates the assumed exogeneity of the $m$ potentially endogenous variables in $Y_2$; the second condition reflects the assumption that the remaining

---

[1]Where, as is common, the validity of multiple instruments is in question, the algorithm also provides a sensible indication as to which of the instruments are the source of any fragility found. R code implementing the algorithm is available from the authors.

[2]As noted above, see Ashley and Parmeter (2015) for a related treatment explicitly allowing the use of (possibly flawed) instruments in 2SLS/GMM estimation; here the focus is on OLS estimation in the absence of credibly valid instruments.

$k$ variables are clearly exogenous. Thus, $Y_2$ and $W_1$ are defined in such a way that we need only concern ourselves with violations of exogeneity for the $m$ variables in $Y_2$.

Letting $\gamma = [\alpha' \ \beta']'$ and $X = [Y_2 \ W_1]$, the structural equation [1] can be written more compactly as:

$$(3) \qquad\qquad Y_1 = X\gamma + \varepsilon.$$

The OLS estimator of $\gamma$ is thus:

$$(4) \qquad\qquad \hat{\gamma}_{OLS} = (X'X)^{-1}X'Y_1.$$

This estimator is consistent, asymptotically efficient and asymptotically normal under the standard assumptions, including (at least asymptotic) exogeneity of **all** the regressors $X$.[3]

When some or all of the variables in $X$ are *not* exogenous, then

$$(5) \qquad\qquad E\left[X_i'\varepsilon_i\right] = n\Sigma_{X\varepsilon} \neq 0.$$

The factor $n$ is introduced here so that $\Sigma_{X\varepsilon}$ can be interpreted as the population covariance vector between the structural error $\varepsilon$ and the $g+k$ supposedly exogenous variables; $\Sigma_{X\varepsilon}$ can thus sensibly be referred to as "the exogeneity flaw covariance vector."

For a given value of the exogeneity flaw covariance vector, then it is easy to show that the modified estimator of $\gamma$,

$$(6) \qquad\qquad \tilde{\gamma} = (X'X)^{-1}\left(X'Y_1 - n\Sigma_{X\varepsilon}\right).$$

is now consistent, asymptotically normal, and asymptotically efficient and that (conditional on the 'flaw' vector, $\Sigma_{X\varepsilon}$) $\tilde{\gamma}$ has asymptotic sampling distribution,

$$(7) \qquad\qquad \sqrt{n}\left(\tilde{\gamma} - \gamma - E_{XX}^{-1}\Sigma_{X\varepsilon}\right) \sim N\left(0, \sigma_\varepsilon^2 E_{XX}^{-1}\right),$$

where $E_{XX}^{-1} = plim \ n^{-1}(X'X)^{-1}$. Thus, obtaining an asymptotically valid $p$-value at which any particular null hypothesis regarding $\gamma$ could be rejected would be straightforward if $\Sigma_{X\varepsilon}$ were known.[4]

## 2.2. **Quantifying the Sensitivity of Inference to Endogeneity in $Y_2$.**

Suppose that a particular null hypothesis regarding $\gamma$ can be rejected – using $\hat{\gamma}_{OLS}$ and its asymptotic sampling distribution, under the assumption that all of the variables in $Y_2$ are exogenous – at, say, the 5% level.[5] This is equivalent to saying that the rejection $p$-value for this null hypothesis

---

[3]These standard assumptions also include that of a homoscedastic and non-autocorrelated error term, a correct specification of the conditional mean of $Y_1$, and full rank of the covariate matrix $X$.

[4]Note that our sensitivity analysis remains squarely within the usual (asymptotic) inference framework for OLS/2SLS/GMM estimation and inference; see Dufour (2003) with regard to finite-sample alternatives.

[5]The analysis would be essentially identical for a rejection at the 1% (or any other) level: the description in this section is made definite for the 5% level solely to enhance the clarity of the exposition. Similarly, the procedure described below can be readily modified to instead analyze the case where the null hypothesis is *not* rejected at the 5% level and the issue is whether this *failure* to reject is due to a flaw in the exogeneity of one of the $g$ variables in question.

is less than 0.05 using the asymptotic sampling distribution of $\tilde{\gamma}$ given in Equation 7 with the exogeneity flaw covariance vector $(\Sigma_{X\varepsilon})$ set equal to zero.

The key issue is how sensitive this 5% rejection of the null hypothesis is to values of $\Sigma_{X\varepsilon}$ which are non-zero, but "plausible." It is straightforward to re-calculate this rejection $p$-value for alternative values of $\Sigma_{X\varepsilon}$, but difficult to have any intuition as to how large such a covariance is likely. In contrast, one might well have some intuition as to how large plausible values of the components of the concomitant *correlation* vector are likely to be. Thus, the crucial issue in a useful sensitivity analysis is to numerically characterize this rejection $p$-value as a function of the first $m$ of these correlations, which are considered to be possibly non-zero.

Converting the covariance vector $\Sigma_{X\varepsilon}$ into the corresponding correlation vector merely involves dividing each of its components by the square toot of the product of the variance of $\varepsilon$ and the variance of the explanatory variable corresponding to this $\Sigma_{X\varepsilon}$ component. Since the columns of $X = [Y_2 \ W_1]$ are observed, it is straightforward to consistently estimate the variance of the corresponding $m$ explanatory variables $(Y_2)$ considered to be possibly endogenous. The model errors, $\varepsilon$, in contrast, are not observed. But the variance of $\varepsilon$ can still be consistently estimated from the fitting errors implied by any posited value for $\Sigma_{X\varepsilon}$.

The sensitivity analysis consequently proceeds by drawing values of the first $m$ (possibly non-zero) components of the exogeneity flaw covariance vector $(\Sigma_{X\varepsilon})$ at random and, conditional on this vector, calculating both the implied values for the concomitant $m$ correlations (between the components of $Y_2$ and $\varepsilon$) and the implied rejection $p$-value for the null hypothesis. Repeating this random drawing a large number of times – denoted $M_{rep}$ below – one can then numerically invert the relationship to delineate the set of correlations between $Y_2$ and $\varepsilon$ for which the null hypothesis rejection $p$-value is no longer less than 0.05. This set is called the "No Longer Rejecting" (or "NLR") set below.[6]

Interest then centers on how large such a correlation needs to be in order to overturn the inference of interest; this is typically quantified by the length of the shortest $m$-vector from the origin (where all $m$ variables are actually exogenous) to this NLR set – i.e., by the length of the minimal-length ray to the "No Longer Rejecting" set. This minimal-length vector is denoted $r_{min}$ below. Clearly, the length of $r_{min}$ lies between zero and one, with a value close to zero indicating that this particular inference is fragile with respect to failures of the exogeneity assumptions, whereas a value of the length of $r_{min}$ which is large (relative to what one might expect to be the case) indicates that this inference is robust in this regard.[7]

---

[6]Ashley and Parmeter (2015) describes a algorithm for calculating this NLR set in detail, for the more general case where a sensitivity analysis with respect to instrument invalidity is being done in the context of linear GMM/IV regression. This algorithm is implemented in R code (available from the authors) and as a Stata batch file (in preparation), but our approach is easily implemented in any matrix-oriented computer language.

[7]Our code uses the Euclidean norm, the square root of the sum of the squares of the $m$ components of the ray to the NLR, as its length measure. While obviously not the only possible choice, this norm emphasizes the importance of the ray components which are largest in magnitude, which likely contributes to descriptive clarity.

Because one is sometimes able to infer the sign of a component of $\Sigma_{X\varepsilon}$ – as in the case where endogeneity is induced by measurement error – the signs on the components of this minimum-length ray from the origin to the NLR set can be very informative also. And, more generally, the absolute and relative sizes of the $m$ components of $r_{min}$ point at which of the $m$ explanatory variable exogeneity assumptions are most crucial to the validity of the inference result.

A word of discussion on how to interpret $r_{min}$ in applied work. Applied econometrics is inherently a combination of objective analysis and judgement. For example, frequentist inferential analysis produces a rejection $p$-value for a particular null hypothesis, based on an application to the data of a specific statistical model; such a result is fundamentally objective. In contrast, the decision as to how to *interpret* a $p$-value of, for example, 0.04 or 0.06 necessarily involves some judgment on the part of the analyst. Analogously, the sensitivity analysis proposed here produces an objective, replicable result – e.g., a computed value for $r_{min}$ – which can be easily and clearly communicated to others. But again the interpretation of this result requires a degree of judgement on the analyst's part: Where $r_{min}$ is notably less than 0.10 it would seem reasonable to conclude that this particular IV inference is fragile with respect to possible instrument flaws. In contrast, if $r_{min}$ exceeds 0.70, say, then it is very sensible to conclude that this IV inference is robust. And – analogous to a null hypothesis rejection $p$-value of, for example 0.07 – an intermediate $r_{min}$ value is indicating that the sensitivity analysis is not giving a clear result, but that a degree of caution is in order.

## 3. Illustration of the Method

Exogeneity is a prime concern in various applied economic milieus. However, it is an especially crucial issue in the applied economic growth literature. The search for valid instruments in this context has resulted in a cottage industry of papers. This literature was recently taken to task by Bazzi and Clemens (2013) for routinely testing individual growth determinants one at a time, a practice which virtually guarantees the production of distorted inferences.

Relatedly, Durlauf, Johnson and Temple (2005, pg. 118) note:

> "... the belief that it is easy to identify valid instrumental variables in the growth context is deeply mistaken. We regard many applications of instrumental variable procedures in the empirical growth literature to be undermined by the failure to address properly the question of whether these instruments are valid i.e., whether they may be plausibly argued to be uncorrelated with the error term in a growth regression. ... Not enough is currently known about the consequences of "small" departures from validity, but it is certainly possible to envisage circumstances under which ordinary least squares would be preferable to instrumental variables ..."

Clearly, if one could demonstrate that key inferential conclusions were robust to potential endogeneity, then the need for valid instruments would be mitigated.

The sensitivity analysis described in the previous section is applied here to the exogeneity assumptions made by Mankiw, Romer, and Weil (1992) – 'MRW' below – in their seminal work examining the role of human capital accumulation in determining a country's growth rate. The

MRW study is foundational in the empirical growth literature, so an assessment of the validity of its inferential results is still economically relevant. Their model was chosen for illustrative use here, however, because MRW use no instruments at all: their parameter estimates and hypothesis test inferences are all obtained using OLS estimation, based on a set of exogeneity assumptions.[8]

The main point of the MRW paper is to test an 'augmented Solow model' in which the logarithm of per capita output in a country is taken to depend on the rate of human capital formation ('$\ln(SCHOOL)$'), on the rate of investment in physical capital ('$\ln(I/GDP)$'), on the population growth rate ('$n$'), and on the growth rate in the level of technology ('$g$'). MRW aggregate $n$, $g$, and the depreciation rate ('$\delta$') into a single regressor, '$\ln(n + g + \delta)$', in their regression model.[9]

MRW assume that all three of these explanatory variables – $\ln(SCHOOL)$, $\ln(I/GDP)$, and $\ln(n + g + \delta)$ – are exogenous and obtain the OLS estimates:

(8)
$$ln(GDP/L)_i = \underset{(1.17)}{6.84} + \underset{(0.07)}{0.65\,ln(SCHOOL)_i} + \underset{(0.13)}{0.69\,ln(I/GDP)_i} - \underset{(0.41)}{1.75\,\ln(n+g+\delta)_i} + \varepsilon_i,$$

with $\overline{R}^2 = 0.779$ and $s^2 = 0.258$ using the data on the 98 countries in their full ('Non-oil') sample.

MRW's essential point is that the null hypothesis $H_o : \beta_{school} = 0$ can be easily rejected, implying that the standard Solow model should be augmented so as to account for changes in human capital. They also test a second null hypothesis, $H_o : \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 1$, which corresponds to constant returns to scale technology. The $p$-value for this linear restriction is 0.388, so this null hypothesis cannot be rejected using their model.[10]

Because Equation 8 is estimated using OLS, however, these inference results are potentially invalid unless all three explanatory variables in this equation are actually exogenous. This set of exogeneity assumptions is not directly testable because $\varepsilon_i$ is not observable; but the sensitivity of these two MRW inference results to potential flaws in these exogeneity assumptions can be easily assessed using the sensitivity analysis procedure proposed here. More specifically, this analysis quantifies how large the correlations between $[\ln(SCHOOL)_i, \ln(I/GDP)_i, \ln(n + g + \delta)_i]$ and the model error ($\varepsilon_i$) would need to be in order to overturn either of these two MRW inference results at the 5% level.

---

[8]As a referee rightly notes, if the model is misspecified, e.g., by failing to capture nonlinearities that exist between growth and capital (either human or physical), then the sensitivity analysis results will be distorted by the resulting inconsistency in the parameter estimation for the original model. That is, we need to assume that the model is correctly specified in order for a sensitivity analysis based on it to work properly. Omitted variables are less of a concern because in this case it is precisely their omission that potentially would invalidate the OLS exogeneity assumption and could lead to invalid inference.

[9]The variable '$\ln(SCHOOL)$' is the average percentage of the working-age population in secondary school during the period 1960-1985; thus, MRW are ignoring other investments in human capital – e.g., health. The difficulties in measuring investment in human capital are described in their Section IIB. Also, the sampling variation in $\ln(n+g+\delta)$ is entirely due to variation in population growth rates, as $g + \delta$ is set to a constant value of 0.05 in the MRW study.

[10]In a natural notation, $\beta_{school}$ denotes the coefficient on $ln(SCHOOL)_i$, etc. MRW obtain other results in addition, but these are their most important inferences.

Table 1 presents the sensitivity analysis results, using $M_{rep} = 25,000$ repetitions, with respect to the null hypothesis $H_o: \beta_{school} = 0$ for four different scenarios: first allowing for possible exogeneity flaws in each of the three explanatory variables separately, and finally allowing for possible flaws in both of the two explanatory variables, $\ln(SCHOOL)_i$ and $\ln(I/GDP)_i$, whose potential endogeneity appears to matter.[11]

The component (or components) of $r_{min}$, the vector corresponding to a ray from the origin to the closest point in the 'No Longer Rejecting' set for $H_o: \beta_{school} = 0.0$ are given in the first row of each column of Table 1.

TABLE 1. Sensitivity Analysis Results on $H_o: \beta_{school} = 0.0$ in the Mankiw, Romer and Weil (1992) Model, Equation 8).

| Variable | $\ln(n+g+\delta)$ | $\ln I/GDP$ | $\ln SCHOOL$ | $\ln SCHOOL$ & $\ln I/GDP$ |
|---|---|---|---|---|
| $r_{min}$ vector | 1.000 | -0.536 | 0.425 | $(0.188, -0.298)$ |
| $r_{min}$ length | 1.000 | 0.536 | 0.425 | 0.353 |

Interpreting these results:

(1) The values calculated for the length of $r_{min}$ in Table 1 range from 0.353 all the way up to 1.000; evidently, quite substantial correlations between MRW's explanatory variables and the model error term are necessary in order to reverse their rejection of $H_o: \beta_{school} = 0$ at the 5% level. We conclude that this MRW inference with regard to the human capital variable – $\ln(SCHOOL)_i$ – is quite robust to flaws in the exogeneity assumptions with regard to all three explanatory variables: $\ln(SCHOOL)_i$, $\ln(I/GDP)_i$, and $\ln(n+g+\delta)_i$.

(2) The result that $r_{min}$ equals one for the sensitivity analysis allowing for correlation between $\ln(n+g+\delta)_i$ and the model errors, $\varepsilon_i$, indicates that this null hypothesis is still rejected at the 5% level for any possible value for this correlation. Thus, this particular MRW inference result is completely robust with respect to possible non-exogeneity in $\ln(n+g+\delta)_i$.

(3) For the $m = 2$ calculation, analyzing the sensitivity of the $H_o: \beta_{school} = 0$ inference to possible non-exogeneity in both $\ln(SCHOOL)_i$ and $\ln(I/GDP)_i$ simultaneously, the magnitudes of the two components (0.188 and -0.298) corresponding to a ray from the origin to the 'No Longer Rejecting' set are roughly equal. This result suggests that the inference is equally robust to non-exogeneity in either of these explanatory variables.

(4) Portela, Alessie and Teulings (2010) argue persuasively that the measurement errors with regard to human capital in data sources based on enrollment rates (e.g. Barro and Lee; 2010) are likely to be negatively correlated with the actual value of human capital. MRW use an amalgam of secondary school enrollment rates and the fraction of the population that is of secondary school age to construct their human capital measure, $SCHOOL_i$. It is thus not unreasonable to assume that the (unobservable) correlation between $SCHOOL_i$ and $\varepsilon_i$

---

[11]The $m = 3$ case can be analyzed also, but it is redundant to report these results, in view of the fact that this MRW inference result is so highly robust with respect to exogeneity flaws in the $\ln(n+g+\delta)_i$ variable.

is not positive. In contrast, our finding that $r_{min}$ must exceed 0.425 (where the sensitivity analysis is on non-exogeneity in $SCHOOL_i$ alone) indicates that it is only strongly positive values of the correlation between $SCHOOL_i$ and $\varepsilon_i$ which overturn the MRW rejection of $H_o: \beta_{school} = 0$. This observation further strengthens our conclusion that this MRW inference is robust to likely non-exogeneity in $SCHOOL_i$.

As noted above, MRW also test the null hypothesis $H_o: \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 1$; this null hypothesis corresponds to constant-returns-to-scale technology and cannot be rejected at the 5% level in their OLS regression. The robustness (or fragility) of this result is of intrinsic economic interest, but it also provides an opportunity to illustrate two additional generic aspects of the sensitivity analysis procedure proposed here: First, that it is just as easy to do the sensitivity analysis with respect to a more complex null hypothesis, and second, that one can also analyze the sensitivity of a 'failure' to reject a null hypothesis.

Table 2 displays the corresponding results with regard to the robustness of this second MRW inference result to flaws in their exogeneity assumptions, again using $M_{rep} = 25,000$ repetitions.

TABLE 2. Sensitivity Analysis Results on $H_o: \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 1$ in the Mankiw, Romer and Weil (1992) Model, Equation 8.

| Variable | $\ln(n + g + \delta)$ | $\ln I/GDP$ | $\ln SCHOOL$ | $\ln(n + g + \delta)$ & $\ln I/GDP$ |
|---|---|---|---|---|
| $r_{min}$ vector | -0.109 | -0.241 | 1.000 | $(-0.236, -0.029)$ |
| $r_{min}$ length | 0.109 | 0.241 | 1.000 | 0.239 |

Note that the MRW failure to reject $H_o: \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 1$ at the 5% level is extremely robust with regard to possible endogeneity in $ln(SCHOOL)_i$, but not quite so robust with regard to possible endogeneity in $\ln(n + g + \delta)_i$ or $ln(I/GDP)_i$ when each is considered separately. Notably, this inference is quite fragile with respect to possible endogeneity in $ln(I/GDP)_i$ when the sensitivity of the $p$-value for this inference with respect to both $\ln(n + g + \delta)_i$ and $ln(I/GDP)_i$ is quantified simultaneously.

## 4. CONCLUDING REMARKS

The primary feature which distinguishes econometrics from statistics is its willingness to confront endogeneity in regression models: where one or more of the explanatory variables in the model are correlated with the model error term, $\varepsilon$. Typical sources of such endogeneity are simultaneity (where an explanatory variable is jointly determined with the dependent variable, commonly because both depend on an omitted explanatory variable) and measurement error (where the observed values of an explanatory variable are corrupted to some degree by noise, which noise then appears also in the error term). Since simultaneity and measurement error are endemic problems in economic regression models – leading to inconsistent parameter estimates – dealing with endogeneity is a central concern for empirical economists.

This paper proposes a sensitivity analysis for flaws in the exogeneity assumptions of a linear structural model. Where some inferences are found to be fragile with respect to one or more of

these exogeneity assumptions, the results of this sensitivity analysis can be used to either temper the discussion of this subset of one's inferential conclusions and/or motivate a search for valid instruments. Where other - or perhaps all - of one's inference results are found to be robust with respect to violations of the exogeneity assumptions, this sensitivity analysis can be used to preclude a search for (potentially problematic) instruments altogether.

## References

[1] Acemoglu, D., S. Johnson, J. Robinson, 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5) pp. 1369-1401.

[2] Angrist, J. and A. Krueger, 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106(4) pp. 979-1014.

[3] Angrist, J. and J.-S. Pischke, 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24(2) pp. 3-30.

[4] Ashley, R., and C. F. Parmeter 2015. "Sensitivity Analysis for Inference in 2SLS Estimation with Possibly-Flawed Instruments," *Empirical Economics* forthcoming. Currently available at `http://ashleymac.econ.vt.edu/Ashley_Parmeter_Credible_GMM.pdf`.

[5] Bazzi, S. and M. Clemens, 2013. "Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth," *American Economic Journal: Macroeconomics*, 5(2) 152-186.

[6] Dufour, J.-M., 2003. "Identification, Weak Instruments and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36(4) pp. 767-808.

[7] Durlauf, S., P. Johnson, and J. Temple, 2005. Growth Econometrics, in P. Aghion and S. Durlauf, eds. *Handbook of Economic Growth*, Vol. 1, Elsevier: USA.

[8] Keane, M., 2010. "A Structural Perspective on the Experimentalist School," *Journal of Economic Perspectives*, 24(2) pp. 47-58.

[9] Leamer, E. E. 2010. "Tantalus on the Road to Asymptopia," *Journal of Economic Perspectives*, 24(2) pp. 31-46.

[10] Mankiw, N. G., Romer, D. and D. N. Weil, 1992. "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107(2) pp. 407-437.

[11] Murray, M. P., 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives*, 20(4) pp. 111-132.

[12] Portela, M. R. Alessie and C. Teulings, 2010. "Measurement Error in Education and Growth Regressions," *The Scandinavian Journal of Economics*, 112(3) pp. 618-638.

[13] Sims, C. A., 2010. "But Economics Is Not an Experimental Science," *Journal of Economic Perspectives*, 24(2) pp. 59-68.

[14] Stock, J. H., 2010. "The Other Transformation in Econometric Practice: Robust Tools for Inference," *Journal of Economic Perspectives*, 24(2) pp. 83-94.