

Ghate, Chetan; Wright, Stephen

**Working Paper**

## The "V-Factor": distribution, timing and correlates of the great Indian growth turnaround

Jena economic research papers, No. 2009,010

**Provided in Cooperation with:**  
Max Planck Institute of Economics

Suggested Citation: Ghate, Chetan; Wright, Stephen (2009) : The "V-Factor": distribution, timing and correlates of the great Indian growth turnaround, Jena economic research papers, No. 2009,010, Universität Jena und Max-Planck-Institut für Ökonomik, Jena

This Version is available at:  
<http://hdl.handle.net/10419/31784>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# JENA ECONOMIC RESEARCH PAPERS



# 2009 – 010

## The “V-Factor”: Distribution, Timing and Correlates of the Great Indian Growth Turnaround

by

**Chetan Ghatе  
Stephen Wright**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact [markus.pasche@uni-jena.de](mailto:markus.pasche@uni-jena.de).

Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

Max Planck Institute of Economics  
Kahlaische Str. 10  
D-07745 Jena  
[www.econ.mpg.de](http://www.econ.mpg.de)

© by the author.

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

January 2009

**Chetan Ghate**  
Indian Statistical Institute

**Stephen Wright**  
University of London

### **Abstract**

Following Bai (2004) and Bai and Ng (2004) we estimate a common factor representation of a panel of output series for India, disaggregated by 15 states and 14 broad industry groups. We find that a single common "V-Factor" accounts for a large part of the significant shift in the cross-sectional distribution of state-sectoral output growth rates since the mid -1980s. The time profile of the V-Factor appears to be closely related to trade liberalization.

JEL-classification: O10, O40, O53, O47

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

Acknowledgement: We thank Amit Sadhukhan for invaluable research assistance. We thank Dr. Savita Sharma and Pronab Sen of the CSO for helpful advice on the data. We thank Gerhard Glomm, Sanghamitra 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, and the 45th Meeting of the Indian Econometric Society (TIES) for useful comments. 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. Chetan Ghate thanks the PPRU Committee for financial assistance related to this project.

Contact: Chetan Ghate, Planning Unit, Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi 110016, India. Tel:91-11-4149-3938; Fax:91-11-41493981. Email: cghate@isid.ac.in

## 1 Introduction

In the past two decades or so there has been a remarkable turnaround in Indian growth. From 1960 to 1985 output per capita in India (measured by real net domestic product<sup>1</sup>) grew by only 1.28% per annum, while on the same measure US output per capital grew at 2.35%, so that Indian and US output levels were steadily diverging. In marked contrast, from 1985 to 2004 Indian output per capita grew at 3.86% per annum, while US per capita growth slowed to 1.71%; thus India has been converging towards US output per capita levels at a more rapid rate than it was diverging in the earlier period.

This turnaround in Indian economic growth has inevitably generated considerable public interest and some academic research with respect to the timing, possible causes, and the unevenly distributed nature of the turnaround.<sup>2</sup> In this paper we present evidence on all three issues.

Our approach exploits the fact that, amongst economies at a 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 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 mid-1980s.

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

---

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

<sup>2</sup>For example, see Rodrik and Subramanian (2005), Virmani (2006), Balakrishnan and Parameswaran (2007), Sen (2007), and Basu (2008).

The strong explanatory power of this common "V-Factor" suggests a single common cause. We show that its profile over time is strongly correlated with the pattern of trade liberalization, as summarized by the effective tariff rate. This mirrors the findings in Jones and Olken (2008) who find that growth takeoffs in the cross country context are largely associated with expansions in international trade. In this respect our analysis 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 observable shift in policy.<sup>3</sup>

But puzzles still remain. The most notable is the distinctly uneven distribution of the growth turnaround across the major states, several of which have shown little or no increase in growth. We examine whether particular state characteristics are associated with the strength of the impact of our "V-Factor". Our results here are less clear-cut, and largely negative, in the sense that at best we appear to be able to rule out some possible explanations that have been proposed in past research. There is some weak evidence, however, that the capacity of a given state to exploit the opportunities presented by trade liberalization is helped by education and urbanization, and hindered by the size of its agricultural and public sectors.

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. In Section 4 we illustrate the link with trade liberalization. In Section 5 we look for correlations between differential impacts of the V-Factor and characteristics of the different Indian States. In Section 6 we attempt to reconcile our results with a standard model of convergence; Section 7 concludes the paper, and an Appendix provides details of data construction and statistical analysis.

---

<sup>3</sup>Rodrik and Subramanian identify a shift in growth in 1980, based on aggregate GDP data. They show that this was well before any directly observable policy changes, and attribute the change to attitudinal shifts on the part of the national government in favor of business (we discuss the evidence for this mechanism in Section 5). 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 more thoroughly in Section 4.

## 2 Sectoral and state-wise shifts in growth

Figures 1 and 2 give two alternative broad-brush pictures of the turnaround in growth.

Figure 1 shows that virtually all sectors of the private sector economy have seen substantial increases in growth over the same two sub-samples, 1960 -1985 and 1985 -2004, referred to in the introduction, albeit from often significantly different initial values.<sup>4</sup> The only exception to this general pattern was agriculture, for which the pickup in growth was very much more modest. Growth in the public sector, in contrast, was more or less identical in both sub-samples.

[Insert Figure 1]

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.<sup>5</sup>

[Insert Figure 2]

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.<sup>6</sup> The remaining

---

<sup>4</sup>All growth rates are shown as growth of sectoral net domestic product per head of total population, since no reliable figures for total 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.

<sup>5</sup>We have made adjustments to output series to allow for changes in state definitions. The sixteen Indian states are: Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Jammu and Kashmir, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal.

<sup>6</sup>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 then had ceased any tendency to converge.

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. Eight states (Andhra Pradesh, Gujarat, 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; a ninth, Madhya Pradesh, managed a very significant shift in growth, but from such a low base that it managed only a modest rate of convergence (partly thanks to a somewhat lower rate of growth in the USA). In the remaining states, however, growth remained at a fairly similar rate to that in the previous sub-period. Within this group four states, Haryana, Punjab, Orissa and Uttar Pradesh did achieve limited convergence (in the last case, at a painfully slow rate); but the remaining three states, Assam, Bihar, and Jammu and Kashmir 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,<sup>7</sup> so differences in growth rates of output growth will have corresponded to significant differences in consumption growth.

### 3 Statistical Analysis

#### 3.1 The dataset

For the purposes of statistical analysis 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. We exclude three series due to clear data problems, leaving 207 series over a balanced panel from 1970 to 2004. We also have a large subset of 139 series

---

<sup>7</sup>Both consumption and output are measured at current prices. Details of data construction for consumption are in the Appendix.

extending back to 1960, which we use below for robustness checking of our results. All series are measured in constant prices per head of the population in the relevant state.<sup>8</sup>

### 3.2 Evidence of common structural shifts?

While the visual evidence in Figures 1 and 2 appears very striking, it at least in principle 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 207 series in our panel over two samples, 1970 to 1985 and 1986 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]

Table 1 shows the results of Kolmogorov-Smirnov tests of the null that both sets of growth rates are drawn from the same distribution. The tests are carried out both on the sub-sample average growth rates, 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.<sup>9</sup> Examination of tests carried out over a range of breakpoints during the 1980s suggest that this result is not simply an artefact of the breakpoint chosen.

---

<sup>8</sup>Full details of data construction are given in the Appendix.

<sup>9</sup>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 probably less reasonable for the test as applied to the annual series. 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.



**Table 1**  
**Kolmogorov-Smirnov Tests for Equality of Distribution**

$H_A$	Functions <sup>10</sup>			
	D Statistic (ss)	D Statistic (ann)	P Values (ss)	P Values (ann)
0	0.2429	.1046	0.000	0.000
1	-0.0048	-0.0011	0.995	.996
Combined K-S	.2429	.1046	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 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 (assumed to be  $I(1)$ ) 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,

$$y_{it} = \beta_{i0} + \beta_{i1}F_{1t} + \dots + \beta_{ik}F_{kt} + u_{it}; i = 1..N \quad (1)$$

$$\Delta F_{jt} = a(L)\varepsilon_{jt}; k = 1..k \quad (2)$$

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

where  $y_{it}$  is log output per capita in state-sector  $i$ , where  $i$  runs from 1 to 207 (i.e., we do not explicitly distinguish between the state and the sector dimension); the  $F_{jt}$  are common factors that are subject to permanent shocks,  $\varepsilon_{jt}$ ; the  $\beta_{ij}$  are factor loadings on the factors; and the  $u_{it}$  capture the remaining idiosyncratic dynamics. The idiosyncratic shocks,  $\omega_{it}$ , may in principle be mutually correlated but Bai (2004) outlines restrictions on

---

<sup>10</sup>The D Statistic (ss) in the second column is based on the sub-sample growth rates: 1961-1985 and 1985-2004. 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 1971 onwards for the annual data. 0 indicates that we test the null against the alternative hypothesis that the second period 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.

the nature of this correlation. 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 idiosyncratic 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. For robustness, we also consider the alternative approach in Bai and Ng (2004) which is consistent even when the idiosyncratic components  $u_{it}$  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 both approaches the processes for the idiosyncratic components in (3) are not estimated directly, but are derived implicitly from the estimated factors, as

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

Bai and Ng (2004) then propose that panel unit root tests be applied to the implied idiosyncratic 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.<sup>11</sup> 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  $k$ .

---

<sup>11</sup>See Bai (2004), p. 145.

**Table 2**  
**Value of  $k$ , the Number of Common Factors, implied by**  
**Information Criteria<sup>12</sup>**

Panel Information Criterion	Estimation in Levels		Estimation in Differences	
$IPC_1(k, k_{\max})$	$1(k_{\max} < 5)$	$2(k_{\max} \geq 5)$	$0(k_{\max} < 5)$	$1(k_{\max} \geq 5)$
$IPC_2(k, k_{\max})$	$1(k_{\max} < 6)$	$2(k_{\max} \geq 6)$	$0(k_{\max} < 6)$	$1(k_{\max} \geq 6)$
$IPC_3(k, k_{\max})$	1		0	

While the results shown are somewhat ambiguous, they are not as ambiguous as they might appear at first glance. 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 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 this latter result would imply that each of the 207 series 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.

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

---

<sup>12</sup>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).

that this confidence appears to be borne out by the explanatory power of the factor representation).

In the Appendix 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.

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.<sup>13</sup>

[Insert Figure 4]

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 however, quite close to being monotonic (and indeed almost piecewise linear) either side of its vertex in the mid-1980s (to be precise, in 1987). In the Appendix we show that the timing of this breakpoint appears to be robust to the inclusion or exclusion of series with high volatility, to adjustments for short-run weather-induced fluctuations, or to a lengthening of the sample backwards with a smaller subset of series.

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 rather less distinct apex (it is closer to being a U-Factor than a V-Factor). For the rest of the paper we shall focus on results based on levels estimation, which appears to be validated by the rejection of unit

---

<sup>13</sup>Since the scale of the factors is irrelevant, all three series are normalised to have zero mean and unit variance.

roots in idiosyncratic components, robustness to which would be the primary rationale for estimation in differences. However results from estimation in differences are mainly very similar.

Finally, as mentioned in the introduction, 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 idiosyncratic components allows in principle for different persistence properties, which allow some series to respond more rapidly to the common permanent shock).

### **3.4 The V-Factor as a representation of growth shifts**

Figures 5 and 6 both illustrate the degree to which the common factor representation captures the key properties of the common shift in growth.

In Figure 3 we showed the strong evidence of a rightward shift in the cross sectional distribution of growth rates. In figures 5 and 6 we show, for each series, the difference in growth rates for each series between the same two sub-samples, alongside the predicted change from the factor model. Not only is the explanatory power quite high (with a correlation coefficient between actual and predicted of 0.77), but, once the data are analyzed in terms of shifting growth rates, as in the charts, it turns out that this explanatory power is essentially entirely due to the V-Factor: a factor model in levels with only a single common "G-Factor" yields a correlation coefficient between actual and predicted insignificantly different from zero (as we would expect since such a model essentially implies nearly constant predicted growth).

The charts both present identical results, differing only in the order in which they are presented. In Figure 5 the results are grouped by sectors, while in Figure 6 they are grouped by states. Each chart also shows the average estimated impact of the V-Factor, by sector (in Figure 5) and by state (in Figure 6).

[Insert Figure 5 and 6]

While the distribution of growth shifts across states within a given sector, or across sectors within a given state, is quite dispersed, 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.

#### **4 The V-Factor and Trade Liberalization**

Having established the common nature of the V-factor, and its ability to account for a large part of the cross-sectional distribution of the shift in growth, it is clearly of interest to examine its timing, and to compare it with the timing of policy shifts. Here our results suggest a resolution of a puzzle identified by Rodrik and Subramanian (2005): based on aggregate data (see Rodrik and Subramanian (2005, Figure 1) they identified a key turning point in growth in the early 1980s (or even late 1970s) which appeared to pre-date major policy changes. However, as Figure 4 shows, in our more disaggregated approach the V-Factor identified by principal components has an apex distinctly later (in the mid-1980s). Figure 7 shows that the time profile of the V-Factor matches very well indeed with the timing of one key policy change: the liberalization of trade policy via tariff reduction (the blue line).

[Insert Figure 7]

Of course 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 independent evidence does suggest it had a particularly important role. On the first point, Rodrik and

Subramanian (2005) examine a range of other possible triggers for the turn-around but none have the same distinctive pattern (the pattern of real exchange changes, for example, reveals a significant depreciation during the 1980s, but a subsequent modest reversal). Additionally 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. But there is also widespread international evidence for the crucial role of openness in economic growth (Temple et al. (2005)) and strong evidence that it has acted as a trigger in economic growth shifts in other countries (Hausmann et al, 2005; Jones & Olken, 2008). Hence, there is a clear economic basis for the strength of the relationship.<sup>14</sup>

Our identification of a turning point in the mid-1980s is in contrast both with the results of Rodrik and Subramanian and with some other recent research based on all-India aggregate figures. Virmani (2006) finds an upward break in growth in the manufacturing sector in 1980-81 and concludes that this is responsible for the structural break in overall growth, which he finds to be in 1981-82. Balakrishnan and Parameswaran (2007) utilize the multiple structural break approach of Bai and Perron (1998) and show that the break in real GDP growth occurs in 1978-79, with the take-off in growth occurring prior to the positive break in manufacturing (which they date to 1982-83). However, as argued by Basu (2008) a common criticism that applies to the existing empirical literature on the Indian growth turnaround is the special nature of the period 1979-1980, which saw a sharp contraction in Indian real GDP growth (-5.2%). Basu concludes that the approach followed by Balakrishnan and Parameswaran (2007) – by essentially fitting linear segments to a fluctuating growth pattern – would find a propensity for the break to appear before 1979-1980. A similar critique applies to the recursive Chow test approach of Virmani (2006). It is notable that as shown in Figure 4, our estimate of the V-factor based on differenced disaggregated data also shows a sharp contraction in 1979-1980; but it then continues to fall, and then bottoms out only in the mid eighties.

---

<sup>14</sup>As we show in the appendix, the apex of the V – in the mid 80s – is fairly robust to a wide variety of data modifications. 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 may conceivably be before de jure changes in the aggregate policy regime.

## 5 V-States vs Non-V States?

While the common nature of the growth turnaround, as identified by the V-Factor, appears to correspond to a common shift in trade policy, the disparate impact of the impact of the V-Factor across the states presents something of a puzzle. We have seen that the V-factor accounts for a large part of the turnaround in growth of the underlying series in our panel. In the Appendix we show that we can derive implied loadings on the V-Factor for each state (i.e., an appropriate weighted average of the  $\beta_{i2}$  in equation (1) taking account of the time-varying sectoral composition of output in each state). While in principle this means that the implied state factor loadings are time-varying, we show in the Appendix that the degree of time variation is quite small. As would be expected, the average state wise factor loadings show a very similar pattern across states to the disparate pattern of growth shifts shown in Figure 2. Since we have been able to identify a strong apparent correlation between the V-factor and shifts in policy, the obvious question arises: why did some states respond much more positively to this shift than others?

Table 3 examines this issue by showing correlations between identifiable state characteristics shortly before the turnaround, and their estimated V-factor loadings.<sup>15</sup> The table provides both negative and some (weaker) positive results. On the negative side, it allows us to dispose of some candidate explanations: a) the turnaround in growth was not restricted to a club of richer states: initial income levels were unrelated to the magnitude of the response to the V-Factor; b) explanations based on differences in key magnitudes in a standard Solow-style growth model (saving, investment population growth rate and level) do not show any systematic differences (the sign of the investment correlation is perverse - the correlation with population growth is of the correct sign but very low); c) the direct contribution of the public sector to the turnaround appears to have been at best weak, and possibly perverse: there was essentially a zero correlation between the initial values of development spending and the subsequent impact of the V-factor; and total public spending actually had a v-factor loading which

---

<sup>15</sup>The data used in this table come from a new dataset with a range of state level and sectoral data from 1960 onwards, which encompasses all data used in this paper. Further details of the dataset and the data definitions in are provided in an earlier working paper version of this paper, Ghate and Wright (2008).



was somewhat negatively correlated with the overall state loading (i.e., "V-states" - those with a positive loading on the V-Factor - tended if anything to have slowdowns in growth of public spending after the turnaround in overall growth).

**Table 3 State Characteristics and State V-Factor Loadings**

<b>State Characteristics</b>	<b>Correlation with average state v-factor loading</b>
<b>A rich state club?</b>	
log Real Output per capita, 1985	<b>-0.07</b>
<b>Solow Variables</b>	
Fixed Investment, % of NSDP, 1981	<b>-0.42</b>
log Population, 1981	<b>0.09</b>
Population Growth Rate, 1971-81	<b>-0.22</b>
<b>The role of Public Spending</b>	
Development spending, % of NSDP, 1981	<b>0.17</b>
Public Spending v-factor loading	<b>-0.22</b>
<b>Supply Side Characteristics</b>	
Share registered manufacturing, 1985	<b>0.18</b>
Electricity generation, kwh per capita, 1981	<b>-0.19</b>
Share of agriculture, %, 1985	<b>-0.68</b>
Literacy Rate, 1981	<b>0.49</b>
Urban Population, %	<b>0.32</b>

The bottom block of the table provides evidence of some proxies for supply-side characteristics of individual states. The table shows that "V-states" tended to be somewhat more literate, somewhat more urbanized, and (the strongest correlation shown) had lower shares of agriculture (which our earlier charts showed was typically relatively little affected by the V-factor); but most of the correlations shown are fairly weak.

A final negative result: Table 3 shows that there was essentially no link between V-factor loadings and the share of registered manufacturing, which played an important role in Rodrik and Subramanian's (2005) explanation of the turnaround. Indeed, if anything the link goes the wrong way, since Figure 5 showed that registered manufacturing was the only sector of the private sector economy with a *negative* average V-factor loading. This suggests that, while registered manufacturing may, in line with Rodrik and

Subramanian's analysis, have been a catalyst for growth in the early 1980s, it was far from being an engine of growth over the longer term.<sup>16</sup> If anything, the very disparate pattern of individual states around the negative average loading for registered manufacturing shown in Figure 5 reveals it as a "separating" sector; i.e., in V-states the registered manufacturing sector was typically a V-Sector, but in non-V states it was "anti-V".

## 6 Trying to make sense of the V-Factor

The V-Factor provides evidence of a highly pervasive, but by no means universal shift in behavior in India during the course of the 1980s. Can we reconcile this evidence with any underlying economic model? To simplify matters we focus on a model of output per capita at the state level, and ignore sectoral issues. Consider a fairly general model of convergence of the form

$$\Delta (y_{it+1} - y_{t+1}^{US}) = \alpha_i \left( y_t^{US} + s_{it} + s_t^{India} - y_{it} \right) + \Delta TFP_{it+1} - \Delta TFP_{t+1}^{US} + \varepsilon_{it+1}, \quad (5)$$

where  $y_{it}$  is log output per capita for state  $i$ , the  $s_{it}$  and  $s_s^{India}$  variables captures factors that determine steady-state output relative to the frontier represented by  $y_t^{US}$ , log output per capita in the United States, for individual states and for India as a whole;  $TFP_{it}$  and  $TFP_t^{US}$  is growth rate of total factor productivity in state  $i$  and in the United States and  $\varepsilon_{it}$  captures short-run cyclical factors.

As noted in relation to the discussion of Figure 2, before the mid-1980s very few Indian states were converging towards the frontier, and if so only marginally. In contrast, since the mid-1980s the majority of states have all been converging (albeit at very different rates), but a few states have continued to fall behind.

The simple framework of (5) offers a range of possible ways of accounting for the all-India pattern; but it is distinctly harder to account for the relative performance of different states within the standard convergence framework.

It seems reasonable to argue that the sum of the last three terms on the right-hand side of (5) is unlikely to provide an adequate explanation

---

<sup>16</sup>Balakrishnan and Parameswaran (2007) also argue that registered manufacturing can only have played a limited role in the turnaround given its small share in GDP (8.7 %).

of longer-term trends. In standard Cobb-Douglas type technology models TFP growth shocks are common across all economies and hence cancel out precisely. But even if they are country specific, such relative shocks might reasonably be assumed to have a stationary distribution. The same applies to the short-term error term,  $\varepsilon_{it+1}$ . Thus we need to look for an explanation somewhere in the first term.

One possible (and rather pessimistic) interpretation of the earlier period was that the bracketed "convergence" term (the term multiplied by  $\alpha_i$ ) was on average close to zero - i.e., that most, or possibly all Indian states were, conditional upon the  $s_{it}$  and  $s_t^{India}$  processes, fairly close to their steady-state values. The downward drift in most states' relative output levels would, according to this interpretation, be interpreted either as a succession of bad relative TFP growth shocks, or possibly (and even more pessimistically) as a downward drift in  $s_t^{India}$ . This pessimistic interpretation is largely consistent with the evidence presented in, for example, the international panel data study by Islam (1995), based on a dataset which largely preceded the Indian turnaround, in which the India-specific level of total factor productivity (conditioning upon proxies for human capital and Solow variables) was estimated at only around 7% of that in the USA.

It is harder to continue the logic of this explanation after the growth turnaround. The evidence of the V-Factor, and its correlation with our measure of trade liberalization, suggests very strongly that the impetus for the turnaround was common across all states, hence it is reasonable to attribute them to changes in the common Indian steady state factor,  $s_t^{India}$ . Given the subsequent dramatic changes in rates of convergence, then, conditional upon a reasonable degree of stability in the other elements on the right-hand side of (5), the implied changes in  $s_t^{India}$  must have been quite dramatic. Rodrik and Subramanian (2005) argue that this is plausible because India was well away from its production possibility frontier.

But since these changes were common across states, the puzzle presented by the differential impact of the V-Factor is why any such shift in  $s_t^{India}$  did not have largely symmetric effects across the states. There is one possible explanation which reconciles both the all-India and state wise evidence. The analysis of these shifts has implicitly assumed that the state-specific rates of convergence,  $\alpha_i$  were both strictly positive and reasonably similar across states. But an alternative explanation would attribute the pattern of the

evidence largely to the  $\alpha_i$  themselves. On this interpretation, and consistent with the arguments of Rodrik and Subramanian, the bracketed expression in the first term was *not* necessarily close to zero in the first period; but failure to converge to the global frontier was largely due to the  $\alpha_i$  being so close to zero that differences between actual and steady state income levels had essentially no impact. The turnaround in growth and its differential pattern would then be attributed to some combination of a common shift in  $s_t^{India}$  and state wise differences in the  $\alpha_i$ . A differential impact of the all-India shock might be attributed to different values of  $\alpha_i$ , with non-V states, by implication, having  $\alpha_i$  values extremely close to zero, thus closing off any convergence response.

But a further possibility is that the differential impact of the V-factor reflects not just differential responses to common shocks to the steady states, but also shocks to the  $\alpha_i$  themselves. One interpretation of convergence is as a process of arbitrage, driven by international differences in factor returns. Even in a frictionless model of convergence, low values of  $\alpha_i$  can reflect low inter-temporal elasticities of substitution, with the limiting case of  $\alpha_i = 0$  corresponding to an elasticity of inter-temporal substitution of precisely zero (Barro and Sala-Martin, 1992; Campbell, 1994). But models with frictions can also generate similar results, even when the true elasticity is positive. On this interpretation, the reforms of the 1980s and thereafter may not just have raised steady-state output levels, but may also have reduced frictions; with some states being better capable of exploiting the implied arbitrage opportunity. All states might in principle ultimately converge on very similar long-run output levels, but with very different rates of convergence.

This does not, of course, provide a full answer to the question of why different states might have such different rates of convergence, but it is at least suggestive of avenues to explore in future research. The correlations presented in Section 5 suggest that a) trade liberalization offered more scope for expansion in non-agricultural sectors of the private economy (where the opportunities for international arbitrage might reasonably be expected to be weaker); b) that the impact was stronger in states with higher levels of education and urbanization; and c) that the greater the role of the public sector in a given state, the weaker the impact of liberalization. It would be a very strong (and very pessimistic) conclusion to draw that these differences should have an impact on any ultimate steady state. But it does seem quite

plausible that they might all have a significant influence on the rate at which convergence to that steady state is achieved.

## 7 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 strong link with trade liberalization. We also argue that registered manufacturing is unlikely to be a cause of the V-Factor. We report some weak evidence that the absorptive capacities of a given state to exploit the opportunities presented by trade liberalization is helped by education and urbanization, and hindered by the size of its agricultural and public sectors. The very different extent to which different states have participated in this turnaround presents a puzzle to standard models of convergence if all states are assumed to converge at roughly the same rate, but is somewhat easier to explain if convergence rates differ.

## 8 Bibliography

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.

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

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

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.

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.

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 fac-

tors in factor models with large datasets", Queen Mary College, University of London, *Working Paper* no. 525

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.

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.

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.

## 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 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.<sup>17</sup>

#### 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

---

<sup>17</sup> These changes mainly affect Bihar and, to a lesser extent, Madhya Pradesh 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.

## **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]

### **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]

## C Robustness Checks for V-Factor Estimates

### C.1 Robustness to changes of sample

As noted in the main paper, our core analysis is carried out on a balanced panel of data for 15 states. However for a subset of states we have a longer run of data. 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. Given our interest in the state wise impact of the V-Factor our results in the main paper focus on results from 1970 onwards with the maximum cross-sectional coverage of the large states; however 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. 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]

### C.2 Robustness to idiosyncratic volatility

Our dataset includes a number of series with very high levels of volatility. Some of this volatility is inherent to the series. Agricultural output, in particular, is inevitably affected by weather conditions. But other volatility may also reflect measurement problems. We illustrate the robustness of our estimation to both forms of volatility by illustrating the impact on the estimate of the first principal component in differences of two amendments to the data.

**Rainfall Adjustment:** We prior-filter the differenced data 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 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

**Exclusion of Outliers:** 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  $(-63\%, 71\%)$ ). This reduces the cross-sectional dimension to 194, as compared to 207 in our base case.

Figure A4 shows the results of each of these experiments, taken in isolation, and in combination. All four estimates are extremely similar, demonstrating that the cross-sectional dimension is sufficiently large that strictly idiosyncratic volatility effectively cancels out in terms of its impact on common factors.

[Insert Figure A4]

### C.3 Data Construction for Figures 4, 5 and 6

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).

For Figures 5 and 6 we construct estimated permanent components consistent with the estimated transitory components as defined in (4), as

$$\widehat{y}_{it}^P = \widehat{\beta}_{i0} + \widehat{\beta}_{i1}\widehat{F}_{1t} + \widehat{\beta}_{i2}\widehat{F}_{2t}, \quad (6)$$

where, given orthogonality of the estimated factors, we can derive the  $\widehat{\beta}_{ij}$  directly by OLS. We then calculate differences between average growth rates (i.e., average log differences) in  $\widehat{y}_{it}^P$  in the samples 1971-1985 vs 1986-2004, and compare these with the same calculations for the underlying  $y_{it}$ . If the V-Factor,  $\widehat{F}_{2t}$ , is excluded from the calculation of permanent components the correlation between actual and predicted differences in average growth rates is -.061, whereas if the V-Factor is included the correlation is 0.77. If

highly volatile series are removed from the dataset, as in Appendix C, the correlations are -.045 and 0.89, respectively.

## D 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.

## **E 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. Our results relating  $\widehat{F}_{2t}$  to other proxies for trade liberalization - such as India's real effective exchange rate (e.g., see Rodrik and Subramanian (2005, Figure 6)) - also suggests that the time profile of the V- factor matches well with a sudden real depreciation of the Indian rupee starting in the mid-eighties (with and without export subsidies). However, latest data show that this is was partially reversed in recent years.

## **F Data Construction and Sources for Table 3**

In the representation in (1) each individual series has a time-invariant factor loading on the V-factor. The implied loading for total output in each state is therefore a weighted average of these individual factor loadings, weighted by shares in state output. State-wise V-Factor loadings are therefore time-varying. However, Figure A5 shows that the average loadings over the same two sub-samples used in the rest of the paper are very similar. The correlations shown in Table 3 are with the average loadings over the full sample. The state characteristics summarized 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).

[Insert Figure A5]

Figure 1

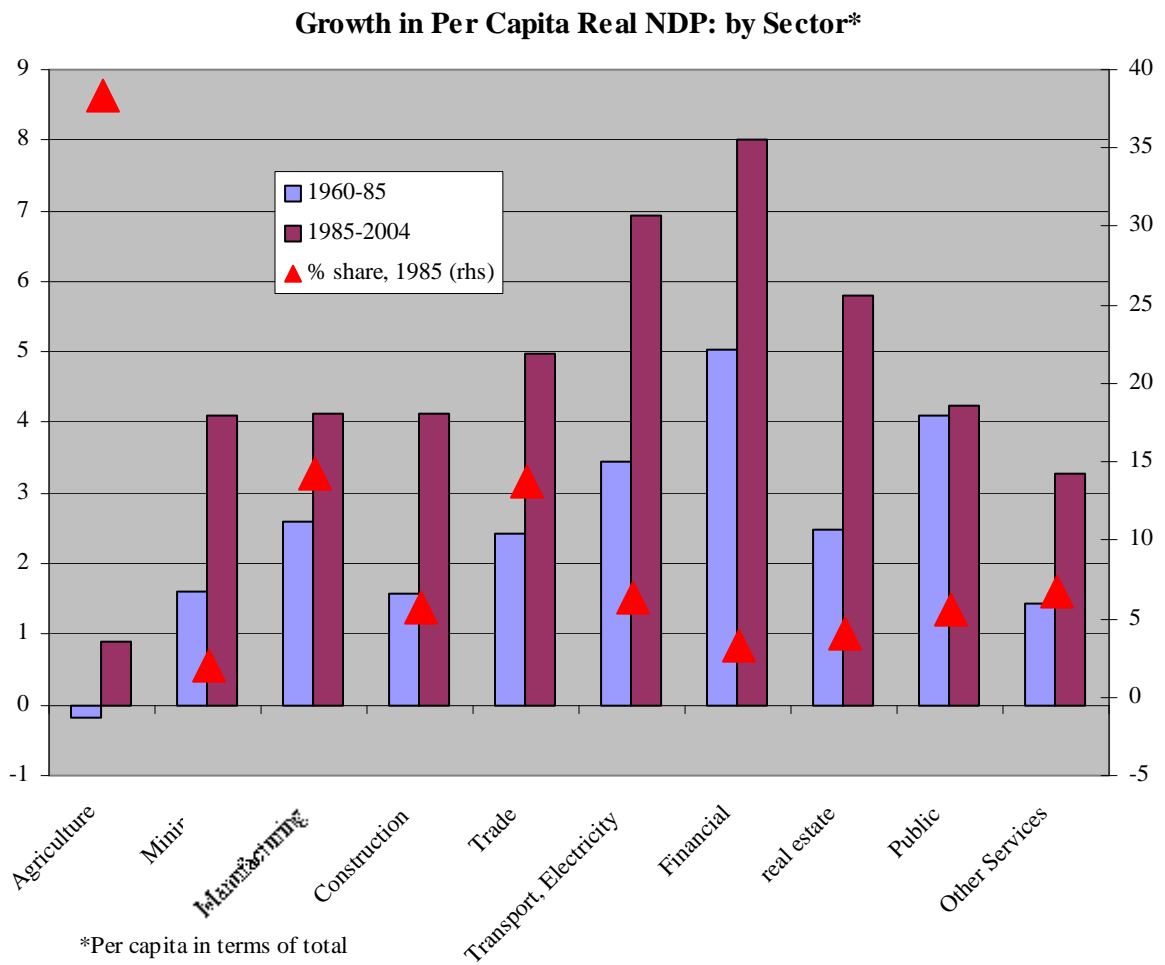


Figure 2

Growth in Per Capita Real NDP, by State

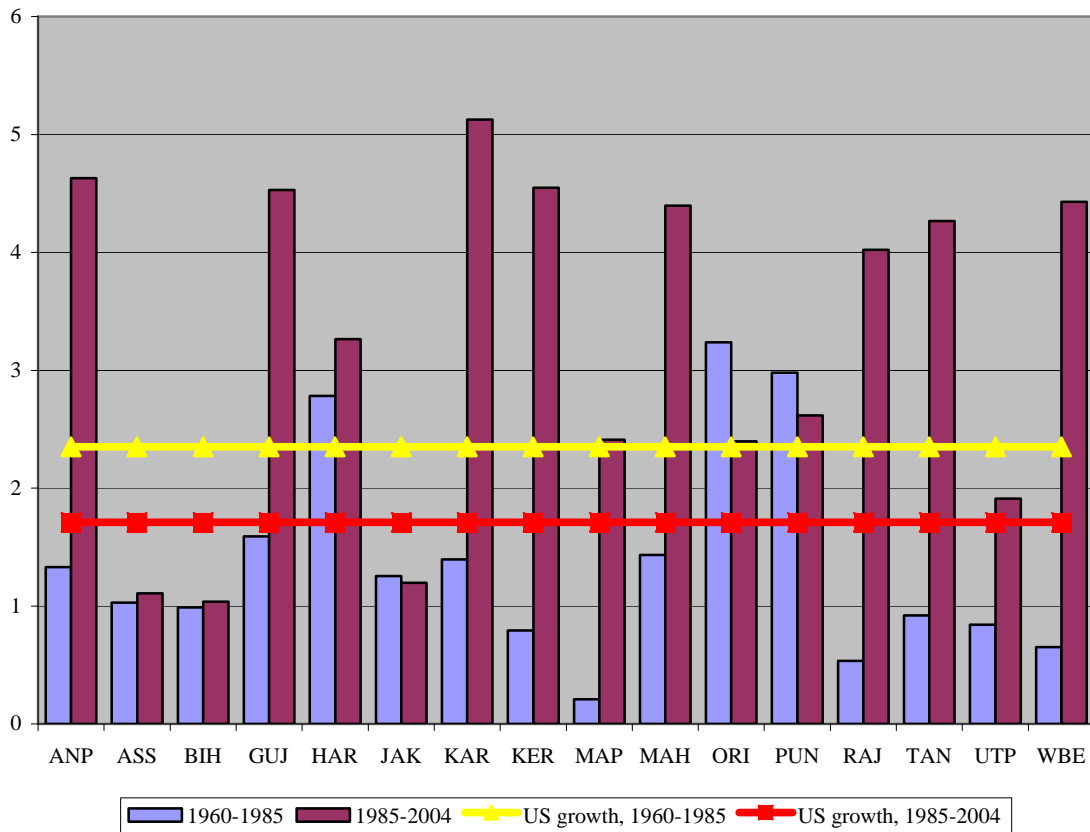


Figure 3

The Distribution of Average Sub-Sample Growth Rates

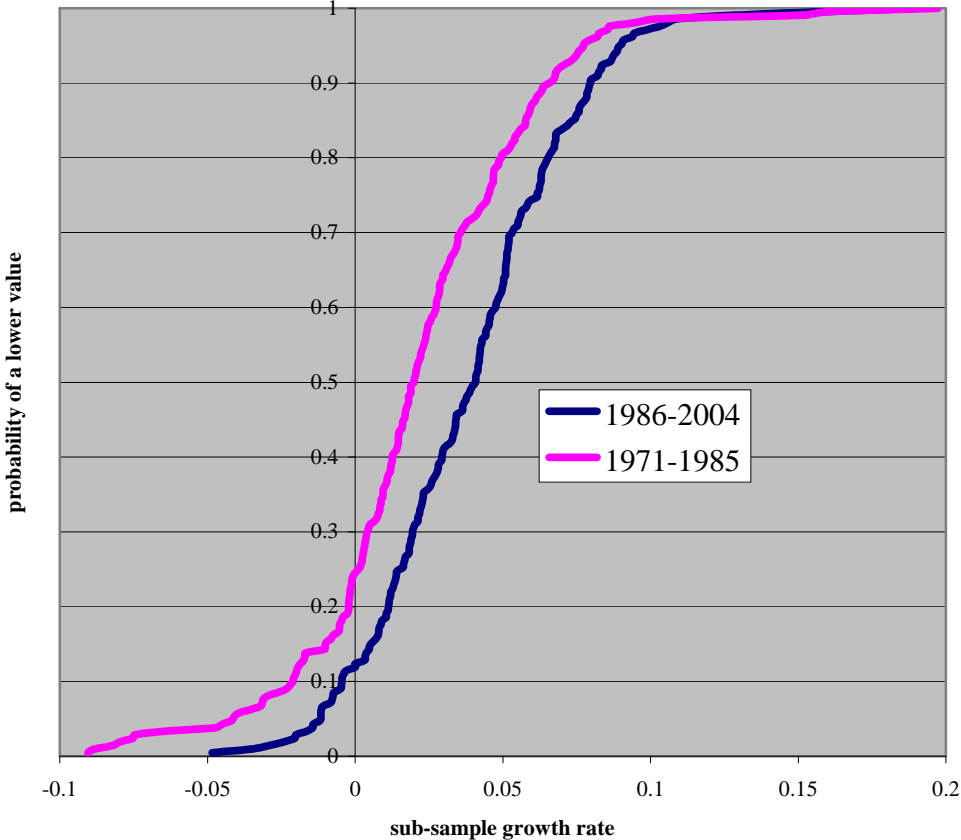




Figure 4

Common Factors Estimated by Principal Components

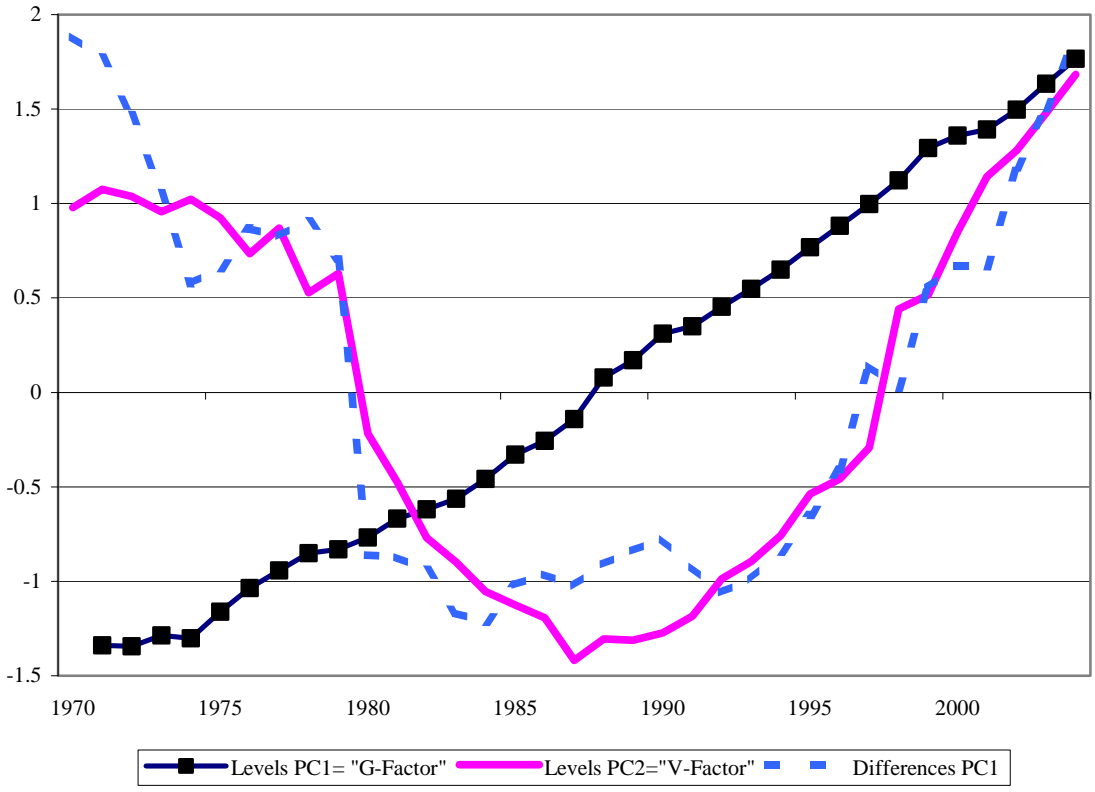


Figure 5

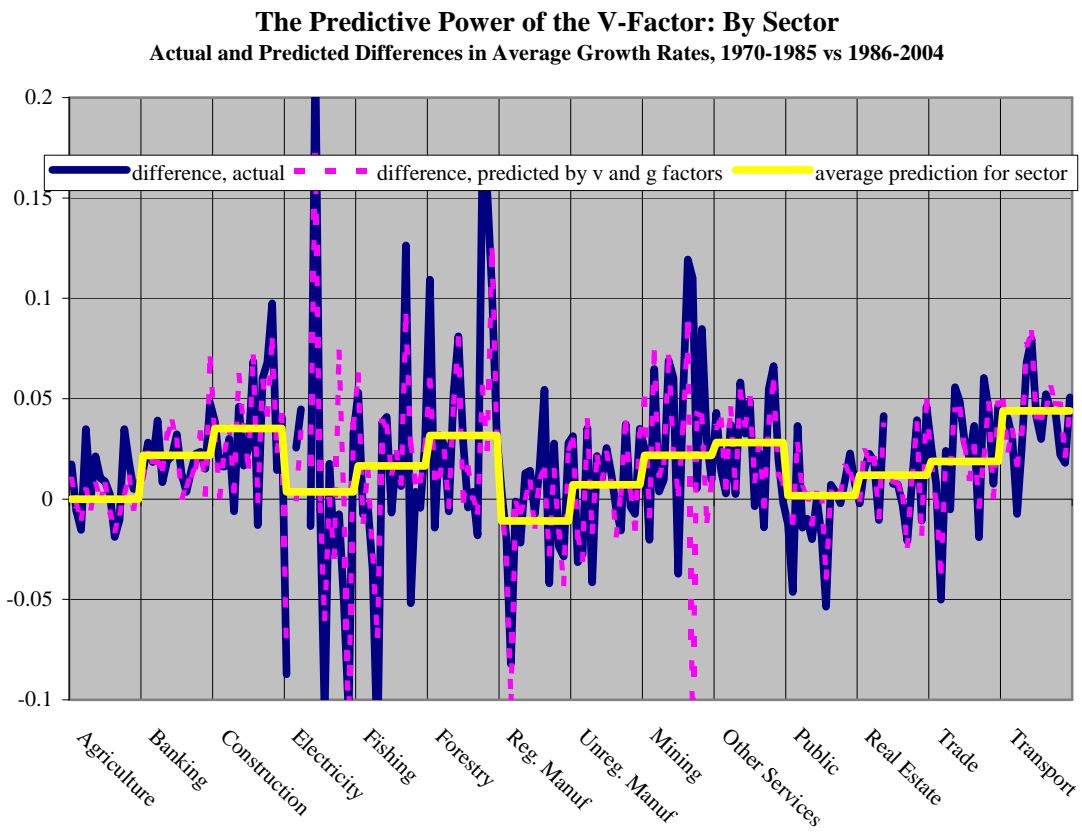


Figure 6

**The Predictive Power of the V-Factor: By State**  
Actual and Predicted Differences in Average Growth Rates, 1970-1985 vs 1986-2004

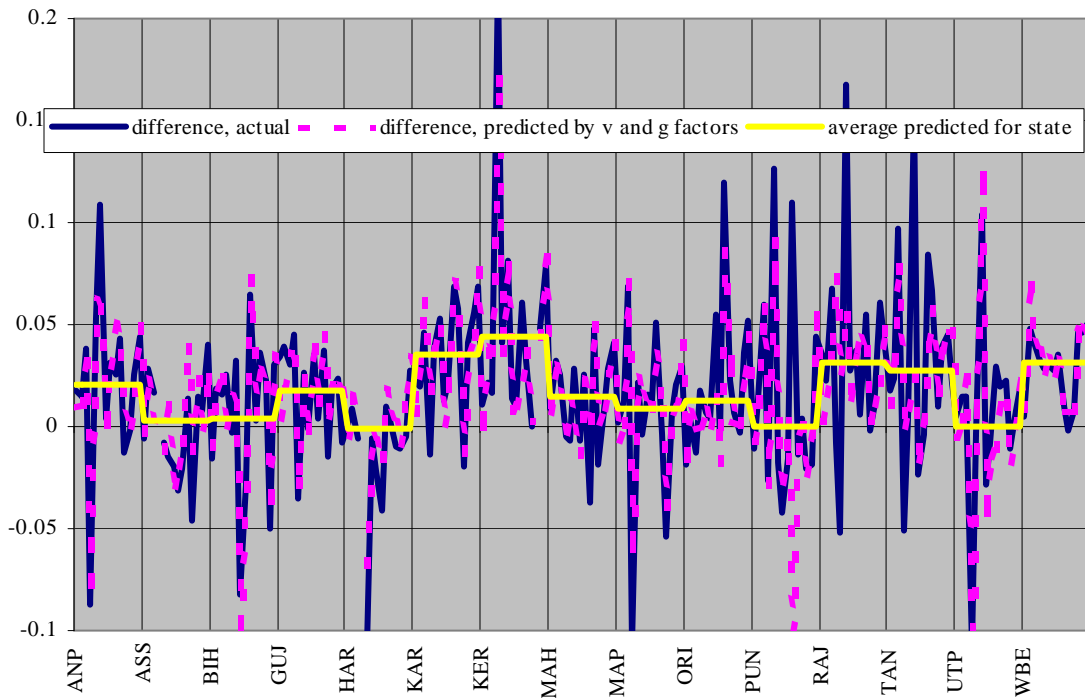
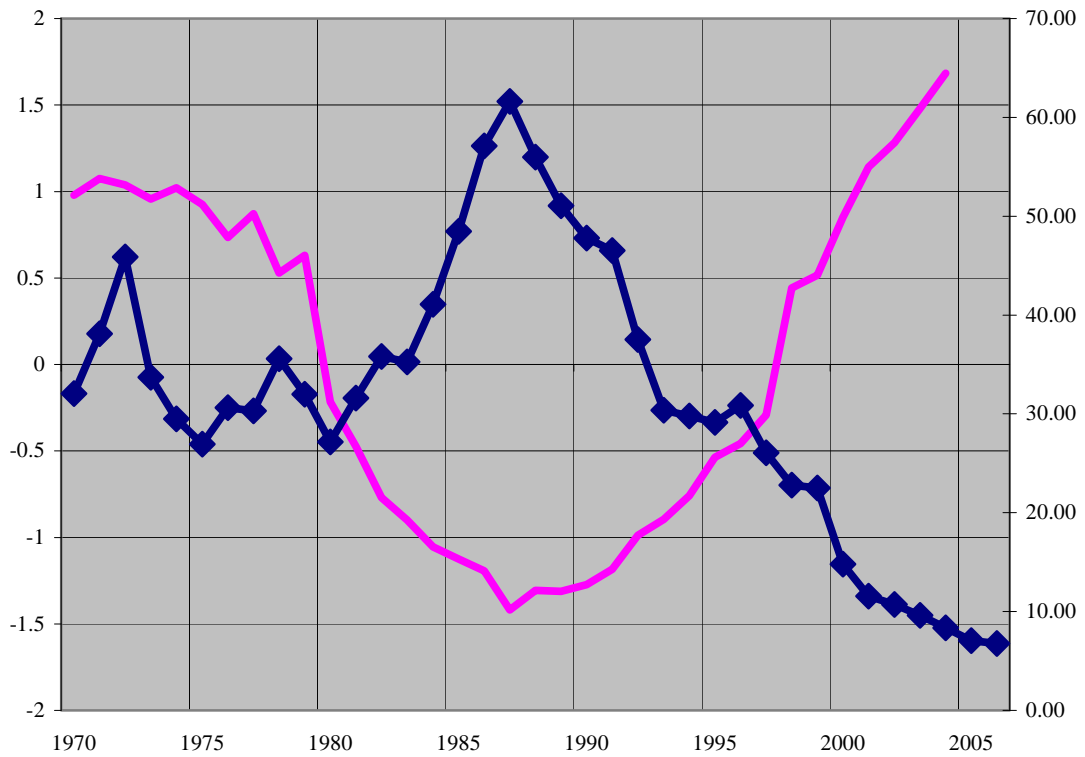


Figure 7

**The V-Factor and Trade Liberalisation**



\* Duty Revenue as % of Total Imports

Figure A1

Ranked ADF Statistics for Estimated Transitory Components from Levels Estimation

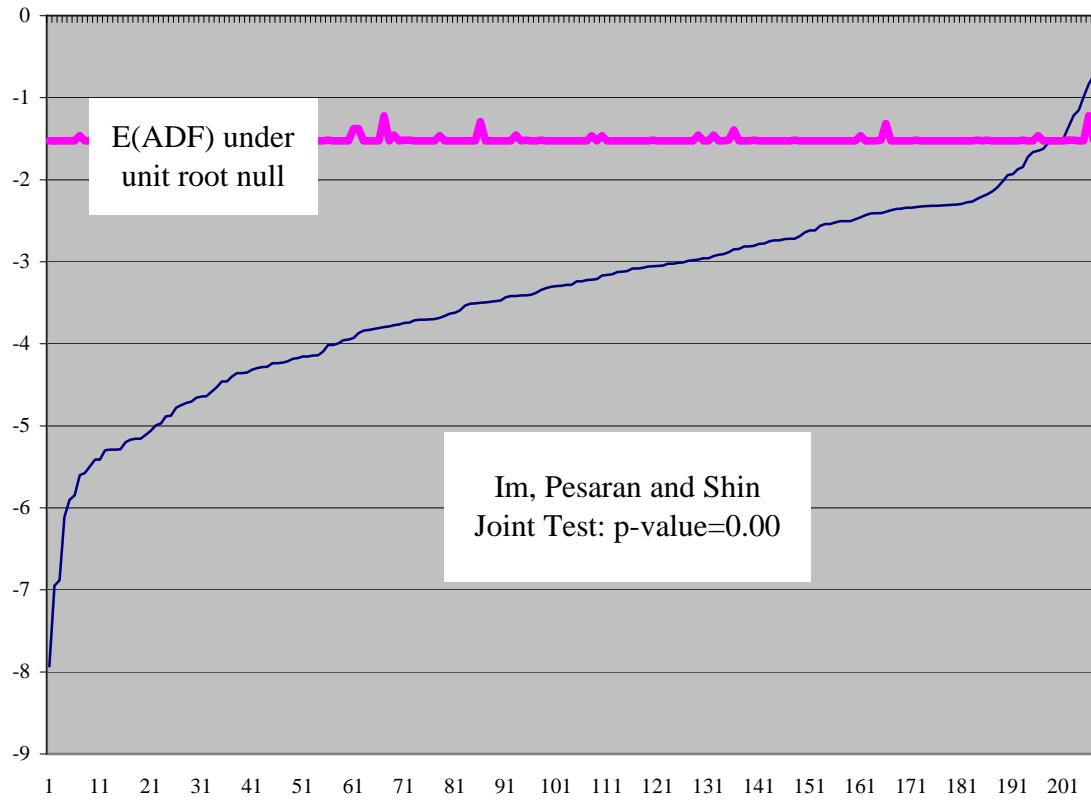


Figure A2

Ranked ADF Statistics for Estimated Transitory Components from Estimation in Differences

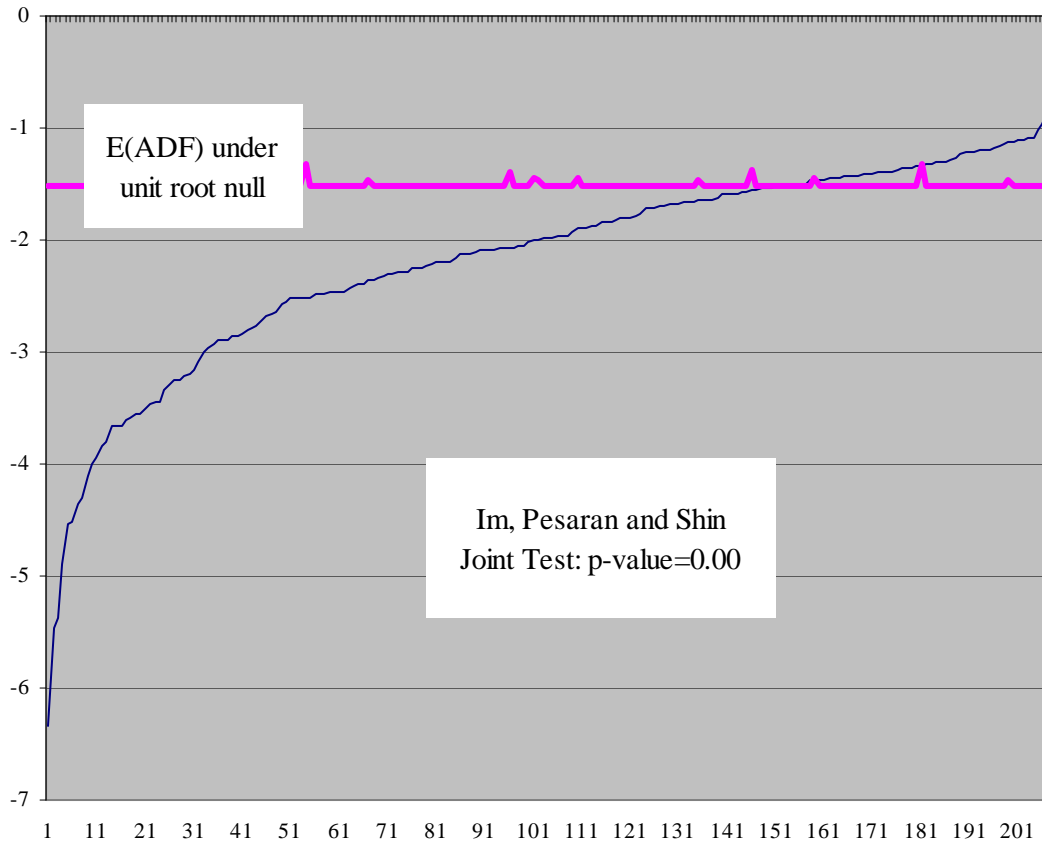


Figure A3

Alternative V-Factor Estimates

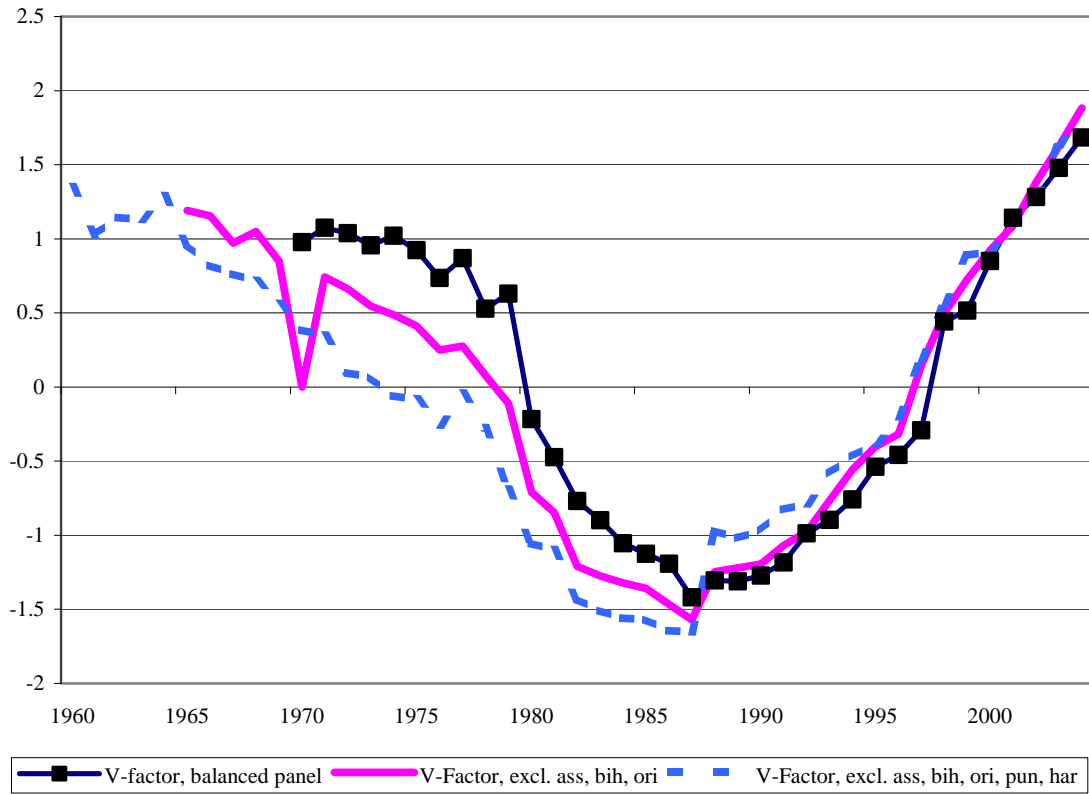


Figure A4

Alternative V-Factor Estimates (Estimation in Differences)

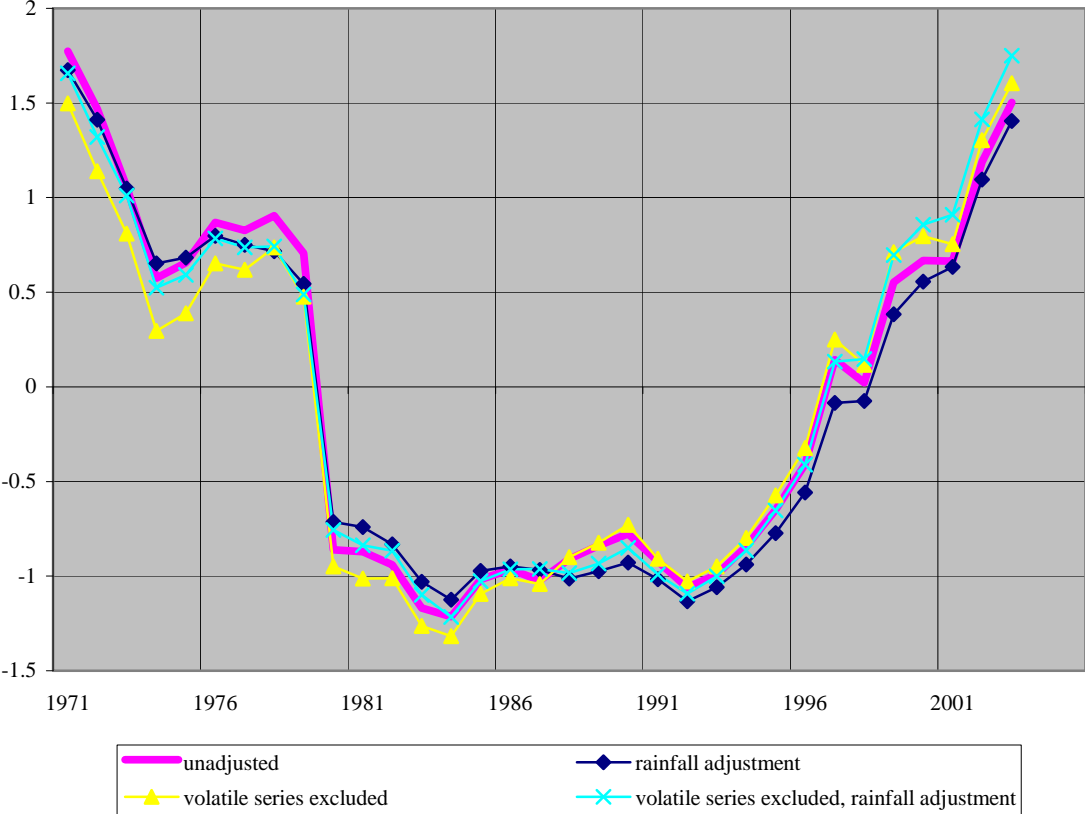




Figure A5

Average Statewise V Factor Loadings

