Supplementary Appendix for "The Specialization Curse"

Contents

A Additional Information for US State Data

A .1 Descriptive Statistics

Table 1 provides the data sources for the US state regressions. Table 2 shows the descriptive statistics for all of the variables used in the US state regressions. Table 3 provides the mean levels of our main variables by state: economic specialization, unemployment, population density, union density, and income per capita.

Table 1: US State Data Sources

Variable	Source
Education (% of Budget) Income per Capita Unemployment Rate Population Density Union Density Democratic Vote Share	Goldberg, Wibbels, and Mvukiyehe (2008) US State Database
% of Population under 19 % of Population over 65	U.S. Census Bureau: Population Division
% of Population Non-White	Surveillance, Epidemiology, and End Results (SEER) Program
Economic Specialization	U.S. Bureau of Economic Analysis: Regional Data

Table 2: US State Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Education (% of Budget)	36.7	7.1	10.1	55.5	1250
Income per Capita	22031	5114	9009	39973	1750
Unemployment Rate	6.372	2.054	2.3	17.4	1100
Population Density	160	226	0.53	1071	1200
Union Density	19.682	8.611	3.3	44.8	1700
Democratic Vote Share	51.482	11.32	20.216	100	467
% of Population under 19	32.1	4	21.9	44.3	1428
% of Population over 65	11.4	2.4	2.1	18.9	1428
% of Population Non-White	14.45	14.35	0.32	75.96	1479
Economic Specialization	0.165	0.068	0.095	0.703	1785

Table 3: Descriptive Statistics for US States Data

State	Avg. Economic Specialization	Unemployment	Pop Density	Union Density	Income Per Capita
Alabama	0.12	7.51	78.78	18.10	17998.63
Alaska	0.37	8.75	0.88	29.24	29869.89
Arizona	0.16	6.55	28.59	12.10	20746.48
Arkansas	0.11	7.10	44.84	12.09	17200.42
California	0.13	7.54	170.70	24.03	26202.47
Colorado	0.13	5.69	30.59	14.76	23605.25
Connecticut	0.17	5.45	655.28	22.86	29525.19
Delaware	0.16	5.60	322.15	21.61	25410.43
Florida	0.16	6.70	211.43	10.96	22467.40
Georgia	0.11	5.92	104.63	10.53	20107.30
Hawaii	0.30	5.05	162.36	29.00	25825.23
Idaho	0.15	6.45	12.05	16.73	19332.65
Illinois	0.13	7.31	207.00	26.99	25367.91
Indiana	0.16	6.49	153.62	28.74	21513.11
Iowa	0.14	5.19	51.10	20.71	21621.55
Kansas	0.12	4.60	29.83	14.81	22170.92
Kentucky	0.11	7.23	92.90	19.49	18378.11
Louisiana	0.23	8.27	94.47	13.12	18524.31
Maine	0.12	6.29	37.80	19.10	19592.53
Maryland	0.16	5.61	457.99	20.40	26047.12
Massachusetts	0.15	6.00	750.66	21.96	25912.42
Michigan	0.21	8.87	163.18	33.27	23600.40
Minnesota	0.13	5.18	53.54	26.07	23408.00
Mississippi	0.12	8.43	54.27	11.48	15693.72
Missouri	0.13	6.15	73.19	20.58	21585.45
Montana	0.15	6.44	5.55	24.50	19719.56
Nebraska	0.16	3.54	20.77	16.16	21661.92
Nevada	0.22	6.52	9.46	24.81	25571.59
New Hampshire	0.16	4.68	111.60	15.54	23358.40
New Jersey	0.13	6.69	1011.06	29.08	27756.49
New Mexico	0.21	7.65	11.81	12.19	18315.40
New York	0.13	7.09	374.34	31.87	26716.01
North Carolina	0.11	5.52	129.55	7.00	19540.46
North Dakota	0.26	4.53	9.42	14.58	19918.86
Ohio	0.16	7.28	265.40	28.34	22751.46
Oklahoma	0.14	5.72	45.48	12.79	19858.23
Oregon	0.17	7.34	28.48	26.40	22216.72
Pennsylvania	0.14	7.18	265.22	28.97	23064.60
Rhode Island	0.16	6.38	928.37	22.63	22910.78
South Carolina	0.13	6.33	110.04	6.09	17953.48
South Dakota	0.24	4.00	9.24	12.57	19314.87
Tennessee	0.11	6.96	115.40	16.91	19251.41
Texas	0.13	6.47	60.80	9.99	21151.99
Utah	0.14	5.29	2.46	16.62	18786.16
Vermont	0.16	5.32	57.85	14.79	20430.34
Virginia	0.11	5.11	146.20	12.06	23035.27
Washington	0.18	7.49	68.02	31.43	23978.29
West Virginia	0.15	9.99	77.82	27.41	17652.82
Wisconsin	0.14	5.70	88.54	26.34	22262.79
Wyoming	0.26	5.47	4.76	16.36	22661.06

A .2 Robustness - Mixed Effects Modeling

We hypothesize that economically specialized regions will provide more subsidies to core economic sectors in lieu of spending on broad-reaching public goods like education. Since we are unable to directly observe a budgetary trade-off between education spending versus industrial subsidies, we start by looking at the association between the diversity of the economic core and education spending across US states. However, rather than estimating the association between economic specialization and total education expenditures, we focus on the share of the government's budget committed to education. This allows us to capture a government's effort towards education relative to other important areas of the budget. An increase in the percentage of the budget devoted to education means that the government is favoring education over other areas that they could be spending on. The dependent variable (Y_{tj}) is the share of the state government's budget spent on education in a given year.

The slow changing nature of economic specialization presents a problem when using cross-sectional time-series data. The typical use of fixed effects to account for time-invariant heterogeneity across states will absorb most of the effects of economic specialization across polities. To get around these obstacles, in line with **?**, we employ a mixed effects approach to estimate the association between economic specialization and education spending both across and within states. This controls for slow changing factors, such as economic specialization, and captures the association between economic specialization and spending within a state across time. To accomplish this, each variable is separated into an across-unit and within-unit component, and used in an OLS regression with random effects at the state-level. The across-unit aspect, or the "fixed effect" aspect, is estimated by taking the mean of each variable. The within-unit aspect of variables (X_{tj}^{W}) is calculated by the following equation:

$$X_{tj}^{W} = X_{tj} - \bar{X}_j$$

Where $\bar{X}_j$ represents the mean level of specialization of state j over t years and X_{tj} represents the level of specialization at year t in state j. The full model can be written as:

$$Y_{tj} = \alpha_j + \beta X_{tj}^{W} + \gamma \bar{X}_j + \zeta \mathbf{Z}_{tj}^{W} + \lambda \bar{\mathbf{Z}}_j + u_j + e_{tj}$$

where X_j represents the level of economic specialization, $\mathbf{Z}_j$ represents a series of control variables, and α_j are random intercepts for each state j. They are modeled as draws from a common normal distribution with an overall mean of μ_α and state-level covariates $\mathbf{c}_j$: $\alpha_j \sim \mathcal{N}(\mu_\alpha + \mathbf{c}_j'\theta, \sigma_j^2)$. The random intercept approach allows us to account for state-level heterogeneity, but still pool information across states (**?**). The level-2 (across state) error is captured by u_j while the level-1 (within state) error is captured by e_{tj}. Separating the variables into their within and between components allows for the correct interpretation between the level of economic specialization of a state and the government's spending on public goods. The mean of economic specialization ($\bar{X}_j$) captures the association between economic specialization and education spending effort across different states, while the within effect of specialization (X_{tj}^W) captures the association between economic specialization and education spending effort within a state over time. In order to alleviate concerns that the mean of economic specialization ($\bar{X}_j$) is simply a proxy for all time-invariant heterogeneity across countries, all control variables include a time-invariant component as well ($\bar{\mathbf{Z}}_j$).

A .3 Robustness - State-Level Results

Table 4 shows the results from our mixed effects models. We find that both across state and within state economic specialization decreases the relative amount of money spent on education by a state. However, the across state effects are much larger in real terms when considering their substantive importance. A move from the 25th percentile to the 75th percentile of economic specialization is roughly associated with a change in education spending from 36% to 29% of the state government's budget. Table 4 also shows that a change in economic specialization *within* a state is negatively associated with education spending. Although the substantive effect is smaller than the across-state effect, the relationship is statistically significant at a level of $p<0.05$ in the reduced-form model.

Table 4: US States Public Education Spending & Economic Specialization – Mixed Effects Model with Random Effects and Robust Clustered Standard Errors

	Dependent Variable:		
	Education Expenditures (as % of Budget)		
	(1) OLS with RE	(2) OLS with RE	(3) OLS with RE
Economic Specialization – Across States	-0.442***	-0.560***	-0.571***
	(0.101)	(0.167)	(0.165)
Economic Specialization – Within States	-0.240**	-0.175*	-0.321***
	(0.0751)	(0.0715)	(0.0895)
% of Population under 19 – Across States		0.270	0.612
		(0.630)	(0.608)
% of Population under 19 – Within States		0.298	0.193
		(0.320)	(0.446)
% of Population over 65 – Across States		-0.261	0.167
		(0.655)	(0.700)
% of Population over 65 – Within States		-0.551	-0.577
		(0.491)	(0.738)
% of Population Non-White – Across States		0.0889	0.0864
		(0.0541)	(0.0535)
% of Population Non-White – Within States		0.228	0.232
		(0.276)	(0.345)
Income Per Capita – Across States		-0.00000705+	-0.000000466
		(0.00000410)	(0.00000546)
Income Per Capita – Within States		0.00000206	0.000000656
		(0.00000146)	(0.00000222)
Unemployment Rate – Across States		-0.0000352	0.00399
		(0.00593)	(0.00708)
Unemployment Rate – Within States		0.00227	0.00252
		(0.00178)	(0.00177)
Population Density – Across States		-0.000102**	-0.000131**
		(0.0000380)	(0.0000444)
Population Density – Within States		-0.000408+	-0.000506+
		(0.000213)	(0.000264)
Core GDP – Across States		0.000000197	0.000000138
		(0.000000210)	(0.000000210)
Core GDP – Within States		-0.000000153	-0.000000170
		(0.000000130)	(0.000000181)
Share of Democratic Vote – Across States			0.00241
			(0.00148)
Share of Democratic Vote – Within States			-0.000289
			(0.000212)
Unionization Rates – Across States			-0.000666
			(0.00159)
Unionization Rates – Within States			-0.0000564
			(0.00119)
Constant	0.408***	0.521	0.130
	(0.0200)	(0.322)	(0.370)
Year Fixed Effects	Yes	Yes	Yes
Observations	1250	950	274
# of States	50	50	50
Within R^2	0.318	0.299	0.419
Between R^2	0.157	0.548	0.610
Overall R^2	0.207	0.505	0.557
σ_u	0.0555	0.0425	0.0447
σ_e	0.0337	0.0253	0.0259

Robust standard errors clustered at the state in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A .4 Robustness - US County-Level Results

One of the main reasons we selected education as our public good of interest is because of data availability. Over any other type of public good, the US states provide the longest and most complete time-series of data on education. Education data are also readily available for the Indian cases since 1990. Focusing on education allowed us to make comparisons across the two countries. However, the scope of our argument is not limited to only education spending, nor is it limited to the state-level. In this section, we outline results from the county-level in the US on two other types of public goods. Similar to our state-level analysis, we employ a mixed effects model to capture the relationship between economic specialization and public goods expenditures. We employ the same method to calculate county-level economic specialization, using a Herfindahl-Hirschman Index of the core economic activity from the BEA. For our dependent variables, we obtained county-level health expenditures and police expenditures from the Local Government Census (conducted by the US Census and collected every five years). Similar to our state-level analysis, we look at expenditures as a percentage of the total county budget.

Table 6 displays our results. Here, we see that economic specialization across counties is negatively, and statistically significantly, associated with the percentage of the county budget spent on health and police. Figure's 1(a) & 1(b) show the post-estimated results of the change in county-level economic specialization. There is a negative relationship between economic specialization and the levels of public goods spending. To get a sense of the magnitude of the impact, the predicted change from a movement from the 25th percentile of economic specialization to the 75th percentile is roughly associated with a 12% change in health expenditures and a 5% change in police expenditures. Table 5 shows the descriptive statistics of the variables used in the analysis.

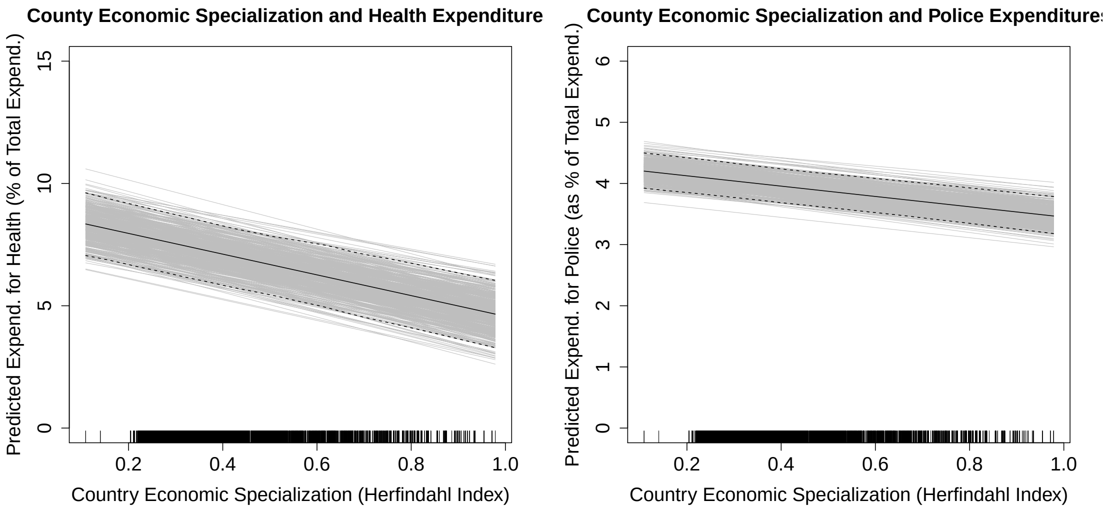
Table 5: US County Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
% of Budget on Health	0.084	0.105	0	0.754	19104
Economic Specialization	0.391	0.165	0	1	15624
% GOP Vote Share	50.673	13.235	1.133	92.882	18599
% of County Black	8.109	0.451	7.697	8.679	15916
Unemployment %	7.296	1.612	5.146	9.477	15916
Normalized HS Graduates Score	-0.003	0.998	-4.504	3.109	18599
Total County GDP	2590009.182	18179113.177	541	879160128	15608
County Population	154249.641	931378.918	78	32486010	15608

Table 6: Economic Specialization and US County Public Goods Spending – Mixed Effects Model with Random Effects and Robust Clustered Standard Errors

	(1) Health Expenditures	(2) Police Expenditures
Economic Specialization – Across Counties	-4.163e-02***	-8.381e-03***
	(1.182e-02)	(1.874e-03)
Economic Specialization – Within County	-1.340e-03	2.155e-04
	(4.742e-03)	(1.161e-03)
GOP Vote – Across Counties	4.749e-04*	-1.699e-04***
	(1.971e-04)	(2.840e-05)
GOP Vote – Within County	-3.548e-04**	4.647e-05*
	(1.212e-04)	(1.924e-05)
% Population Black – Across Counties	3.981e-01	1.687e-01
	(5.897e-01)	(1.452e-01)
% Population Black – Within County	1.146e-02	1.987e-03
	(7.824e-03)	(1.466e-03)
% Population Unemployed – Across Counties	3.552e-01	1.041e-01
	(4.447e-01)	(1.128e-01)
% Population Unemployed – Within County	1.423e-03	-6.922e-04
	(2.311e-03)	(4.409e-04)
High School Graduates – Across Counties	2.890e-03	5.733e-03***
	(2.602e-03)	(3.993e-04)
High School Graduates – Within County	-6.833e-03*	3.908e-05
	(2.986e-03)	(5.644e-04)
County Income – Across Counties	-1.886e-09	-8.157e-10
	(1.479e-09)	(5.533e-10)
County Income – Within County	2.430e-10	2.049e-11
	(1.840e-10)	(6.428e-11)
County Population – Across Counties	2.339e-08	2.461e-08*
	(2.549e-08)	(1.029e-08)
County Population – Within County	-2.495e-08	1.281e-09
	(1.554e-08)	(4.506e-09)
Constant	-5.514e+00	-2.029e+00
	(7.777e+00)	(1.944e+00)
Observations	15216	15216

Robust standard errors clustered at the county-level in parentheses
Year and State Fixed Effects
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



(a) US County – Health Expenditures

(b) US County – Police Expenditures

Figure 1: Economic Specialization & Public Goods Expenditures at the County-Level

B Economic Specialization & Education Expenditures: Cross-National Evidence

We find support for a negative relationship between economic specialization and education spending at the national level. We demonstrate this by compiling a dataset of 145 countries over the time period 1970-2000. For our measure of economic specialization, we calculate a Herfindahl-Hirschman Index index using industrial data (4-digit level) from the World Trade Flows Database. Data for control variables are from the World Bank Development Indicators, Quality of Government Dataset, and Penn World Tables. Table 7 shows the descriptive statistics for all observations that had a non-missing value in our dependent variable: education expenditures.

Table 7: Economic Specialization Analysis Around the World: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Log Education Expenditures	3.28	0.36	0.52	4.46	2818
Economic Specialization	0.14	0.19	0.005	0.96	1638
GDP Per Capita	11989	12303	241	109883	2785
% of Population Youth (under 18)	31.16	10.7	11.51	50.04	2776
% of Population Rural	43.96	23.63	0	95.89	2818
Population (in thousands)	35559	120702	40	1252735	2603

Table 8 displays our results using a mixed effects model with random effects and robust clustered standard errors at the country-level. Consistent with our hypothesis and the findings discussed in the paper, we confirm that economic specialization is a negative and statistically significant predictor of national education spending at a level of $p<0.05$ or better. This holds after controlling for variables such as a country's GDP per capita, total population, the youth share of the population, and the rural share of the population.[1] Figure 2 displays post-estimated results. A move from the 25th percentile to the 75th percentile of economic specialization is roughly associated with a drop in education spending from 27% to 24% of the national budget. This provides additional support for our hypothesis and demonstrates that the link between economic specialization and public goods provision exists at national and subnational levels.

[1]Including other control variables often results in list-wise deletion of a large number of countries due to missing data.

Table 8: Country-Level Economic Specialization and Public Goods Spending – Mixed Effects Model with Random Effects and Robust Clustered Standard Errors

	(1) OLS with RE	(2) OLS with RE	(3) OLS with RE	(4) OLS with RE
Economic Specialization – Across Countries	-0.362*	-0.381*	-0.461*	-0.592**
	(0.168)	(0.175)	(0.214)	(0.227)
Economic Specialization – Within Countries	-0.0554	-0.0879	0.0320	0.0976
	(0.161)	(0.174)	(0.185)	(0.142)
GDP Per Capita – Across Countries		-0.00000114	0.000000549	0.00000276
		(0.00000285)	(0.00000378)	(0.00000418)
GDP Per Capita – Within Countries		0.00000225	-0.00000394	-0.00000596
		(0.00000350)	(0.00000498)	(0.00000438)
% of Youth Population – Across Countries			0.00414	0.00416
			(0.00431)	(0.00459)
% of Youth Population – Within Countries			-0.0107	-0.00620
			(0.00669)	(0.00715)
% of Population Rural – Across Countries				0.00186
				(0.00172)
% of Population Rural – Within Countries				-0.0109*
				(0.00485)
Total Population – Across Countries				-1.98e-10
				(2.53e-10)
Total Population – Within Countries				-1.42e-09+
				(7.29e-10)
Constant	3.286***	3.299***	3.153***	3.045***
	(0.0341)	(0.0514)	(0.153)	(0.163)
Observations	1638	1617	1596	1576
# of Countries	145	145	142	140
Within R^2	0.000923	0.00240	0.0145	0.0524
Between R^2	0.0366	0.0376	0.0348	0.0299
Overall R^2	0.0317	0.0288	0.0436	0.0629
σ_u	0.302	0.304	0.310	0.309
σ_e	0.216	0.216	0.214	0.210

Robust standard errors clustered at the country-level in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

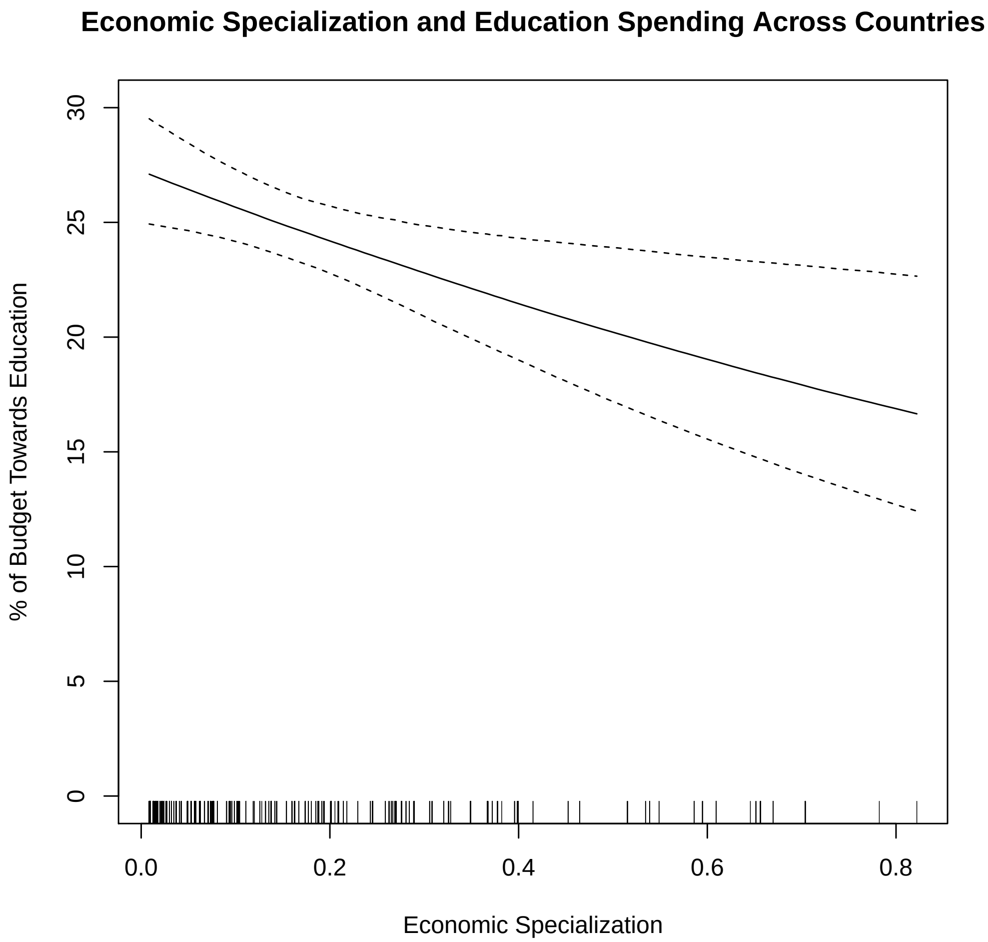


Figure 2: Economic Specialization and Education Spending Across Countries

C Additional Information for the India Analysis

C .1 Map of India's Post-bifurcation States (as of 2001)

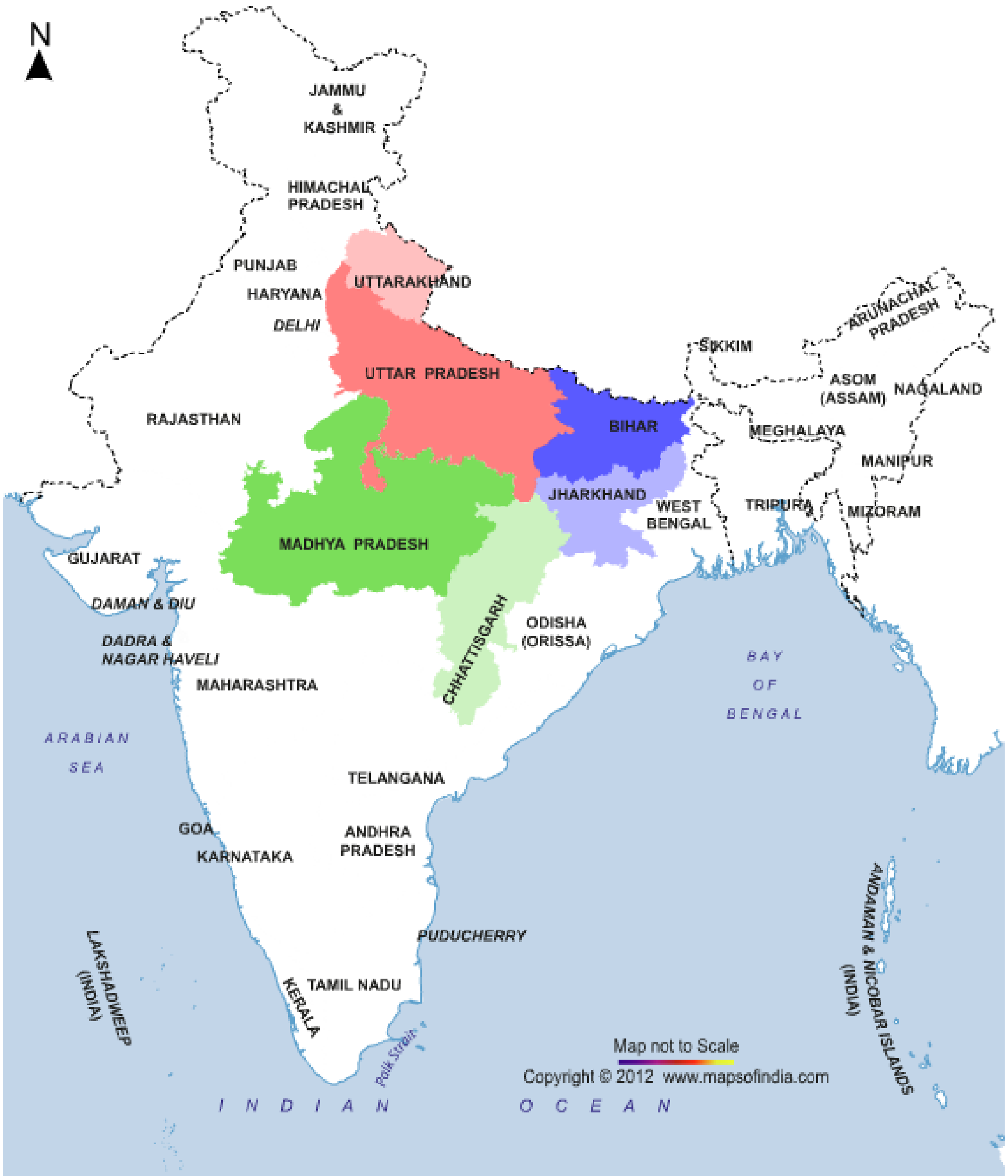


Figure 3: India's Post-bifurcation States (as of 2001)

C .2 Data Sources and Additional Notes for India Analysis

- Data on gross state domestic product (factor cost by industry of origin) are from the Indian Ministry of Statistics and Programme Implementation (Central Statistical Organisation) and Indiastat.com.
- In the public finance data, education spending includes education, sports, art and culture. Data are available from Reserve Bank of India (http://www.rbi.org.in) and Indiastat.com.
- For the breakpoint analysis conducted in this section, we use all available state education spending data (total of 27 Indian states) to find where the most likely break in time-series data occurs (along with a confidence interval). We do not restrict the algorithm on how many breaks to find. When multiple break points are identified, as is common with time-series data, we take the breakpoint with the highest F-score value.
- Urban population and Scheduled Caste data are taken from 2001 India Census.
- Education spending data referenced in discussion of Uttarakhand are available from Indiastat.com "State-wise Eleventh Finance Commission (EFC) Grants for Upgradation of Elementary Education in India."
- Agriculture spending data are available from the Reserve Bank of India (http://www.rbi.org.in).
- Electricity subsidy data are available from Indiastat.com "Selected State-wise Uncovered Subsidy on Electricity for Various Categories of Consumers in India" (multiple years).
- NAIS data are available from Indiastat.com "Selected State/Season-wise State Government Share in Premium Subsidy under National Agriculture Insurance Scheme (NAIS) in India" (various years). Note that state governments provide other types of agricultural subsidies but the NAIS data are the only annual time-series subsidy data available for all six states. Other agricultural subsidy data are restricted to two or three years of annual data, which prevents us from viewing trends in states over time.
- Power subsidies data are available from Indiastat.com "Region/State-wise Subsidies Received by Power Utilities and Distribution Companies in India (2003-04 to 2005-06)."
- Investment data are available from Indiastat.com "State-wise Investment Intentions (IEMs + LOIs + DILs)" (various years).

C .3 Measuring State-Level Economic Specialization in India

In order to estimate the ratio of agricultural output to industrial output and changes in economic specialization at the state-level over time, we rely on a few sources of data. For the manufacturing component (factories), we use data on the total value of output (Rs) as recorded in the *Annual Survey of Industries* (multiple years), which are compiled by the Central Statistics Office (Government of India). For the agricultural component, we use state-level estimates of agricultural (and allied activities) total value of output (Rs), which are compiled by the Central Statistics Office (Government of India). For the mining production output component, we use value of output data (Rs) from Indiastat.com "State-wise Value of Mineral Production by Mineral Groups in India" (multiple years). This allows us to look at total output in core sectors that were examined in the United States analysis. This encompasses the following sectors: Agriculture/Farm, Livestock, Forestry, Fisheries, Mining/Quarrying; and Manufacturing of the following: Food Products and Beverages, Tobacco Products, Textiles, Tanning and Leather Products, Wood Products, Furniture, Paper and Paper Products, Publishing and Printing, Coke/Refined Petroleum/Fuel, Chemicals, Rubber and Plastics, Non-metallic Minerals, Basic Metals, Fabricated Metals, Machinery and Equipment, Electrical Machinery, Motor Vehicles, Transport Equipment, Computers/Electronics, Pharmaceuticals/Medicine.

A few caveats apply when using these data to make inferences about patterns in economic specialization across Indian states. First, note that the Indian government altered the sampling design used to estimate the manufacturing/factory component more than once during the 2000s. Second, some sectoral output data contains service sector activities, which we defined to be "non-core" and were not included in the US state-level measures of specialization. This will inflate measures of total output value in some areas. Finally, the data are unable to account for unregistered firms or many micro-level enterprises, which comprise a sizeable portion of economic activities across sectors and states. Estimating this contribution would be difficult, if not impossible. These issues notwithstanding, we believe that an examination of the above data is the only way to look at changes in economic specialization across Indian states in a manner that is consistent with the approach used in our US states analysis.

References

Bartels, Brandon. 2008. "Beyond 'Fixed Versus Random Effects': A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data." *The Society for Political Methodology Working Paper*: http://home.gwu.edu/~bartels/cluster.pdf.

Gelman, Andrew, and Jennifer Hill. 2008. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.