Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Supplementary Materials

Appendix to

Technological Catch-up to the National and Regional Frontier: Firm-level Evidence for India

Shubin Yang, Sandra Lancheros, Chris Milner

Table A1. Definition of variables and summary statistics

Variable	Definition	Mean and Standard deviation		
		1999-2010	1999-2006	2007-2010
lnTFP	Logarithm of total factor productivity (TFP) calculated using the Levinsohn and Petrin (2003) approach.	1.75 (0.97)	1.76 (0.97)	1.73 (0.98)
TFP-National Frontier	TFP of the firm with the highest TFP in the country for each industry at a given time period.	4.67 (0.78)	4.56 (0.81)	4.88 (0.66)
TFP-Regional Frontier	TFP of the firm with the highest TFP in each state and each industry at a given time period.	3.08 (1.01)	3.04 (1.00)	3.16 (1.03)
TFP Growth	TFP growth rate	-0.01 (0.60)	0.01 (0.58)	-0.03 (0.61)
National Gap	Productivity distance to the National Frontier.	3.11 (1.12)	3.02 (1.13)	3.27 (1.07)
Regional Gap	Productivity distance to the Regional Frontier.	2.11 (1.17)	2.07 (1.18)	2.16 (1.16)
Age	Logarithm of firm's age.	2.94 (0.71)	2.91 (0.72)	3.00 (0.70)
Size	Logarithm of total asset.	5.26 (1.67)	5.20 (1.62)	5.36 (1.75)
Continuous exporters	Dummy variable equal to 1 if the firm exported every year during the sample period (or if it started to export and remained exporting during the period of analysis), zero otherwise.	0.30 (0.46)	0.32 (0.47)	0.28 (0.45)
Non-continuous exporters	Dummy variable equal to 1 if the firm exported in some years of the period of analysis, but not in other years; zero otherwise.	0.37 (0.48)	0.37 (0.48)	0.36 (0.48)
Non-exporters	Dummy variable if the firm remained domestic during the period of analysis, zero otherwise.	0.33 (0.47)	0.31 (0.46)	0.37 (0.48)
OFDI firms	Dummy variable equal to one if the firm exported and invested abroad during the sample period, zero otherwise.	0.17 (0.37)	0.17 (0.38)	0.16 (0.37)

Note: This table reports the average and standard deviations (in parentheses) of the main variables used in the empirical analysis. The sample includes 43,913 observations in the manufacturing sector over the period 1999-2010. The table also presents a summary statistics for the subperiods before (1999-2006) and after (2007-2010) the global financial crisis. Prowess reports data for the fiscal year running from April to March each year, hence we consider 1999-2006 as the pre-crisis period in order to fully eliminate the influence of the global financial crisis. All monetary variables used to calculate TFP, as well as firms' total assets were deflated using the wholesale price index based on fiscal year 1993-1994.
Source: Authors' calculations using the dataset.

Table A2. Mean and standard deviation of the main variables by type of firm

Before and after the global financial crisis

	1999-2006				2007-2010			
	TFP	TFP Growth	National Gap	Regional Gap	TFP	TFP Growth	National Gap	Regional Gap
National Frontiers	4.56 (0.82)	0.36 (0.76)			4.88 (0.66)	0.35 (0.54)		
Regional Frontiers	3.04 (1.00)	0.18 (0.52)			3.16 (1.03)	0.17 (0.59)		
Lagging Firms	1.71 (0.93)	-0.00 (0.59)			1.68 (0.94)	-0.04 (0.61)		
Continuous exporters	1.90 (0.82)	0.01 (0.46)	2.87 (1.01)	1.96 (1.07)	1.79 (0.85)	-0.04 (0.49)	3.16 (0.96)	2.10 (1.10)
Non-continuous exporters	1.78 (0.95)	0.00 (0.60)	2.98 (1.11)	2.06 (1.16)	1.74 (0.99)	-0.03 (0.64)	3.25 (1.08)	2.18 (1.15)
Non-exporters	1.59 (1.11)	0.02 (0.69)	3.22 (1.25)	2.21 (1.28)	1.67 (1.05)	-0.01 (0.67)	3.37 (1.13)	2.20 (1.22)
Continuous exporters-OFDI	2.04 (0.80)	0.00 (0.42)	2.78 (0.97)	1.87 (1.04)	1.89 (0.87)	-0.05 (0.49)	3.08 (0.98)	2.03 (1.12)
Non-continuous exporters-OFDI	1.96 (0.97)	0.01 (0.56)	2.75 (1.11)	1.87 (1.14)	1.85 (1.00)	-0.10 (0.61)	3.11 (1.09)	2.08 (1.16)

Note: This table reports the average and standard deviations (in parentheses) of the main variables of interest. The sample includes 43,913 observations in manufacturing over the period 1999-2006 and 2007-2010, respectively. Prowess reports data for the fiscal year running from April to March each year, hence we consider 1999-2006 as the pre-crisis period in order to fully eliminate the influence of the global financial crisis.

Source: Authors' calculations using the dataset.

Table A3. Firms' productivity convergence: baseline results

	Convergence to the national frontier			Convergence to the regional frontier			Convergence to both frontiers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	GMM	FE	OLS	GMM	FE	OLS	GMM	FE
National Gap (t-1)	0.21***	0.25***	0.41***				0.15***	0.19***	0.25***
	(0.01)	(0.07)	(0.01)				(0.01)	(0.05)	(0.01)
Regional Gap (t-1)				0.18***	0.26***	0.39***	0.08***	0.10***	0.23***
				(0.01)	(0.06)	(0.01)	(0.01)	(0.03)	(0.01)
Constant	yes	yes	yes	yes	yes	yes	yes	yes	yes
Firm's characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes		yes	yes		yes	yes	
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Industry dummies	yes	yes		yes	yes		yes	yes	
Autocorrelation of ε_{ijt} (p-value)		0.11			0.35			0.26	
Hansen test (p-value)		0.92			0.69			0.84	
Observations		33,477			33,477			33477	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are displayed in parentheses. The dependent variable is the annual TFP growth rate. Firm' characteristics include the size and age of the firm. All regressions are estimated on non-frontier Indian manufacturing firms over the period of 1999-2010, using OLS, GMM, and fixed effect respectively. GMM estimations are performed using the two-step system GMM estimator with firm-clustered and Windmeijer (2005)-corrected standard errors (in parentheses). First differences of the potential endogenous productivity gap variables (dated t-2 and longer) are used as instruments in the level equation, and level values of these variables (dated t-2 and longer) are used as instruments in the differenced equation. Following Roodman (2009)'s, we collapse these instruments to avoid instrument proliferation, which might overfit our endogenous regressors, failing to remove their endogenous component Roodman (2009, p.128). This table shows the untransformed sample size after FE and GMM, as reported by the corresponding STATA commands.

Table A4. Firms' productivity convergence across groups of regions

	Gujarat		Maharashtra		Rest of states	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	FE	OLS	FE	OLS	FE
National Gap (t-1)	0.07***	0.20***	0.05***	0.16***	0.13***	0.29***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Regional Gap (t-1)	0.17***	0.29***	0.16***	0.31***	0.07***	0.22***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
National Gap (t-1) * common	0.00	-0.03	0.01	-0.03	0.05***	-0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
common	0.05	0.11	-0.06	0.04	-0.15***	0.03
	(0.10)	(0.09)	(0.05)	(0.05)	(0.04)	(0.04)
Constant	yes	yes	yes	yes	yes	yes
Firm' characteristics	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes
Industry dummies	yes		yes		yes	
State dummies					yes	
Observations	3,668	3,668	8,945	8,945	20,864	20,864

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are displayed in parentheses. The dependent variable is the annual TFP growth rate. Firm' characteristics include the size and age of the firm. All regressions are estimated on non-frontier Indian manufacturing firms over the period of 1999-2010, using OLS and FE respectively. The common frontier is defined as a dummy variable that takes the value of 1 if the national frontier is the same regional frontier, and zero otherwise.

Table A5. Firms' productivity convergence: robustness tests

	Controlling for selection		Top 5 national frontiers		Excluding new firms		Imputing employment		Period 1999-2006	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
National Gap (t-1)	0.15***	0.25***	0.21***	0.47***	0.14***	0.24***	0.15***	0.30***	0.16***	0.33***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Regional Gap (t-1)	0.08***	0.23***	0.04***	0.10***	0.07***	0.22***	0.07***	0.26***	0.07***	0.24***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Size	-0.01***	-0.02*	0.03***	0.01	0.02***	-0.003	0.02***	-0.08***	0.03***	0.05**
	(0.003)	(0.01)	(0.002)	(0.01)	(0.002)	(0.01)	(0.002)	(0.01)	(0.003)	(0.02)
Age	-0.10***	-0.09**	-0.02***	-0.03	-0.00	0.03	0.00	0.06	-0.02***	-0.05
	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.05)	(0.01)	(0.05)	(0.01)	(0.07)
IMR	-0.97***	-0.90***								
	(0.06)	(0.07)								
Constant	0.12**	-0.49***	-0.75***	-1.32***	-0.67***	-1.15***	-0.94***	-1.10***	-0.77***	-1.51***
	(0.05)	(0.13)	(0.03)	(0.14)	(0.03)	(0.15)	(0.04)	(0.14)	(0.05)	(0.22)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Industry dummies	yes		yes		yes		yes		yes	
State dummies	yes		yes		yes		yes		yes	
Observations	33462	33462	33325	33325	31024	31024	33441	33441	19403	19403

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are displayed in parentheses. The dependent variable is the annual TFP growth rate. The national productivity frontier is defined as the firm with the highest TFP in the industry in a year, and the regional frontier is the firm with the highest TFP in an industry-year-state basis. All regressions are estimated on non-frontier Indian manufacturing firms using OLS and fixed effect.