*Econ Journal Watch,*
*Volume 1, Number 2,*
*August 2004, pp 244-259.*

# Reply to Chang and Ram: Statistical Adequacy and the Reliability of Inference

JEFF EDWARDS* AND ANYA MCGUIRK**

**FIRST, WE WOULD LIKE TO APOLOGIZE TO JIH CHANG AND RATI** Ram for the tone of our Comment; we in no way meant to be denigrating. When submitting the paper we had no clue that such offense would be taken. On receiving the response from Chang and Ram and rereading what we said, we can see why offense was taken. If we could reword our Comment as it goes to electronic press, we would.

Rather than continue the mud slinging which we initiated, we would like to step back and give some perspective on our Comment and attempt to clarify our main points, which for the most part were misinterpreted by Chang and Ram. We then reflect on the whole discussion and consider what, if anything, Chang and Ram could conclude even if their analysis were statistically justified. A simple look at some confidence intervals indicates that nothing at all can be concluded regarding the differences found in the inequality-income profiles of high- and low- growth countries. The model we ultimately propose based on statistical grounds also finds no statistically significant differences between high- and low- growth countries.

We begin by noting that like many other economists, we are skeptical of, and disappointed with, the current state of published empirical work in economics. Empirical results often seem to be sensitive to model

* Department of Economics, Texas Tech University.
** Department of Statistics and Department of Agricultural and Applied Economics, Virginia Tech.

specification, few theories have ever been abandoned based on evidence from regression analysis, and conflicting results seem to co-exist happily (some even utilizing the same data). In their *EDCC* paper, Chang and Ram illustrate that the inequality and growth literature is not immune from these problems: "the overall empirical evidence does not show a clear picture" and "studies that did consider the growth-inequality nexus have reported somewhat contradictory results" (Chang and Ram 2000, 787). Unfortunately, this is status quo in almost all areas of economic research.

Many have recognized this unsatisfactory state of affairs. For example, almost 30 years ago, Leontief noted, "In no other field of empirical inquiry has so massive and sophisticated statistical machinery been used with such indifferent results" (1971, 3). A multitude of reasons have been given for the apparent unreliability and non-decisiveness of empirical evidence. At least some of the problem can be attributed to the non-systematic way in which applied econometric studies are carried out.[1] Pagan (1984) cynically describes the "typical" approach to econometrics.

> Four steps almost completely describe it: a model is postulated, data gathered, a regression run, some t-statistics or simulation performance provided and another 'empirical regularity was forged.' (Pagan 1984, 103)

Although the state of affairs described by Pagan is, we hope, somewhat exaggerated, it is true that the way econometrics is currently taught provides little guidance in how to systematically choose a model or to fix (re-specify) empirical models if statistical or theoretical "problems" with the initial specification are encountered.

The probabilistic reduction approach, proposed and championed by Spanos (1986, 1995, and 1999) provides a well-needed systematic alternative to the approach characterized by Pagan. The foundation of Spanos' approach is the principle of *statistical adequacy*. This principle asserts that to evaluate any theory empirically, that theory must be viewed in the context of a valid statistical model—a model whose underlying assumptions are appropriate for the data being analyzed. Otherwise, test statistics will not have their anticipated distributions (and associated error probabilities), and inferences drawn using those statistics will be unreliable and likely misleading.

[1] We could go on about other aspects of this problem, but we must limit the scope of this response. See for example, Leamer (1983), Spanos (1995), Tomek (1993, 1997), and Hendry (1993).

Note that "inferences drawn" include both inferences regarding hypotheses formally tested, as well as, *predictions* made from the model.

Our comment on the Chang and Ram article was written to illustrate potential problems and pitfalls with the way empirical work is often carried out and to illustrate that "statistical" results may not be as clear-cut as they first seem if one digs a little deeper, and examines more carefully the statistical properties of their data, and the relevance of the model estimated.

On the basis of the regression model we re-examine in our comment, Chang and Ram test Kuznets's hypothesis (though not their primary focus) *and* examine simulated/predicted gini coefficient profiles for high- and low growth regimes. The quality—reliability and properties—of the inferences drawn from *both* these activities depends on the validity of the statistical assumptions underlying the model estimated for the data they are using.

Like many researchers, Chang and Ram chose to estimate the simplest, most widely used model in this field of work. This simple model selection criterion is very common in applied work. Unfortunately the popularity and simplicity of a regression model does not guarantee that the model assumptions are reasonable for the data being analyzed and, thus, that the analysis carried out using the model will lead to reliable inferences. It is in this sense that we claimed that the model estimated by Chang and Ram is *ad hoc*. They simply chose the model, subjected the model to the data, and drew conclusions regarding the Kuznets hypothesis using the reported t-statistics, and regarding the differences between high- and low-growth countries, using predictions from the model. Without attempting to assess whether the statistical assumptions underlying this model are valid for the data chosen, we have no way of assessing whether or not, a la Pagan, 'another empirical regularity was forged.'

The readily testable assumptions underlying Chang and Ram's model:

$$INEQ_i = a_3 + b_3(LY_i) + c_3(LY_i)^2 + a_{33}D + b_{33}(D \times LY_i) + c_{33}\left[D \times (LY_i)^2\right] + u_{3i}, \qquad [1]$$

which forms the basis of their analysis, are:

(i) functional form is correct: $E(u_{3i}) = 0$

(ii) $Var(u_{3i}) = \sigma^2$: homoskedasticity--$Var(u_{3i})$ is not a function of the regressors

(iii) $(a_3, b_3, c_3, a_{33}, b_{33}, c_{33}, \sigma^2)$ are constant over the index i

(iv) $E(u_{3i}u_{3j}) = 0 \;\forall i \neq j$, and

(v) $u_{3i} \sim N\,(\cdot,\cdot)$. [2]

Although not mentioned in our Comment, we performed a battery of misspecification tests designed to examine the appropriateness of each of these assumptions.[3] Note that in order to evaluate whether the parameters varied over the index i (assumption iii), and whether the errors were autocorrelated (assumption iv), we ordered the data three different ways—by growth rate, by location (region), and by GDP.[4] The results of these tests suggested that almost all of the testable assumptions underlying the simple model above seemed reasonable for the data being analyzed. The one assumption found to be clearly inappropriate for these data was parameter stability; misspecification tests and graphical analysis both indicated that the intercept of the regression differed by region (see our Comment for details).[5] When equation [1], supplemented with regional dummies, was subjected to the complete battery of misspecification tests, we found no evidence against any of the testable model (statistical) assumptions. Because, based on the misspecification tests we conducted, the new model is adequate from a statistical standpoint, we can use it to examine any hypotheses or theories we are interested in and for prediction—there are well-known results about the relevant test statistics and the properties of these forecasts, etc.[6]

An F-test of the significance of all the regional dummies (base was South America) indicated that, as a group, these dummies were significantly different from zero (p-value 0.004)—hence the intercept differs by region.[7]

---

[2] Note assumptions (i)-(v) are encompassed by $u_{3i} \sim NIID(0,\sigma^2)$ --the notation used in our Comment. We used this notation so we could avoid the details we now feel we should reveal to make our general point.

[3] We performed both the individual and joint misspecification tests suggested in McGuirk, Driscoll, and Alwang (1993), and McGuirk, Driscoll, Alwang, and Huang (1995).

[4] To shorten our discussion we write GDP rather than real GDP per capita, though we are referring to the latter.

[5] Violation of the parameter stability assumption implies OLS estimators are biased and inconsistent.

[6] Unfortunately we can never conclude that this is "*the correct*" model to use—it is possible, for example, that another researcher, conducting a different set of misspecification tests, could find evidence against the model assumptions. It is also possible that, say, another model, with a different functional form could pass the same battery of tests conducted here. In this case, we would try to distinguish between these models, to determine which is most appropriate for these data using other criteria.

[7] Note that the F-test reported here is different than the one reported in the Comment. For the F-test here we have the North American Countries split into USA/Canada and Central

However, when all the dummies were included, the only variables significant at the 5% level were the Central American (CA) and European (E) dummies; the variables LY and $LY^2$ were significant at the 10% level. This evidence, in conjunction with the graphical evidence, led us to test whether or not we even needed to include the non-significant regional variables at all. An F-test of the hypothesis that all but the CA and E dummies were equal to zero yielded: $F(3,38)=0.101$ (p-value=0.96) and, thus, we dropped the remaining regional effects. The main effect of deleting these extra regional variables was an increase in the t-statistics associated with the variables involving the high-growth dummy. When this simpler model was subjected to the complete battery of misspecification tests, we found no evidence against any of the testable model (statistical) assumptions. Thus, we concluded that this simpler model was also adequate from a statistical point of view (see Table 1).[8]

**Table 1**

| **Edwards and McGuirk Results** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Constant | LY | $LY^2$ | D | D*LY | D*$LY^2$ | Europe | C. Amer. |
| -121.45<br>(-1.09) | 42.10<br>(1.45) | -2.68<br>(-1.44) | -260.86<br>(-1.44) | 59.03<br>(1.31) | -3.31<br>(-1.20) | -9.48*<br>(-3.08) | 11.57*<br>(3.44) |
| t-statistics in parenthesis<br>The number of observations is 48<br>The unadjusted $R^2$ is 0.61<br>The adjusted $R^2$ is 0.54<br>* $P \leq 0.10$ | | | | | | | |

The question that remains is whether or not all of this additional analysis actually changes substantively any of the conclusions of Chang and Ram. Chang and Ram's test of the Kuznets hypothesis and the income-inequality profiles analyzed were obtained using a model whose assumptions were not all reasonable for the data being analyzed—thus,

America. In the Comment we refer to the test using the un-split group. Though the regression with the un-split group also seemed to be statistically adequate, we realized, after communicating with Chang and Ram, that splitting this group not only made sense theoretically but also improved the evidence for the model assumptions. For a discussion regarding the issues raised by Chang and Ram regarding 'specification search' or data mining and when it is a legitimate practice, see Spanos (2000).

[8] This is the model reported in Table 3 of our Comment except--as pointed out by Chang and Ram—we made a typographical error when reporting our constant. Our results are identical to those reported by Chang and Ram.

there is the potential that we will be misled by their model results.[9] For completeness, the final Chang and Ram model is summarized in Table 2.

**Table 2**

| **Chang and Ram Final Results** | | | | | |
|---|---|---|---|---|---|
| Constant | LY | $LY^2$ | D | D*LY | $D*LY^2$ |
| -212.77<br>(-1.62) | 65.37*<br>(1.92) | -4.11*<br>(-1.87) | -193.79<br>(-0.89) | 43.88<br>(0.82) | -2.529<br>(-0.77) |
| t-statistics in parenthesis<br>The number of observations is 48<br>The unadjusted $R^2$ is 0.40<br>The adjusted $R^2$ is 0.33<br>* $P \leq 0.10$ | | | | | |

We begin by comparing the forecasts of the two models, as this was the primary focus of the *EDCC* paper (see Chang and Ram's Response). In figure 5 of our Comment we attempted to illustrate how our model predictions for high- and low-growth countries differed from those in Chang and Ram. To make this plot, we subtracted out the European (E) and Central American (CA) regional differences and then graphed the model predictions. Thus, as stated in our comment, we graphed our model predictions for all the *other* high- and low-growth countries, and for E and CA countries *after* having subtracted out their specific regional effects.[10] This is a very different exercise than *not* modeling the regional differences at all. If we ignored the regional effects we would get the Chang and Ram predictions.

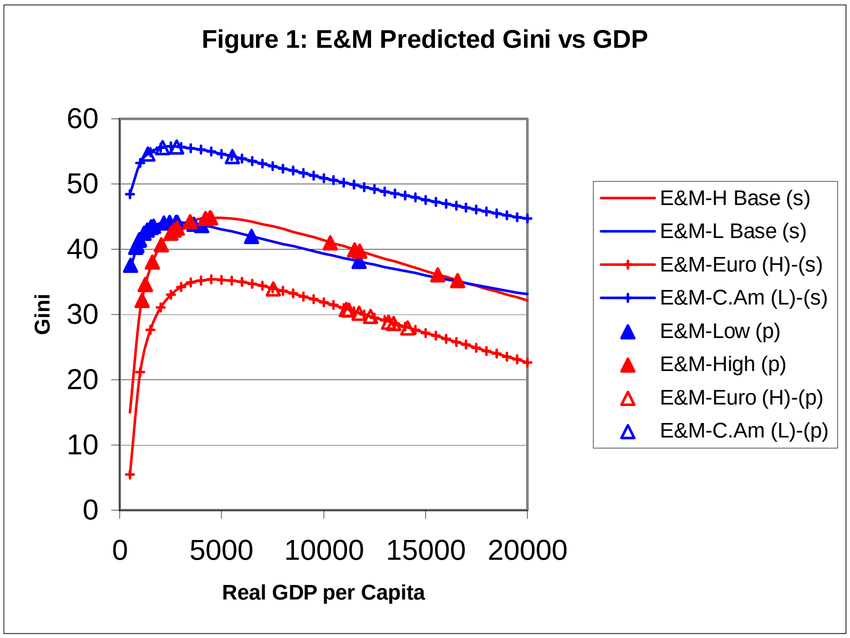
We now make another attempt to illustrate the differences in predictions. As pointed out by Chang and Ram, the 8 E countries are all high-growth and the 4 CA countries are low-growth. Thus, our model essentially has 2 high-growth groups (of 8 and 17 countries respectively) and 2 low-growth groups (of 4 and 19 countries respectively) and as a consequence we get (smooth) prediction lines for all 4 groups. The E and

[9] Chang and Ram are right in their Response when they say that the presence or absence of the inverted U does not affect comparisons of the predicted income-inequality profiles. The quality and reliability of the different predictions (and thus, comparisons)—since they are based on a regression model—*will* depend on the relevance of the model's underlying statistical assumptions for the data being analyzed.

[10] In hindsight, we should have just graphed the simple predictions for our base groups. The resulting graph would have been *identical* to the one presented, except it would have had 8 fewer triangles and 4 fewer dots.

CA country predictions differ from their particular base group predictions by a simple vertical shift.

In Figure 1, we illustrate our predictions for these 4 groups, with the predictions associated with actual GDP levels designated with a triangle.[11] Figure 2 illustrates the comparable predictions from the Chang and Ram model. Obviously their predictions for the E and CA countries are captured in their high- and low-growth groups respectively.



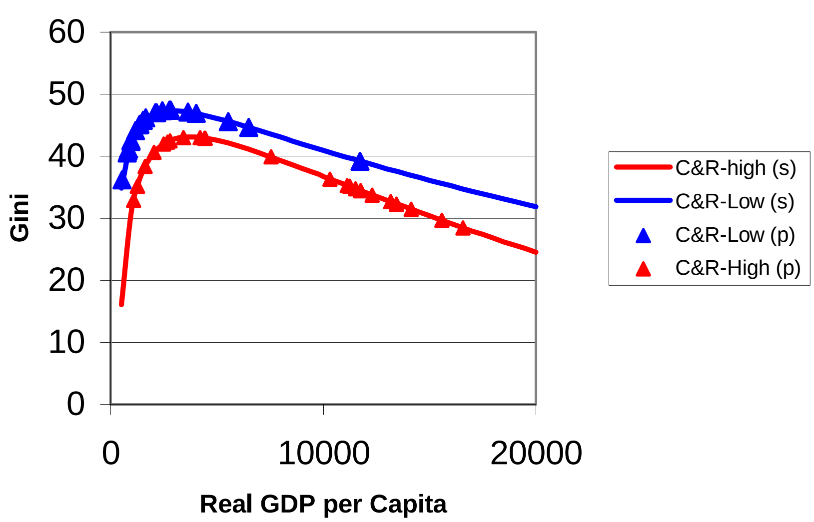

There are several points we would like to make regarding these figures and the associated results. First, by allowing for E and CA regional effects, our model predicts little difference in the inequality-GDP relationship between the 17 high- and 19 low-growth base countries. Based on our model results (Table 1), these differences are, *at best*, marginally significant from a statistical point of view.[12] Further, the differences

[11] Indeed the base group lines in Figure 1 are identical to the 2 lines graphed in Figure 5 of our Comment. Note also that our new Figure 1 is essentially Figure 1 in Chang and Ram's Response augmented with predictions of the 4 groups at all income levels.

[12] Prediction confidence intervals (derived later) are needed to assess the significance of these differences.

illustrated in the figure seem economically insignificant, particularly in comparison to those predicted by Chang and Ram, except, perhaps, at the very lowest GDP levels. In contrast to the Chang and Ram predictions, our model also does not predict that higher growth countries have lower inequality at all income levels for these base group countries.

**Figure 2: C&R Predicted Gini vs GDP**




Second, our predictions for CA and E *are* consistent with the story painted by Chang and Ram in the sense that the low-growth CA countries do have higher predicted levels of inequality relative to the high growth countries with similar GDP levels. Similarly, the high-growth E countries have much lower predicted levels of inequality compared to the (few) low-growth countries with similar GDP levels. However, the fact that we needed the 2 regional fixed effects, *in addition* to the growth dummy variables, indicates that it must be more than just differences in speed of growth that explain the positions of these CA and E curves. The increase in the individual t-statistics on the high-growth dummy variables in our model relative to those of Chang and Ram seems to provide evidence that the inclusion of the CA and E dummies has not detracted from our ability to capture differences between the growth groups and that there is something else driving these results. Certainly, additional research is needed to understand what these regional variables are actually capturing.

We now shift to consider whether or not the data provide support for the Kuznets hypothesis (though this is a minor focus for Chang and Ram). Before doing so, it is instructive to look at the data we are attempting to model.

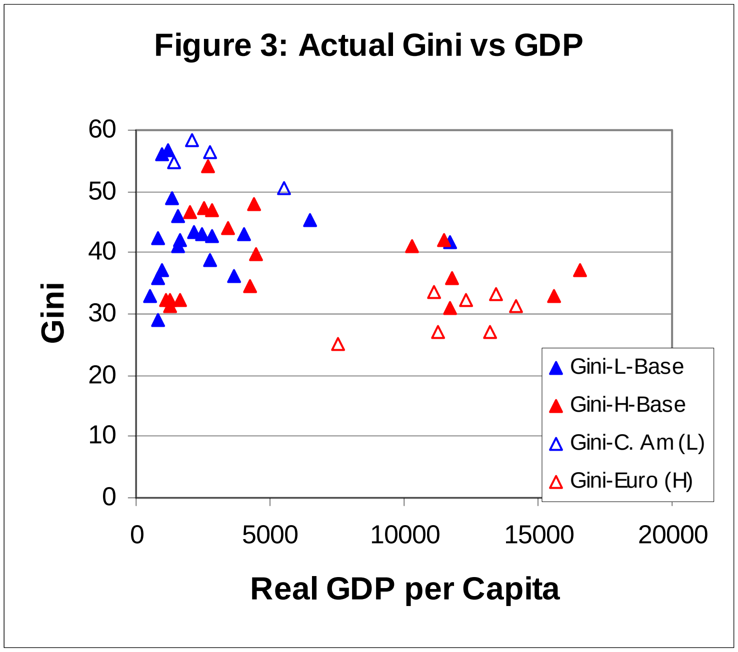



Figure 3 provides a cross-plot of the gini coefficient data versus GDP; we delineate the data by group to facilitate our discussion.[13] There are several interesting things that can be learned from this graph—without using any statistics. First, the low-growth countries almost all have very low GDP levels, and at any given (low) level of GDP, there is a strikingly large (economically significant) range of gini coefficients. The high-growth countries, on the other hand, have a much larger range of GDP levels, and

[13] Figure 6 in our Comment is the graph of these same data; unfortunately, in the Comment we used the same symbol (dot) for all the observations. As a consequence, we confused Chang and Ram and misled ourselves somewhat. By delineating the data by group we can see more clearly what is happening with these data (see text).

though there is also a range of gini coefficients at each level of GDP, it is not too hard to visualize that an inverted U (Kuznets-curve) *could* possibly fit these data.[14] This simple "reading" of the data contradicts some of the findings of both the Chang and Ram and Edwards and McGuirk models. The Chang and Ram model seems to support the hypothesis for low-growth countries, while the evidence is much weaker for the high-growth countries. Chang and Ram attribute the weaker high-growth country evidence to collinearity, and conclude that both group results support the Kuznets hypothesis. Our model, on the other hand, seems to provide no evidence in its favor for either group. More evidence regarding the relevance of the Kuznets hypothesis is obtained in the Chang and Ram Response (point 4), when they estimate 2 separate regressions for the high- and low-growth groups. This time they find evidence that the Kuznets hypothesis holds for the high-growth group and not for the low group; these new findings are consistent with Figure 3. Given that the separate regressions estimated are less restricted versions of our statistically adequate model, we know that the t-statistics obtained should be distributed student's-t as the model assumptions seem relevant for the data.[15] The puzzle then is why these new separate regression results seem to differ from those of our statistically adequate model. After considering this question, we realized that the best, most direct, way to examine the Kuznets hypothesis for the two groups is to estimate the following (equivalent) form of the basic model:

$$INEQ_i = a_3 D_L + b_3 (LY_i \times D_L) + c_3 (LY_i^2 \times D_L) + a_{33} D_H + b_{33} (D_H \times LY_i) + c_{33} \left[ D_H \times LY_i^2 \right] + u_{3i},$$

where $D_L$ ($D_H$) is a dummy variable equal to 1 for the low (high) growth countries and zero otherwise. Of course, our formulation includes the two regional dummies and the Chang and Ram formulation does not. The estimation results are presented in Table 3.

[14] The left leg of the inverted U would begin at the (GDP, Gini) point of approximately (1200, 32) or so, reach a maximum at approximately (2500, 51) and then decrease (much more slowly than it rose) through the remaining red triangles.

[15] Under point 5 in their Response, Chang and Ram estimate two more regression models and talk about evidence for the Kuznets hypothesis. Of course nothing should be concluded from these models until their statistical validity is assessed. If Chang and Ram really believe that the income-inequality relationship differs by growth rate, these regressions will, by definition, be misspecified and thus, should not be used to test any hypothesis.

**Table 3**

| Edwards and McGuirk Model Results by Growth Group | | | | | | | |
|---|---|---|---|---|---|---|---|
| $D_L$ | $D_L \times LY$ | $D_L \times LY^2$ | $D_H$ | $D_H \times LY$ | $D_H \times LY^2$ | Euro | CA |
| -121.4<br>(-1.09) | 42.10<br>(1.45) | -2.67<br>(-1.44) | -382.3*<br>(-2.67) | 101.12*<br>(2.95) | -5.99*<br>(-2.94) | -9.48<br>(-3.08) | 11.57<br>(3.44) |
| t-statistics in parenthesis<br>The number of observations is 48<br>The unadjusted $R^2$ is 0.61<br>The adjusted $R^2$ is 0.54<br>* $P \leq 0.10$ | | | | | | | |

| Chang and Ram Model Results by Growth Group | | | | | |
|---|---|---|---|---|---|
| $D_L$ | $D_L \times LY$ | $D_L \times LY^2$ | $D_H$ | $D_H$*LY | $D_H$*$LY^2$ |
| -212.77<br>(-1.62) | 65.37*<br>(1.92) | -4.11*<br>(-1.88) | -406.6*<br>(-2.36) | 109.24*<br>(2.64) | -6.64*<br>(-2.71) |
| t-statistics in parenthesis<br>The number of observations is 48<br>The unadjusted $R^2$ is 0.40<br>The adjusted $R^2$ is 0.33<br>* $P \leq 0.10$ | | | | | |

Indeed, our results are very close to those obtained from the separate regressions in the Chang and Ram Response. This is reassuring, as the only difference between the two specifications is that our pooled regression assumes that the conditional variance is the same for both groups. The Table 3 results indicate that the Edwards and McGuirk and Chang and Ram models both reveal evidence of a Kuznets inverted U curve for the high-growth country data. Thus, we were wrong in concluding that our results did not support the hypothesis for this group. [16] Further, it is reassuring that our eyes did not deceive us as we looked at the data in Figure 3. Interestingly, we find more support for Kuznets than Chang and Ram for the high-growth countries—even with our regional dummies. Again, this suggests that by including the regional dummies we are not detracting from our ability to pick up differences by growth rate.

For the low-growth countries, the Edwards and McGuirk model indicates that there is little to no support for the Kuznets hypothesis, while the Chang and Ram model indicates support. In terms of the Kuznets hypothesis, this is the main difference between the two model formulations.

[16] Thank goodness the examination of the Kuznets hypothesis was only a minor part of Chang and Ram's paper!

Once again, it is reassuring that our statistically adequate model results agree with what we see in Figure 3.

Before concluding the Kuznets discussion, we should note that the fact that we find evidence for Kuznets in the high-growth group and not in the low-growth group does not necessarily imply that the structure of the income-inequality profiles differ significantly by growth level. After looking at the data in figure 3, we believe that our model is unable to find a significant relationship between gini and GDP for the low-growth group, and a significant difference between the high- and low-growth countries, because of the tremendous "noise" in the low-growth data (the huge range of gini coefficients over a relatively small range of GDP).

To summarize, thus far we can say the following: Based on a model that was found to be statistically adequate, we find—using point estimate/predictions only—that high-growth countries do not necessarily have lower inequality levels for a given GDP than low-growth countries. Further, in contrast to what we claimed in our comment, there appears to be evidence of the Kuznets hypothesis for high-growth countries and not for low-growth countries. With the exception of the high-growth country evidence for the Kuznets hypothesis, these findings contradict those of Chang and Ram and corroborate the findings of our Comment.

Before concluding this Reply we want to address the final aspect of our complaints that Chang and Ram 'force' the data through a quadratic functional form and that their functional form/model is *ad hoc*. The first aspect of this criticism has been the focus of the Reply up to this point: you cannot just specify a model, fit the data, and run some t-tests—you must investigate the statistical validity of this model. The second aspect—the overall lack of significance of most of the variables in the models and their impact on the predictions—has not been addressed. What we were thinking is this: *if*, for example, the income terms are really not significantly different from zero, it is not really fair to use the (insignificant) parameter estimates to predict the relationship between gini and income and make a big deal out of these predictions. For example, suppose there really is no relationship between gini and income at all—the true parameters are zero. If we use a quadratic model, say, to capture the relationship between gini and income our parameter estimates will never be 0.0 (with probability 1)—and any predictions from this model will indicate some sort of quadratic relationship between gini and income.[17] Notice that like Chang and Ram, our model, though statistically adequate, also has very few significant variables; the parameter estimates on the two regional fixed effects are the only ones statistically different than 0 (at $\alpha$=0.05; see Table 1). While it is legitimate to use our model to show how Chang and Ram's conclusions, based on point predictions, regarding high- and low-growth countries, no longer seem to hold, we cannot conclude much else using these results until we consider confidence intervals on these predictions. The impact of the lack of significant variables and fit of the overall model will show up in these confidence intervals.

We begin this last investigation, by simply ignoring the issues raised above regarding the statistical adequacy of the Chang and Ram model. That is, we take the Chang and Ram results at face value and derive confidence intervals for the low and high-growth country predictions presented in Figure 1 and Table 3 of their *EDCC* paper. In Figure 4 we re-graph their predictions along with the associated 95% confidence intervals.[18] As illustrated, the 95% confidence intervals are huge; the low-growth predictions are very near the middle of the high-growth confidence intervals and the high-growth predictions are very near the middle of the low-growth confidence intervals. The conclusions of Chang and Ram:

> the estimates and the simulations show a statistically significant and quantitatively substantial structural difference between the two groups, and the high-growth scenario is characterized by lower inequality at all income levels. Moreover, the high-growth advantage seems particularly large at low-income levels, (Chang and Ram 2000, 795)

are not substantiated with their own empirical evidence.

[17] Yes, we do know that one can get different shapes to this curve depending on the parameter signs and magnitude—we might be idiots but not total idiots—Chang and Ram misinterpreted our complaints.

[18] The prediction confidence intervals were made using the usual formula for forecast error variance (see, for example, Davidson and MacKinnon 2004, 104).

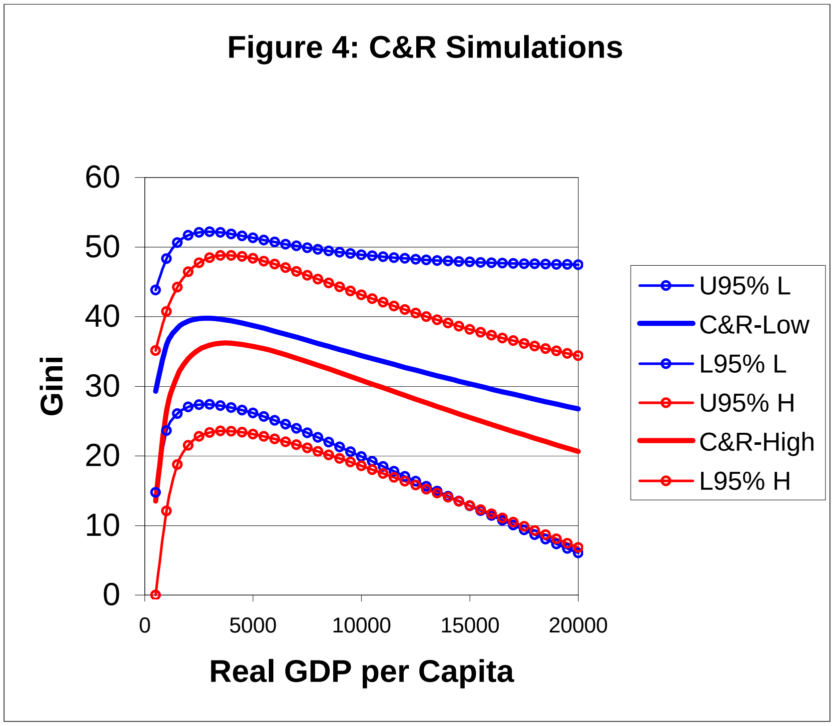


Figure 4: C&R Simulations

Figure 4 illustrates quite clearly that one cannot talk about differences between the high- and low-growth countries with any confidence. The intervals obtained with our model, though we may have more confidence in their statistical validity, are almost as large as those of Chang and Ram. Thus, based on these data and our models, we can only conclude that there is no significant difference between the predicted high- and low-growth country inequality-GDP profiles.

## REFERENCES

**Chang, Jih Y., and Rati Ram**. 2000. Level of Development, Rate of Economic Growth, and Income Inequality. *Economic Development and Cultural Change* 48(4): 787-799.

**Chang, Jih Y., and Rati Ram**. 2004. Response to Edwards and McGuirk: Income Level, Economic Growth, and Inequality: Flawed Methodology and Inaccurate Inference. *Econ Journal Watch* 1(2): 235-243.

**Davidson, R. and James G. MacKinnon**. 2004. *Econometric Theory and Methods*. Oxford: Oxford University Press.

**Edwards, Jeff, and Anya McGuirk**. 2004. Kuznets Curveball: Missing the Regional Strike Zone. *Econ Journal Watch* 1(2): 222-234.

**Hendry, David F**. 1993. *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

**Leamer, E. E**. 1983. Let's Take the Con out of Econometrics. *American Economic Review* 73(1): 31-43.

**Leontief, W**. 1971. Theoretical Assumptions and Nonobserved Facts. *American Economic Review* 61(1): 1-7.

**McGuirk, Anya M., Paul J. Driscoll, and Jeff Alwang**. 1993. Misspecification Testing: A Comprehensive Approach. *American Journal of Agricultural Economics*. 75(4): 1044-1055.

**McGuirk, Anya M., Paul J. Driscoll, Jeffrey Alwang, and Huilin Huang**. 1995. System Misspecification Testing and Structural Change in the Demand for Meats. *Journal of Agricultural and Resource Economics*. 20(1): 1-21.

**Pagan, A**. 1984. Model Evaluation by Variable Addition. In *Econometrics and Quantitative Economics*, ed. D.F. Hendry and K.F. Wallis. Oxford: Blackwell, 103-133.

**Spanos, Aris**. 1986. *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

**Spanos, Aris**. 1995. On Theory Testing in Econometrics: modeling with non-experimental data. *Journal of Econometrics* 67(1): 189-226.

**Spanos, Aris**. 1999. *Probability Theory and Statistical Inference: Econometric modeling with observational data.* Cambridge: Cambridge University Press.

**Spanos, Aris**. 2000. Revisiting Data Mining: 'Hunting' with or without a License. *Journal of Economic Methodology* 7(2): 231-264.

**Tomek, William G**. 1993. Confirmation and Replication in Empirical Econometrics: A Step Toward Improved Scholarship. *American Journal of Agricultural Economics* 75(special 75th Anniversary Issue): 6-14.

**Tomek, William G**. 1997. *Structural Econometric Models: Past and Future*. Havlicek Memorial Lecture, Department of Agricultural Economics, Ohio State University.