The "V-factor": Distribution, timing and correlates of the great Indian growth turnaround ☆

Chetan Ghate [a],[*], Stephen Wright [b]

[a] Planning Unit, Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi 110016, India
[b] Department of Economics, Birkbeck College, University of London, Malet Street, London W1E 7HX, UK

ABSTRACT

We analyze a panel of output series for India, disaggregated by 15 states and 14 broad industry groups. Using principal components (Bai, 2004; Bai and 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 since the 2nd half of the 1980s. The timing of the turnaround implied by the V-factor is more closely related to the pattern of policy reforms than has been found in previous research. Regression-based analysis also provides some insights into the uneven distribution of the turnaround across Indian states.

© 2011 Elsevier B.V. All rights reserved.

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
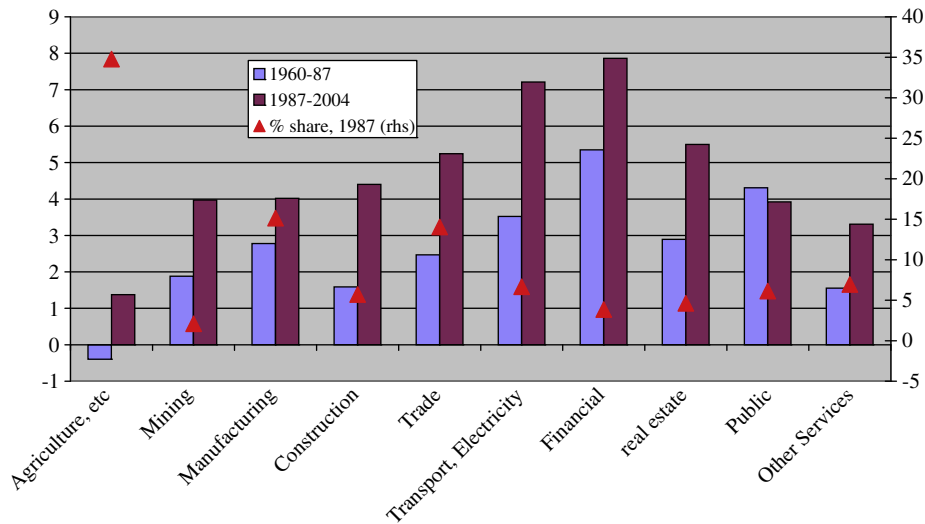
Fig. 1. Growth in per capita real NDP: by sector*.

*Per capita in terms of total population.

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 a minimum 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 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 appear 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 Panagariya (2004) and Pursell (1992)). 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]

The remainder of the paper is structured as follows. In Section 2 we 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 use regression analysis to examine whether state characteristics can account for the very disparate performance across the states noted above. Section 6 concludes the paper. A web appendix provides details of data construction and statistical analysis.[5]

2. Sectoral and state-wise shifts in growth

Figs. 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.[6] Fig. 1 shows that virtually all sectors of the private sector economy have seen substantial increases in growth, albeit from often significantly different initial values.[7] Growth in the public sector, in contrast, actually slowed somewhat between the two sub-samples.

When the economy is divided into states, rather than sectors, the pattern is distinctly more disparate. Fig. 2 shows output growth in the same two sub-samples for the 16 major states, which collectively represent 97% of the Indian population.[8]

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. Fig. 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.[9] The remaining states were all growing

---

[3] Rodrik and Subramanian identify a shift in growth in 1980, based on aggregate GDP data. Balakrishnan and Parameswaran (2007) and Virmani (2006) 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.

[5] Downloadable from www.ems.bbk.ac.uk/faculty/wright.

[6] 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.

[7] Full details of data transformations are provided in Appendix A. 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.

[8] 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.

[9] 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.
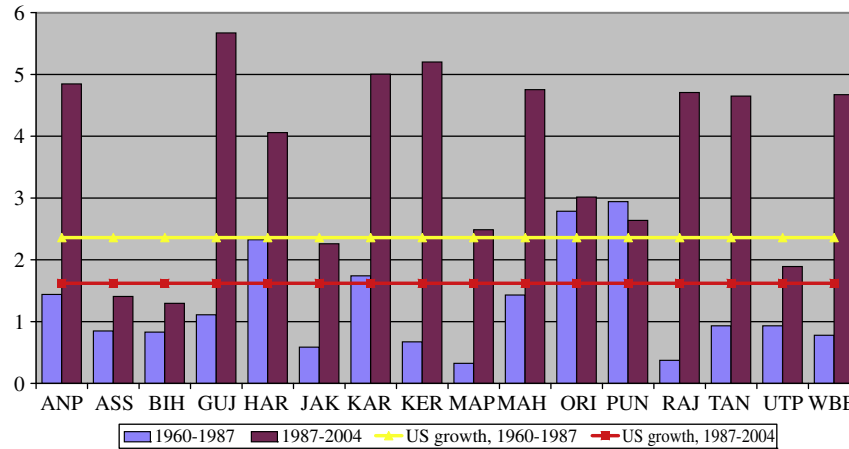
Fig. 2. Growth in per capita real NDP, by state.

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 and 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,[10] so differences in growth rates of output growth will have corresponded to significant differences in consumption growth.

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 Fig. 2, excluding Jammu and Kashmir) we have a sectoral break-down into fourteen broad industrial sectors, from 1970 to 2004; for a sub-set 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.[11]

3.2. Evidence of common structural shifts?

While the visual evidence in Figs. 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. Fig. 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.

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, as shown on the chart; 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, thus greatly increasing the number of observations. 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.[12] Examination of tests carried out over a range of sub samples suggests that this result is not simply an artifact of the breakpoint chosen.[13,14]

---

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

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

[12] 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.

[13] 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.

[14] 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.
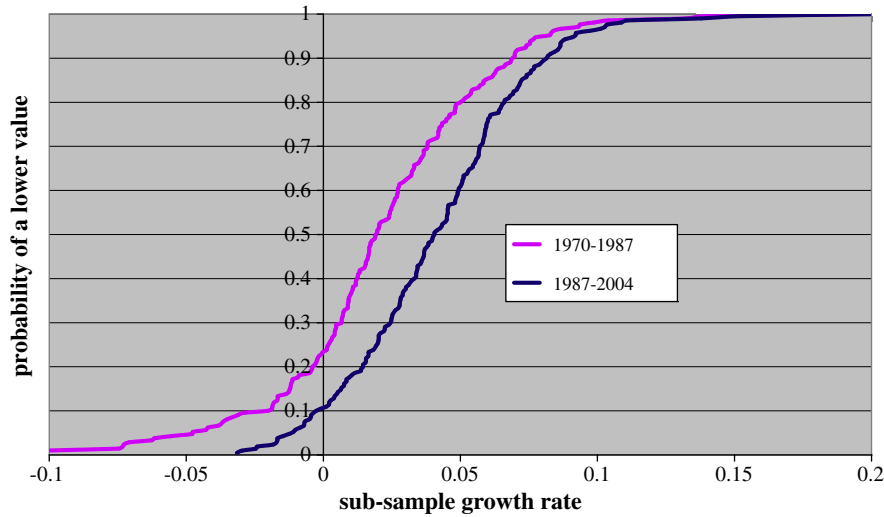
Fig. 3. The distribution of average sub-sample growth rates.

Table 1
KS Tests for Equality of Distribution Functions from 1970-2004.

HA	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	0.995
Combined K-S	0.2714	0.1114	0.000	0.000

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 Fig. 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.

$$y_{it} = \beta_{i0} + \beta_{i1}F_{1t} + \ldots + \beta_{ik}F_{kt} + u_{it}; i = 1..N \tag{1}$$

$$\Delta F_{kt} = a_k(L)\varepsilon_{kt}; \ k = 1..k \tag{2}$$

$$u_{it} = b_i(L)\omega_{it}; \ i = 1..N, \tag{3}$$

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} captures the remaining transitory dynamics. We assume that the $b_i(L)$ 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 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.[15] 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 component 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 $a_k(L)$ or $b_i(L)$. Principal components provide estimates $\hat{F}_{kt}$ of the factors and factor loadings $\hat{\beta}_{ik}$,[16] and the transitory components in Eq. (3) are derived from these estimates, as:

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

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 used to derive an estimate of the average of the variances of the idiosyncratic components; this in turn feeds into the penalty function.[17] 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.[18]

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. 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 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.

---

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

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

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

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

Table 2
Value of k, the number of common factors, implied by information criteria.

		Panel information criteria											
		Estimation in levels						Estimation in differences					
		IPC1		IPC2		IPC3		IPC1		IPC2		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	<∞
1960–2004	139	2	≥5	2	≥5	2	≥10	1	≥3	1	≥5	0	<∞
1965–2004	166	1	<4	1	<4	1	<10	0	<4	0	<5	0	<∞
1965–2004	166	2	≥4	2	≥4	2	≥10	1	≥4	1	≥5	0	<∞
1970–2004	207	1	<5	1	<6	1	<12	0	<5	0	<6	0	<∞
1970–2004	207	2	≥5	2	≥6	2	≥12	1	≥5	1	≥6	0	<∞

A more significant form of ambiguity is that, for low values of k_{max} (and, in the case of the most conservative criterion, 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. However we have a number of reasons to prefer representations with an additional factor in each case, and the 2 factor levels representation in particular:

- 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, in Appendix B we construct the implied transitory components, using Eq. (3), from the levels models with both one and two factors, and from the single factor differences model. The null that each of the resulting series contains a unit root is strongly rejected in all three cases; but the assumption that *all* transitory components are stationary (which is much harder to test directly) appears to be particularly well-supported by the data in levels with two factors.
- Third, and most crucially, we have already seen very strong evidence of a common shift in the distribution of growth rates from the Kolmogorov–Smirnov tests shown in Table 1. Implicitly this is strong evidence against both the zero-factor differences representation and the single factor levels representation. The former representation is, by construction, incapable of representing a permanent common growth shift. And we show below (in Section 3.5) that, while the single factor levels representation could in principle represent such a shift, it cannot do so in practice, given the properties of the single common factor.[19]

We therefore focus our attention on the results from estimation in levels with two factors, and, as a robustness check, 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).

3.4. Factor estimates: the "V-factor" and the "G-factor"

To illustrate the nature of the results, Fig. 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,[20] 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, Fig. A3).

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".[21] The second component, which captures shifts in growth, we term the "V-factor". Fig. 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 minimum 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. (In Section 3.6 we discuss some further issues relating to the date of the turnaround).

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 minimum (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 breakpoints, and less robust. For the rest of the paper we therefore focus on results based on levels estimation with two factors.

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

Figs. 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 Figs. 1 and 2 we showed the strong evidence of a shift in the cross-sectional distribution of both sectoral and state growth rates. In Figs. 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 (Fig. 5) and by state (Fig. 6).[22] The charts show that the two common factors alone provide a good parsimonious representation of the observed

---

[19] The zero factor differences representation would imply that the growth rate of each series could be represented by a process with a fixed unconditional mean, thus common growth shifts can, by construction, at best be transitory in any such representation. The single factor levels representation *could* in principle imply permanent growth shifts if this was a property of the single common factor in this representation; but as we show in the next section, it is not.

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

[21] Note that if we estimate levels model with a single common factor, the resulting estimate is identical to the G-factor estimated in the two factor model, since, by construction, factors estimated by principal components are mutually orthogonal.

[22] 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

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

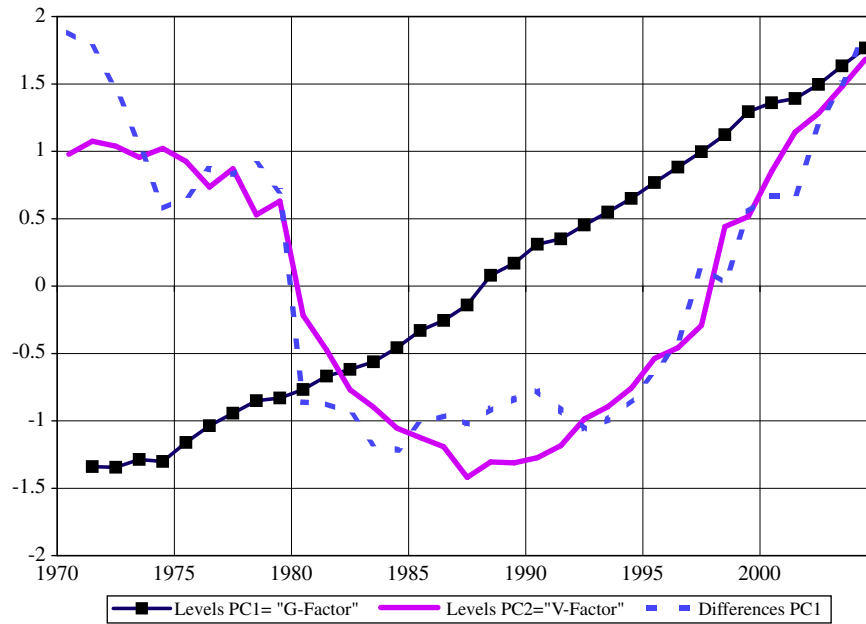Figs. 5 and 6 then show unweighted averages, across sectors and states respectively, of the D_i and the $\hat{D}_i$.

Fig. 4. Common factors estimated by principal components.

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 the single common "G-factor" yields a correlation coefficient between actual and fitted insignificantly different from zero (as we would expect, given that the estimated G-factor, as shown in Fig. 4, must essentially imply nearly constant predicted growth for each series in this representation).

Figs. 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 Figs. 1 and 2 (with the discrepancies largely due to weighting differences since the averages shown in Figs. 5

and 6 are simple averages across states and sectors of very different sizes).

Thus Fig. 5 confirms the message of Fig. 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 Fig. 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 Fig. 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
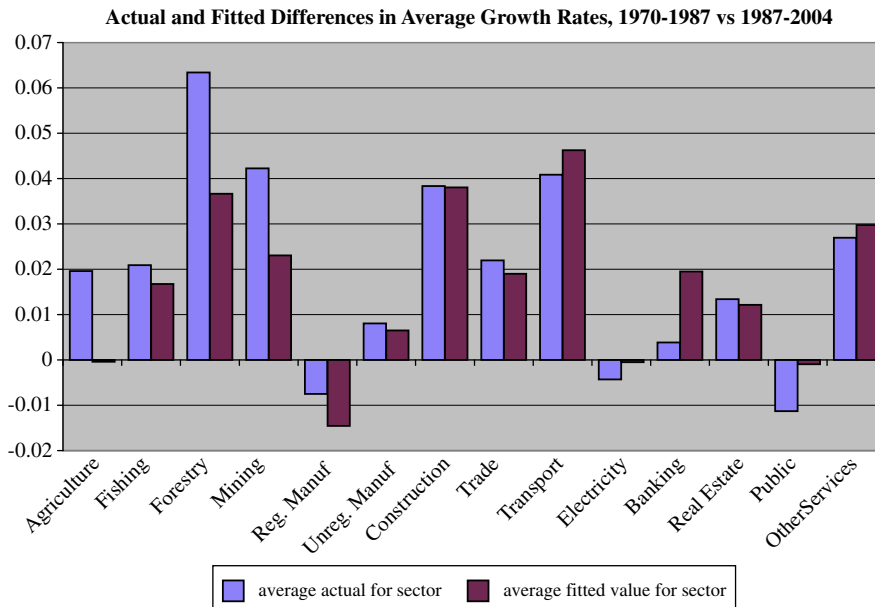

Fig. 5. The V-factor as a representation of growth shifts: by sector.

Actual and Fitted Differences in Average Growth Rates, 1970-1987 vs 1987-2004
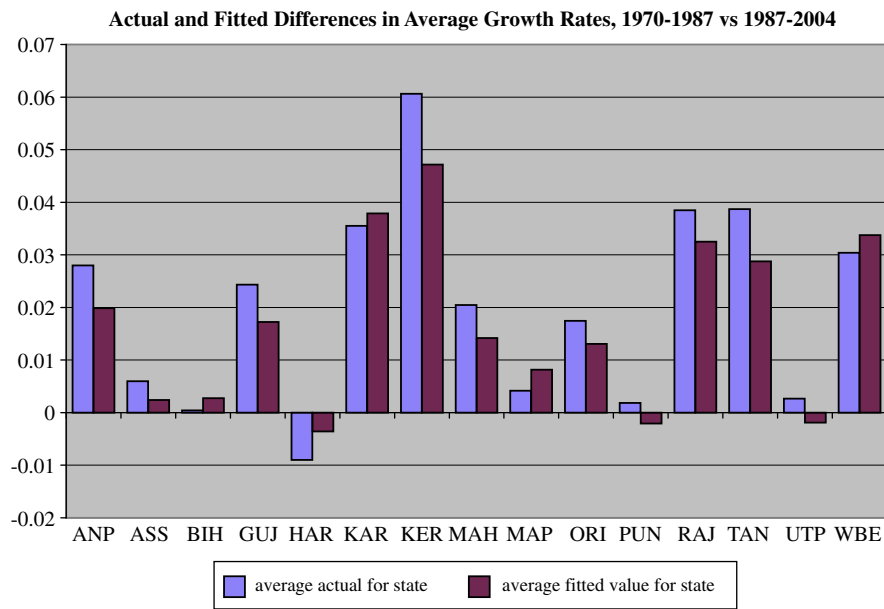

Fig. 6. The V-factor as a representation of growth shifts: by state.

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 turn-around 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 minimum 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 a minimum 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. Thus, 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%. We can therefore 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 Fig. 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?

Fig. 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. 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 tariffs 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 liberalization. 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
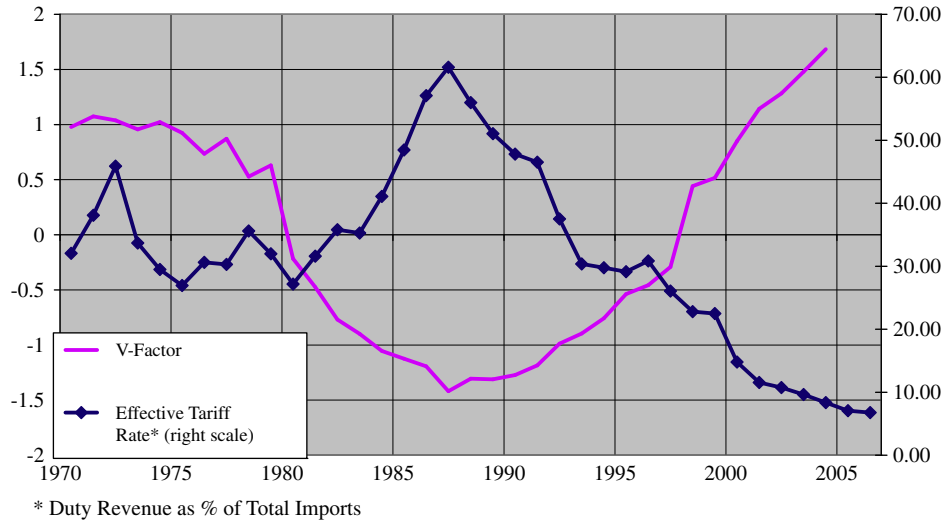
* Duty Revenue as % of Total Imports

Fig. 7. The V-factor and trade liberalization.

the proportion of products banned/restricted, limited or canalized. This shows a modest fall through the 1980s, but much steeper falls thereafter. Thus, Fig. 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.[23]

We emphasize our results on the tariff rates because they represent a clear-cut and measurable change in policy, 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 Fig. 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 Fig. 7.[24] Fig. 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.[25] Finally, in terms of non-trade policy indicators, there was a significant relaxing of the "License Raj", that imposed a wide range of state controls on the manufacturing sector in particular, 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 acting 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.

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 Fig. 2) is quite striking. In this section we use our panel dataset to investigate whether this disparate performance can be captured by observable state characteristics. We find that it can; however our results reveal less about the role of individual indicators.

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 6), in the latter half of the 1980s. In 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).

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 Figs. 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)).[26]

---

[23] 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 minimum 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.

[24] 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.

[25] 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).

---

[26] 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 Figs. 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.

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	Coefficient estimates (p-values in parentheses)			
	(1)	(2)	(3)	(4)
State dummies	all	all	none	none
Sector dummies	all	none	all	all
Share of agriculture, 1987				−0.0014 (0.03)
Share of reg. manufacturing, 1987				−0.0036 (0.01)
Real state income per capita, 1987				0.02 (0.41)
% urban population, 1981				0.011 (0.62)
Literacy rate, 1981				0.006 (0.84)
Average rainfall, 1983–1987				0.0060 (0.47)
Aghion et al's pro-worker dummy				0.0011 (0.82)
Landlocked dummy				−0.0145 (0.14)
Population, 1981				−0.0198 (0.33)
Population growth, 1971–1981				0.542 (0.73)
Development spending, % of NDP, 1981				0.071 (0.27)
Regression diagnostics				
Observations	207	207	207	207
R-bar-squared	0.307	0.115	0.170	0.317
s.e.	0.036	0.041	0.039	0.036
intraclass residual correlation (states)	−0.035	−0.012	0.192	−0.034
intraclass residual correlation (sectors)	−0.099	0.158	−0.118	−0.100
Tests of implied restrictions on Eq. (1) (p-values)				
Likelihood ratio (chi-squared)	n/a	0.000	0.000	0.564
Likelihood ratio (F-test)	n/a	0.000	0.000	0.999
Wald (F-test)	n/a	0.000	0.000	0.463

In regression (4) 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.[27] The overall goodness of fit barely differs from the benchmark regression (1) and the implied restrictions are easily accepted: i.e., the state-level regressors jointly span all significant variation across states.[28,29] 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.[30]

Thus regression (4) can only reveal a limited amount about the role of individual regressors.

- One strongly significant individual effect 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 is a negative impact of the share of registered manufacturing. This result directly contradicts those of Rodrik and 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, Fig. 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 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.[31]
- The remaining state characteristics are all individually significant in regression (4), although collectively they do have some explanatory power (in terms of improved R^2 and information criteria) over and above that due to the two significant regressors. This suggests that further investigation of the role of state-wise factors in the participation in the Indian growth turnaround would be worthwhile.[32]
- Finally, our regression results suggest that the role of public sector output in the turnaround was quite distinctive. Fig. 5 showed that overall it was the slowest growing sector (reflecting this, its sector dummy is significantly negative in regressions (1), (3) and (4)). 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.[33]

---

[27] 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.

[28] While not immediately obvious (indeed we are very grateful to one of the referees for pointing it out) Eq. (2) to (4) are all nested within the specification in Eq. (1), since if we had the same number of linearly independent state-level regressors as there are state dummies, the fit of the equation would be identical. The implied restrictions can therefore be tested either by likelihood ratio tests on the restricted versus unrestricted models, or Wald tests on the unrestricted model. Both are reported in Table 3, and give virtually identical results.

[29] 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.

[30] If each individual state-level characteristic is regressed on the remaining characteristics, the minimum R^2 is above 0.8, and some are very close to unity.

[31] 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 state 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 labor laws, private and public firms cannot fire their employees, and so inefficient labor continues to be employed (see Bhattacharjea, 2006), leading to output losses.

[32] Since the refereeing process for this paper was completed we have carried out a preliminary investigation of the role of these more marginal indicators using robustness analysis along the lines of Sala-i-Martin (1997); these are summarized in Ghate and Wright (2011).

[33] This is presumably 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).

6. Conclusions

In their international study of growth accelerations, Hausmann et al. (2005, p. 328) conclude that:

"It would appear that growth accelerations are caused predominant-ly 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."

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 se-ries 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 re-forms 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 turn-around, and some evidence that sectors where government interven-tion remained significant (most notably in registered manufacturing) participated less in the turnaround.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jdeveco.2011.09.002.

References

Aghion, P., Burgess, R., Redding, S., Zilibotti, F., 2008. The unequal effects of liberaliza-tion: evidence from dismantling the License Raj in India. The American Economic Review 98 (4), 1397–1412.

Angrist, J., Pischke, J., 2009. Mostly Harmless Econometrics. Princeton, New Jersey.

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

Bai, Jushan, Ng, Serena, 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.

Bai, Jushan, Ng, Serena, 2004. A PANIC attack on unit roots and cointegration. Econo-metrica 72 (4), 1127–1177.

Balakrishnan, Pulapre, Parameswaran, M., 2007. Understanding economic growth in India: a prerequisite. Economic and Political Weekly 14, 2915–2922 July.

Basu, Kaushik, 2008. The enigma of India's arrival: a review of Arvind Virmani's propel-ling india: from social stagnation to global power. Journal of Economic Literature 46 (2), 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.

Das, Deb Kusum, 2003. Quantifying trade barriers: has protection declined substantial-ly in Indian manufacturing? ICRIER Working Paper, No. 105.

Ghate, Chetan, 2008. Understanding divergence in India: a political economy approach. Journal of Economic Policy Reform 11, 1–9.

Ghate, Chetan, Wright, Stephen, 2008. The "V-Factor": distribution, timing, and correlates of the Great Indian growth turnaround. DIW Working Paper No. 783, Berlin.

Ghate, Chetan, Wright, Stephen, 2011. Correlates of statewise participation in the great Indian growth turnaround: some preliminary robustness results. Birkbeck College Working Paper 1104. http://www.ems.bbk.ac.uk/faculty/wright.

Harrison, A., 1996. Openness and growth: a time series, cross country analysis for de-veloping countries. Journal of Development Economics 48 (2), 419–447.

Hausmann, R., Pritchett, L., Rodrik, Dani, 2005. Growth accelerations. Journal of Eco-nomic Growth 10, 303–329.

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.

Lee, J.W., 1995. Capital goods imports and long run growth. Journal of Development Economics 48 (1), 91–110.

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

Panagariya, Arvind, 2004. India's trade reform. India Policy Forum 1, 1–57.

Pursell, Garry., 1992. Trade policy in India. In: Salvatore, Dominick (Ed.), National Trade Policies. Greenwood Press, New York.

Rodriguez, F., Rodrik, D., 2001. Trade policy and economic growth: a skeptic's guide to the cross-national evidence. In: Bernanke, B., Rogoff, K. (Eds.), NBER Macroeco-nomics Annual. National Bureau of Economic Research, Cambridge, MA.

Rodrik, Dani, Subramanian, Arvind, 2005. From Hindu growth to productivity surge: the mystery of the Indian growth transition. IMF Staff Papers 52 (2), 193–228.

Sala-i-Martin, X., 1997. I just ran two million regressions. The American Economic Re-view 87 (2), 178–183.

Sen, Kunal, 2007. Why did the elephant start to Trot? India's. Economic and Political Weekly 27, 37–47 October.

Virmani, Arvind, 2006. India's economic growth history: fluctuations, trends, break-points, and phases. Indian Economic Review 41 (1), 81–103.

Yanikkaya, H., 2003. Trade openness and economic growth: a cross country empirical investigation. Journal of Development Economics 72 (1), 57–89.

Further reading

Bai, J., Perron, Pierre, 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.

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

Gregory, Clark, Wolcott, Susan, 2003. One polity, many countries: economic growth in India, 1873–2000. In: Dani, Rodrik (Ed.), In Search of Prosperity: Analytical Narra-tives on Economic Growth. Princeton University Press, New Jersey.

Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53–74.

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

Shin, Y., Snell, A., 2006. Mean group tests for stationarity in heterogeneous panels. The Econometrics Journal 9, 123–158.