Institutions, Trade, and Growth: Revisiting the Evidence

David Dollar and Aart Kraay
The World Bank

January 2003

Abstract: Several recent papers have attempted to identify the partial effects of trade integration and institutional quality on long-run growth, using the geographical determinants of trade and the historical determinants of institutions as instruments. In this paper we show that many of the specifications in these papers are weakly identified despite the apparently good performance of the instruments in first-stage regressions. Consequently, we argue that the cross-country variation in institutions, trade, and their geographical and historical determinants is not very informative about the partial effects of these variables on long-run growth.

JEL Classification Codes: F10, F43, O11, O17

1818 H Street NW, Washington, DC 20433, **ddollar@worldbank.org**. **akraay@worldbank.org**. We would like to thank Romain Wacziarg and Jeff Sachs for helpful comments The opinions expressed here are the authors' and do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

Two recent papers have proposed novel instruments and have used them to separately identify strong causal effects of trade and institutions on economic development. Frankel and Romer (1999, FR) show that the geographically determined component of trade as a fraction of GDP exerts a strong positive effect on growth in the very long run, while Acemoglu, Johnson and Robinson (2001, AJR) show that the historically determined component of current institutional quality exerts strong effects on development in a similar framework.

An open question, however, concerns the partial effects of institutions and trade on growth in the long run. Several very recent papers have investigated this question, but come to strikingly different conclusions. Alcálá and Ciccone (2002, AC) estimate a variety of regressions of GDP per worker on trade as a share of GDP and measures of institutional quality. They instrument for these two variables using the FR instrument as well as linguistic origins, and find that trade is quite significant, while institutional quality is occasionally but not consistently significant. Dollar and Kraay (2002, DK) estimate a specification very similar to that of AC. In contrast with AC, DK argue informally that the cross-country variation in the data is uninformative about the relative contribution of institutions and trade Finally, Rodrik, Subramanian and Trebbi (2002, RST) estimate a variety specifications similar to those of AC and DK in an effort to sort out the geographical, institutional, and trade-related determinants of development. In contrast with AC and DK, RST conclude that institutions matter while trade does not.[1]

The objective of this paper is to show that empirical specifications which seek to isolate the partial effects of institutions and trade can suffer from serious identification problems. The issue, which we discussed informally in our previous paper, is straightforward. In many of the specifications in these papers, the historical and geographical instruments proposed in the previous literature tend to have good

[1] In a related paper, Easterly and Levine (2002) attempt to disentangle the geographical and institutional determinants of development. In a number of specifications they control for various measures of policy, including trade policy. They generally find that institutions dominate geography and trade. Since their paper considers trade policy rather than trade itself, we do not consider it further here.

explanatory power for *both* endogenous variables, institutions and trade. Consequently, the fitted values of institutions and trade from first-stage regressions are highly correlated with each other, and these specifications are at best weakly identified, even though first-stage regressions have apparently good explanatory power. In short, although the historical and geographical instruments for institutions and trade proposed by AJR and FR perform well in their respective univariate applications, they are not strong enough to identify the partial effects of these variables.

We document the identification problem using the partial R-squared diagnostic measures suggested by Shea (1997) , and the minimum eigenvalue test of the null hypothesis of weak instruments suggested by Stock and Yogo (2001). In most cases, both diagnostic statistics as well as the formal tests verify the informal observation in DK that these models are not very well identified. Consequently, inferences about the structural parameters of interest, i.e. the causal effects of institutions and trade on long-run growth, cannot reliably be based on conventional t-statistics.[2] We therefore compute confidence intervals for these structural coefficients of interest using three methods suggested in the literature on inference with weak instruments: the AR statistic proposed by Anderson and Rubin (1949), the conditional likelihood ratio statistic proposed by Moreira (2002), and the S-statistic proposed by Startz, Zivot and Nelson (2001). These methods which take into account the low strength of the instruments produce substantially larger confidence intervals than those based on standard asymptotic theory in the specifications we consider. In many cases, confidence intervals for the parameters of interest cover (nearly) the entire real line. These results serve to formalize the informal intuition in our previous paper that the cross-country variation in trade, institutions, and their historical and geographical determinants is not very informative about their relative importance for growth in the long run.

Finally, in our previous paper we also investigated the relative contributions of institutions and trade over shorter horizons by looking at changes in decadal growth rates over the past four decades. We apply the same weak instrument diagnostics to these specifications. Although we find that dynamic regressions controlling only for trade and initial income are well-identified, the regressions which add institutional quality

[2] See Staiger and Stock (1997) for a seminal paper, and Stock, Wright, and Yogo (2002) for a recent survey of weak instruments.

are poorly identified when institutional quality is taken as endogenous, but are reasonably well-identified when institutional quality is treated as exogenous.

The empirical examples that we discuss in this paper are specific to the recent debate on the relative importance of institutions and trade for long-run growth. We think however that our results have broader implications for the cross-country empirical growth literature, which has only in recent years begun to take seriously the endogeneity of many of the key determinants of growth. Our results here highlight the more general difficulty in finding exogenous and independent sources of variation required to identify the partial effects of such determinants of growth.

The remainder of this paper proceeds as follows. In Section 2 we review some of the main empirical specifications that have been advanced to show the effects of institutions and trade on long-run growth. In Section 3 we describe and implement diagnostic tests for the presence of weak instruments, and show that most of the empirical specifications we consider in Section 2 are in fact weakly identified. In Section 4 we describe and implement three approaches to inference with weak instruments, and show that existing estimates of the partial effects of institutions and trade are much less precisely estimated than conventional t-statistics would suggest. In Section 5 we examine the strength of the identification of the decadal regressions from our previous paper. Section 6 concludes.

2. Empirical Background

All of the papers discussed above estimate variants on the following linear instrumental variables regression:

(1) $$\ln(\text{GDP Per Capita})_j = \beta_0 + \beta_1 \cdot \text{Institutions}_j + \beta_2 \cdot \text{Trade}_j + \beta_3{}'\text{Controls}_j + u_j$$

Cross-country differences in GDP per capita today primarily reflect differences in growth over the very long run since a date in the distant past when initial incomes were not very different across counties. Consequently, this specification can be thought of as capturing the determinants of growth in the very long run. Institutions and trade are both treated as endogenous, and are instrumented using their deep historical and geographical determinants. Each of the papers discussed above differs somewhat in the measures of institutions and trade used, the control variables included in the regression, and the choice of instruments. Table 1 summarizes the key differences across the main specifications of interest in each of these papers.

The top part of Tables 2 and 3 report some of the main results of these papers. To conserve on space, we report only the coefficients on the main endogenous variables of interest: institutions, trade, and their instruments in the first-stage regressions. Table 2 reports selected results from DK, while Table 3 reports selected results from RST. We do not report results from AC as we were unable to obtain their dataset at the time of writing. As a benchmark, the first two columns of Table 2 present regressions from DK in the spirit of the original AJR and FR papers, estimating the effects of institutions and trade separately, without controlling for the other variable. Specification 2.1 shows the effects of institutions only, analogous to the basic specification of AJR. The main differences with AJR is that DK use linguistic origins rather than settler mortality to instrument for institutional quality, and use a different measure of institutional quality. Nevertheless, the results are broadly similar to that of AJR. As the OLS regressions show, institutional quality is strongly positively correlated with per capita incomes. The IV regressions show a statistically and economically significant effect of institutions on per capita income, with a slope coefficient larger than the corresponding OLS estimate. Specification 2.2 shows the analog of the main FR specification. Here there are again only two differences with the original paper: the trade/GDP ratio is measured at PPP, and the only included exogenous variable is the logarithm of population. Again the results are very similar to those of the original paper, with a large and significant effect of trade on growth. In both specifications 2.1 and 2.2, the first-stage regressions of institutions and trade on their respective instruments perform quite well, with comfortably large F-statistics of 13.3 and 28.8.

Specification 2.3 reports the main DK result on the partial effects of institutions and trade. Both institutions and trade enter positively in the OLS regression, with t-statistics greater than 2. However, in the IV regression only institutional quality has a t-statistic greater than 2, while the coefficient on trade falls to near zero and the t-statistic falls to 0.3. In our other paper, we showed however that this apparent result on the importance of institutions relative to trade was quite fragile, and that both variables were insignificant once four influential observations were discarded. Here however we focus on this main specification, and show more formally below that the apparent significance of institutional quality vanishes once we take weak identification into account.

Table 3 presents the baseline RST specification, and three variants. RST regress log per capita GDP on distance from the equator, the same institutional quality variable as in DK, and trade as a fraction of GDP in local currency units.[3] In their preferred 80-country specification, they use settler mortality from AJR and distance from FR as instruments. The first column of Table 3 replicates their basic specification. To conserve on space, we again suppress the coefficient on distance from the equator (and all other included exogenous variables). The IV estimate of the effect of institutions is positive, large, and highly significant using a conventional t-test. In contrast, trade enters negatively and insignificantly.

RST go on to provide a number of robustness checks on this basic variant. A subset of these are reported in the remaining columns of Table 3. In specification 3.2, they replace the local currency trade share with the PPP trade share in order to make their results more comparable with DK and AC. In specification 3.3 they return to local currency trade shares, but add measures of domestic market size (the logarithms of area and population) in order to make their results more comparable with the original FR paper which included these variables. In specification 3.4 we report one of several specifications produced by RST including a variety of other variables capturing different channels through which geography might matter for development. We focus on a cluster of three variables they identify as jointly significant, consisting of land area in the tropics, number of frost-days per year, and the incidence of malaria. In each of these robustness checks, they find that institutions enter the IV specification positively and with

[3] In their paper, RST argue that this is preferable to using PPP-adjusted trade shares as do AC and DK. The sketch a simple model in which an increase in productivity which raises the relative price of non-tradeables relative to the benchmark country will also raise the PPP-adjusted trade ratio. This is because the PPP-adjusted GDP will fall to reflect the higher price of non-tradeables. Since the increase in productivity raises income, this can introduce a spurious correlation between income and the PPP-trade share. While the choice of denominator for the trade ratio is not the focus of this paper, we note in passing that we are not entirely convinced by this argument. First, it as long as these other sources of productivity differences are uncorrelated with the FR instrument, this potential bias will not matter for the IV coefficient estimates. Second, at a more intuitive level, if we want to capture the size of the traded sector relative to overall economic activity, it seems natural to use an internationally-comparable measure of overall economic activity in constructing trade shares.

t-statistics in excess of two. In contrast, trade enters negatively in three of the four specifications, and in all cases the t-statistics are well below 2. Based on these findings, as well as the basic DK finding in specification 2.3, they conclude in the title of their paper that "institutions rule". Below we aregue that these results could be interpreted more cautiously once we take the strength of the identification into account.

3. Are the Partial Effects of Institutions and Trade Identified?

The key question in this paper is the extent to which the empirical strategy of the papers discussed in the previous section is a good one for capturing the partial effects of institutions and trade on growth in the very long run. Our focus here is on one shortcoming of this approach -- that existing instruments at best only weakly identify the structural parameters of interest. In order to review the tools on which we base this conclusion, some further notation is useful. The structural model in Equation (1) can be written in the usual matrix notation as

(2) $$y = X\beta + u$$

where y is an nx1 vector of observations on the dependent variable, X is an nxk matrix of endogenous regressors, β is a kx1 vector of structural coefficients, and u is an nx1 vector of disturbances.[4] Let Z represent an nxq matrix of instruments. Then the reduced form consists of the following k+1 regressions of y and the columns of X on Z:

(3) $$\begin{aligned} y &= Z\theta + v \\ X &= Z\Gamma + e \end{aligned}$$

where θ is an qx1 vector and Γ is a qxk matrix of reduced form coefficients, and v and e are nx1 and nxk matrices of reduced-form disturbances. Let σ and Σ denote the variance of v_j and the covariance matrix of e_j respectively, and let Ω denote the covariance matrix of $(v_i, e_i')'$.

[4] In what follows we suppress the additional exogenous control variables in the regression, by interpreting y, X, and the instruments as the residuals from a regression of these variables on the included exogenous control variables.

In order for the historical and geographical variables in Z to be valid instruments, they need to be independent of u, which we assume to be the case, i.e. E[Z'u]=0. In order for the structural parameters β to be identified, the two textbook order and rank conditions must also be satisfied, i.e. the number of instruments is greater than the number of endogenous variables (q>k), and the matrix Γ must be full rank (rank(Γ)=q). Using this notation, the instrumental variables (IV) estimator of β is

$\hat{\beta}_{IV} = \left(X'P_Z X\right)^{-1} X'P_Z y$ where $P_Z \equiv Z\left(Z'Z\right)^{-1} Z$ for any matrix Z.

A rapidly-growing recent literature on weak instruments has pointed out that the validity of the usual approximation that $\hat{\beta}_{IV}$ is normally distributed in finite samples depends crucially on the strength of the instruments. In the case of only one endogenous variable, the instruments are weak if they jointly have poor explanatory power for the endogenous variable in the reduced-form equation for X. Nelson and Startz (1990), Staiger and Stock (1997) show that when the instruments are weak in this sense, the IV estimator is biased towards the mean of the OLS estimator, and is not normally distributed. Thus inferences about the structural parameters based on usual t-tests can be highly misleading.

When there are multiple endogenous variables, and so more than one first-stage regression, it is less obvious how to gauge the strength of the instruments and the quality of the normal approximation. An important point, however, is that it is not sufficient for the F-statistics in the individual first-stage regressions to be large.[5]

Shea (1997) has proposed a simple diagnostic statistic for determining the strength of instruments when there are multiple endogenous variables. If X_1 and X_2 are two endogenous variables, then the instruments are strong if the component of the fitted value of X_1 from the first-stage regressions that is orthogonal to the fitted value of X_2 is highly correlated with the component of X_1 that is orthogonal to X_2 . In the case of a single endogenous variable, this criterion is simply the R-squared from the first-stage

[5] Shea (1997) gives a simple example. In the case of two instruments Z_1 and Z_2 for the two endogenous variables, X_1 and X_2, if Z_1 is is a good predictor of X_1 and of X_2, but Z_2 is uncorrelated with both endogenous variables, then the first-stage F-statistics will be large but the model will not be identified since there is in effect only one instrument for the two endogenous variables.

regression. While this statistic is intuitive and easily calculated, results on its distribution are not provided in Shea (1997) and so this statistic does not lend itself well to formal tests of the null hypothesis of weak instruments.

More generally, Stock, Wright and Yogo (2002) point out that the validity of the normal approximation depends on the concentration matrix $\mu = \Sigma^{-\frac{1}{2}\prime}\Gamma' Z' Z\Gamma\Sigma^{-\frac{1}{2}}$ being large in the sense of the smallest eigenvalue of this matrix being large. Note that the concentration matrix can be small for a variety of reasons. All instruments could be nearly uncorrelated with the endogenous variables, i.e. all the elements of Γ are small relative to the variances of the error terms in the first-stage regressions. The instruments could also be very highly correlated with each other so that Z'Z is not full rank, i.e. there are essentially fewer instruments than endogenous variables. Finally concentration matrix could also be small if the rank condition for identification fails, i.e. Γ is not full rank. This can occur if, as is the case in some of the applications considered here, each of the instruments has strong explanatory power for both endogenous variables.

Stock and Yogo (2001) develop a test of the null hypothesis that the instruments are weak based on this idea. In particular, they provide critical values for a test of the null hypothesis of weak instruments based on the minimum eigenvalue of $\Gamma = \hat{\Sigma}^{-1/2\prime} X' P_Z X \hat{\Sigma}^{-1/2}$, where $\hat{\Sigma} = X'(I - P_Z)X/(n-q)$ is a consistent estimate of the covariance matrix of errors from the first-stage regressions.[6] In the version of the test we implement below, the researcher specifies a null hypothesis that the instruments are weak if the maximum size distortion of Wald tests about parameters of interest under the null of weak instruments is equal to some pre-specified value. Stock and Yogo (2001) then tabulate the critical values of the minimum eigenvalue of G for which the null of weak instruments should be rejected.

[6] When the instruments are weak, they show that $E[G] = \mu^2 / q + I$. One drawback of this test in our current application is that it is derived under the assumption of iid homoskedastic error terms in the structural equation. The same is true for the approaches to inference discussed below. We are unaware of tests for weak instruments which relax this assumption, and it is difficult to say how important deviations from this assumption are for the results we obtain.

With these tools in hand we can investigate the strength of the instruments in the specifications in Tables 2 and 3. Consider first the core DK specification in regression 2.3. In the first-stage regressions, we see that the linguistic origin variables and the FR instrument are highly significant predictors of both endogenous variables. In our previous paper we noted that this implied a very high correlation between the fitted values of institutions and trade in the second-stage regression, which made the IV estimates highly sensitive to a few influential observations. Table 2 reports that this correlation of fitted values from the first-stage regression is 0.74, indicating strong collinearity in the second-stage regression. From this we concluded informally that the cross-sectional variation in institutions, trade and incomes was not very informative about the partial effects of the former two on the latter.

The weak instrument diagnostics discussed in the previous subjection help us to make this point more precisely. Consider first the Shea (1997) partial R-squared statistics. We find in our specification that these statistics are very small at only 0.02 and 0.04 respectively. In contrast, the R-squareds from the first-stage regressions in the first two univariate specifications in Table 2 were 0.14 and 0.29. This provides a first indication of the idea that while linguistic origin and distance are reasonable instruments when the effects of institutions and trade are estimated separately, they are quite poor instruments when the effects of the two are considered jointly.

The middle panel of Table 2 also reports the eigenvalues of the matrix of estimated coefficients in the first-stage regression, $\hat{\Gamma}$, as well as the eigenvalues of G. The first observation here is that the ratio of the largest to the smallest eigenvalues of $\hat{\Gamma}$, i.e. the condition number of $\hat{\Gamma}$, is nearly 100 and the smallest eigenvalue is small at 0.025. This indicates that the matrix of estimated coefficients in the first-stage regression is very close to not being of full rank, contributing to weak identification of the parameters of interest. The smallest eigenvalue of G is 1.36. To interpret this value, suppose we were to define the set of weak instruments very narrowly as only those instruments that would lead to size distortions of 25% or more (i.e. the size of a nominal 5% test about a parameter of interest were actually 25%), we would not reject the null hypothesis that the instruments in this application are weak according to this stringent criterion, since the critical value provided by Stock and Yogo is 5.75.

The middle panel of Table 3 also reports the same diagnostic statistics for selected specifications from RST. In all of these specifications, the first-stage regressions perform well: seven of the eight first-stage F-statistics are larger than 10, and even the smallest is still respectable at 8. However, these F-statistics are not sufficient to diagnose the strength of the instruments. Closer inspection of the first-stage regressions shows the same tell-tale signs of potential problems as in the DK specification in Table 2. In seven of the eight first-stage regressions in Table 3, the instrument for institutions (settler mortality) has strong explanatory power for both institutions and trade, while the instrument for trade (the FR constructed trade share) is a significant predictor of both institutions and trade at the 5% level in seven of the eight specifications as well.

Table 3 also presents the same diagnostic statistics on instrument quality as in Table 2. The Shea partial R-squared statistics are small in most cases, although not as extremely so as in the basic DK specification. In four of the eight first-stage regressions, the partial R-squareds are less than 0.1, and in seven out of eight they are less than 0.2. More formally, in each of the four specifications, we cannot reject the null hypothesis that the instruments are weak. However, the size of the set of weak instruments under the null is different in the four specifications. In specifications 2 and 4, the minimum eigenvalues of G are 1.95 and 3.24, and we cannot reject even the most conservative set of weak instruments tabulated in Stock and Yogo (2001). In specifications 1 and 3, the minimum eigenvalues of G are 6.2 and 6.8, and we cannot reject the null that the instruments fall in the larger set of weak instruments that would lead to size distortions of at least 5 percent, i.e. a test with nominal size of 5% would in fact have a true size of 10%.

These results suggest that there are reasons to question inferences about the importance of trade and institutions based on conventional t-statistics in the main specifications considered by RST as well. In the next section of the paper we discuss approaches to inference that are valid even when the instruments are weak, and show that the confidence intervals for the parameters of interest are much larger than simple t-statistics would suggest.

4 How Large are the Partial Effects of Institutions and Trade?

In this section we briefly describe three approaches to constructing valid confidence intervals when instruments are weak. The first is based on the Anderson-Rubin statistic:

$$\text{(4)} \qquad AR(\beta_0) = \frac{(y - X\beta_0)' P_Z (y - X\beta_0)/q}{(y - X\beta_0)'(I - P_Z)(y - X\beta_0)/(n-q)}$$

which was first proposed by Anderson and Rubin (1949). Under very general conditions, including for any arbitrary concentration matrix μ, Staiger and Stock (1997) show that q times the AR statistic converges in distribution to a $\chi^2(q)$ random variable. Since the asymptotic distribution of the AR statistic does not depend on any nuisance parameters even when the instruments are arbitrarily weak, it is pivotal and can be used for inference. Wang and Zivot (1998) suggest the use of this result to form valid $(1-\alpha)\cdot 100\%$ confidence sets for the elements of β by "inverting" this statistic, i.e. by finding the set of values of β_0 such that $AR(\beta_0) \le \frac{\chi^2_\alpha(q)}{q}$ where $\chi^2_\alpha(q)$ denotes the $(1-\alpha)\cdot 100$ percentile of a $\chi^2(q)$ distribution. In the case of two endogenous variables, the joint confidence sets for the two elements of β can take a variety of shapes. Confidence intervals for the individual elements of β can be obtained by projecting these joint confidence sets on the appropriate axes.[7]

Figure 1 shows an example of two such 95% confidence sets, and the associated confidence intervals. The top panel shows the case where the 95% confidence set is the area inside an ellipse, corresponding to the coefficients on institutions and trade in RST specification 3.1. The confidence intervals for the coefficients on trade and institutions are obtained by projecting the height and the width of the ellipse on the corresponding axes. The bottom panel shows a case where the 95% confidence set is the area inside a hyperbola, corresponding to the DK specification

[7] Kleibergen (2002) develops an alternative test which corrects the deficiency of the AR statistic which is that it has low power when the number of instruments is large. In our applications with at most three instruments this is not a concern, and inferences based on Kleibergen's K-statistic are almost identical to those based on the AR statistic.

2.3. Projecting this on the two axes shows that the 95% confidence intervals for the coefficients on institutions and trade are the entire real line. That is, the data is so uninformative due to weak identification that it is consistent with any value of these parameters. We will discuss the specific cases to which these confidence sets apply below when we present the empirical results in more detail.

Moreira (2002) has proposed a conditional likelihood ratio statistic as the basis for inference when instruments are weak. Moreira observes that it is possible to write the usual likelihood ratio statistic for the linear IV model as a function of two orthogonal sufficient statistics: $S = Z'Yb_0$ and $T = Z'Y\Omega^{-1}A_0$ where $Y \equiv (y, X)$, $b_0 \equiv (1, -\beta_0')'$, and $A_0 = (\beta_0, I_k)$. His key observation is that since S is a function only of data and the parameters of interest, the likelihood ratio statistic conditional on the value of T observed in the data can be used to make inferences about the parameters of interest. In particular, Moreira shows that the likelihood ratio statistic can be written as:

$$(5) \qquad LR = \bar{S}'\bar{S} - \bar{\lambda}_{min}$$

where $\bar{S} \equiv (Z'Z)^{-1}S(b_0'\Omega b_0)^{-1/2}$ and $\bar{T} \equiv (Z'Z)^{-1}T(A_0'\Omega^{-1}A_0)^{-1/2}$ are the standardized versions of S and T, and $\bar{\lambda}_{min}$ denotes the smallest eigenvalue of $(\bar{S},\bar{T})'(\bar{S},\bar{T})$. Confidence intervals for β can be constructed by inverting this statistic in the same way as the AR statistic. The LR statistic does not have a known distribution on which to base critical values. However, Moreira (2002) suggests a strategy for simulating this distribution empirically to obtain the necessary critical values, which we adopt here.[8]

The final approach to inference with weak instruments is based on the S-statistic proposed by Startz, Nelson and Zivot (2001). This statistic is defined as:

$$(6) \qquad S^2 = \Psi_i(\hat{\lambda})^2 / \left(\frac{\partial \Psi_i(\hat{\lambda})}{\partial \hat{\lambda}}' V[\hat{\lambda}] \frac{\partial \Psi_i(\hat{\lambda})}{\partial \hat{\lambda}} \right)$$

---

[8] In particular, since $\bar{S}$ follows a standard normal distribution, the distribution of LR conditional on the value of $\bar{T}$ observed in the data can be simulated by repeatedly taking draws from the distribution of $\bar{S}$ and evaluating LR at the observed value of $\bar{T}$.

where $\hat{\lambda}$ is a vector containing all of the reduced-form coefficient estimates, $\Psi_i(\hat{\lambda}) = \frac{\beta_{i,IV} - \beta_{i0}}{\sqrt{\left(\left(\hat{\Gamma}' Z' Z \hat{\Gamma}\right)^{-1}\right)_{ii}}}$; $\beta_{i,IV}$ is the IV estimate of the i^{th} element of β; and $\beta_{i,0}$ is its value under the null hypothesis. The S-statistic is the square of the standardized value of $\Psi_i(\hat{\lambda})$, and rejects the null when it is large. To provide an intuition for this statistic, Startz, Nelson and Zivot (2001) note that the numerator of $\Psi_i(\hat{\lambda})$ is the distance between the IV estimate and its hypothesized value. The denominator is the middle part of the concentration matrix, and is small when identification is weak. Thus the S-statistic will reject the null if the coefficient estimate is far from the null, or if identification is weak.

Startz, Nelson and Zivot (2001) show that the S statistic converges to a $\chi^2(1)$ random variable when the instruments are good, and is bounded above by a $\chi^2(q-k+1)$ random variable when the instruments are weak. To construct confidence intervals for the ith element of β, we need only find the set of hypothesized null values for which the S-statistic is less than the appropriate critical value. Figure 2 illustrates one such confidence set, for the coefficient on institutions in the RST specification 2.1. Note that the confidence set need not be the usual closed interval, but can consist of two disconnected intervals. A further convenient property of this approach to inference is that it is possible to construct confidence intervals for a single parameter directly, and that projection methods needed for the AR and LR statistic are not required in this case.

The bottom panels of Tables 2 and 3 present confidence intervals constructed using these three approaches to inference that are robust to weak instruments. For reference purposes, we first report 95% confidence intervals for the coefficients on institutions and trade based on the usual Wald statistics, i.e. the estimated coefficient plus or minus two estimated standard errors. In the remaining rows we report the 95% confidence intervals based on the AR, LR, and S-statistics, respectively.

Consider first the main DK specification in the third column of Table 1. Conventional confidence intervals for the coefficients on institutions and trade are (0.57, 1.95) and (-1.00, 1.35) respectively, suggesting a positive and significant effect of institutions and an insignificant effect of trade. Confidence intervals that are robust to weak instruments lead to quite different conclusions. All three statistics give confidence intervals that consist of the entire real line for both the coefficients on institutions and on trade. This finding formalizes the informal observation in our previous paper that the cross-country variation in institutions, trade, and their geographical and historical determinants is not sufficiently informative to pin down the partial effects on growth of either of these endogenous variables.

The RST results are less stark, but still show that inferences based on the usual t-statistics can be misleading. In the basic RST specification 3.1, the 95% confidence intervals for institutions based on the AR and LR statistics are much wider than those based on the usual Wald statistic, but still contain only positive values.The confidence interval for institutions based on the S-statistic consists of two disjoint intervals, $(-\infty, -1.77)$ and $(1.67, \infty)$. The former two tests suggest that the partial effect of institutions is positive, but much less tightly estimated than a conventional t-test would suggest. The third statistic indicates that the coefficient on institutions is in fact very imprecisely estimated, in that the data are consistent with a wide range of both positive and negative values for this coefficient. The intervals for the coefficient on trade based on the AR, LR and S-statistics are also much larger that those based on conventional t-statistics, but mostly tend to include negative values. Overall, the substantially-larger confidence intervals based on weak instrument asymptotics suggest that the partial effects of institutions and trade are not very tightly estimated.

The remaining specifications consider some of the robustness checks in RST. First, RST argue that their finding on the relative importance of trade is robust to using PPP trade shares to measure trade. The results in specification 3.2 suggest that this equation is also weakly identified. In the bottom panel of Table 3 we see that the AR, LR and S-statistics all produce very large confidence intervals in the form of two disjoint sets which exclude only what might be thought of as a priori reasonable values for these parameters in the general vicinity of the OLS estimates. The results of this specification are therefore more similar to those in the basic DK finding, in which the partial effects of institutions and trade are essentially unidentified. The same is true in specification 3.4 with additional geographical control variables. The correct 95% confidence intervals span nearly the entire real line, indicating that the data are consistent with virtually any partial effect of institutions and trade. In specification 2.3 the AR and LR statistics give confidence intervals that are similar (but substantially wider) to conventional ones, while those based on the S-statistic encompass a very large range of parameter values.

To sum up, we have seen that the basic DK conclusion that the cross-sectional data in their specification is uninformative about the partial effects of institutions and trade is supported by the more formal weak instrument asymptotic approximations used here. The same techniques also suggest that the RST findings on the relative importance of institutions may be less robust than meets the eye. In their specifications using PPP trade shares and adding additional geographical controls, the data appears consistent with virtually any values of the parameters of interest, due to the weak identification of these equations. In their basic specification, the different approaches to inference lead to different precise conclusions. Qualitatively however they all lead to much larger confidence intervals than those based on t-statistics, especially in the case of the S-statistic.

5. Decadal Regression Results

How might we proceed from the negative findings above that the partial effects of institutions and trade are not identified in cross-country regressions of levels of per capita GDP on instrumented institutions and trade? One possibility which we explored in our previous paper was to instead look at decadal changes in growth, trade and institutional quality in the hopes that the partial effects of the latter two variables would be better identified in this framework relying on internal instruments. Of course, a drawback of this approach is that by construction it cannot pick up the very long-run effects of institutions and trade emphasized in the levels regressions above. Moreover, while we showed that there is non-trivial variation in measures of trade and institutional quality even at decadal frequencies, we are less sure how informative this variation is about true unobserved changes in these variables. In our previous paper we found that changes in trade volumes were a significant predictor of changes in decadal growth rates. However, the effects of changes in five measures of institutional quality on changes in growth were very imprecisely estimated and rarely statistically significant across specifications.

In this section we deploy the tools descriped above to verify the strength of the instruments in the decadal regressions of our previous paper. We began with the following standard cross-country growth regression:

$$(7) \qquad y_{ct} = \beta_0 + \beta_1 \cdot y_{c,t-k} + \beta_2 ' X_{ct} + \eta_c + \gamma_t + v_{ct}$$

where y_{ct} is log-level of per capita GDP in country c at time t, $y_{c,t-k}$ is its lag k years ago (k=10 years in our application using decadal data) and X_{ct} is a set of control variables which are measured as averages over the decade between t-k and t. We will consider trade volumes and measures of institutional quality among the variables in X. Subtracting lagged income from both sides of the equation gives the more conventional formulation in which the dependent variable is growth, regressed on initial income and a set of control variables. The disturbance term in the regression consists of an unobserved country effect that is constant over time, η_c, an unobserved period effect that is common across countries, γ_t, and a component that varies across both countries and years which we assume to be uncorrelated over time, v_{ct}.

We adopted the approach of Caselli, Esquivel and Lefort (1996), which is to estimate equation (7) in differences, using appropriate lags of the right-hand side variables as instruments. In particular, we estimate the following regression:

$$(8) \qquad y_{ct} - y_{c,t-k} = \beta_1 \cdot \left(y_{c,t-k} - y_{c,t-2k}\right) + \beta_2 ' \left(X_{ct} - X_{c,t-k}\right) + \left(\gamma_t - \gamma_{t-k}\right) + \left(v_{ct} - v_{c,t-k}\right)$$

This is nothing more than a regression of growth on lagged growth, and on changes in the set of explanatory variables. Or, subtracting lagged growth from both sides of the equation, we have changes in growth from one decade to the next as a function of initial growth and changes in the explanatory variables. Our identifying assumption is that while trade volumes and institutional quality may be correlated with the contemporaneous and lagged shocks to GDP growth ($E[X_{ct} \cdot v_{c,t-s}] \neq 0$ for $s \geq 0$), they are uncorrelated with future shocks to GDP growth, ($E[X_{ct} \cdot v_{c,t+s}] = 0$ for $s > 0$). We therefore instrument for lagged growth using the log-level of per capita income at the beginning of decade t-2k, and we instrument for the changes in trade and institutional quality using their levels at the beginning of decade t-k.

The first column of Table 4 reports the results of this specification when X contains only the share of trade in GDP. The OLS and IV estimates replicate those in our previous paper. We find a positive effect of trade on growth at this decadal frequency, with a coefficient that is economically sensible and has a t-statistic greater than 2. In the first-stage regressions, we find that while only initial income is a significant predictor of growth, both initial income and initial trade are significant predictors of trade. This suggests some concerns about the strength of the identification. The diagnostic statistics for the presence of weak instruments paint a somewhat mixed picture as well. The Shea partial R-squared for growth is very small, while the partial R-squared for trade is a respectable 0.25. On the other hand, the Stock-Yogo test statistic for the null of weak instruments is large enough that we can reject the null. The bottom panel reports conventional Wald-based confidence intervals as well as intervals based on the S-statistic.[9] While the confidence interval for the coefficient on initial income is very large, the confidence interval for the coefficient on trade is essentially the same as that based on the usual Wald statistic. This gives us some confidence in the robustness of this basic finding on the effects of trade on growth at decadal frequencies, even though the coefficient on initial income appears to be less well identified.

The middle panel of Table 4 reports results adding one of the five time-varying measures of institutional quality we considered in our previous paper. Following Clague, Keefer, Knack and Olson (1999) we proxy for the strength of property rights protection using what they refer to as “contract-intensive money”, i.e. the share of currency in circulation held inside the banking system. In the middle panel of Table 4, we enter this variable and treat it as endogenous. The top part of the table again replicates the results of our previous paper. We find that the coefficient on trade does not change very much relative to the previous case, while the measure of institutional quality enters negatively with a very large standard error. However, the diagnostic statistics for the presence of weak instruments strongly point to identification problems in this specification. The Shea partial R-squared statistics are tiny, as is the minimum eigenvalue of the concentration matrix. Not surprisingly, in this case confidence intervals based on the S-statistic cover

[9] In this section we report only the confidence intervals based on the S-statistic because of its convenient property that projection methods are not required. This is most useful in the middle specification of Table 4 where we have three endogenous variables, and ensures that the results are comparable across these three specifications.

the entire real line, suggesting that this specification is entirely uninformative about all of the parameters of interest.

The reason for the poor performance of the IV estimator is however somewhat different from the levels regressions discussed before. In the levels regressions the problem was that the instruments for trade also explained institutional quality while the instruments for institutional quality also predicted trade well. This is not the case here. In our previous paper we noted that initial levels of trade predict subsequent changes in trade but not changes in institutions. Conversely initial levels of institutions predict changes in institutions but not changes in trade. However, the problem here is that initial institutions also predicts the endogenous change in initial income, while initial levels of per capita income that are supposed to instrument for the change in initial income also explain changes in trade and changes in institutions. This is why the identification of this specification breaks down.[10]

We obtain rather more encouraging results when we treat institutions as exogenous. In our previous paper we noted that the large estimated standard errors on the coefficient on institutional quality in the dynamic regressions might be due to weak instruments. We then also discussed results in which we treated institutions as exogenous, which if anything were likely to exaggerate the effects of institutions on growth. In these specifications we continued to fail to find a significant impact of institutions on growth, while trade continued to enter positively and significantly. In the final panel of Table 4 we revisit this alternative specification. The results are quite similar to those in the first panel where institutional quality was omitted entirely. The main difference now however is that the Stock-Yogo test favours the null hypothesis of weak instruments. Closer inspection of the first-stage regressions suggests that this may be because initial levels of income are predicting both lagged growth and changes in trade. When we compute appropriate confidence intervals using the S-statistic, we

[10] RST argue that the significance of trade in the decadal regressions in DK is not robust to the inclusion of selected regional dummies and their interactions with decadal dummies. For example, for the specification using contract-intensive money as a proxy for institutional quality, when these variables are estimated, nothing is significant in the IV regression. In light of the results above, this is perhaps not very surprising. Since these additional exogenous variables are likely to be good predictors of all three endogenous variables, this is likely to further weaken the performance of the instruments. Moreover, given that these additional dummy variables are not significant in this specification, we are not sure that it is a good strategy to include them and undermine any hopes of identification.

find once again that the coefficient on initial income is not very precisely estimated. Nevertheless we do find that the coefficient on trade is fairly precisely estimated, with a confidence interval very similar to that based on a conventional Wald test.

Overall these results suggest that it is also difficult to identify the partial causal effects of institutions on growth in this shorter-run framework. We do not appear to have sufficiently informative variation in the decadal data to identify the partial effects of institutions and trade treating both as endogenous. However, the specification which includes institutions but treats this variable as exogenous still serves a useful purpose, since it helps to control for a spurious correlation of changes in trade with changes in institutional quality, even if we cannot properly identify the partial effects of the latter.

6. Conclusions

In this paper we have argued that existing attempts to isolate the partial effects of institutions and trade on growth in the long run suffer from serious identification problems. The reason for this is simple -- existing historical and geographical instruments in the literature tend to have strong predictive power for *both* institutions and trade. As a result, these specifications are only weakly identified, despite the apparent good performance of first-stage regressions.

How do we move forward from such a negative result? One possibility is to rely more on the time-series variation in institutions, trade and growth, in the hopes that internal instruments will be less weak than in the cross-sectional applications we have described. This is the approach we followed in our previous paper. However, this approach appears to yield mixed results regarding the strength of the instruments. If we treat institutions as exogenous, we are at least able to uncover a significant partial association between trade and growth which survives the inclusion of a variety of proxies for institutional quality. However, if we allow institutions to be endogenous, we find that the model is too weakly identified to be able to sharply estimate any of the parameters of interest.

Clearly simple cross-country linear instrumental variables regressions, either in levels or in decadal differences, cannot provide definitive answers about questions as complex as the interacting roles of institutions and trade for growth. At best this approach can summarize interesting partial correlations in the data while at least partially controlling for obvious endogeneity problems. Combining this approach with more microeconometric evidence, case studies, and descriptions of historical episodes is likely to constitute fruitful research agenda. In the meantime, the empirical examples we provide here illustrate a broader problem in the cross-country empirical growth literature, which only in recent years has begun to take serious the problem of endogeneity of many of the key determinants of growth. Finding compelling instruments for these endogenous variables is difficult to begin with. If in addition we are interested in sorting out the partial effects of multiple endogenous determinants of growth, the search for instruments is further complicated by the need to find exogenous *and* independent sources of variation in order to identify partial effects of interest.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review.*

Alcalá, Francisco and Antonio Ciccone (2001). "Trade and Productivity". CEPR Discussion Paper No. 3095.

Anderson, T.W. and H. Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". Annals of Mathematical Statistics. 20:46-63.

Caselli, Francesco, Gerardo Esquivel and Fernando Lefort (1996). "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics." *Journal of Economic Growth*, 1: 363-389.

Clague, Christopher, Philip Keefer, Stephen Knack, and Mancur Olson (1999). "Contract-Intensive Money: Contract Enforcement, Property Rights, and Economic Performance." *Journal of Economic Growth*, 4:185-211.

Dollar, David and Aart Kraay (2002). "Institutions, Trade and Growth". Paper Prepared for the Carnegie Rochester Conference on Public Policy.

Frankel, Jeffrey A. and David Romer (1999). "Does Trade Cause Growth?" *The American Economic Review*, (June) 379-399.

Easterly, William and Ross Levine (2002). "Tropics, Germs, and Crops: How Endowments Influence Economic Development". Paper Prepared for the Carnegie Rochester Conference on Public Policy.

Kleibergen, Frank (2002). "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regressions". Econometrica.

Moreira, Marcelo (2002). "A Conditional Likelihood Ratio Test for Structural Models". Manuscript, Harvard University.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi (2002). "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development". NBER Working Paper No. 9305.

Shea, John (1997). "Instrument Relevance in Multivariate Linear Models: A Simple Measure". Review of Economics and Statistics. 79:348-352.

Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression With Weak Instruments". Econometrica. 65:557-586.

Startz, Richard, Charles R. Nelson and Eric Zivot (2001). "Improved Inference for the Instrumental Variable Estimator". Manuscript, University of Washington.

Stock, James H. and Motohiro Yogo (2001). "Testing for Weak Instruments in Linear IV Regression". Paper Prepared for Festschrift in Honour of Thomas Rothenberg.

Stock, James H, Jonathan H. Wright, and Motohiro Yogo (2002). "A Survey of Weak Instrument and Weak Identification in Generalized Method of Moments". Journal of Business and Economic Statistics. 20(4):528-529.

Wang, Jiahui and Eric Zivot (1998). "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments". Econometrica. 66(6):1389-1404.

Table 1 -- Summary of Estimates of Effects of Institutions and Trade on Growth

Paper	Dependent Variable	Institutions Variable	Trade Variable	Instruments	Included Exogenous Variables
Frankel and Romer (1999), Table 3, Column 2	Real GDP Per Capita at PPP, 1985	Excluded	Trade/GDP in current local currency units	Trade predicted by gravity model	Area, Population
Acemoglu, Johnson and Robinson (2001), Table 4, Column 1	Real GDP Per Capita at PPP, 1995	International Country Risk Guide	Excluded	Settler mortality	None, but several other specifications include a number of other exogenous control variables
Alcala and Ciccone (2002), Table 6, Column 3	Real GDP Per Worker at PPP, 1985	International Country Risk Guide	Trade/GDP at PPP	Population, log(Population), Area, Trade predicted by gravity model, log(Trade predicted by gravity model), fraction of population speaking English, fraction of population speaking major European language	log(Area), regional dummies, distance from equator
Dollar and Kraay (2002), Table 1, Column 6	Real GDP Per Capita at PPP, 1995	Rule of Law from Kaufmann, Kraay and Zoido-Lobaton (2002)	Trade/GDP at PPP	Trade predicted by gravity model, fraction of population speaking English, fraction of population speaking major European language	log(Population)
Rodrik, Subramanian and Trebbi (2002), Table 2, middle panel	Real GDP Per Capita at PPP, 1995	Rule of Law from Kaufmann, Kraay and Zoido-Lobaton (2002)	Trade/GDP in current local currency units	Trade predicted by gravity model, Settler mortality	Distance from equator

Table 2 -- Dollar-Kraay (2002) Specification

	2.1 -- Institutions Only		2.2 -- Trade Only		2.3 -- Institutions and Trade	
Endogenous RHS Variables	Institutions	Trade	Institutions	Trade	Institutions	Trade
OLS and IV Estimates						
OLS	0.985			1.077	0.792	0.405
	(0.044)			(0.088)	(0.075)	(0.102)
IV	1.310			1.623	1.258	0.178
	(0.141)			(0.406)	(0.349)	(0.595)
Included Exogenous Variables			lnpop		lnpop	
First-Stage Regressions						
Ln(FrankelRomer)				0.386	0.643	0.466
				(0.109)	(0.122)	(0.102)
Engfrac	0.868				0.944	0.368
	(0.356)				(0.348)	(0.292)
Eurfrac	0.500				0.691	0.363
	(0.247)				(0.215)	(0.181)
F-Statistic	13.350			28.800	13.540	16.770
R-Squared	0.139			0.285	0.296	0.342
# Observations	134		134		134	
Weak Instrument Diagnostics						
Shea Partial R-Squared	0.139			0.285	0.039	0.023
Corr(Fitted Values)					0.740	
Eigenvalues of Π					0.025, 2.239	
Eigenvalues of Γ					0.669, 27.191	
95% Confidence Intervals						
Wald					(0.57,1.95)	(-1.00,1.35)
AR					$(-\infty,\infty)$	$(-\infty,\infty)$
LR					$(-\infty,\infty)$	$(-\infty,\infty)$
S					$(-\infty,\infty)$	$(-\infty,\infty)$

Note: Dependent variable in all specifications is ln(Real GDP per capita at PPP) in 1995. Heteroskedasticity-consistent standard errors in parentheses.

Table 3 -- Rodrik, Subramanian and Trebbi (2002) Specifications

	3.1 -- Institutions and Trade LC Trade Share		3.2 -- Institutions and Trade PPP Trade Share	
Endogenous RHS Variables	Institutions	Trade	Institutions	Trade
OLS and IV Estimates				
OLS	0.745	0.259	0.548	0.474
	(0.109)	(0.161)	(0.115)	(0.101)
IV	2.126	-0.513	2.860	-0.909
	(0.589)	(0.420)	(1.261)	(0.972)
Included Exogenous Variables	DISTEQ		DISTEQ	
First-Stage Regressions				
Ln(FrankelRomer)	0.222	0.588	0.222	0.491
	(0.104)	(0.057)	(0.104)	(0.110)
Engfrac				
Eurfrac				
LnSetMor	-0.264	-0.133	-0.264	-0.212
	(0.081)	(0.045)	(0.081)	(0.074)
F-Statistic	17.000	36.130	17.000	8.030
R-Squared	0.413	0.594	0.413	0.264
# Observations	80		72	
Weak Instrument Diagnostics				
Shea Partial R-Squared	0.110	0.324	0.052	0.083
Corr(Fitted Values)	0.202		0.667	
Eigenvalues of Π	0.035, 0.447		0.015, 0.380	
Eigenvalues of Γ	6.273, 40.834		1.792, 13.210	
95% Confidence Intervals				
Wald	(0.95, 3.30)	(-1.35, 0.33)	(0.34, 5.38)	(-2.85, 1.03)
AR	(1.44, 7.86)	(-3.81, 0.15)	(-∞, -7.26) U (1.79, ∞)	(-∞, 0.28) U (5.83, ∞)
LR	(1.44, 7.86)	(-4.14, 0.15)	(-∞, -7.26) U (1.79, ∞)	(-∞, 0.28) U (5.83, ∞)
S	(-∞, -1.77) U (1.67, ∞)	(-∞, -0.18)	(-∞, 0.20) U (2.32, ∞)	(-∞, -0.11) U (1.07, ∞)

Note: Dependent variable in all specifications is ln(Real GDP per capita at PPP) in 1995. Heteroskedasticity-consistent standard errors in parentheses.

Table 3, cont'd -- Rodrik, Subramanian and Trebbi (2002) Specifications

	3.3 -- Institutions and Trade		**3.4 -- Institutions and Trade**	
	LC Trade Share		**LC Trade Share, Add Area**	
	Add Area, Pop		**Pop, Tropics, Frost, Malaria**	
Endogenous RHS Variables	Institutions	Trade	Institutions	Trade
OLS and IV Estimates				
OLS	0.737	0.304	0.479	0.282
	(0.111)	(0.184)	(0.133)	(1.70)
IV	1.905	0.351	1.659	-0.502
	(0.423)	(0.988)	(0.822)	(0.601)
Included Exogenous	DISTEQ, LNAREA, LNPOP		DISTEQ, LNAREA, LNPOP	
Variables			TROPICS, FROST, MALARIA	
First-Stage Regressions				
Ln(FrankelRomer)				
	-0.471	0.288	0.330	0.605
Engfrac	(0.249)	(0.124)	(0.142)	(0.088)
Eurfrac				
LnSetMor	-0.178	-0.104	-0.202	-0.063
	(0.076)	(0.049)	(0.102)	(0.065)
F-Statistic	12.120	24.690	19.920	10.320
R-Squared	0.481	0.625	0.525	0.566
# Observations	80		70	
Weak Instrument Diagnostics				
Shea Partial R-Squared	0.169	0.089	0.062	0.196
Corr(Fitted Values)	0.252		0.040	
Eigenvalues of Π	0.032, 0.315		0.021, 0.499	
Eigenvalues of Γ	3.035, 9.708		3.148, 21.222	
95% Confidence Intervals				
Wald	(0.98, 2.83)	(-1.62, 2.32)	(0.02, 3.30)	(-1.70, 0.70)
AR	(1.04, 4.31)	(-4.57, 9.37)	$(-\infty, -7.46) \cup (0.80, \infty)$	$(-\infty, 0.10) \cup (4.51, \infty)$
LR	(1.04, 4.49)	(-4.57, 11.02)	$(-\infty, -7.46) \cup (0.80, \infty)$	$(-\infty, 0.10) \cup (4.51, \infty)$
S	$(1.39, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, 1.74) \cup (3.66, \infty)$

Note: Dependent variable in all specifications is ln(Real GDP per capita at PPP) in 1995. Heteroskedasticity-consistent standard errors in parentheses.

Table 4 -- Decadal Regressions from Dollar and Kraay (2002)

	4.1 Trade Only		4.2 Institutions and Trade, Institutions Endogenous			4.3 Institutions and Trade, Institutions Exogenous	
Endogenous RHS Variable are Decade-Average Growth in:	Initial Income	Trade	Initial Income	Trade	Institutions	Initial Income	Trade
OLS and IV Estimates							
OLS	0.308	0.182	0.28	0.166	0.529	0.28	0.166
	(0.057)	(0.048)	(0.065)	(0.048)	(0.223)	(0.065)	(0.048)
IV	0.732	0.249	0.649	0.227	-0.156	0.832	0.209
	(0.246)	(0.108)	(0.905)	(0.113)	(1.679)	(0.366)	(0.107)
Included Exogenous Variables	Decadal dummies		Decadal dummies			Decadal dummies, Institutions	
First-Stage Regressions							
Level at beginning of previous decade of:							
ln(Per Capita GDP)	0.004	0.012	-0.001	0.01	0.002	0.036	0.012
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
ln(Trade/GDP)	0.002	-0.014	0.001	-0.019	-0.001	0.003	-0.018
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Institutions			0.034	0.021	-0.022		
			(0.010)	(0.016)	(0.005)		
F-Statistic	19.74	20.16	14.56	17.79	12.39	12.94	19.04
R-Squared	0.215	0.272	0.252	0.347	0.273	0.238	0.344
# Observations	274		193			193	
Weak Instrument Diagnostics							
Shea Partial R-Squared	0.053	0.247	0.005	0.01	0.021	0.034	0.295
Corr(Fitted Values)	-0.18		r12=-0.25, r13=0.06, r23=-0.12			-0.29	
Eigenvalues of Π	0.0002, 0.0003		1x10E-6, 0.0004, 0.002			0.00002, 0.0005	
Eigenvalues of Γ	7.718, 43.934		0.169, 18.759, 58.527			3.366, 39.878	
95% Confidence Intervals							
Wald	(0.25, 1.22)	(0.04, 0.46)	(-1.14, 2.43)	(0, 0.45)	(-3.47, 3.16)	(0.11, 1.55)	(0, 0.42)
S	(0.23, ∞)	(0.04, 0.46)	(-∞, ∞)	(-∞, ∞)	(-∞, ∞)	(-∞, ∞)	(0, 0.43)

Note: Dependent variable is decadal average annual growth. Heteroskedasticity-consistent standard errors in parentheses.

Figure 1: Confidence Sets Based on AR Statistic

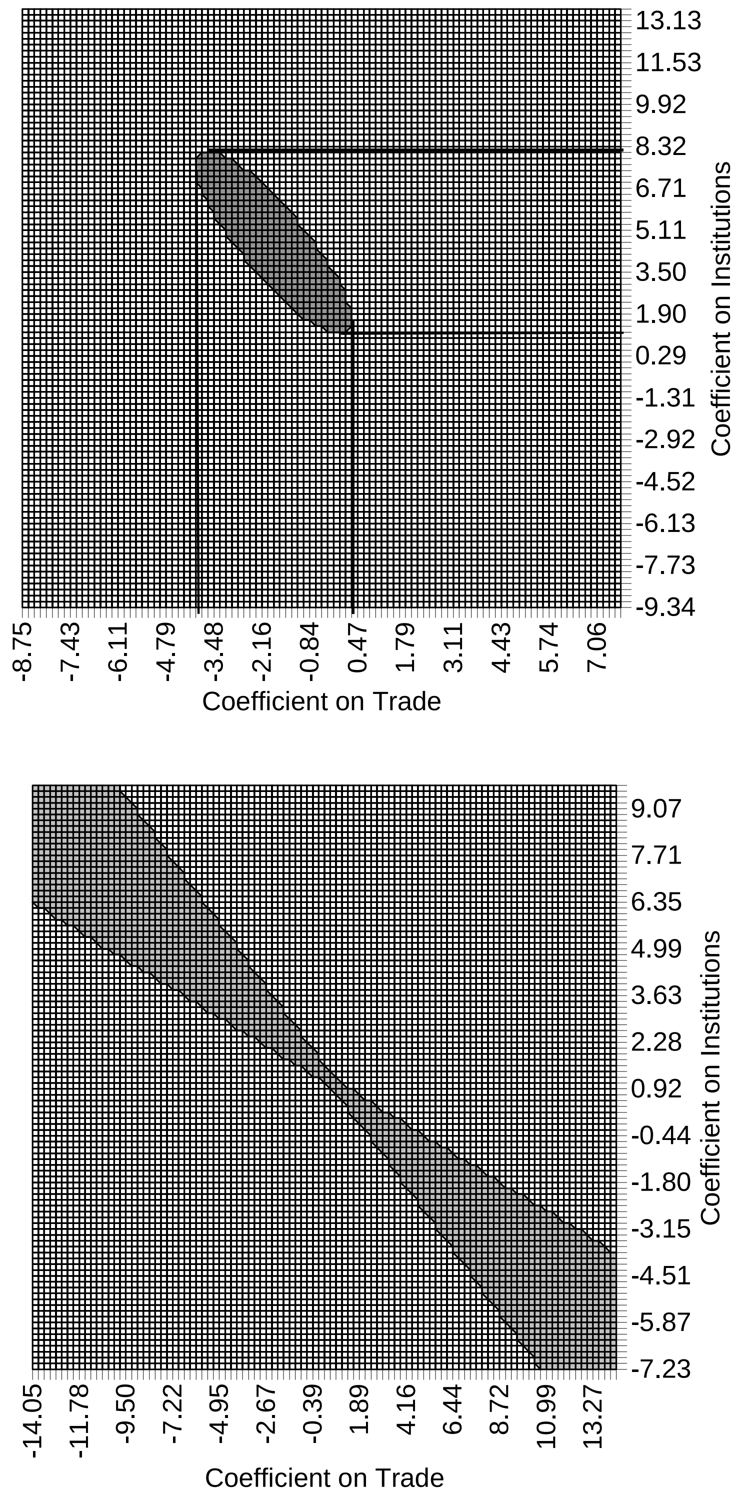

Figure 2: Confidence Intervals Based on S Statistic