The "V-Factor": Distribution, Timing and Correlates of the Great Indian Growth Turnaround*

Chetan Ghate[†] and Stephen Wright[‡]

May 5, 2010

Abstract

We anaylse a panel of output series for India, disaggregated by 15 states and 14 sectors. Using principal components (Bai, 2004; Bai & Ng, 2004) we find that a single common "V-Factor" captures well the significant shift in the cross-sectional distribution of state-sectoral output growth rates. The implied timing of the turnaround, in the second half of the 1980s, contrasts with previous research which has typically identified a significantly earlier turning point, but appears more consistent with the timing of policy reforms. We also provide some insights into the uneven distribution of the turnaround across Indian states.

JEL classifications: O10, O40, O53, O47,

*We are extremely grateful to Amit Sadhukhan for research assistance during the course of the project. We thank Dr. Savita Sharma and Pronab Sen of the Indian Central Statistical Office for helpful advice on the data. The co-editor, William Easterly, and two referees gave invaluable comments on our first draft. We also thank Gerhard Glomm, Sanghamitra Das, Samarjit Das, Abhiroop Mukhopadhyay, George Kapetanios, Ron Smith, and seminar participants at ICRIER, DIW Berlin, the Max Planck Institute - Jena, ISI Delhi, JNU, Institute of Economic Growth, the Delhi School of Economics, the 45th Meeting of the Indian Econometric Society (TIES), Jadavpur University, Claremont Graduate University, and Indiana University (Bloomington). Stephen Wright is grateful to the Indian Statistical Institute, Delhi, and the EGP group at the Max Planck Institute - Jena for hospitality during research visits in 2007 and 2008. Both authors are very grateful to the PPRU Committee for financial assistance related to this project.

†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

Keywords: Indian Economic Growth; Factor Models; Principal Components; Convergence; Divergence; Indian States.

1 Introduction

In the past two decades or so there has been a remarkable turnaround in Indian growth. From 1960 to 1987 output per capita in India (measured by real net domestic product[1]) grew by only 1.31% per annum, while on the same measure US output per capita grew at 2.36%, so that Indian and US output levels were steadily diverging. In marked contrast, from 1987 to 2004 Indian output per capita grew at 4.12% per annum, while US per capita growth slowed to 1.62%; thus India has been converging towards US output per capita levels at a more rapid rate than it was diverging in the earlier period. However a notable feature of the turnaround has been the distinctly uneven distribution of the growth turnaround across the major states, several of which have shown little or no increase in growth.

The turnaround in Indian economic growth has inevitably generated considerable public interest and some academic research with respect to its timing, possible causes, and unevenly distributed nature.[2] In this paper we present evidence on all three issues.

Our approach exploits the fact that, amongst economies at similar income levels, India's economy is unusually well provided with data. We utilize a new panel dataset, disaggregated into 15 major states and, within each state, into 14 broad industrial sectors, over the sample 1970-2004; we can also extend the dataset back a further ten years for a subset of ten states. We first show that the shift in growth has been highly pervasive across the Indian economy, in that there has been a shift in the cross-sectional distribution of growth rates of output per capita that is highly significant in statistical terms. We then use principal components analysis (following Bai and Ng, 2002; 2004 and Bai, 2004) to derive a common factor representation of the dataset. We show that a single common factor provides a powerful and parsimonious account of the distributional shift. This common factor is V-shaped, with an apex in the second half of the 1980s.

A significant advantage of this approach is that we do not need to impose a particular date for the turnaround in growth. Nor do we need to impose

[1] Throughout this paper we use net domestic product as our measure of output since the longest and most consistent output measures for India at both state and sectoral levels are on this basis. State-wise GDP data are only available from 1980.

[2] For example, see Rodrik and Subramanian (2005), Virmani (2006), Balakrishnan and Parameswaran (2007), Sen (2007), and Basu (2008).

that it be a deterministic shift, as in standard econometric representations of structural breaks; nor even that all series participate in the shift at identical dates.

The strong explanatory power of this common "V-Factor" suggests a single common cause. Our results appears to resolve the puzzle discussed by Rodrik and Subramanian (2005), who, along with other researchers, had concluded that the turnaround in growth came in the late 1970s or early 1980s, well before any significant observable shift in policy.[3] We find a later turnaround, in the second half of the 1980s, which is much more consistent with what is known about the pattern of liberalization (see Pursell (1992) and Panagariya (2004)). In particular, we show that the time profile of the V-Factor is strongly correlated with the pattern of trade liberalization, as summarized by the effective tariff rate. We emphasize our results on the tariff rate because it is the closest thing we have to an indicator of a true trade policy measure, rather than of an endogenous response to policy. But we also provide evidence on other trade and non-trade indicators that are consistent with the time profile of the V-factor.[4]

In a final section we use regression analysis to examine whether particular state characteristics can explain the very disparate performance across the states noted abve. There is some evidence that the capacity of a given state to exploit the opportunities presented by policy liberalization is helped by education and transport links, and hindered by the size of its agricultural and public sectors.

The remainder of the paper is structured as follows. In Section 2 we

[3]Rodrik and Subramanian identify a shift in growth in 1980, based on aggregate GDP data. Virmani (2006) and Balakrishnan and Parameswaran (2007) also identify shifts in the late 1970s/ early 1980s, but Basu (2008) identifies weaknesses in the methodology employed. We discuss the contrast between our results and earlier research at various points in the paper.

[4]Given the large body of literature that shows that the link between trade policy and economic growth is largely inconclusive, caution needs to be applied in interpreting our results. The openness debate is still active, particularly after the influential study of Rodriguez and Rodrik (2001) which showed that there is little conclusive evidence supporting a positive link between trade policy and economic growth. Harrison's (1996) review of the empirical work in this area prior to 1992 reports that, while in general, there is a positive association between openness measures and growth, these results are sensitive to a change in specification and on the choice of time aggregation. Yanikkaya (2003) shows that the measure of openness matters. Lee (1995) builds an endogenous growth model in which import intensity in the composition of capital increases growth directly by improving productivity. He finds that the import of capital goods, not total imports, is the key factor that links trade to economic growth.

provide some summary evidence of growth shifts at the sectoral and state levels. In Section 3 we carry out the statistical analysis and derive the factor representation. We examine the evidence for a shift in the second half of the 1980s, and contrast this with the results from earlier studies. In Section 4 we compare the path of the V-Factor with what we know about shifts in policy. In Section 5 we examine the differential impact of the turnaround across different sectors and states. Section 6 concludes the paper, and appendices provides details of data construction and statistical analysis.

2 Sectoral and state-wise shifts in growth

Figures 1 and 2 give two alternative broad-brush pictures of the turnaround in growth. We compare average sub-sample growth rates before and after 1987.[5] Figure 1 shows that virtually all sectors of the private sector economy have seen substantial increases in growth, albeit from often significantly different initial values.[6] Growth in the public sector, in contrast, actually slowed somewhat between the two sub-samples.

[Insert Figure 1, Sectoral and Statewise Data for Agg Economy mar 09.xls]

When the economy is divided into states, rather than sectors, the pattern is distinctly more disparate. Figure 2 shows output growth in the same two sub-samples for the 16 major states, which collectively represent 97% of the Indian population.[7]

[Insert Figure 2, Sectoral and Statewise Data for Agg Economy mar 09.xls]

[5] In our formal statistical analysis below we shall present the evidence for this particular year as a breakpoint, but the broad profile we present here is not sensitive to the precise sub-samples chosen.

[6] All growth rates are shown as growth of sectoral net domestic product per head of total state population, since no reliable figures for state-sectoral employment are available. The list of sectors shown is exhaustive - but some of the smaller sectors we include in our statistical analysis have been absorbed into broader definitions.

[7] We have made adjustments to output series to allow for changes in state definitions. The sixteen states are: Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Jammu and Kashmir, Kerala, Karnataka, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal.

The chart displays very clear dividing lines, both across time and across states, which are most revealing if expressed in terms of convergence towards the global frontier, which as in our discussion at the start of this paper, we proxy by the USA. Figure 2 also shows growth rates of the equivalent measure of US output per capita over the same sub-samples. Using this as the benchmark, only three Indian states, Haryana, Punjab and Orissa, showed any tendency to even marginal convergence in the first sub-period: they would be better described as just holding their own.[8] The remaining states were all growing less rapidly than the frontier - indeed some, like Madhya Pradesh, were barely growing at all - so that almost all were actually diverging systematically from the global frontier.

For the majority of states the contrast in the second period could hardly be any more striking. Nine states (Andhra Pradesh, Gujarat, Haryana, Karnataka, Kerala, Maharashtra, Rajasthan, Tamil Nadu and West Bengal) had per capita growth rates in the neighborhood of 4% to 5%, and were thus unambiguously converging; two others, Madhya Pradesh and Jammu & Kashmir, achieved significant shifts in growth, but from such a low base that they were still at best barely converging (partly due to a somewhat lower rate of growth in the USA). In the remaining states, however, growth remained at a similar rate to that in the previous sub-period. Within this group three states, Punjab, Orissa and Uttar Pradesh did achieve modest rates of convergence; but Assam and Bihar continued to lose ground.

Since Indian citizens live in states rather than industrial sectors, this very disparate pattern has significant welfare implications. While we have only imperfect data on state wise consumption (and this only on an infrequent basis over time), such data that can be constructed suggest a strong link with state wise output. In 2004, for example, the cross-sectional correlation coefficient in logs between estimated state consumption per capita and net state output per capita was 0.88,[9] so differences in growth rates of output growth will have corresponded to significant differences in consumption growth.

[8] Of these three states, closer inspection of the data shows that the fastest growing state, Orissa, had shown extremely rapid growth during the 1960s, but thereafter showed no tendency to converge.

[9] Both consumption and output are measured at current prices. Details of data construction for consumption are in Appendix A.4.

3 Statistical Analysis

3.1 The dataset

We analyze a panel dataset of output per capita series broken down both by state and by sector. For fifteen major states (the same group shown in Figure 2, excluding Jammu & Kashmir) we have a sectoral breakdown into fourteen broad industrial sectors, from 1970 to 2004; for a subset of 12 states (also excluding Assam, Bihar and Orissa) we have the same sectoral breakdown from 1965, and for 10 states (also excluding Haryana and Punjab) from 1960. We eliminate three series due to clear data problems, leaving 207 series over a balanced panel from 1970 to 2004, 166 series from 1965, and 139 series from 1960. All series are measured in constant prices per head of the population in the relevant state.[10]

3.2 Evidence of common structural shifts?

While the visual evidence in Figures 1 and 2 appears very striking, at least in principle it is possible that this pattern could emerge from shifts in a relatively small number of the underlying series in our dataset. However, examination of the full dataset shows the pervasive nature of the shift. Figure 3 shows the observed distribution of average log growth rates of all series in the panel with the maximum cross-sectional dimension (207 series) over two samples, 1970 to 1987 and 1987 to 2004. The visual evidence of a clear systematic rightward shift in the cross-sectional distribution is strongly supported by statistical testing.

[Insert Figure 3. distribution of sub-sample growth rates, march 2010.xls]

Table 1 shows the results of Kolmogorov-Smirnov (KS) tests of the null that both sets of growth rates are drawn from the same distribution. The tests are carried out using two sets of data: sub-sample average growth rates of sector specific output from 1970 to 1987, and 1987 to 2004; and annual sectoral growth rates, i.e., each observation of the annual growth rate of a given series from 1970 onwards is considered as a separate observation,

[10]Full details of data construction are given in Appendix A.

thus greatly increasing the number of observations. As shown in the chart, and on the underlying annual series: both show equally strong rejections of the null against the alternative that the distribution in the second sub-sample stochastically dominates that in the first. Thus without putting any structure on the underlying data generating process being assumed, there is strong statistical evidence of some form of common shift in growth that is pervasive across the cross-sectional distribution.[11] Examination of tests carried out over a range of sub samples suggest that this result is not simply an artefact of the breakpoint chosen.[12]

Table 1
KS Tests for Equality of Distribution Functions from 1970 - 2004[13]

H_A	D Statistic (ss)	D Statistic (ann)	P Values (ss)	P Values (ann)
0	0.2714	0.1114	0.000	0.000
1	0.000	-0.0011	1.000	.995
Combined K-S	0.2714	0.1114	0.000	0.000

[11] The null assumes independence of all observations, which in the panel context implies both serial and cross-sectional independence. The former assumption is reasonable in the context of average growth rates since the underlying annual figures have only low temporal persistence which essentially disappears across sub-samples; [it is less justifiable for the test as applied to the annual series, hence these should be interpreted with caution.] The cross-sectional independence assumption is precisely the element in the null hypothesis that we are interested in rejecting, since its violation implies a common element to the shift.

[12] We report some of these results here. We have a balanced panel for a subset of 12 states from 1965 onwards, and for 10 states from 1960 onwards. Using sub-sample averages for the states with data from 1965 onwards (i.e., sub-sample average growth rates based on 1965-87 and 1987-2004)), the D-statistic for the combined K-S test is .3214 with a P-value of 0.000. Using sub-sample averages for the states with data from 1960 onwards (i.e., sub-sample average growth rates based on 1960-1987 and 1987-2004)), the D-statistic for the combined K-S test is .3857 with a P-value of 0.000. Both results strongly reject the null of equality of distributions when the breakpoint is 1987. The results of other breakpoint tests are available from the authors on request.

[13] The D Statistic (ss) in the second column is based on the sub-sample growth rates: 1970-1987 and 1987-2004. The D-statistic (ann) in the third column is for annual growth rates (i.e., using each observation of the annual growth rate of a given series as a separate observation, thus greatly increasing the number of observations). To ensure that we have a balanced panel, we have only used data from 1970 onwards. 0 indicates that we test the null against the alternative hypothesis that the second period stochastically dominates the first. 1 indicates a test against the alternative that the first period dominates the second. Combined K-S is a test against the general alternative that the two distributions are not equal.

3.3 A Common Factor Representation

We can put more structure on the shifts identified in the previous section by assuming that the dataset can be given a common factor representation, on the assumption that the factors will capture the common element in the shift in the distribution shown in Figure 3. This approach has the advantage that we need make no prior assumptions on the timing of such shifts.

Following Bai (2004) and Bai and Ng (2002; 2004), we assume that longer-term trends in the underlying output series can be captured by a relatively small number of common factors that determine permanent (i.e., unit root) movements, i.e., a representation of the form,

$$\begin{aligned} y_{it} &= \beta_{i0} + \beta_{i1} F_{1t} + ... + \beta_{ik} F_{kt} + u_{it}; i = 1..N && (1) \\ \Delta F_{kt} &= a(L)\,\varepsilon_{kt}; \quad k = 1..k && (2) \\ u_{it} &= b(L)\,\omega_{it}; \quad i = 1..N, && (3) \end{aligned}$$

where y_{it} is log output per capita in state-sector i (i.e., we do not explicitly distinguish between the state and the sector dimension); the F_{kt} are common factors that are subject to permanent shocks, ε_{kt}; the β_{ik} are factor loadings on the factors; and the u_{it} capture the remaining transitory dynamics. We assume that both $a(L)$ and $b(L)$ are stationary polynomials in the lag operator (defined such that for any variable x_t, $Lx_t = x_{t-1}$), so that (consistent with Bai, 2004) the factors are at most $I(1)$ and the transitory components are $I(0)$).

Bai (2004) shows that as long as the u_{it} are $I(0)$, then consistent estimates of the common factors (or rotations thereof), and of the factor loadings, can be derived from the application of static principal components analysis.[14] For robustness, we also consider the alternative approach in Bai and Ng (2004) which is consistent even when the u_{it} are non-stationary. In this approach principal components analysis is applied to first differenced data, and the resulting factors are cumulated. In both approaches information criteria originally proposed in Bai and Ng (2002) provide consistent estimates of r, the true number of common factors; Bai (2004) derives modified versions of these criteria for estimation in levels.

In neither approach is it necessary to estimate the parameters in $a(L)$ or

[14] The transitory shocks, ω_{it}, may in principle be mutually correlated but Bai (2004) outlines restrictions on the nature of this correlation.

$b(L)$. Principal components provides estimates $\widehat{F}_{kt}$ of the factors and factor loadings $\widehat{\beta}_{ik}$,[15] and the transitory components in (3) are derived from these estimates, as

$$\widehat{u}_{it} = y_{it} - \left(\widehat{\beta}_{i0} + \widehat{\beta}_{i1}\widehat{F}_{1t} + ... + \widehat{\beta}_{ik}\widehat{F}_{kt}\right). \tag{4}$$

Bai and Ng (2004) then propose that panel unit root tests be applied to the implied transitory components to check the validity of the stationarity assumption, on the assumption that cross-sectional dependence has been largely or entirely captured by the common factor representation.

In Table 2 we show the results of using Bai and Ng's information criteria to identify k, the number of common factors in our dataset, which minimizes the relevant information criterion. The additional argument for each criterion, $k_{\max}$ is the maximum value of k considered, which is also used to derive an estimate of the average of the variances of the idiosyncratic components which feeds into the penalty function.[16] As in Bai (2004) and in a number of subsequent studies (see, for example, Kapetanios, 2004), the value of k identified by information criteria is known to be sensitive to the value of $k_{\max}$ chosen, with a lower value of $k_{\max}$ usually resulting in a lower estimate of k : Table 2 shows that this feature is also clearly evident in our dataset.

Table 2

Value of k, the Number of Common Factors, implied by Information Criteria[17]

		Panel Information Criteria											
		Estimation in Levels						Estimation in Differences					
		IPC_1		IPC_2		IPC_3		IPC_1		IPC_2		$IPC3$	
Sample	N	k	$k_{\max}$	k	$k_{\max}$	k	$k_{\max}$	k	$k_{\max}$	k	$k_{\max}$	k	$k_{\max}$
1960-2004	139	1	<5	1	<5	1	<10	0	<3	0	<5	0	$<\infty$
1960-2004	139	2	≥ 5	2	≥ 5	2	≥ 10	1	≥ 3	1	≥ 5	0	$<\infty$
1965-2004	166	1	<4	1	<4	1	<10	0	<4	0	<5	0	$<\infty$
1965-2004	166	2	≥ 4	2	≥ 4	2	≥ 10	1	≥ 4	1	≥ 5	0	$<\infty$
1970-2004	207	1	<5	1	<6	1	<12	0	<5	0	<6	0	$<\infty$
1970-2004	207	2	≥ 5	2	≥ 6	2	≥ 12	1	≥ 5	1	≥ 6	0	$<\infty$

[15] Hence, given that we also estimate means for each series, the total number of parameters estimated is $3N$.

[16] See Bai (2004), p. 145.

[17] Information criteria for estimation in levels are as defined in Bai (2004) equation (12), which are modified versions of the criteria in Bai and Ng (2002).

The table shows a clear contrast between the number of factors identified by estimation in levels, compared to estimation in differences, with levels estimation always implying one more factor. But this is to be expected. Since most series in our dataset are strongly trending, we would expect that the first principal component in levels would be dominated by this trend element (as indeed our results show below), with the second principal component picking up common stochastic shifts in trends. In contrast, for estimation in differences all deterministic trend growth in levels is extracted by demeaning the differenced data before extracting principal components, so that the first principal component in differences can play the same role in picking up common shifts as does the second principal component in levels.

A more significant form of ambiguity is that, for low values of $k_{\max}$ (and, in the case of IPC_3 for estimation in differences, for all values of $k_{\max}$) the information criteria suggest only a single common factor in levels, and no common factor in differences. Taken at face value the a representation with no common factors in differences would imply that each of the series in the panel was simply an independent unit root process. However we would argue strongly that this possibility can be dismissed on two grounds: first, the Bai and Ng information criteria are known to yield ambiguous results, and to have low power to distinguish common factors in relatively noisy processes (Kapetanios, 2004); second, and more crucially, we have already seen very strong evidence for a common shift in the distribution of growth rates, in Table 1: the rejection of a common distribution by the Kolmogorov-Smirnov test is thus indirectly a rejection of a zero-factor representation in differences.

We therefore focus our attention on the results from estimation in levels with two factors, and from estimation in differences with a single factor. In contrast with some previous studies, we do not find that the estimated value of k rises further as we increase $k_{\max}$, hence we can feel reasonably confident that such a low order factor representation will be sufficient (we shall see that this confidence appears to be borne out by the explanatory power of the factor representation).

In Appendix B we show that if we construct the implied transitory components from the two factor levels model and the single factor differences model, in both cases panel unit root tests strongly reject the unit root null, thus the assumption of stationary idiosyncratic components appears consistent with the data.

3.4 Factor Estimates: the "V-Factor" and the "G-Factor"

To illustrate the nature of the results, Figure 4 shows the two common factors derived from the first two principal components from estimation in levels, alongside the single common factor derived by cumulating the first principal component from estimation in differences,[18] over the sample period 1970-2004, which gives the maximum cross-sectional dimension of 207. Results for the longer samples, with smaller cross sections, are very similar (see Appendix, Figure A3).

[Insert Figure 4, pc comparisons march 2010.xls]

As discussed above, the first common factor from levels estimation is very close to being a deterministic trend; the different factor loadings of individual series on this component thus proxy for nearly constant deterministic growth rates. We therefore term this component the "G-Factor". The second component, which captures shifts in growth, we term the "V-Factor". Figure 4 shows that the pattern of the V-Factor closely parallels the pattern of divergence from the global frontier during the period of the "Hindu Rate of Growth", followed by subsequent convergence, as discussed in the Introduction. Factor loadings of individual series on the V-Factor capture the extent to which each series has participated in the turnaround. The profile of the V-Factor is quite close to being monotonic either side of its vertex in the second half of the 1980s. In Appendix D we show that the timing of this breakpoint is unaffected by a lengthening of the sample backwards with a smaller subset of states; it also appears to be robust, to within a year or at most two years, to the inclusion or exclusion of series using a range of criteria.

The chart also shows the single common factor derived from estimation in differences. For most of the sample it shows a very similar pattern, albeit with a less distinct apex (it is closer to being a U-Factor than a V-Factor). This weaker identification of the turnaround is consistent with Monte Carlo evidence presented in Appendix G. This suggests that estimation in differences is systematically both significantly less reliable in identifying common

[18] Since the scale of the factors is irrelevant, all three series are normalised to have zero mean and unit variance.

breakpoints, and less robust. For the rest of the paper we therefore focus on results based on levels estimation.

As noted at the start of the paper, a very significant advantage of this representation is that we do not need to impose a particular date for the turnaround in growth. Nor do we need to impose that it be a deterministic process (as in standard econometric representations of structural breaks); nor even that all series participate in the shift at identical dates (since the representation of the transitory components for individual series allows in principle for different persistence properties, which allow some series to respond more rapidly to the common permanent shock).

3.5 The V-Factor as a representation of growth shifts

Figures 5 and 6 provide a summary illustration of the extent to which the common factor representation captures the key properties of the common shift in growth.

In Figures 1 and 2 we showed the strong evidence of a shift in growth rates in the cross sectional distribution of both sectoral and state growth rates. In Figures 5 and 6 we aggregate up the fitted values for the change in growth rates in individual series from our factor representation (where the fitted values for each series are solely driven by the two factors, weighted by their factor loadings) and compare them with the average actual change in growth rates, by sector (Figure 5) and by state (Figure 6).[19] The charts show that the two common factors alone provide a good parsimonious representation of the observed growth shifts (the correlation coefficient between actual and fitted values is 0.83 for sectoral averages, 0.96 for state averages, and 0.82 for all series taken together). Furthermore, this explanatory power is essentially entirely due to the V-Factor: a factor model in levels with only

[19] For individual series, the actual change in (log) growth is defined by

$$D_i = \frac{y_{i,2004} - y_{i,1987}}{17} - \frac{y_{i,1987} - y_{i,1980}}{17}$$

while the fitted change in growth is defined by

$$\widehat{D}_i = \sum_{k=1}^{2} \beta_{ik} \left(\frac{F_{k,2004} - F_{k,1987}}{17} - \frac{F_{k,1987} - F_{k,1980}}{17} \right)$$

Figures 5 and 6 then show unweighted averages, across sectors and states respectively, of the D_i and the $\widehat{D}_i$.

a single common "G-Factor" yields a correlation coefficient between actual and fitted insignificantly different from zero (as we would expect since such a model essentially implies nearly constant growth).

[Insert Figure 5 actual and permanent correlations, mar 09.xls and Figure 6, actual and permanent correlations, ranked by state,mar 09.xls]

Figures 5 and 6 make clear that the impact of the V- factor is highly pervasive but at the same time by no means universal, or indeed universally positive. The average impact on both sectors and states more or less corresponds to the summary pictures of sectoral and state wise growth shifts shown in Figures 1 and 2 (with the discrepancies largely due to weighting differences since the averages shown in Figures 5 and 6 are simple averages across states and sectors of very different sizes).

Thus Figure 5 confirms the message of Figure 1 that, on average (i.e., across the 15 states), almost all of the 14 sectors analyzed have been positively affected by the common shift in growth (we discuss the exceptions below). But Figure 6 also shows the disparate performance across states, with basically the same group of states being left out of the pickup in growth, at least in terms of its average effect, as illustrated in Figure 2.

3.6 How precisely can we date the turnaround?

The V-Factor estimated by our preferred technique of principal components in levels has a turning point in 1987. We show in Appendix D that, [to within a year or at most two], this date emerges consistently from the dataset, whichever sample is chosen, and whether or not volatile series are excluded from the panel. This result is in contrast with a range of past studies that concluded, on the basis of aggregate data, that the turnaround occurred distinctly earlier: Rodrik and Subramanian (2005) identify a breakpoint in the early 1980s or late 1970s; Virmani(2006) in 1980-81 (manufacturing) and 1981-82 (total GDP); while Balakrishnan and Parameswaran (2007)) identify a breakpoint as early as 1978-9.

An obvious question therefore arises: how much statistical significance should we place on our results? In Appendix G we carry out a simulation study that sheds some light on this issue. We simulate artificial samples of data that are calibrated to have similar properties to the actual dataset, in terms both of the typical growth path of the component series, their dispersion, and, most crucially, the proportion of the variance of the total dataset that is captured by a representation with a simulated G-Factor and V-Factor. In Table G1 in the Appendix we show that in such simulated datasets our preferred estimation procedure correctly identifies the "true" breakpoint, to within one year either side, in between two thirds and three quarters of our simulations, depending on the specification.

Thus our estimation technique is (unsurprisingly) by no means 100% accurate in identifying the timing of breakpoints, implying that we should be cautious in placing too much emphasis on the significance of any particular year. In Appendix D we also present evidence that suggests that the sharpness of the apex in the V-Factor in 1987 may arise from short-term volatility in a relatively small number of series within agriculture, forestry and fishing; once these are excluded the V-factor has a somewhat smoother profile, with an apex a year or so later. Nonetheless, the simulations suggest that the technique is *sufficiently* accurate that it should allow us to discriminate fairly well between breakpoints as distant in time as those we find in our actual dataset, and those identified in past research. When we simulate a dataset of 139 series starting in 1960 (as in our longer sample of ten states), in which the *true* breakpoint is in 1979, our simulations show that the probability of identifying a breakpoint in 1987 or later, as in our dataset, is only around 3%. Thus we would conclude that our finding of a breakpoint at some point in the second half of the 1980s (with a reasonably well identified central estimate of 1987) is both robust and significantly different from the results of past research.

How can we reconcile our results with those from past research? Basu (2008) notes the crucial role of a single year, 1979-80 (largely due to a sharp fall, then sharp recovery, in agricultural output) in affecting inferences based on aggregate data. This year also shows up strongly in our disaggregated approach, however our results are much less affected by this particular year, since agriculture is weighted equally with all other sectors. As shown in Figure 4, both our estimates of the V-factor show a sharp fall in 1979-1980; but then continue to fall, only reversing this decline in the second half of the 1980s. The later turnaround captured by the V-Factor is thus representative of a shift that was much more pervasive throughout the economy.

4 The V-Factor and Economic Policy

The contrast between our results on the timing of the turnaround and those of earlier research is of particular interest, since it suggests a resolution of a puzzle discussed by Rodrik and Subramanian (2005): while they, in line with most other research, identified a turning point in the late 1970s or early 1980s, this appeared significantly to pre-date major policy changes. Is the later turning point we identify in the V-Factor more consistent with what we know about the timing of economic policy?

Figure 7 shows that the time path of the V-Factor matches very well indeed the timing of one key policy change: the liberalization of trade policy via tariff reduction (the blue line). While the gradual liberalization of trade policy began as early as the late 1970s, these changes were pretty minimal until the mid eighties (Pursell (1992) and Panagariya (2004)), and consisted entirely of a gradual relaxation of quantitative controls. In particular, in 1980, imports were divided into three categories: banned, restricted, and Open General License (OGL) with the goods in the last category not requiring any license. The OGL list kept expanding over time. Initially, the OGL only had 79 capital goods. By 1988, 1170 capital goods and 949 intermediate goods were covered. By 1990, 30% of all imports were covered (Panagariya, 2004). However, countering this, until the mid-1980s there were significant increases in tarrifs on goods that had been banned or restricted earlier. The tariffs on goods in the restricted list also increased. Panagariya (2004) attributes this to the government capturing the quota rents - implying that protection became more efficient, but without any clear-cut overall liberalisation. This version of events is consistent with Das's (2003) data on the import coverage ratio (a proxy for non-tariff barriers) in manufacturing, which measures the proportion of products banned/restricted, limited or canalised. This shows a modest fall through the 1980s, but much steeper falls thereafter. Thus, Figure 7 suggests that either the *net* effect of these changes was negative until tariff rates themselves started to fall, or that there were lags, or some combination of the two.[20]

[20] Since reforms have announcement effects (i.e., once an economy wide reform is announced, forward looking investors would modify their investment decisions prior to the actual legislative enactment of the reform), the apex of the V might conceivably be before de jure changes in the aggregate policy regime. Panagariya and Pursell do suggest that reforms had been progressing for several years, so we do not really need to plead anticipation.

[Insert Figure 7 V Factor and Duties, April 2010.xls]

We emphasize our results on the tariff rates because they represent a clear-cut and measurable change in policy, rather than an endogenous response to policy changes, and therefore tell the most useful story in terms of causality. However, we have also examined a series of other policy indicators (both trade and non-trade) and their time profile relative to the V-factor. Some changes such as quota liberalizations applied primarily to registered manufacturing which the evidence of Figure 5 suggests was actually negatively affected by the V-factor. Variables such as the log openness ratio (exports + imports as percentage of GDP) also exhibit a fairly sharp increase in 1987. The time profile of duties as a percentage of GDP also exhibits a sharp decline in the mid 1980s, falling 13% between 1985 and 1991, supporting the time profile of the effective tariff rate in Figure 7.[21] Figure 6 in Rodrik and Subramanian (2005) is particularly noteworthy. India's real effective exchange rate (REER) shows a marked real depreciation of more than 40% in the second of half of the 1980s (see Rodrik and Subramanian, 2005, p.210), with the export subsidy adjusted REER showing even a more marked decline in 1987. The real depreciation would have had a significant short term growth effect (see Rodrik and Subramanian, 2005, p. 211), and the timing of the shift is also broadly consistent with the time profile of the V-factor.[22] Finally, in terms of non-trade policy indicators, there is a significant relaxing of the licence raj during the 1980s and 1990s (Aghion et al., 2008). A third of three digit industries were exempt from licensing in 1985 (Aghion et al., 2008, p.1398). Since the licensing system was acts as a barrier to entry, de-licensing would result in a sizeable re-allocation of industrial production from states with pro-worker labor institutions to states with pro-employer institutions, accentuating the importance of labor regulation in determining the trajectory of industrial activity (and increases in output) in India. We provide some evidence below, in Section 5, that this may at least partially explain the uneven impact of the turnaround across

[21] Both the effective tariff rate and duties as a percentage of appear consistent with other evidence derived from tariff rates, rather than revenue: for example the five-yearly estimates of the effective rate of protection calculated by Deb Kusum Das (2003), based on manufacturing tariffs, show a rise in the second half of the 1980s relative to the first half, but a sharp decline thereafter.

[22] The deeper and more systematic liberalization a few years later in 1991, in which there was a reduction of tariffs on most goods (other than consumer goods) further sustained the shift in trend growth (Panagariya, 2004).

the states.

In sum, the progressive reduction in tariffs was not the only policy change introduced during the period of liberalization, but both the strength of the link with the V-Factor and other evidence on trade and non-trade policy indicators does suggest it had a particularly important role.

5 Participation in the turnaround: some regression results

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 Figure 2) is quite striking. In this section we provide some regression-based evidence that sheds at some light on the issue (although, as we shall show, some puzzles remain).

The factor representation both identifies strong evidence of a common element in the growth turnaround, and provides at least a reasonably reliable estimate of its timing (as discussed in Section 3.6), in the latter half of the 1980s. In Table 3 we present some evidence on the determinants of the 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).

For purposes of comparison, the first three columns report regressions where the only regressors are dummy variables for each sector and state. Consistent with the evidence of 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)).[23]

In regressions (4) and (5) we investigate whether identifiable state char-

[23] The predicted change in the growth rate for each series in the panel in regression (1) is thus the sum of the sector and state dummy. Given the power of the V-Factor as a representation of the common element in the growth shift, as demonstrated in Figures 5 and 6, it is unsurprising that this predicted value is strongly correlated with the factor loading of each component series on the V-Factor. Regression results where the dependent variable is the state-sector factor loading are accordingly very similar.

acteristics can account for the disparate performance across the states. We retain the sectoral dummies, but include up to 11 different state characteristics (all either time-invariant, or measured just before the turnaround), in place of the state dummies.[24] Most are (unsurprisingly) individually at best only marginally significant, but the overall goodness of fit is comparable to that of the benchmark regression (1) with both sets of dummies. In regression (5) we narrow the list of state-level regressors down by a standard specification search, retaining only those that have notional p-values of 10% or less. In equation (6) we do a similar specification search for the sectoral dummies, retaining ony those with p-values of 10% or less (the coefficients on this subset of sectoral dummies are shown at the foot of the table). Finally, in regression (7), we include both this subset of sectoral dummies, and the 6 state dummies which had p-values of 10% or less in regressions (1) and (2). The table shows that, once state characteristics are included, the state dummies are both individually and jointly insignificantly different from zero: thus the state-level regressors, at a minimum, span the informational content of the state dummies.[25]

The actual statistical significance of this subset of regressors is questionable in terms of classical hypothesis testing, given the obvious data mining critique. It should also be borne in mind that the available indicators may be imperfect proxies for unobservable state characteristics; and given that the indicators are quite strongly mutually correlated, it would not be wise to place excessive weight on any given indicator. Nonetheless both the list of included and excluded regressors sheds some interesting light on the disparate impact of the turnaround. Our main findings can be listed as follows.

- The most significant individual effect is a negative impact of the sec-

[24] We cannot include a full set of both state dummies and state characteristics, since in a cross sectional regression the resulting matrix of regressors would be singular. Note that there are no obvious sectoral regressors that would allow us to carry out a similar exercise across the sectors.

[25] All regressions report intraclass residual correlation coefficients, as an indication of whether clustering is likely to lead to OLS standard errors understating true standard errors, when these correlations are positive (see Angrist and Pischke, 2009). All are close to zero, and negative, with the exception of regressions (2) and (3), in each of which one set of dummies is excluded, which leads to a modestly positive intraclass correlation for the class for which the dummies is omitted. Thus it appears that the sector dummies, which are retained throughout, are sufficient to capture any intraclass correlation within states, so that uncorrected standard errors can be used. If White robust standard errors are used all state level regressors in regressions (5) and (6) remain significant at the 5% level, with the exception of literacy.

toral 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.

- Two other state characteristics appear to have played a similar role. On the positive side, a higher level of literacy appears to have boosted state-level growth (albeit with only a marginal degree of notional significance); on the negative side, landlocked states appear to have been less able to participate in the turnaround. This latter 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 Section 4.

- The regression results also appear to offer support to the results of 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; the negative coefficient on their dummy variable in regressions (5) to (7) is consistent with these results.

- In contrast, our regression results appear directly to contradict 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 regressions suggest that, far from having a positive effect on subsequent growth, a high share of registered manufacturing in any state just before the turnaround actually appears to have had a significantly *negative* effect on growth in that state. Furthermore, Figure 5 showed that registered manufacturing was one of the very few sectors that actually grew less rapidly on average after 1987: Table 3 shows that this difference, as measured by the sector dummy, 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.[26]

- The role of public sector output in the turnaround is quite distinctive. Figure 5 showed that overall it was the slowest growing sector (with a significantly negative sector dummy, shown in Table 3). But there is also an interesting contrast between our regression based results and the role of the V-factor. For all other sectors, more rapidly growing states tended to have higher growth across all sectors: hence for any given sector, correlations across states between V-factor loadings for that sector and the state dummies derived from our regressions are all positive, and mostly strongly so. But this is not the case for the public sector: indeed the correlation is marginally negative, suggesting that if anything states where non-public output grew more rapidly tended to have less rapid growth of the public sector.[27]

- Finally it is worth noting the state-wise regressors that do *not* appear to have any explanatory power. 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), and demographic and climatic variables, as well as total development spending as % of NDP (the latter providing further evidence of the limited role of the public sector).[28]

[26] []We can only really speculate about the explanation for the negative correlation. Our best guess is that it ties up with the negative role of the public sector in general. Panagariya (2004) makes the forceful point that even in recent years government intervention in registered manufacturing remains extensive. If the bulk of the capital in the manufacturing sector is owned by the public sector, this makes it immobile (Marathe, 1986), keeping capital-output ratios inefficiently high. Also, if there are restrictive labour laws, private and public firms cannot fire their employees, and so inefficient labour continues to be employed (see Bhattacharjea, 2006), leading to output losses.

[27] This is because the recorded output of the public sector is largely driven by public sector wages. In an earlier working paper version of this paper, Ghate and Wright (2008) show that many Indian states since the mid 1980s increased public spending on manpower (employment) and decreased capital investments. Such revenue expenditures are symptoms of poverty, as supply constrained economies reduce capital investments and raised public sector wages for political reasons (see Ghate, 2008).

[28] 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.

6 Conclusions

In their international study of growth accelerations, Hausmann, Pritchett and Rodrik (2005, p. 328) conclude that:

> "It would appear that growth accelerations are caused predominantly by idiosyncratic, and often small-scale, changes. The search for the common elements in these idiosyncratic determinants—to the extent that there are any—is an obvious area for future research."

We believe that this paper provides evidence of such common factors in the context of the Indian economy; we hope that the techniques we employ may inform future investigations both of the Indian and other economies.

We have presented evidence of a common "V-Factor", derived from principal components of a panel of Indian output per capita series disaggregated by state and by sector, that appears to capture well a systematic and pervasive shift in growth rates during the 1980s. The timing of the V-Factor is more consistent with the history of Indian policy reform than previous studies, such as Rodrik and Subramanian (2005), that have dated the turnaround to the beginning of the 1980s or even earlier. Our results suggest a particularly important role for trade liberalization. We also provide some evidence that the capacity of a given state to exploit the opportunities presented by policy reforms were helped by education and transport links, and hindered by the size of its agricultural sector. We find no evidence that public sector output or development spending played any role in the turnaround, and some evidence that sectors where government intervention remained significant (most notably in registered manufacturing) participated less in the turnaround.

7 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.

Angrist, J and Pischke, J (2009), *Mostly Harmless Econometrics*, Princeton, New Jersey.

Bai, J. and Pierre Perron. (1998). "Estimating and Testing Linear Models with Multiple Structural Changes". *Econometrica*, Vol. 66(1), pages 47-78.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics*, 122, pages 137-183.

Bai, Jushan & Ng, Serena (2002). "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, vol. 70(1), pages 191-221.

Bai, Jushan & Ng, Serena (2004). "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, vol. 72(4), pages 1127-1177.

Balakrishnan, Pulapre, and M. Parameswaran. (2007). "Understanding Economic Growth in India: A Prerequisite", *Economic and Political Weekly*, July 14, pages 2915-2922.

Barro, Robert and Xavier Sala-i-Martin (1992). "Convergence", *Journal of Political Economy*, 100(2), pages 223-251.

Basu, Kaushik (2008). "The Enigma of India's arrival: A Review of Arvind Virmani's Propelling India: From Social Stagnation to Global Power", *Journal of Economic Literature*, Vol. 46(2), pages 396-406.

Bhattacharjea, A. (2006). "Labor Market Regulation and Industrial Performance in India: A Critical Review of the Empirical Evidence." CDE Working Paper No. 141.

Campbell, John (1994) "Inspecting the Mechanism", *Journal of Monetary Economics,* 33(3), pages 463-506.

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.

Das, Deb Kusum. (2003). "Quantifying Trade Barriers: Has protection declined substantially in Indian Manufacturing?", ICRIER Working Paper, No. 105.

Economic and Political Weekly Research Foundation State dataset, 2005, Mumbai.

Ghate, Chetan. (2008). "Understanding Divergence in India: A political economy approach," *Journal of Economic Policy Reform*, Vol. 11, pages 1-9.

Ghate, Chetan and Stephen Wright (2008). "The "V-Factor": Distribution, Timing, and Correlates of the Great Indian Growth Turnaround", DIW Working Paper No.783, Berlin.

Hausmann, R., Pritchett, L., and Dani Rodrik, (2005), Growth Accelerations, *Journal of Economic Growth,* Vol.10, pages 303-329.

Harrison, A. (1996). "Openness and Growth: A Time Series, Cross Country Analysis for Developing Countries," *Journal of Development Economics*, 48(2), pages 419-447.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, pages 53-74.

Islam, Nazrul (1995) "Growth Empirics: A Panel Data Approach" *The Quarterly Journal of Economics*, Vol. 110, No. 4. (Nov., 1995), pages 1127-1170.

Lee, J.W. (1995). "Capital Goods Imports and Long Run Growth," *Journal of Development Economics*, 48(1), pages 91-110.

Jones, Benjamin and Olken, Benjamin (2008) "The Anatomy of Start-Stop Growth", *Review of Economics and Statistics,* August 2008, 90(3) pages 582-587.

Kapetanios, G. (2004) "A new method of determining the number of factors in factor models with large datasets", Queen Mary College, University of London, *Working Paper* no. 525.

Marathe, S. (1986). "Regulation and Deregulation". Sage Publications, New Delhi, India.

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. (2004). India's Trade Reform. *India Policy Forum*, Vol, 1, pages, 1-57.

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

Pursell, Garry. (1992). "Trade Policy in India," In *National Trade Policies*, (ed) Dominick Salvatore, New York: Greenwood Press.

Temple, J.R.W. (2005). Growth Econometrics. In P. Aghion and S.N. Durlauf (eds.), Handbook of Economic Growth, Volume 1A, North- Holland:Amsterdam, 2005, pages 555-677.

Rodriguez, F., and D. Rodrik (2001). Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence. In: Bernanke,B., Rogoff, K. (Eds.), *NBER Macroeconomics Annual.* National Bureau of Economic Research, Cambridge, MA.

Sen, Kunal. (2007). "Why did the Elephant start to Trot? India's Growth Acceleration Re-examined." *Economic and Political Weekly*, October 27, pages 37-47.

Virmani, Arvind. (2006). "India's Economic Growth History: Fluctuations, Trends, Breakpoints, and Phases", *Indian Economic Review*, Vol. 41(1), pages 81-103.

Yanikkaya, H. (2003). "Trade Openness and Economic Growth: A Cross Country Empirical Investigation," *Journal of Development Economics*, Vol. 72(1), pages 57-89.

Appendix

A Data Sources and Definitions

A.1 Figure 1

Source: Net State Domestic Product (NSDP) is from the Economic Political Weekly Research Foundation (2005) dataset on Indian states. The sectoral definitions and sectors are:"Agriculture" includes agriculture, forestry and fishing; "Mining"; "Manufacturing includes registered and unregistered manufacturing; "Construction"; "Trade" includes trade, hotels and restaurants; "Transport, Electricity" include Transport, Storage and Communication plus Electricity, Gas & Water; "Banking" includes Financing, Insurance, Business Services; "Real Estate"; "Public" includes Public Administration and Defence; and, "Other Services".

All series are at constant 93-94 prices projected back using earlier base years.

A.2 Figure 2

Source: The Net State Domestic Product data have been assembled from various tables in the EPW Research Foundation (2005) dataset, the most comprehensive and up to date dataset on Indian states. The observations have been spliced so that all states have real NSDP figures in constant 1993-1994 prices, divided by state population (interpolated between census dates). Our method of splicing ensures that our measures of state RNSDP are largely immunized from the impact of various changes in state definition.[29]

A.3 Panel dataset Used in Section 3

Our core dataset contains output per capita data for 15 major states (the same list of states as for Figure 2, excluding Jammu and Kashmir) using data from the EPW Research Foundation, for fourteen sectoral headings. All data have been spliced so that the underlying sectoral data are in constant 1993-1994 prices, converted into per capita terms using total state

[29] These changes mainly affect Bihar and, to a lesser extent, Madya Pradhesh and Assam. Details of precise methodology are available from the authors.

population as for Figure 2. The sectoral series for each state are: 1)Agriculture, 2)Forestry and Logging, 3)Fishing, 4) Mining and Quarrying, 5) Registered Manufacturing 6) Unregistered Manufacturing, 7) Construction, 8) Electricity, Gas and Water Supply, 9) Transport, Storage and Communication, 10) Trade, Hotels and Restaurants, 11)Banking and Insurance, 12) Real Estate, 13) Public Administration, 14) Other Services.

We eliminate three series from the panel due to clear errors: published data for Electricity, Gas and Water are negative in some years for Assam and Haryana; and published data for real estate in Kerala have clear discontinuities. We also investigate below the implications of omitting some other series that may contain rogue observations.

If we exclude data for Assam, Bihar and Orissa we have a full sectoral breakdown for the remaining 12 states from 1965; if we also exclude Haryana and Punjab we have data for the remaining 10 states from 1960.

A.4 Consumption

To calculate aggregate nominal consumption expenditures by states, we generated a pseudo-panel by utilizing data from various NSS rounds which provide data on nominal monthly mean per capita rural consumption and nominal monthly mean per capita urban consumption These numbers were multiplied by 12 to generate annual figures, and then multiplied by observations for rural and urban population shares. The population data are tabulated from Census figures, with a common compound growth rate applied across decadal observations to impute annual observations for each state. We cross check these figures with population figures obtained by simple extrapolation: (NRSDP/PCNRSDP)*10000000. Both the census figures and extrapolated figures are consistent with each other. Rural Population and Urban Population proportions are then obtained from various rounds of the NSS surveys to give us a full series of rural and urban annual population figures from 1960 - 2005.

To calculate aggregate real consumption expenditures by states, we followed a similar procedure. We generated a pseudo-panel by utilizing data from various NSS rounds on real monthly mean per capita rural consumption (at 1973-74 all India rural prices), real monthly mean per capita urban consumption (at 1973-74 all India urban prices), and population data.

Aggregate annual rural consumption (in crore) is given by: real monthly mean per capita rural consumption $\times 12 \times$ rural population for a given state in a given year.

Aggregate annual urban consumption (in crore) is given by: real monthly mean per capita urban consumption $\times$ 12 $\times$ urban population for a given state in a given year.

Total state (nominal) real consumption expenditures (in crore) is given by: Aggregate (Nominal) Real Rural Consumption + Aggregate (Nominal) Real Urban Consumption / 10000000.

B Panel Unit Root Tests for Implied Transitory Components

B.1 In levels

Figure A1 plots ranked ADF statistics for each of the transitory components calculated as in (4) using the first two principal components in levels, and reports the panel unit root as in Im, Pesaran and Shin (2003), which allows for heterogeneity of auto-regressive coefficients under the alternative. The pooled test strongly rejects the null, and as the chart shows, a very high proportion (97%) of test statistics lie below the expected value under the unit root null. In contrast, if the same procedure is applied to the underlying series in levels, the corresponding proportion is 53%, with a p-value on the joint test of 1.00.

[Insert Figure A1. panel adf charts may 2010]

B.2 In differences

When we carry out the procedure in differences as in Bai and Ng (2004) and calculate the implied transitory components by cumulation the rejection of the unit root null on the joint test remains highly significant despite a somewhat higher proportion of more marginal individual test statistics (73% of which were below the mean value under the unit root null).

[Insert Figure A2, panel adf charts may 2010]

C Data Construction for Figure 4

For Figure 4, we let $\widehat{F}_{1t}$ and $\widehat{F}_{2t}$ be the first and second principal components respectively, (normalized to have zero mean and unit variance, these are the "G-Factor" and "V-Factor" as defined in Figure 4) derived from the sample autocorrelation matrix of y_{it} (or equivalently, from the autocovariance matrix of the series after demeaning and rescaling to have unit sample variance). The series $PC1$ is the cumulated first principal component extracted by the same method from the panel of differenced data as in Bai and Ng (2004).

D Robustness Checks for V-Factor Estimates

D.1 Robustness to changes of time sample

As noted in the main paper, our core analysis is carried out on a balanced panel of data for 15 states. However, as discussed in Appendix A.3, for a subset of ten states we have a longer run of data, back to 1960. A natural robustness check for the dating of the turnaround in the V-Factor is to use the longer datasets, despite the reduction in the cross-sectional dimension (in Appendix G we show that simulation evidence that the gains from increasing T appear to more than offset the losses from decreasing N). Figure A3 shows the results of this experiment. The two alternative estimates of the V-Factor have an identical timing of their apex, and extremely similar paths thereafter. There are somewhat greater differences in earlier years but overall the profiles of all three estimates appear reassuringly similar. It is striking how robust the estimates are both to the inclusion of the additional years and the exclusion of a subset of states.

[Insert Figure A3, pc comparisons march 2010.xls]

D.2 Robustness to changes of cross-sectional sample

As a further robustness check we also investigate, in our panel from 1970 onwards, the impact of removing certain categories of series from the estimation of the principal components. Table D1 and Figure A4 summarise the impact of these changes.

Table D1 lists the exclusions from the cross-section. The first four exclude data based on state characteristics; the next three exclude series by broad industry type. We also show the impact of excluding series with high levels of volatility, and, for comparison, the impact of prior-filtering data for the short-term impact of fluctuations in rainfall (see next section). The table also shows N, the cross-sectional dimension, the correlation, across the cross-section, between actual changes in growth rates and fitted values implied by the estimated V-Factor and G-Factor, as discussed in Section 3.5, as well as showing the year in which the estimated V-Factor reaches its minimum

[include Figure A4. pcs with restricted samples, april 2010.xls]

The first notable feature illustrated by Figure A4 is how similar the broad profiles of the estimated V-Factors are after all these adjustments (as in all other comparisons the estimates are all normalized to have unit mean and variance), despite significant differences in sample both in terms of the change in N, and in terms of the characteristics of the series. All estimates also provide similarly good representations of the shift in growth.

The second notable feature is that, while adjustments for more volatile series have only a modest impact on longer term properties of the estimated V-Factor, they do (unsurprisingly) have some influence on short-run movements. Figure A4 makes it clear that the sharpness of the apex in 1987 for the estimated V-Factor using the full cross-section is reduced, or disappears entirely, in any sample that excludes agriculture, forestry and fishing, in particular, and that as a result for these reduced cross section the apex occurs a year or, at most, two years later. In the light of our simulation results, discussed below in Appendix G, which show that the true apex is only reasonably well estimated to within a year or two either side, this should not be viewed as surprising.

Table D1. Impact on estimated V-Factors of excluding series from the panel

Adjustment	N	$\text{corr}\left(D_i, \widehat{D}_i\right)$ [30]	Apex of V-Factor
excl. top 4 states by % share of agricultural sector	152	0.73	1987
excl top 4 states by income per capita	152	0.73	1987
excl. 7 landlocked states	111	0.78	1987
excl. 4 southern states	152	0.72	1989
only production industries[31]	118	0.81	1987
excl. agriculture, forestry and fishing	162	0.81	1989
only service industries[32]	89	0.83	1988
excl. 13 most volatile series[33]	194	0.81	1989
adjusting for rainfall	207	0.80	1987

D.3 Robustness to rainfall adjustment

As an additional check to adjust for short-run volatility, we prior-filter the data in first differenced form by regressing on a constant and the change in log rainfall over the previous year, and then replace each of the underlying series with the cumulated error from this regression. In the case of agricultural output in particular we find strongly significant positive impacts of rainfall changes, and hence a reduction in the remaining volatility of the series. The impact of rainfall on other sectors is typically less significant. Figure A4 and Table D1 again show that the impact of the adjustment on the V-Factor estimate is very small.

E Policy Indicators and Data Construction and Sources for Figure 7

The V-Factor is equal to $\widehat{F}_{2t}$ as in Figure 4. The effective tariff rate is constructed consistently with Rodrik and Subramanian (2005, Figure 4.) The central government customs duties collection (in crore) and imports (in crore) are from the Reserve Bank of India statistical tables. The effective tariff rate is approximated as Customs Duties Collection/Imports. The Real

[30] As defined in footnote 12.

[31] This includes agriculture, forestry, fishing, mining, registered and unregistered manufacturing, electricity gas and water.

[32] This includes transport, trade, banking,, public, other services, and real estate.

[33] We exclude all series with observations in log differences that lie outside the range $(-1, 1)$ (in percentage terms this corresponds to those with percentage changes in output lying outside the range $(-63\%, 71\%)$).

Exchange Rate data (REER) and the log openness ratio was assembled from the Reserve Bank of India (RBI) database on the Indian Economy. Duties as a percentage of GDP is defined as customs duty collection (in crore) / GDP at factor cost (in crore). This was also obtained from the RBI dataset. See www.rbi.org.in.

F Data Construction and Sources for State-level Regressors inTable 3

The pro-worker dummy is taken from Aghion et al (2008).

The dummy for landlocked states is equal to unity for all series for Assam, Bihar, Haryana, Madhya Pradesh, Punjab, Rajasthan, Uttar Pradesh, and is zero otherwise

The other state characteristics used in the regressions in Table 3 are taken from a new panel dataset for Indian states assembled by the authors comprising roughly 200 regional economic and social indicators for Indian states. A detailed description of the variables in this dataset, and the data used in Table 3, is available in the data appendix in an earlier working paper version of this paper; Ghate and Wright (2008).

G Simulation Methodology

We simulate a system with an underlying common structural shift which is a parameterised version of (1) to (3), as follows

$$
\begin{aligned}
y_{it} &= \beta_{i0} + \beta_{i1} F_{1t} + \beta_{i2} F_{kt} + u_{it}; i = 1..N && (5) \\
\Delta F_{kt} &= g_{k1} + \varepsilon_{kt}; t \leq t_b && \\
&= g_{k2} + \varepsilon_{kt}; t > t_b; k = 1, 2 && (6) \\
u_{it} &= \gamma_{i1} Q_{1t} + \gamma_{i2} Q_{2t} + r_{it} && (7) \\
Q_{jt} &= \rho_j Q_{jt-1} + \xi_{jt}; \quad j = 1, 2 && (8) \\
r_{it} &= \rho_i r_{it-1} + \omega_{it}; \quad i = 1..N, && (9)
\end{aligned}
$$

In (1) we simulate each of the N series as a sum of factor loadings on two $I(1)$ factors, plus a persistent residual component. The two $I(1)$ factors, F_{1t} (the simulated "G-Factor") and F_{2t} (the simulated "V-Factor") are modelled in (6) as drifting random walks with shifts in growth rates at the break point t_b. The transitory components u_{it} are then in turn driven by two common stationary factors, Q_{1t} and Q_{2t} which capture any remaining mutual correlation in the y_{it} after extraction of the two permanent components, plus a strictly idiosyncratic component, r_{it}. The Q_{jt} are modelled in (8) as stationary AR(1) processes without shifts (we examine below the impact of including or excluding these additional stationary factors). We estimate the process for the two permanent and two stationary factors from the time series properties of the first four principal components of the dataset. The data point to a highly significant shift in growth at $t_b = 1987$ for the "V-Factor" ($g_{21} < 0$; $g_{22} > 0$); with a smaller, but still significant shift for the "G-Factor ($0 < g_{11} < g_{12}$). While conventional tests of significance are suspect due to a data mining critique, the primary objective is to simulate a null model where there is a structural shift in growth that also matches the broad properties of our dataset. The estimation procedure for the factor processes is thus for purposes of calibration, rather than to carry out any direct hypothesis testing. The correlation matrix of the vector of estimated factor innovations $\begin{bmatrix} \widehat{\varepsilon}_t' & \widehat{\xi}_t' \end{bmatrix}'$ is close to diagonal in the data so we simulate the four factor innovations as orthogonal processes.

The factor loadings $\{\{\beta_{ik}\}, \{\gamma_{ij}\}\}$ are calibrated to match (subject to minor modifications noted below) those of the estimated factor loadings on the principal components in the data. Each element is modelled as an independent draw from a normal distribution with mean and standard deviation given by the cross-sectional mean and standard deviation of the loadings on each of the principal components in the data. The simulated orthogonality of the factor loadings that results from this methodology is consistent with the orthogonality (by construction) of factor loadings derived by the method of principal components.

Finally in (9) we model the residual idiosyncratic components, the r_{it} as AR(1) processes with mutually uncorrelated innovations. The $\{\rho_i\}$ and the $\{\sigma_i\}$, (where $\sigma_i = E\left(\omega_{it}^2\right)$) are modelled as independent draws from uniform distributions that approximate the key cross-sectional properties of these parameters in our dataset. We draw from a uniform, rather than normal distribution, since we need to impose bounds on both sets of parameters, such that $\rho_i \in (-1, 1)$, $\sigma \in (0, \infty)$. We calibrate these distributions to match the cross-sectional means and standard deviations of the estimated parameters in our dataset, subject to these inequalities.

Reassuringly the simulation methodology gives a generally good match of the key properties of the dataset. We make only two minor modifications to ensure that the simulated contribution of the two nonstationary factors to the total variance in the dataset is on average (across simulations) equal to that in the data (since we do not wish to over- or understate the importance of these two factors in our simulations). This is achieved by raising $\overline{\beta_{i1}}$, the cross-sectional mean loading on the "G-Factor" from 0.0266 in the data to 0.032 in the simulations (this ensures a match of the average contribution of the first factor in the simulations), and by reducing $\sigma(\beta_{i2})$, the cross-sectional standard deviation of the loadings on the "V-Factor" from 0.030 in the data to 0.025 (this ensures a match of the average contribution of the second factor in the simulations).[34] Given the approximations involved in our simulations (in particular the distributional assumptions for the parameters), the magnitude of the changes required is reassuringly modest.

Table G1 summarizes the key results of our simulations. The first row shows our base case. In each artificial sample we simulate a balanced panel of 207 series all starting in 1970, where the true break year, t_b is set at 1987, in line with the profile of the V-Factor shown in Figure 4 in the main paper. The results show that if the true data generating process has the same breakpoint, the 2nd principal component in levels would identify the breakpoint in the true V-Factor (simulated as F_{2t}) to within $\pm$ 1 year in 60% of replications.[35]; in comparison the cumulated 1st principal component in differences has an equivalent percentage of only 32%. Both approaches are somewhat biased: i.e., if the true breakpoint year were 1987, on average both approaches would estimate it to be 1988. But this bias is to be expected since it arises from the AR(1) processes assumed for the u_{it}, such that the mean lag from the impact of a shift in the factors, given by $\rho_i/(1-\rho_i)$ is always positive. Based on our dataset, ρ_i ranges from -.15 to .67, hence the simulated mean lags range from zero to roughly 2, hence a bias of around one year is to be expected.

[34] The mean loading on the V-factor is close to zero in the data, and we retain this feature in the simulations.

[35] Note that the proportions shown in the table are when the minimum of the estimated component matches that of F_{2t}. This does not always match the true breakpoint, since, given random variation in the simulated F_{2t}, it does not always reach a minimum in the "true" breakpoint year.

The second row of the table shows that if we simulate a smaller cross section, over a longer sample (as in Figure A3), the loss of precision from a lower cross-section appears to be more than offset by the gain in precision from a longer sample.[36]

The third row of the table shows the impact of excluding the impact of the two additional stationary factors. Using both techniques there is a clear improvement, unsurprisingly so, since all remaining variation in the y_{it} is due to the mutually orthogonal u_{it} terms. The improvement in the performance of the approach in differences is particularly marked, but it remains less reliable than the levels approach; albeit only marginally so. The much greater sensitivity to the exclusion of the stationary factors does however indicate a lack of robustness of this approach (we show below that this conclusion is further strongly reinforced by the comparative performance of the two approaches with a stochastic breakpoint).

This improvement in identification of breakpoints in the smaller cross-section over a longer sample is clearly a helpful result in itself, but all the more so if we wish to distinguish between the break point of 1987 identified in our dataset and the earlier breakpoints identified in past research. We note in the main paper that some studies have concluded that there was a break point as early as the late 1970s. In the fourth and fifth rows of the table we simulate an alternative data generating process consistent with this earlier breakpoint. With the shorter sample and a larger cross-section neither of the two approaches would be very successful in identifying such an early breakpoint (i.e. only 9 years into the sample); however the fourth row of the table shows that with a longer sample but a lower cross-section the earlier break point would still be reasonably well estimated. We can use this simulated DGP to assess the probability of estimating a break point in 1987 (as in our dataset), or later, if the true breakpoint were in 1979: using principal components in levels this occurs in only 3% of simulations, suggesting that the technique we use can discriminate well between an earlier and a later breakpoint.

A more general way of assessing how well the two alternative techniques perform in identifying breakpoints is summarized in the last two rows of Table G1 and in Table G2. These show the results of allowing the breakpoint

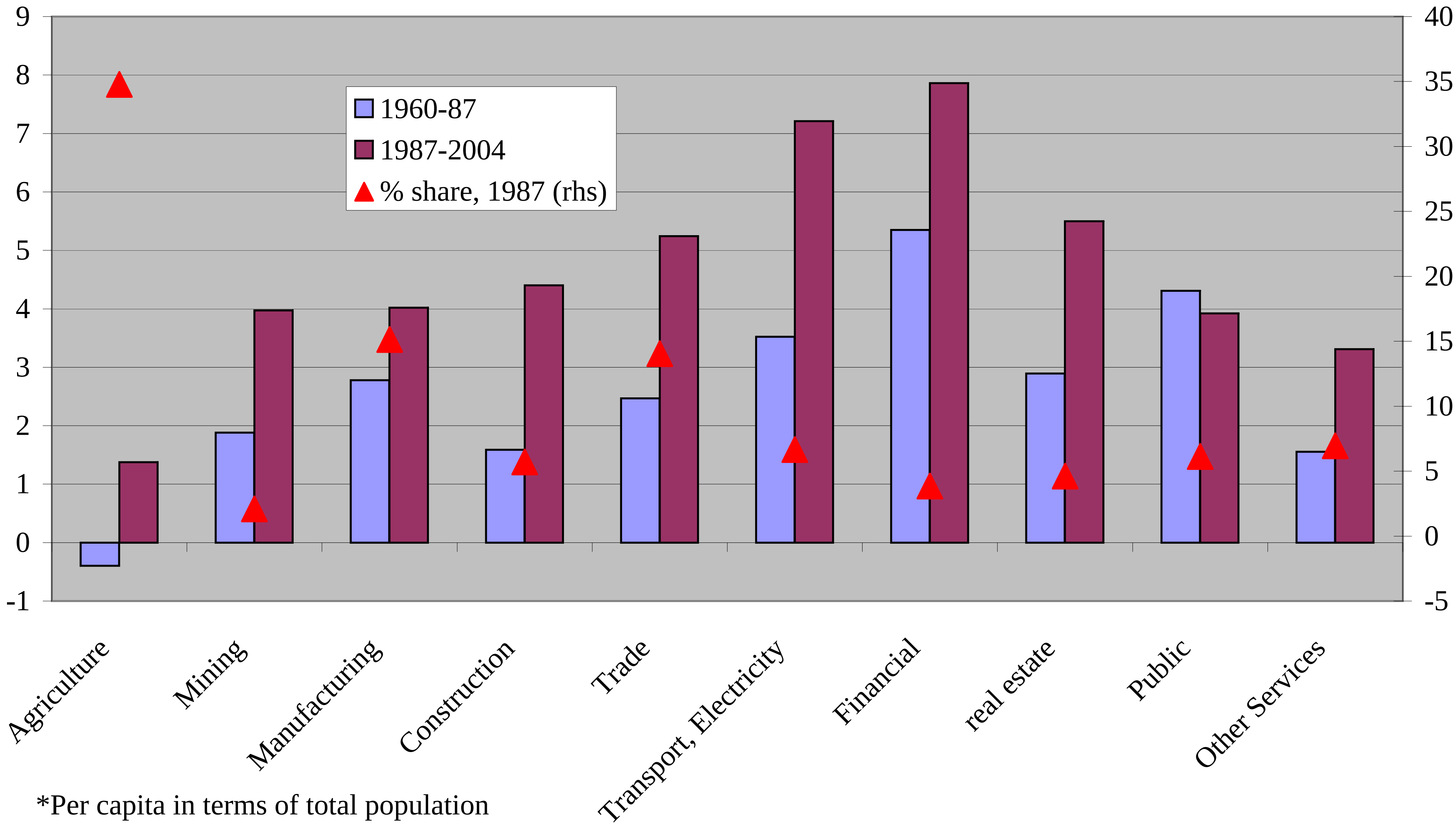
[36] If we increase T and decrease N separately the impacts are, as would be expected to improve and decrease precision respectively.

to be a random variable across simulations. The true breakpoint t_b is drawn for each simulation as a uniform random variable ranging between 1982 and 1992. The precision with which the breakpoint is estimated by both techniques falls somewhat, but the proportions of simulations in which the estimated breakpoint is within a year of the true breakpoint are quite similar. Table G2 shows that using the levels approach the estimated breakpoint is quite strongly positively correlated with the true breakpoint across the simulations (with correlation coefficient 0.7) but that it does not typically move one for one: essentially there is some bias (albeit not especially strong) towards finding a breakpoint at or near the mid-point of the sample. In contrast Table G2 shows that the estimated breakpoint using the differences approach is only weakly correlated with the true breakpoint across different simulations.

Finally we note that the comparative properties of the simulations summarized above, which focus (for obvious reasons) on the identification of the breakpoint, are not dependent on the assumption that the deterministic component of the "V-Factor" is precisely V-shaped. We have also experimented with an alternative DGP in which the second factor is roughly "U"-shaped - i.e., closer to the shape identified by the differences approach in our dataset, as illustrated in Figure 4. The ranking of the two approaches, expressed in terms of the correlation between the estimated principal component and the true factor, remains the same in all cases. When the true factor is a "U"- rather than a "V"-factor this property is captured fairly well in the majority of simulations by the levels approach: i.e. there is no bias in estimation towards finding "V"- as opposed to "U"-Factors.

Thus we can feel reasonably confident that, even if the breakpoint of the true V-Factor cannot be precisely identified in our dataset, it seems likely to have occurred within a year or two of the estimated breakpoint of 1987. Furthermore, it does appear that the turnaround was relatively rapid: thus a "V"-Factor representation does appear valid.

Figure 1. Growth in Per Capita Real NDP: by Sector*



*Per capita in terms of total population

Figure 2. Growth in Per Capita Real NDP, by State

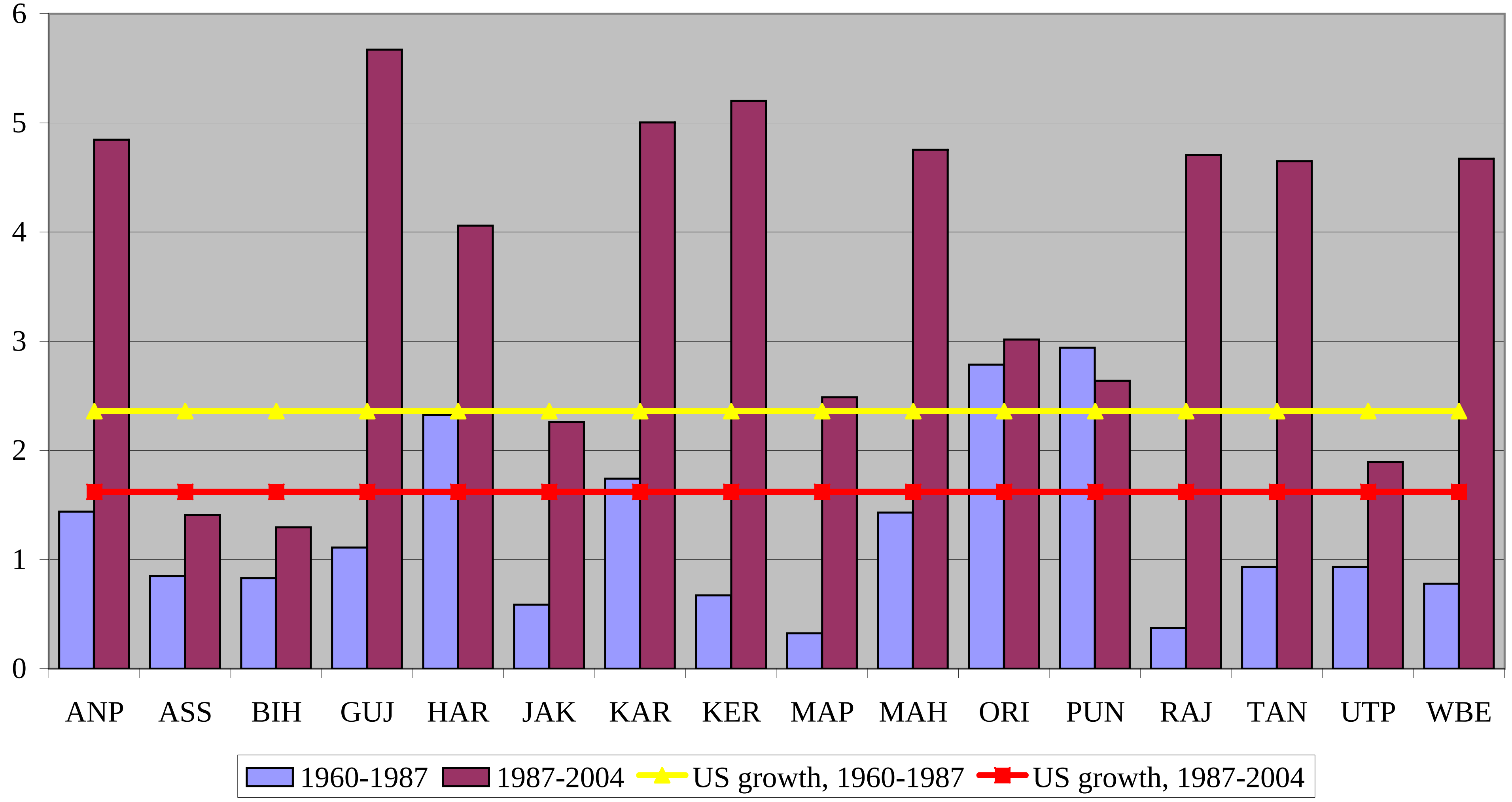


Figure 3. The Distribution of Average Sub-Sample Growth Rates

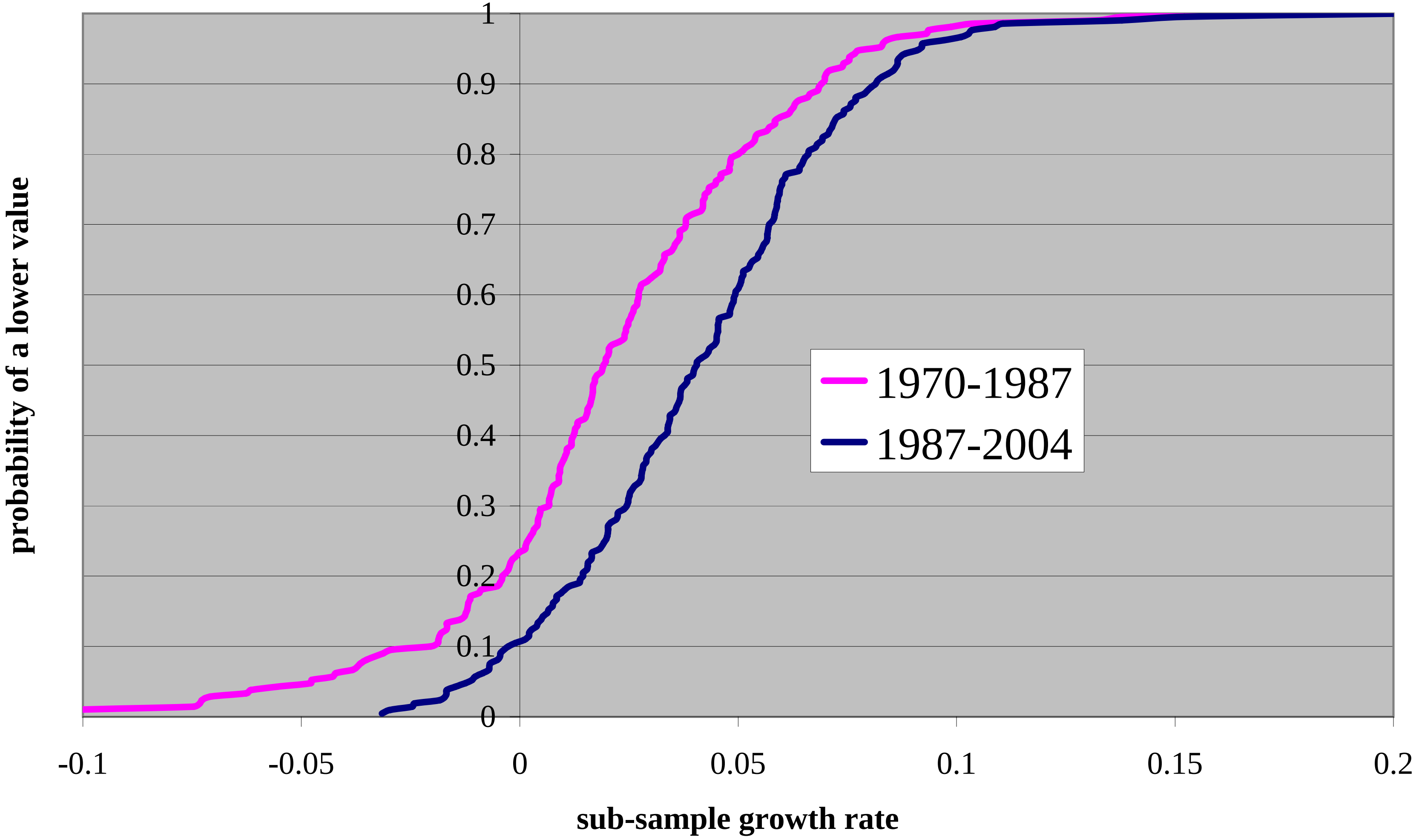


Figure 4. Common Factors Estimated by Principal Components

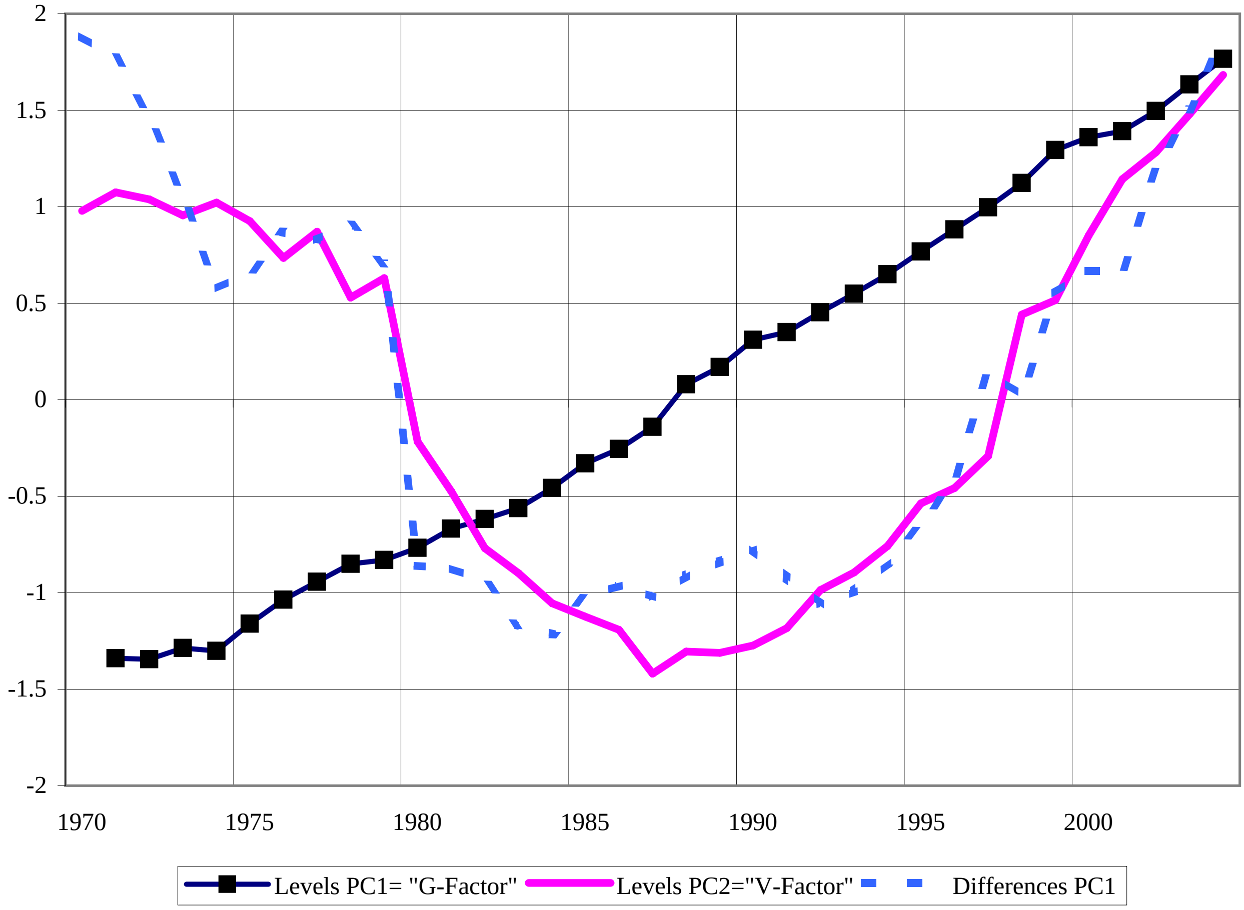


Figure 5. The V-Factor as a Representation of Growth Shifts: By Sector

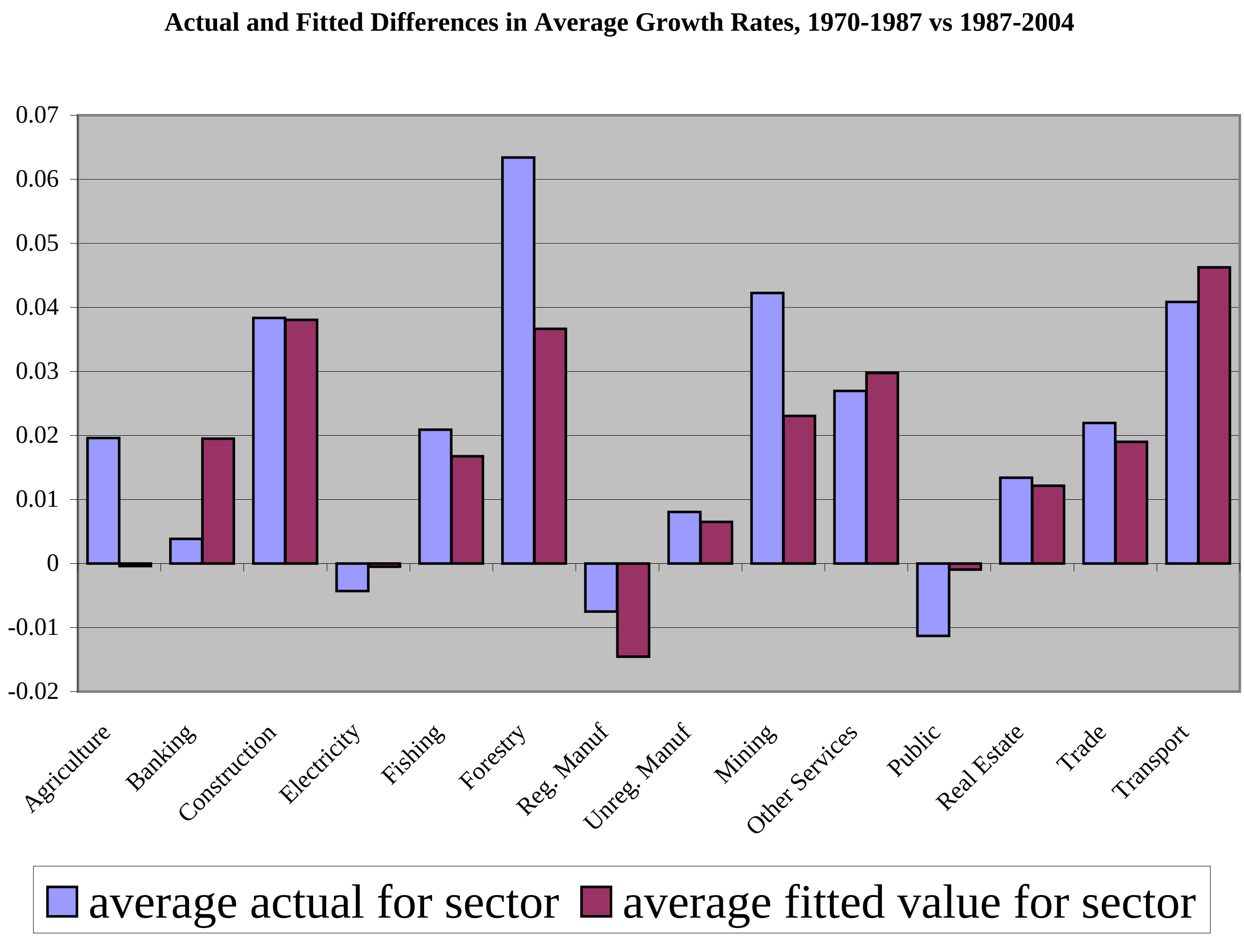


Figure 6. The V-Factor as a Representation of Growth Shifts: By State

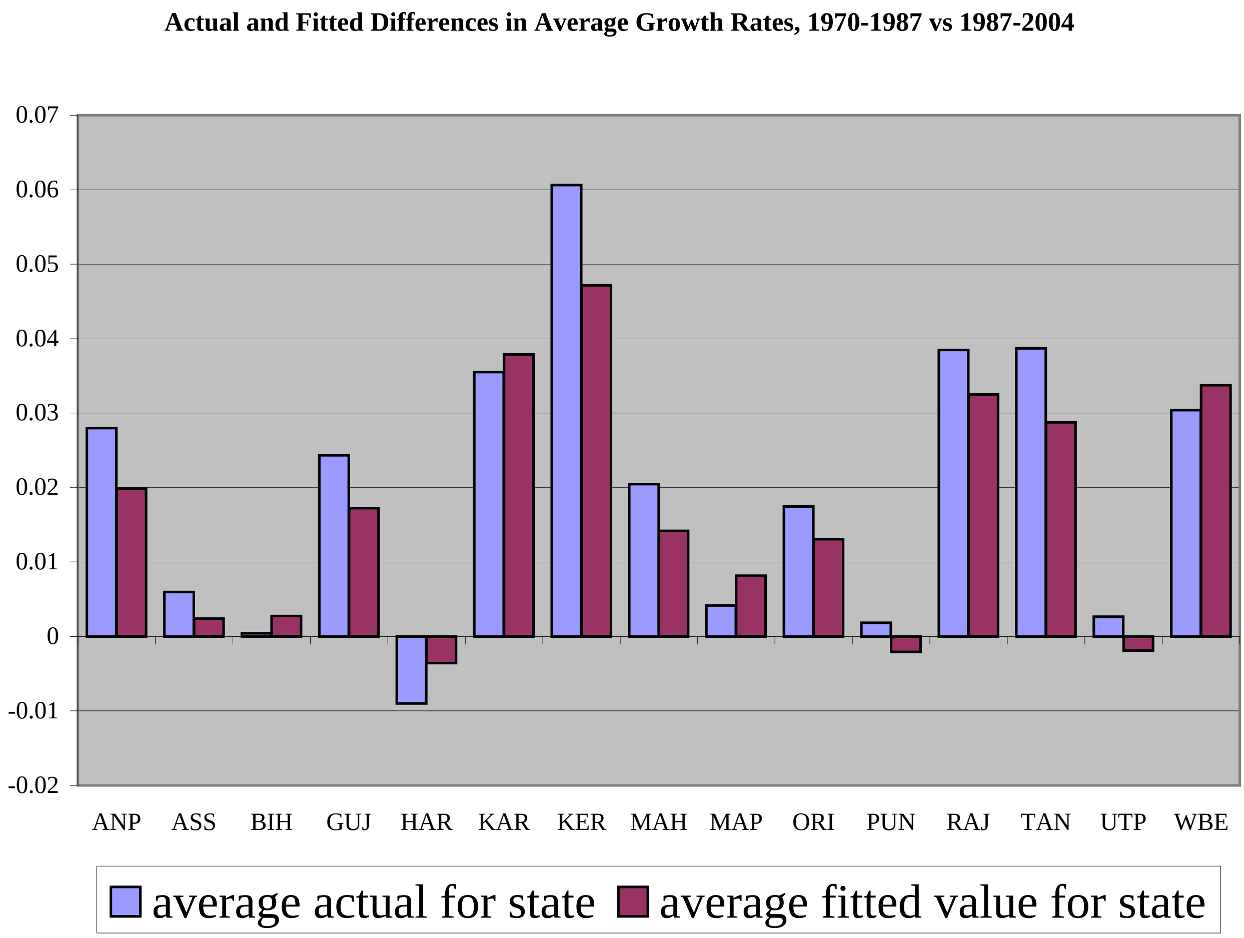
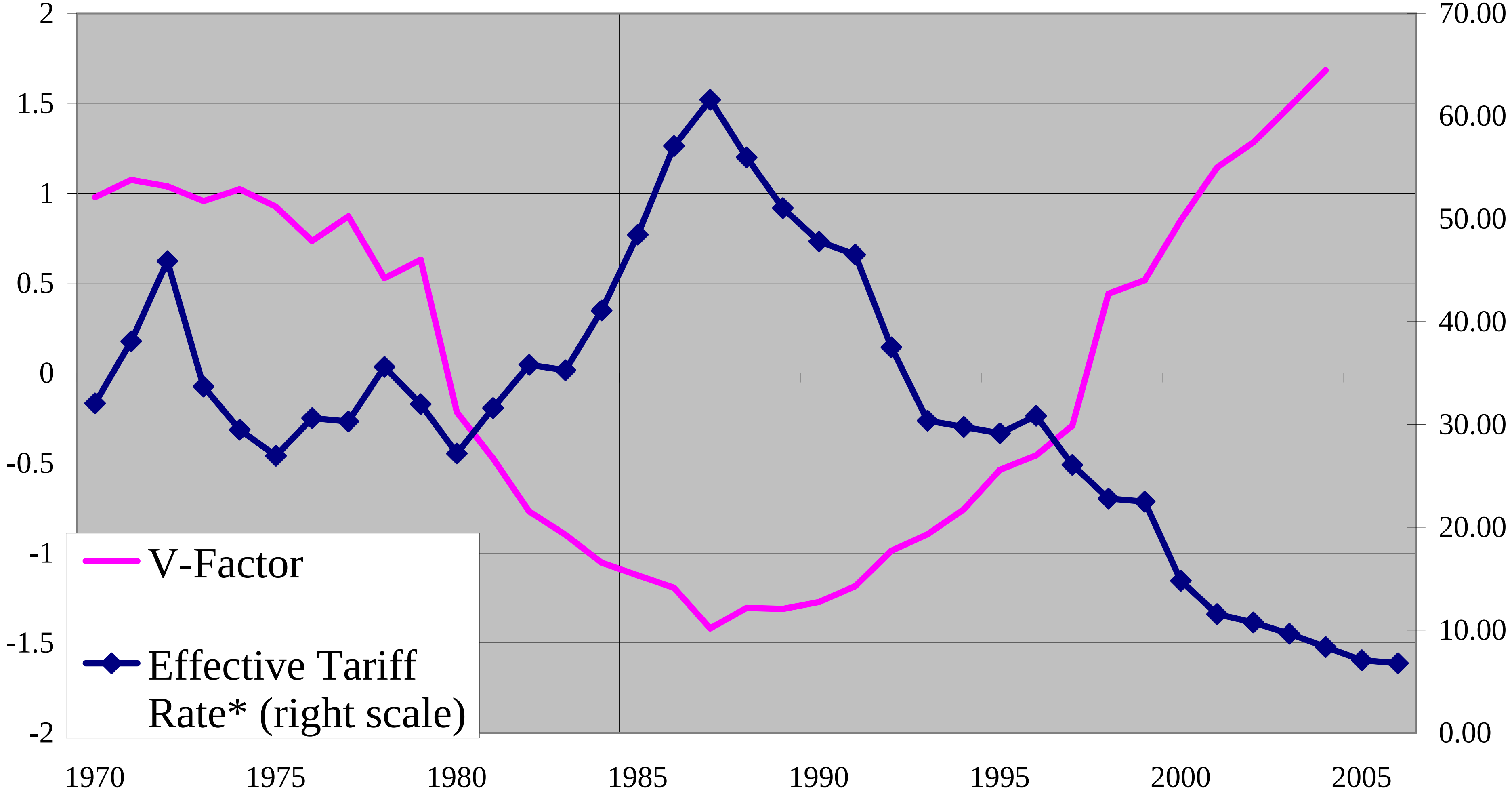


Figure 7. The V-Factor and Trade Liberalisation



* Duty Revenue as % of Total Imports

Figure A1
Ranked ADF Statistics for Transitory Components from Levels Estimation

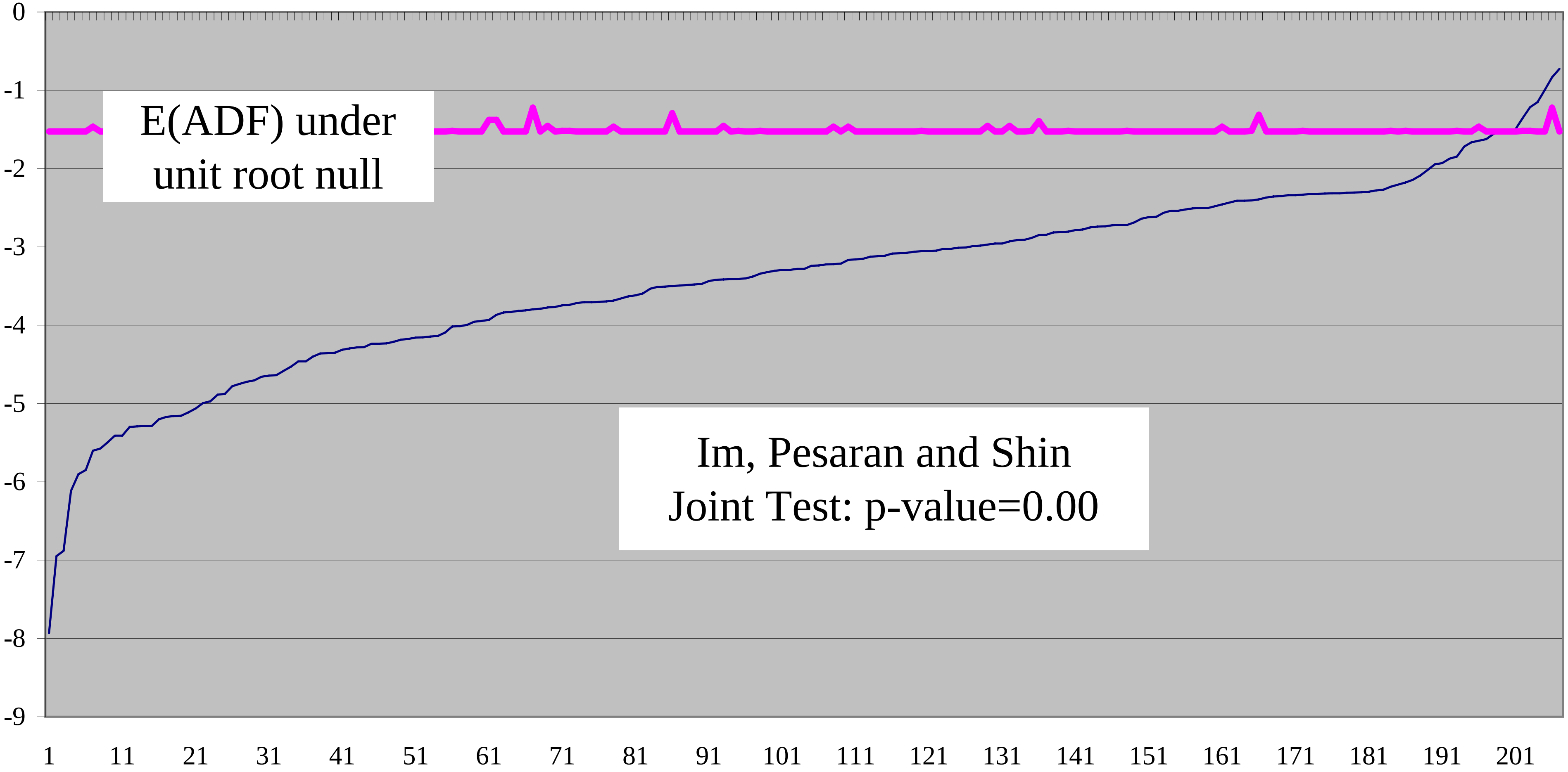


Figure A2
Ranked ADF Statistics for Transitory Components from Estimation in Differences

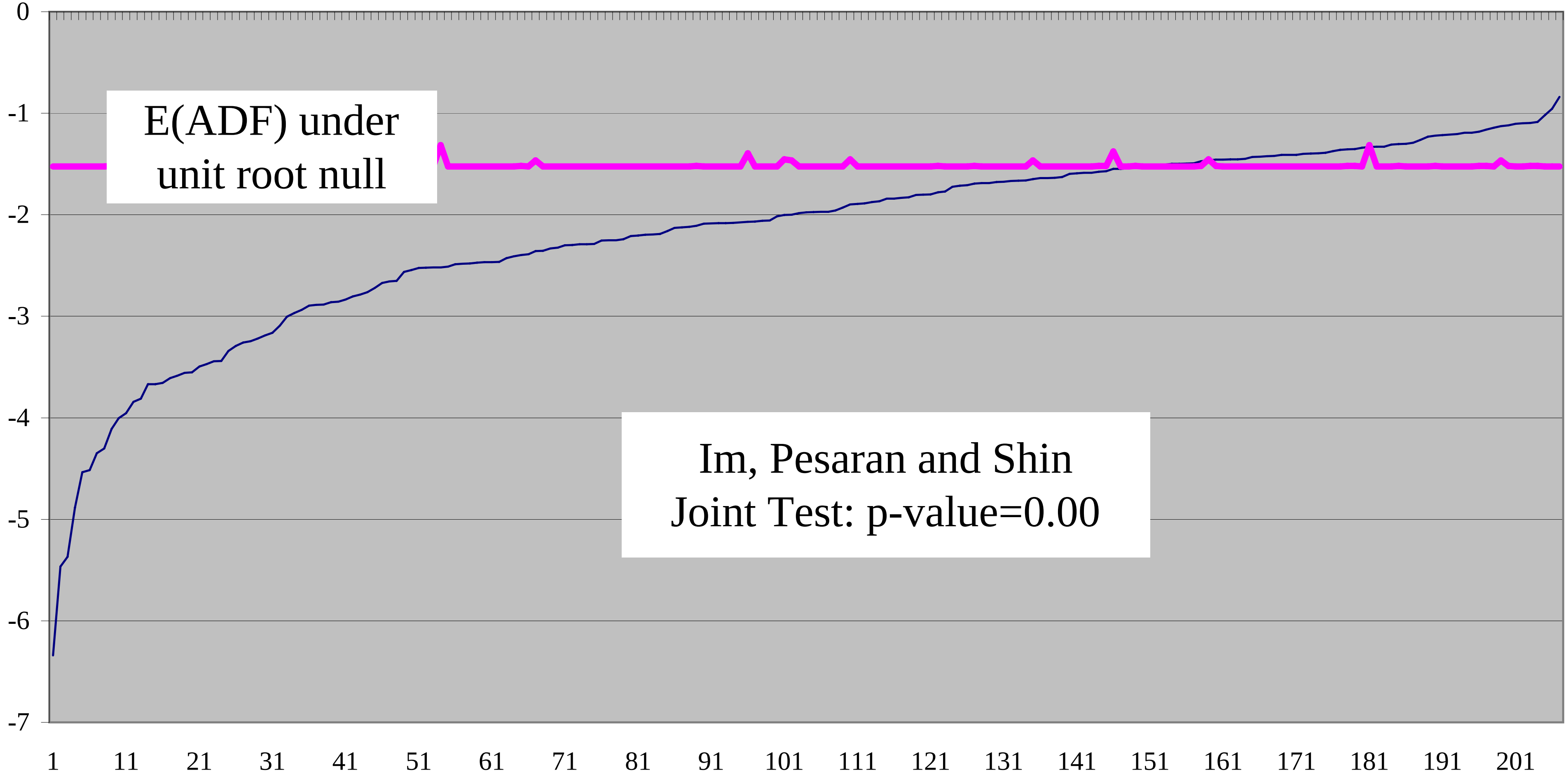


Figure A3
Alternative V Factor Estimates

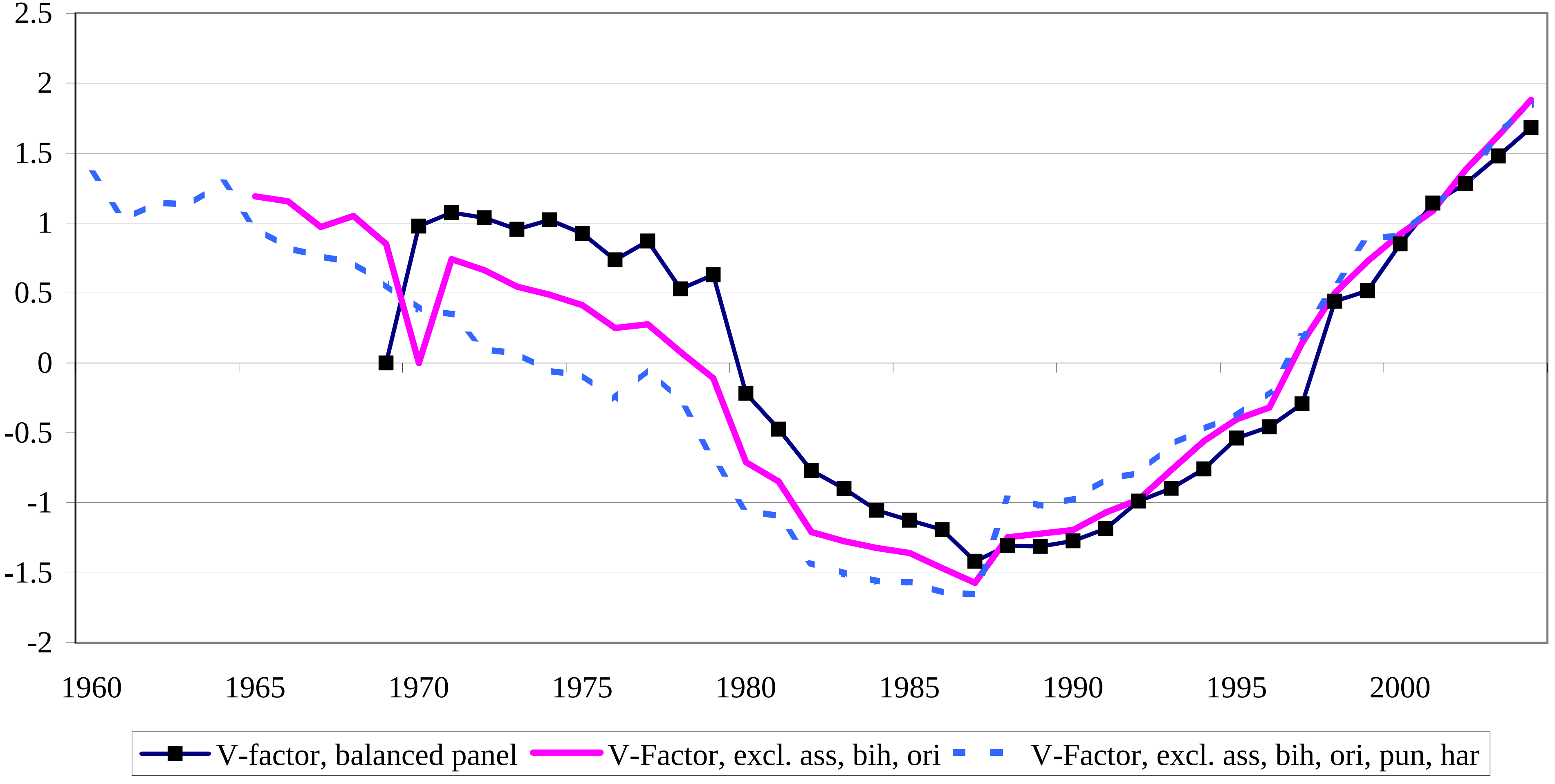


Figure A4. Impact on estimated V-Factors of Excluding Series from Panel

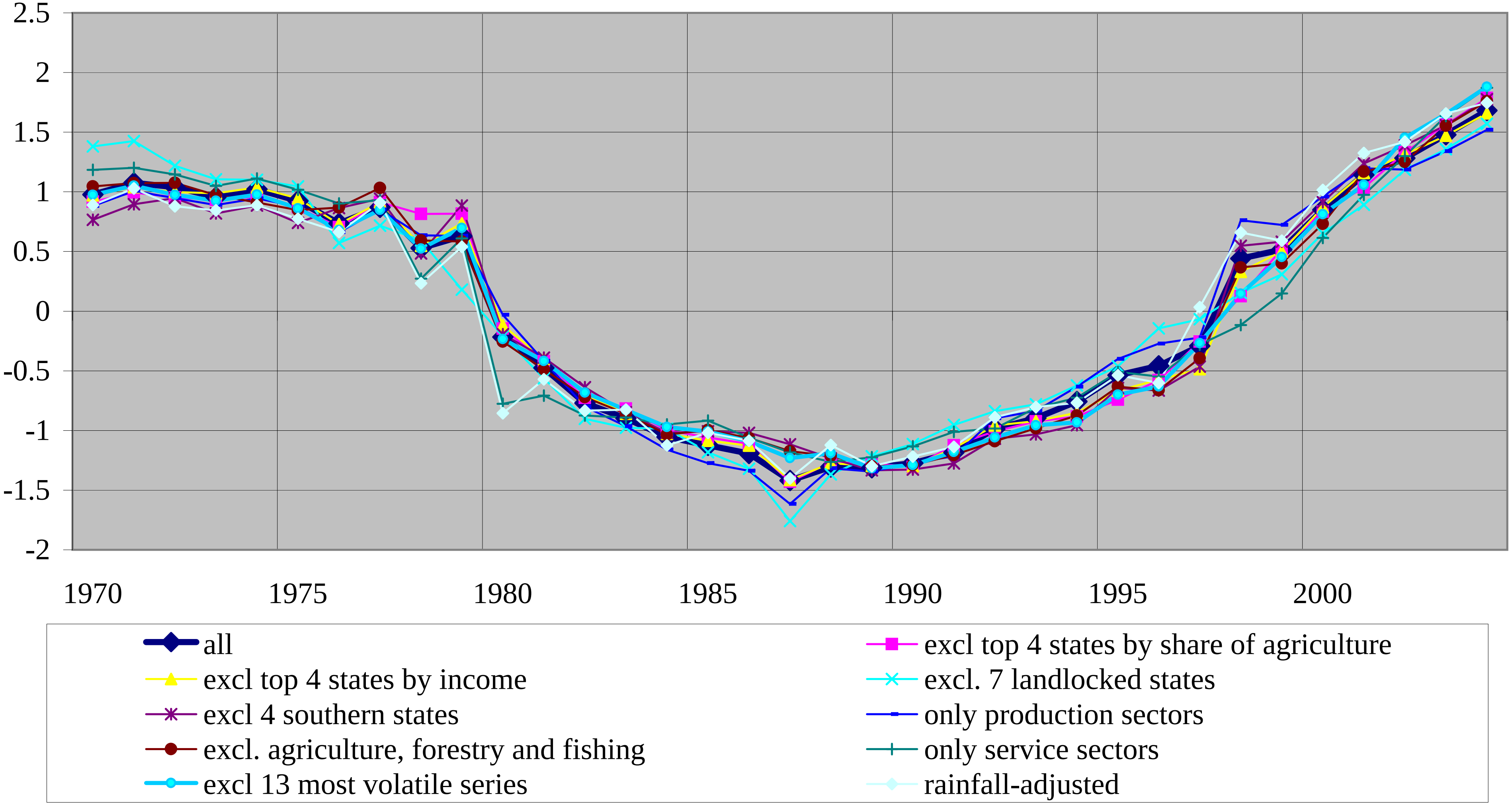


Table 3: State Characteristics and the Growth Turnaround: Cross-sectional Regression Results							
Dependent variable: Change in average log growth in state-sectoral real NDP per capita between 1970-87 and 1987-2004							
regressors (p-values in brackets)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
state dummies	all	all	none	none	none	none	some**
sector dummies	all	none	all	all	all	some*	some*
share of agriculture, 1987				-0.0014 (0.03)	-0.0014 (0.00)	-0.0014 (0.00)	-0.0017 (0.08)
share of reg. manufacturing, 1987				-0.0036 (0.01)	-0.030 (0.00)	-0.030 (0.00)	-0.033 (0.19)
real state income per capita, 1987				0.02 (0.41)			
% urban population, 1981				0.011 (0.62)			
literacy rate, 1981				0.006 (0.84)	0.021 (0.09)	0.021 (0.09)	0.021 (0.61)
average rainfall, 1983-1987				0.0060 (0.47)			
Aghion et al's pro-worker dummy				0.0011 (0.82)	-0.0037 (0.08)	-0.0037 (0.08)	-0.0037 (0.10)
landlocked dummy				-0.0145 (0.14)	-0.0136 (0.05)	-0.0136 (0.05)	-0.0136 (0.24)
population, 1981				-0.0198 (0.33)			
population growth, 1971-1981				0.542 (0.73)			
development spending, % of NDP, 1981				0.071 (0.27)			
Observations	207	207	207	207	207	207	207
R-bar-squared	0.307	0.115	0.170	0.310	0.322	0.327	0.314
s.e.	0.036	0.041	0.039	0.036	0.036	0.035	0.035
joint significance tests (p-values)							
state dummies	0.000	0.000	n/a	n/a	n/a	n/a	0.893
sector dummies	0.000	n/a	0.000	0.000	0.000	0.000	0.000
*** sector dummies included in regression (6)**							
	forestry	mining	reg manuf.	construction	transport	public	elec,gas, water
coefficients	0.047	0.025	-0.024	0.022	0.024	-0.028	-0.019
p-values	0.00	0.01	0.01	0.03	0.01	0.00	0.05
**** state dummies included in regression (7)**							
	BIH	HAR	KER	MAP	PUN	UTP	
coefficients	0.001	-0.001	0.001	-0.015	0.004	-0.004	
p-values	0.97	0.98	0.98	0.21	0.81	0.73	

Table G1. Estimating common breakpoints by principal components: some simulation results

				Levels Approach			Differences Approach		
Start year	N	break point	stationary factors? ("1"=yes)	s.d.	bias	% correct +or- 1 year	s.d.	bias	% correct +or- 1 year
1970	207	1987	1	2.7	-1.0	60%	5.7	-1.0	32%
1960	**139**	1987	1	2.2	0.1	74%	8.0	1.6	24%
1970	207	1987	**0**	1.4	-0.9	72%	2.2	-1.2	64%
1970	207	**1979**	1	5.3	-4.3	30%	6.8	-4.3	26%
1960	**139**	**1979**	1	3.5	-1.9	55%	6.5	-1.2	33%
1970	207	**random**	1	2.8	-0.8	64%	6.3	-0.2	32%
1960	**139**	**random**	1	2.5	0.4	69%	8.3	2.0	34%

Table G2 Systematic properties of estimated breakpoints when the true breakpoint is a random variable

	Levels Approach	Differences Approach
Correlation with true breakpoint	0.716611341	0.289595116
Slope coefficient on true breakpoint	0.8528597	0.166217551