Correlates of statewise participation in the great Indian growth turnaround: some preliminary robustness results

Chetan Ghate* and Stephen Wright†

September 2, 2011

1 Introduction

This short note provides some supplementary analysis to the regressions in Section 5 of Ghate and Wright (forthcoming), that was carried out after the refereeing process for that paper was completed, and hence could not be included in the published version. It is not a free-standing paper, but is intended to be read in conjunction with the published paper.

2 Participation in the turnaround: some preliminary robustness results

In Ghate and Wright (forthcoming), henceforth GW we showed that while the common nature of the growth turnaround, as identified by the V-Factor, appears to correspond fairly well to observable shifts in India-wide economic policy, the quite disparate impact of the turnaround across the states (as illustrated in GW Figure 2) was striking. In GW Section 5 we used our panel dataset to attempt to provide some regression-based evidence that sheds at least some light on this issue. However while we found some evidence of

*Corresponding Author. Address: Planning Unit, Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi 110016, India. Tel:91-11-4149-3938; Fax:91-11-41493981. E-mail:cghate@isid.ac.in

†Department of Economics, Birkbeck College, University of London, Malet Street, London W1E 7HX, UK. s.wright@bbk.ac.uk

collective explanatory power, there was a limit to how much we could say about individual indicators using conventional regression analysis.

In Table 1 (the first four columns of which are identical to GW Table 3) we present some evidence on the correlates of the state-wise distribution of the turnaround in growth after our best estimate of a breakpoint, in 1987, across both states and sectors. The table summarizes cross-sectional regressions in which the dependent variable is the change in average log growth across these two sub-samples, for each of the 207 series in our largest panel (running from 1970 to 2004). The first three columns report regressions where the only regressors are dummy variables for each sector and state. Consistent with the evidence of GW Figures 1 and 2, there is strong evidence for significant differences across both sectors and states, whether both are included (as in regression (1)) or just state dummies (in regression (2)) or just sector dummies (in regression (3)).

In the remaining columns of Table 3 we investigate whether identifiable state characteristics can account for the disparate performance across the states. We retain the sectoral dummies, but include 11 different state characteristics (all either time-invariant, or measured just before the turnaround), in place of the state dummies. In regression (4), which includes all 11 indicators, the overall goodness of fit barely differs from the benchmark regression (1) and the implied restrictions are easily accepted: ie, the state-level regressors jointly span all significant variation across states.[1]

However, most individual regressors in regression (4) are statistically insignificant. This is unsurprising since we have nearly as many regressors as states, and the regressors are mostly quite strongly mutually correlated.

To provide some assessment of the robustness of the relationship between individual indicators and participation in the turnaround, we follow the approach suggested by Sala-i-Martin (1997) in relation to cross-sectional growth regressions, where it is well-known that the number of potential regressors far exceeds the number of regressors that can feasibly be included in any given growth regression. We examine the distribution of coefficients on individual state-level regressors, when included in all possible regressions alongside a subset of other regressors. If a large part (or all) of the distribution of the resulting coefficients lies to the right or left of zero, and the

[1]For a more detailed discussion of regression diagnostics, etc, see Ghate & Wright (forthcoming).

coefficient is on average statistically significant, we follow Sala-i-Martin in taking this as evidence that the indicator has a robust relationship with the growth turnaround in individual states.

Each of the regressions carried out in this exercise includes 5 regressors: the first is the variable of interest; the second and third are always the shares of agriculture and registered manufacturing (both of which are individually significant in regression (4)); while the remaining two regressors are picked from the set of remaining 8 regressors. We carry out a regression for every combination of two out of eight possible regressors: thus we run 28 regressions per indicator. (This is a rather modest number compared to the 30,856 regressions per indicator - 2 million in total - run by Sala-i-Martin). Given the strong mutual correlations between our regressors, 5 regressors virtually always captures the great majority of the state-wise variation: the implied F-test of the restrictions against equation (1) is not rejected at the 5% level for more than 9 out of 10 such regressions.

As a summary indicator of robustness we use the unweighted average CDF(0) proposed by Sala-i-Martin, where a number close to unity implies robustness irrespective of sign, while a number close to 0.5 indicates a roughly equal number of positive and negative coefficients. We use the unweighted average of individual CDF(0) values for each equation, which does not require us to assume normality. It is also more appropriate when, as here, most of the indicators are likely to be endogenous. The pitfalls of likelihood-based model averaging, as in Doppelhofer, Miller and Sala-i-Martin (2004), which can lead to a very high weight being placed on a very small number of regressions, have been analysed by Ciccone and Jarocinski (2010); but some at least of these problems had indeed already been alluded to in Sala-i-Martin's (1997) original paper, leading him to give more prominence to the unweighted average CDF(0) statistic.

The last two columns of Table 1 show the results, which help to shed some light on the disparate impact of the turnaround:

- One strongly significant individual effect, both in regression (4) and in terms of overall robustness is a negative impact of the sectoral share of agriculture in any given state. Note that this impact does *not* reflect any direct effect of the resulting high weight of agriculture in dampening growth of state NDP (given the relatively low growth rate of agriculture), since the regression results give each sector an equal weight. Rather it suggests that the mere fact that a state was predominantly agricultural was itself an obstacle to that state's participation in the turnaround in growth across all sectors.

- The only other individually significant coefficient in regression (4), which also appears to be extremely robust, is a negative impact of the share of registered manufacturing. This result directly contradicts those of Rodrik & Subramanian (2005). They posited that the impetus for the turnaround (which, it will be recalled, they dated significantly earlier), was a shift to a pro-business orientation, which they instrumented in their regressions by the share of registered manufacturing in aggregate state level data. Our results suggest that, far from having a positive effect on subsequent growth, a high share of registered manufacturing in any state just before our later estimated turnaround date actually appears to have had a significantly *negative* effect on growth in that state. Furthermore, GW Figure 5 showed that registered manufacturing was one of the very few sectors that actually grew less rapidly on average after 1987: this difference, as measured by a negative coefficient on the sector dummy in regression (3), is strongly significant. The fact that registered manufacturing appears to have played a significantly negative role in the turnaround is clearly more striking than if it simply played no role at all.

- The remaining state characteristics are all statistically insignificant in regression (4), but their CDF(0) values suggest a quite disparate degree of robustness when included in regressions with fewer regressors.

- On the positive side, literacy appears to be quite robustly correlated with participation in the growth turnaround; so to a lesser extent, does the degree of urbanisation. On the negative side, both landlocked and highly populated states appear to have been less able to participate in the turnaround. The former relationship is consistent with the well-documented problems with India's transportation system (Panagariya, Chapter 18, 2008), and the apparent link between the timing of the turnaround and the time profile of trade liberalization discussed in GW Section 4.

- The robustness results also appear to offer some, albeit limited, support to Aghion et al's (2008) firm-level analysis of the impact of the dismantling of the "Licence Raj". They found that in states where employment legislation was pro-worker (as proxied by a qualitative dummy variable), firms were less likely to be able to benefit from the reforms. In our regressions the coefficient on their dummy variable is always negative, but on average not strongly significant.

- It is also worth noting the state-wise regressors that do *not* appear to have a robust relationship with participation in the turnaround (indicated by CDF(0) values close to 0.5). These include state level income per capita in 1987 (thus counteracting claims that have been made that the turnaround has been restricted to a club of richer states), population growth and rainfall, as well as total development spending as % of NDP.[2]

3 A cross-check: general-to-specific modelling

It is interesting to note that the results of our robustness exercise deliver similar, though not identical results to a simpler econometric approach that has tended to fall out of favour in recent years: namely, an iterative process of general-to-specific modelling. In this approach (often referred to as the "LSE Approach" (Hoover & Perez, 1999) insignificant regressors are progressively eliminated until all remaining regressors are significant at some chosen significance level. This approach can been subjected to a data-mining based critique on the grounds that p-values in the final regression cannot be interpreted in classical fashion, since they have arisen in a "Darwinian" process of directed search. On the other hand, as noted by Hoover and Perez, this form of directed search should, in the limit, with sufficiently large datasets, converge on the true model (if such a true model exists). They report simulations (admittedly based on time series, rather than cross-sectional regressions) that for a quite range of different data generating processes, suggest that the true size of t-statistics in equations that have arisen from general-to-specific modelling is actually quite close to the nominal size (ie if a k% significance level is chosen, roughly k% of regressors will be included

[2] Wolcott and Clark (2003) also find that several disaggregated, though measurable, dimensions of state development spending on physical and social insfrastructure have little connection with economic growth in Indian states.

erroneously).

Regression (5) in Table 1 reports the results of an exercise of this type for our dataset. We set the threshold significance level for the marginal regressor relatively high, at 10%. Comparing the regression results with the robustness results in the remaining two columns shows that all the regressors included also have reasonably high CDF(0) values, and indeed for included regressors there is at least a rough correspondence between p-values and $1-\text{CDF}(0)$. Where the two approaches do differ is that the robustness exercise gives fairly high CDF(0) values for two indicators - urbanisation and population (level) that were eliminated from regression (5) in the testing down exercise. But this turns out to be fairly readily explicable, and indeed casts some further light on the robustness results. Both of these variables turn out to be highly collinear with other regressors (when regressed on the remaining regressors the equations have R^2 values of 0.9 and 0.95 respectively), thus when some of these are omitted in the regressions in the robustness exercise, the significance of both indicators is boosted.

4 Conclusions

The results summarised in this note should only be viewed as preliminary. The contrast between the robustness exercise and the model selection exercise is also a reminder that all we are looking at is correlations with participation in the growth turnaround, not necessarily the true determinants. Thus when a given indicator appears as robustly significant across a wide range of specifications, this may simply mean that it is more reliably correlated with the true, unobservable, determinants. Nonetheless our results are suggestive of future research avenues that might be pursued in investigating the ongoing - and very important - puzzle of why participation in the Indian growth turnaround has been so unevenly distributed across different states.

5 Bibliography

Aghion, P, Burgess, R, Redding, S and Zilibotti, F (2008) "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India," *American Economic Review*, 98(4), pages 1397-1412.

Ciccone, A and Jarocinski, M “Determinants of Economic Growth: Will

Data Tell?" *American Economic Journal: Macroeconomics,* 2(4), 222–246

Clark, Gregory and Susan Wolcott. (2003). "One Polity, Many Countries: Economic Growth in India, 1873-2000." In *In Search of Prosperity: Analytical Narratives on Economic Growth, (*Ed.) Dani Rodrik, Princeton University Press, New Jersey.

Doppelhofer, G , Miller R and Sala-i-Martin, X (2004) "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review*, 94(4): 813–35.

World Bank. 2004. World Development Indicators 2004. CD-ROM. Washington, DC: World Bank.

Ghate, Chetan and Stephen Wright (forthcoming). "The "V-Factor": Distribution, Timing, and Correlates of the Great Indian Growth Turnaround", *Journal of Development Economics*

Hoover, K and Perez, S (1999) "Data mining reconsidered: encompassing and the general-to-specific approach to specification search" *Econometrics Journal 2,*pp 167-191

Rodrik, Dani and Arvind Subramanian. (2005). "From Hindu Growth to Productivity Surge: The Mystery of the Indian Growth Transition." *IMF Staff Papers*, Volume 52(2), pages 193-228

Panagariya, Arvind. (2008). *India: The Emerging Giant.* Oxford University Press, New York.

Sala-i-Martin, X (1997) "I just ran two million regressions", *American Economic Review* Volume 87(2) pp178-183

Table 1: State Characteristics and the Growth Turnaround: Cross-sectional Regression Results

Dependent variable: Change in average log growth in state-sectoral real NDP per capita between 1970-87 and 1987-2004

	Coefficient estimates (*p* -values in parentheses)						
Regressors	(1)	(2)	(3)	(4)	(5)		
state dummies	all	all	none	none	none	Robustness Measures*	
sector dummies	all	none	all	all	all	CDF(0)	Dominant Sign
share of agriculture, 1987				-0.0014 (0.03)	-0.0014 (0.00)	1.000	negative
share of reg. manufacturing, 1987				-0.0036 (0.01)	-0.030 (0.00)	0.999	negative
real state income per capita, 1987				0.02 (0.41)		0.555	negative
% urban population, 1981				0.011 (0.62)		0.923	positive
literacy rate, 1981				0.006 (0.84)	0.021 (0.09)	0.963	positive
average rainfall, 1983-1987				0.0060 (0.47)		0.593	positive
Aghion et al's pro-worker dummy				0.0011 (0.82)	-0.0037 (0.08)	0.840	negative
landlocked dummy				-0.0145 (0.14)	-0.0136 (0.05)	0.980	negative
population, 1981				-0.0198 (0.33)		0.980	negative
population growth, 1971-1981				0.542 (0.73)		0.758	negative
development spending, % of NDP, 1981				0.071 (0.27)		0.786	positive
Regression Diagnostics							
Observations	207	207	207	207	207		
R-bar-squared	0.307	0.115	0.170	0.310	0.322		
s.e.	0.036	0.041	0.039	0.036	0.036		
intraclass residual correlation (states)	-0.035	-0.012	0.192	-0.034	-0.030		
intraclass residual correlation (sectors)	-0.099	0.158	-0.118	-0.100	-0.100		
Tests of implied restrictions on Equation 1 (*p* -values)							
Likelihood Ratio (Chi-Squared)	n/a	0.000	0.000	0.564	0.824		
Likelihood Ratio (F-Test)	n/a	0.000	0.000	0.999	0.051		
Wald (F-Test)	n/a	0.000	0.000	0.463	0.817		

* Unweighted average of individual CDF(0) (Sala-i-Martin, 1997) values in all possible regressions including variable, shares of agriculture and registered manufacturing, and two other regressors. A high value of CDF(0) indicates robustness, irrespective of sign.