Online Appendix

This online appendix provides supplemental material to the paper, *Banking and Growth: Evidence from a Regression Discontinuity Analysis.*

1. Additional Policy Details
2. Regression Discontinuity Validity and Robustness Checks
3. Additional Results
4. Additional Data Appendix
5. Theoretical Framework Appendix

1 Additional Policy Details

1.1 Branch closures and location shifting

Banks were very limited in their ability to move branches between markets. Branches were not allowed to shift outside otherwise unbanked centers. Given that a location was served by another commercial bank branch (other than an RRB), a branch could only shift to centers in the same or lower population group classification. Relocated branches could only move to markets of the same or lower population classification. In the case of branches in under banked districts, they could only shift to centers within under banked districts.

1.2 Policy and ABEP timing

Although the reform became effective immediately upon its release, banks were essentially allowed a year long grace period to construct their first ABEP, with an implicit deadline for September 2006. Several banks, many of them from the private sector, waited close to the full year to submit their ABEP, during which time they were able to receive licenses in a disaggregated manner. The histogram of branch license dates for a large private sector bank, shown in figure 1, demonstrates this pattern.[1] Although annual branch expansion plans may not be observed directly, the large spikes in branch licenses set approximately a year apart are consistent with ABEPs. The figure shows the licenses from the first likely ABEP for this bank were granted in July 2006, roughly one year after the reform implementation. Similar patterns are identified for many private sector banks. Banks could then delay their branch openings another year by exploiting the year long validity of licenses. Though either quick or delayed entry could be strategically optimal, the empirical evidence suggests most private sector banks chose to delay entry for nearly two years in many locations of induced entry.

1.2.1 LEAD Banks and SAA

While the reform became official in September 2005, events leading up to its release likely provided signals as to its impending introduction. In a speech from December 2002, the Deputy Governor of the RBI pointed to the high share of bank investments in government

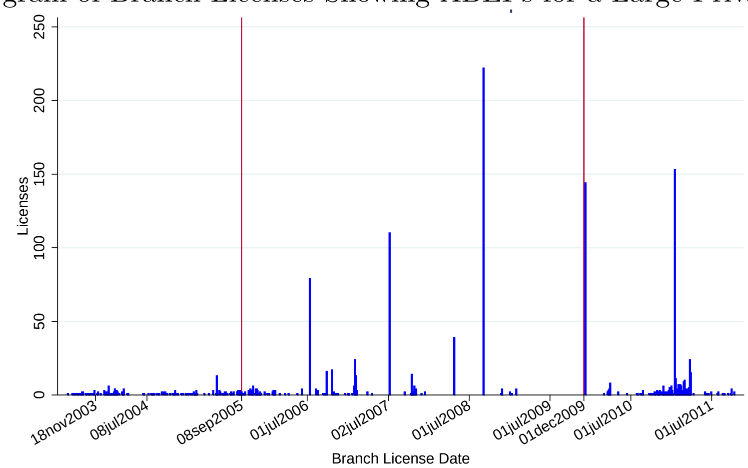
[1] Known acquisitions of branches from other banks have been excluded for the histogram analysis.

securities, 39% relative to the regulatory minimum of 25%, encouraging banks to expand their commercial lending particularly in small manufacturing and agriculture (Mohan, 2002). The following November, the Vyas Committee was commissioned to investigate the flow of capital to agricultural activities. They met with several commercial banks during their investigation. In April 2004, they released an interim report followed by the final report in June, suggesting revisions to the service area approach (SAA) and encouraging greater lending by private and public sector banks. The report included a rough map identifying areas underserved by the formal banking sector.[2]

The LEAD banking scheme was in operation prior to 2005, by which one bank was assigned to each development block and made responsible for meeting agreed levels of branching and banking services. These banks were typically selected from the set of government owned banks. The service area approach (SAA) meanwhile, partitioned rural areas between banks for the purpose of implementing development objectives.

The SAA program was subsequently discontinued, allowing all banks to freely apply for entry and operate in rural areas. The official list of under banked districts released in 2005 almost exactly matches selection based on district average population per branch relative to the national average, consistent with the language in the report. Thus, aspects of the Vyas Committee report could have provided solid signals to banks of the forthcoming reform. Figure 2 provides a timeline of policy events since 2003.

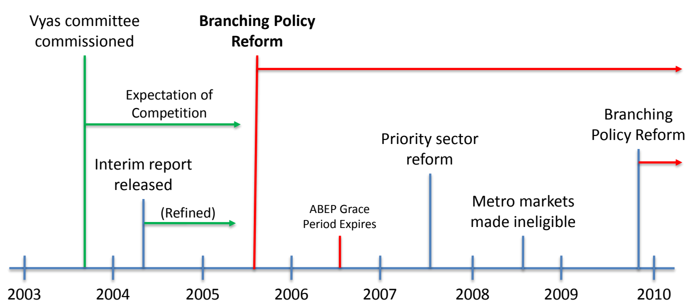
Figure 1: Histogram of Branch Licenses Showing ABEPs for a Large Private Sector Bank



Note: Branch license dates are from the MOF. Bin widths are set to 4 days. Though annual branch expansion plans (ABEP) may not be observed directly, the large spikes in branch licenses set approximately a year apart after 2005 are consistent with licenses issued through ABEPs. The dates of Master Circular releases are shown, with vertical red lines at the 2005 policy reform and the subsequent reform in December 2009. Branches acquired through mergers and acquisitions are excluded.

[2] Some areas identified on the map were described as places where the "branch network of commercial bank[s] [is] below the national average (Vyas Committee, 2004)."

Figure 2: Policy Time Line for Bank Branching and Related Reforms



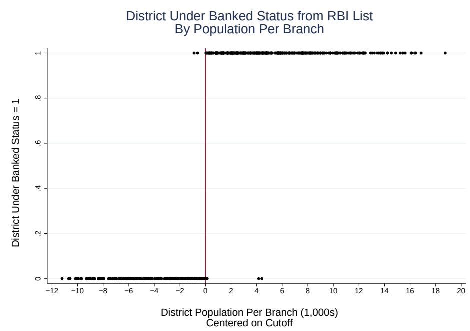
Note: The diagram identifies major events in the regulatory environment affecting the commercial banking sector in India for the time frame of analysis.

2 Regression Discontinuity Validity and Robustness Checks

2.1 Assignment to under banked status

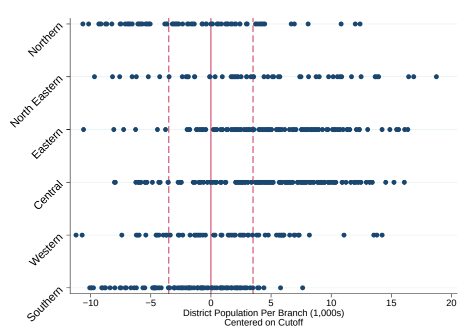
Figure 3 shows district under banked status from the 2006 list of under banked districts plotted against district population per branch around the national average. According to the rule, districts to the right of the cutoff should be assigned to under banked status, as is broadly confirmed in the figure. The even distribution of districts around the cutoff holds within regions as well, shown in figure 4. While the Central, Eastern and North Eastern regions of India are relatively less banked than the North, South and West, each region has districts falling near the cutoff on both sides.

Figure 3: Under Banked Status by District Population Per Branch



Note: The dots report the under banked status of a district, taking a value equal to one if the district appeared on the list of under banked districts in the 2006 RBI MC on Branching Authorisation Policy, and zero otherwise. The forcing variable, district population per branch centered on the national average, is on the x-axis scaled to thousands of persons per branch. Values to the right of the cutoff are predicted to have under banked status. 369 districts of 572 have under banked status, with 5 incorrect predictions based on the rule.

Figure 4: Geographic Distribution of District Population Per Branch across RBI Regions

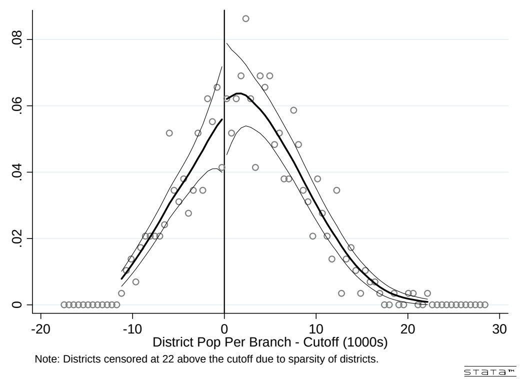


Note: Districts are reported within their RBI region, plotted horizontally by a dot according to their population per branch that is centered on the cutoff. A solid vertical line denotes the threshold, with dashed vertical lines indicating a bandwidth of 3,500 persons per branch, the same used throughout the analysis.

2.2 McCrary Test

The visual McCrary test shows a lack of manipulation of he running variable around the cutoff and highlights that the peak of the running variable density occurs near the cutoff. The effects we find are driven by districts falling close to the highest density and are not driven by the tails.

Figure 5: Visual McCrary Test



Note: The graph plots a density of districts along the forcing variable, district population per branch, centered on the cutoff. The discontinuity estimate in the log difference in height is 6.6 with a standard error of 22. I fail to reject the null hypothesis of continuity at the cutoff, suggesting a lack of manipulation.

2.3 Continuity tests for baseline values regression results

Table 1: Continuity tests for Baseline Values at the Cutoff

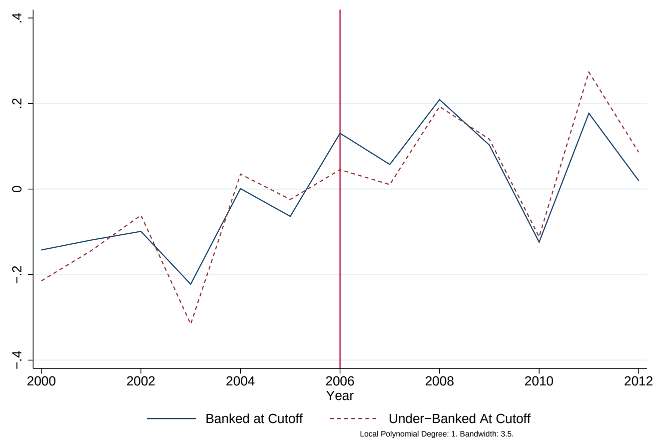
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Population	Pop Share Top 4 Centers	Sched Caste Tribe Pop	Pct Literate	Pct Pop Working	Share Workers in Agri	PrivBranches2000
Conventional	0.839	0.0135	-1.436	0.0114	-0.0114	0.0321	0.192
	[35.38]	[0.0400]	[8.483]	[0.0219]	[0.0197]	[0.0482]	[3.026]
Bias-corrected	16.01	0.0218	0.265	0.0187	-0.0129	0.0363	0.567
	[35.38]	[0.0400]	[8.483]	[0.0219]	[0.0197]	[0.0482]	[3.026]
Robust	16.01	0.0218	0.265	0.0187	-0.0129	0.0363	0.567
	[42.75]	[0.0461]	[9.840]	[0.0261]	[0.0235]	[0.0591]	[3.527]
Bandwidth	3.500	3.500	3.500	3.500	3.500	3.500	3.500
N_Banked	95	95	95	95	95	95	95
N_UBanked	122	122	122	122	122	122	122
DepMean	176.7	0.221	45.24	0.553	0.421	0.550	7.198

Note: Estimated using local linear regressions with no controls. Bandwidths are set 3.5 thousand persons per branch and estimated using a triangular kernel.

2.3.1 Rainfall

Figure 6 presents the estimated discontinuities in the percentage deviations of rainfall measures from their mean levels for districts.[3] As anticipated, deviations in rainfall do not show significant discontinuities at the cutoff. This suggests the response from credit and agricultural performance is not discontinuously effected by exogenous productivity shocks around the cutoff in the years considered.

Figure 6: Discontinuity from Reduced Form: Deviations in Monsoon Rainfall



Note: The dependent variable is the percentage deviation of annual monsoon rainfall from the time series average, 1999-2012 by district. The deviation is calculated for each location of measurement separately, and then averaged across locations in a district. Monsoon rainfall is considered total rainfall for June-September.District Average Percentage Deviation from Mean. Estimated using local linear regressions. Bandwidths are set 3.5 thousand persons per branch and estimated using a uniform kernel. The figure plots the estimated intercepts at the cutoff from the estimation of the RD equation repeated annually. The red dashed line provides the estimated intercept from approaching the threshold along the under banked side. The solid blue line reports the corresponding intercept approaching from the banked side. The distance between the two, reported for each year, shows the estimated discontinuity at the threshold. A solid line between the two points indicates an estimated discontinuity with statistical significance of at least the 10% level. The thin vertical line at 2006 represents the first estimation made following the reform.

[3] Specifically, for each geographic location where rainfall is measured, I construct a mean for the time series and the annual percentage deviation from that mean. Then I use the average of these measurements from locations within the district boundaries as the value for analysis. For agricultural purposes, rainfall occurring during the monsoon season is the most relevant. Therefore, I use annual measures of rain falling during June, July, August and September. The results are largely unchanged by using total annual rainfall instead.

2.4 Robustness test: NREGA Roll Out Around the Cutoff

To confound results, there must be a discontinuous break in the implementation of the NREGA program and disbursement of benefits at the "under banked" cutoff determined by the branching policy reform used in this study's regression discontinuity. I show that no discontinuity can be detected for any implementation phase. The implementation of NREGA occurred in three stages, with 200 districts selected to begin the program in the fiscal year April 2006 through March 2007, with 130 new districts introduced in 2007-8 and the remaining 263 districts introduced in 2008-9. Zimmermann (2012) and Klonner and Oldiges (2014) analyze the effect of NREGA using these rollout phases and provide background on the program. Of particular importance to the current analysis, NREGA benefits were distributed through bank accounts. NREGA incentives encourage an increase in the demand for formal banking and its geographic reach, but there is no evidence of differential effect at the cut-off.

Districts were assigned to the various roll-out phases based on a composite index on district "backwardness" from the National Planning Commission (2003).[4] As a robustness check, I test whether a discontinuity in phase assignment can be detected at the cutoff. A significant discontinuity would suggest a correlation with the NREGA program. The test fails to reject the null hypothesis of continuity at the cutoff for all three phases. Thus, NREGA phase assignment, and therefore its benefits, would be unexpected to differ at the cutoff.

Some districts were not assigned a backwardness index number. To check that the omitted districts are not disproportionately from one side of the cutoff or the other, I repeat the McCrary test only including districts missing the composite index value. I fail to reject the null hypothesis of continuity in the density of districts at the cutoff with the discontinuity estimate in the log difference in height at -31 and a standard error of 38.

Table 2: NREGA Discontinuity in District Phase Assignment

	(1)	(2)	(3)
VARIABLES	Phase_1	Phase_2	Phase_3
Conventional	-0.0648	0.0145	0.0503
	[0.119]	[0.0909]	[0.135]
Bias-Corrected	-0.121	0.0710	0.0497
	[0.119]	[0.0909]	[0.135]
Robust	-0.121	0.0710	0.0497
	[0.139]	[0.109]	[0.160]
Bandwidth	3.500	3.500	3.500
N_Banked	93	93	93
N_UBanked	121	121	121
DepMean	0.285	0.201	0.514

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

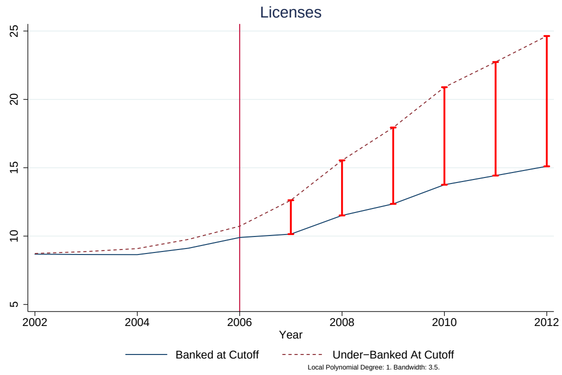
Note: Reduced form estimated using local quadratic regressions with controls for district population and its square. Bandwidths are set to 3.5 thousand persons per branch and estimated using a triangular kernel. NREGA was rolled out in 3 phases between 2006 and 2009 based on some measure of expected program need by district.

[4] In analysis not shown, the district composite index of backwardness is shown to be smooth and continuous at the under banked cutoff. Further, persons per branch shows a positive correlation with the index. This reaffirms that banks typicaly avoid branching in troubled areas.

3 Additional Results

3.1 Discontinuity Figures

Figure 7: Discontinuity from Reduced Form: Operating Private Bank Licenses



Note: Estimated using local linear regressions with controls for district population and its square, the percent of workers in agriculture and the pre-randomization 2001 value of the dependent variable. Bandwidths are set 3.5 thousand persons per branch and estimated using a uniform kernel. The figure plots the estimated intercepts at the cutoff from the estimation of the RD equation repeated annually. The red dashed line provides the estimated intercept from approaching the threshold along the under banked side. The solid blue line reports the corresponding intercept approaching from the banked side. The distance between the two, reported for each year, shows the estimated discontinuity at the threshold. A solid line between the two points indicates an estimated discontinuity with statistical significance of at least the 10% level. The thin vertical line at 2006 represents the first estimation made following the reform.

3.2 Regression Tables: Bias-corrected and Robust Estimates from RD

The main text reports conventional estimates of the regression discontinuities and their standard errors except where noted for the nighttime lights analysis. The set of discontinuities are consistent across the conventional and bias-corrected estimates in most cases, with some of the closest measures found from bank branching responses shown in table 3 below. Levels of statistical significance also generally match up across conventional, bias-corrected and "robust" measures. The greatest departure from this pattern in the set of analysis reported in the paper is in the discontinuity estimates for the percentage change in average district nighttime lights. Measures of all three estimate methodologies are shown below in table 6. For the nighttime lights, local quadratic regressions provided better fits to the data over linear ones.

For some aggregate measures on credit, bias-corrected estimates of the discontinuity are larger and show statistical significance in earlier years. Table 4 provides an example. However, analysis focusing on narrower applications of credit, such as credit amounts to direct agriculture in rural and semi-urban areas shown in table 5 exhibit greater consistency. From the visual RD of the data for several years, a few exceptional outcomes in aggregate credit measures very close to the cutoff appear to exaggerate the discontinuity when using a quadratic fit. Those instances may drive the difference observed between measures. Contact the author for estimates from the three methodologies for other analysis appearing in the main text.

Table 3: Fuzzy RD: Private Bank Operating Branches

VARIABLES	(1) 2002	(2) 2003	(3) 2004	(4) 2005	(5) 2006	(6) 2007	(7) 2008	(8) 2009	(9) 2010	(10) 2011	(11) 2012
Conventional	0.181	0.343	0.577	0.644	0.719	1.270	3.262**	4.840***	7.051***	9.219***	10.58***
	[0.152]	[0.320]	[0.557]	[0.641]	[0.865]	[1.005]	[1.279]	[1.653]	[2.159]	[2.718]	[3.102]
Bias-corrected	0.166	0.383	1.036*	1.036	1.139	1.468	3.754***	5.158***	7.414***	9.730***	11.11***
	[0.152]	[0.320]	[0.557]	[0.641]	[0.865]	[1.005]	[1.279]	[1.653]	[2.159]	[2.718]	[3.102]
Robust	0.166	0.383	1.036	1.036	1.139	1.468	3.754**	5.158***	7.414***	9.730***	11.11***
	[0.180]	[0.376]	[0.665]	[0.771]	[1.027]	[1.179]	[1.507]	[1.956]	[2.558]	[3.214]	[3.660]
Bandwidth	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500
N_Banked	95	94	94	94	94	94	94	94	94	94	94
N_UBanked	122	122	122	122	122	122	122	122	122	122	122
DepMean	8.636	8.801	9.125	9.597	10.34	10.87	12.25	14.42	16.19	17.91	20.00

Note: Standard errors in brackets. ***p<0.01, **p<0.05,*p<0.1 for all tables. Estimated using local linear regressions with controls for district population and its square, the percent of workers in agriculture, a control for monsoon rainfall and the pre-randomization 2001 value of the dependent variable. Bandwidths are set 3.5 thousand persons per branch and estimated using a triangular kernel. Under banked status is instrumented for with predicted under banked assignment.

Table 4: Fuzzy RD: Private Sector Banks Credit Accounts in Aggregate

VARIABLES	(1) 2002	(2) 2003	(3) 2004	(4) 2005	(5) 2006	(6) 2007	(7) 2008	(8) 2009	(9) 2010	(10) 2011
Conventional	-0.186	0.386	1.695	3.652**	6.220*	10.57***	8.679*	8.432*	13.16**	18.54*
	[0.599]	[0.804]	[1.123]	[1.528]	[3.426]	[3.399]	[5.098]	[4.521]	[6.219]	[10.26]
Bias-corrected	-0.209	0.797	2.539**	4.172***	7.271**	12.18***	9.761*	9.950**	14.82**	19.66*
	[0.599]	[0.804]	[1.123]	[1.528]	[3.426]	[3.399]	[5.098]	[4.521]	[6.219]	[10.26]
Robust	-0.209	0.797	2.539*	4.172**	7.271*	12.18***	9.761	9.950*	14.82**	19.66*
	[0.698]	[0.955]	[1.307]	[1.835]	[4.047]	[4.228]	[5.988]	[5.302]	[7.279]	[11.92]
Bandwidth	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500
N_Banked	95	94	94	94	94	94	92	94	94	94
N_UBanked	122	122	121	122	122	122	122	122	122	122
DepMean	5.067	5.484	6.470	8.800	12.83	13.77	16.14	17.78	22.82	25.80

Note: Accounts reported in thousands. Estimated using local linear regressions with controls for district population and its square, the percent of workers in agriculture, a control for monsoon rainfall and the pre-randomization 2001 value of the dependent variable. Bandwidths are set 3.5 thousand persons per branch and estimated using a triangular kernel. Under banked status is instrumented for with predicted under banked assignment.

Table 5: Fuzzy RD: Percentage Change in Private Credit Amount Direct to Agriculture in Rural and Semi-Urban Areas

VARIABLES	(1) 2002	(2) 2003	(3) 2004	(4) 2005	(5) 2006	(6) 2007	(7) 2008	(8) 2009	(9) 2010	(10) 2011
Conventional	0.0552	0.143	0.216	0.709**	1.029**	0.830*	0.908	1.840***	1.445**	1.755***
	[0.107]	[0.166]	[0.212]	[0.330]	[0.414]	[0.503]	[0.603]	[0.620]	[0.647]	[0.613]
Bias-corrected	0.0866	0.253	0.298	0.931***	1.273***	1.038**	1.195**	2.103***	1.713***	1.923***
	[0.107]	[0.166]	[0.212]	[0.330]	[0.414]	[0.503]	[0.603]	[0.620]	[0.647]	[0.613]
Robust	0.0866	0.253	0.298	0.931**	1.273**	1.038*	1.195*	2.103***	1.713**	1.923***
	[0.131]	[0.197]	[0.247]	[0.395]	[0.504]	[0.617]	[0.722]	[0.751]	[0.791]	[0.739]
Bandwidth	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500
N_Banked	95	94	94	94	94	94	92	94	94	94
N_UBanked	122	122	121	122	122	122	122	122	122	122
DepMean	-0.0700	0.0481	0.164	0.433	0.550	0.964	1.488	1.419	1.953	2.376

Note: Percentage change is approximated using difference in logs relative the value reported in 2001. Estimated using local linear regressions with controls for district population and its square, the percent of workers in agriculture and a control for monsoon rainfall. Bandwidths are set 3.5 thousand persons per branch and estimated using a triangular kernel.

Table 6: Difference in Log Mean District Light from 2004

Fuzzy RD Estimated Annually, Instrumenting for Under Banked Status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	2005	2006	2007	2008	2009	2010	2011	2012
Conventional	-0.0373	0.00436	0.113***	0.112**	0.0479	0.126	0.111*	0.119
	[0.0227]	[0.0325]	[0.0415]	[0.0460]	[0.0778]	[0.0773]	[0.0652]	[0.105]
Bias-corrected	-0.0430*	0.00773	0.129***	0.128***	0.0567	0.151*	0.141**	0.148
	[0.0227]	[0.0325]	[0.0415]	[0.0460]	[0.0778]	[0.0773]	[0.0652]	[0.105]
Robust	-0.0430*	0.00773	0.129***	0.128**	0.0567	0.151*	0.141**	0.148
	[0.0244]	[0.0358]	[0.0445]	[0.0501]	[0.0849]	[0.0845]	[0.0700]	[0.114]
Bandwidth	3.500	3.500	3.500	3.500	3.500	3.500	3.500	3.500
N_Banked	95	95	95	95	95	95	95	95
N_UBanked	122	122	122	122	122	122	122	122
DepMean	-0.139	-0.0805	0.114	0.0722	0.0266	0.355	0.219	0.297

Fuzzy RD Instrumenting for Private Bank Branches, Pre-reform and Post-Reform

	(1)	(2)
VARIABLES	preref	postref
Conventional	-0.0264	0.00508
	[0.274]	[0.00373]
Bias-corrected	-0.0455	0.0115***
	[0.274]	[0.00373]
Robust	-0.0455	0.0115***
	[0.320]	[0.00444]
Bandwidth	3.500	3.500
N_Banked	94	658
N_UBanked	122	854
DepMean	-0.139	0.143

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Reduced form estimated using local quadratic regressions with controls for district population and its square. Bandwidths are set to 3.5 thousand persons per branch and estimated using a triangular kernel. The fuzzy regression discontinuity is estimated using local linear regressions. The number of operating private bank branches is instrumented with predicted under banked assignment. Controls include district population and its square, the percent of workers in agriculture, and a control for monsoon rainfall. Pre-reform year is 2005 using 2004 as the base year for the approximate percentage change. Post-reform years are 2006-2012.

3.3 Individual Crop Outcomes

Figure 8 shows the reduced form regression discontinuity analysis for yield and output for four important crops in India, rice, wheat, cotton and onion. I present discontinuity analysis for crop yield (tonnes per hectare of cultivated land) and output (tonnes). Each specification controls for the district averaged percentage deviations of rainfall and its lag, district population and its square, the percent of workers in agriculture and a pre-randomization value of the dependent variable. The effect on yield is measured with greater precision for rice and wheat relative to output, while the opposite is true for cotton and onion. Positive effects from the reform emerge in 2005 with statistical significance for rice, and for the bias-corrected estimate on wheat. The effect is observed through 2008, though precision is lost in 2008 for rice. After 2008 the effect diminishes in magnitude and is estimated with statistical significance only in the biased corrected estimates for the yield of wheat. Contact the author for tables of estimates across conventional and bias-corrected methodologies. The data on output shows a similar pattern but with less precision.[5]

The empirical evidence from the individual crop and index analysis supports a causal effect of the banking reform on agricultural outcomes. The effects appear to be nuanced, however, with the yield of certain crops improving, while the output of crops such as cotton and onion increased more in treated districts with the reform. The greatest response for onion yield, however, occurred in years 2009-2010, corresponding to the expansion of lending to indirect agricultural loans. Indirect agriculture loans include funds for storage facilities, which are anecdotally suggested to be important for the onion market. Further, the increase in yields of rice appear to be driven by a small expansion of output in underbanked districts with the reform, but a significant decrease in the share of total cultivated land dedicated to rice production in those districts. Concurrently, the share of land for cotton expanded. Hence, the availability of credit appears to have affected cropping patterns in addition to an expansion of yields for revenue important crops.

[5] Estimates for other crops are omitted for brevity, with many yielding null results.

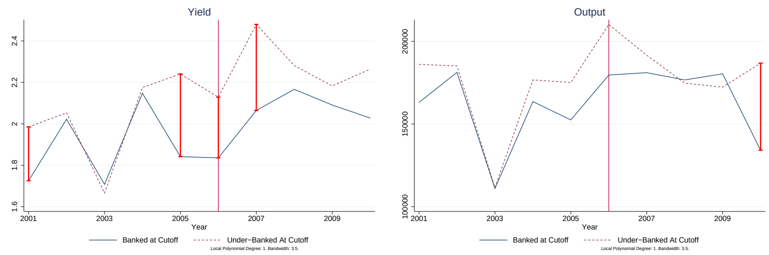
Table 7: Summary Statistics: Crops

	Banked, Pre-reform			Banked, Post-reform			Under Banked, Pre-reform			Under Banked, Post-reform		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Cotton												
Area	403	32,656	53,321	349	31,406	56,677	619	31,351	64,876	471	37,076	75,472
Output	403	59,959	127,462	349	100,347	229,598	619	41,581	89,199	471	86,119	203,562
Productivity	403	1.61	0.98	349	2.12	1.38	619	1.35	0.84	471	1.55	1.28
Maize												
Area	560	11,945	20,923	470	15,124	28,518	968	16,400	32,962	761	16,688	36,426
Output	560	27,988	57,175	470	48,069	103,819	968	28,449	64,162	761	34,070	87,053
Productivity	560	1.87	1.19	470	2.38	2.24	968	1.49	0.84	761	1.76	1.35
Onion												
Area	431	1,527	3,714	342	2,036	5,455	743	1,074	2,489	510	1,485	4,019
Output	431	13,885	29,608	342	17,539	36,355	743	14,587	51,185	510	24,189	99,249
Productivity	431	11.71	7.93	342	12.03	8.58	743	11.34	7.48	510	11.38	7.92
Potato												
Area	351	2,028	4,026	303	2,303	6,024	674	3,014	9,512	587	3,694	12,041
Output	351	28,503	44,128	303	27,843	43,051	674	67,058	248,196	587	71,627	286,377
Productivity	351	13.75	7.51	303	12.93	7.79	674	12.64	7.55	587	11.76	8.19
Rice												
Area	667	64,626	82,739	544	67,299	85,705	1017	88,839	104,258	784	100,968	120,405
Output	667	173,077	285,059	544	194,407	303,283	1017	160,160	221,919	784	197,829	266,243
Productivity	667	2.30	1.01	544	2.51	1.10	1017	1.61	0.87	784	1.81	0.94
Sesamum												
Area	573	3,245	6,935	460	2,790	4,742	908	4,826	11,359	749	5,919	15,535
Output	573	1,220	3,198	460	1,119	2,212	908	1,805	5,529	749	2,032	6,103
Productivity	573	0.35	0.23	460	0.38	0.25	908	0.32	0.22	749	0.35	0.24
Sugarcane												
Area	523	12,161	23,096	419	11,554	22,413	907	8,554	25,972	711	8,866	27,790
Output	523	955,008	1,797,426	419	902,855	1,738,094	907	590,206	1,786,733	711	588,924	1,878,506
Productivity	523	70.26	35.51	419	67.35	39.47	907	53.13	26.72	711	55.86	30.25
Tobacco												
Area	166	7,958	16,242	176	8,267	17,829	258	454	1,647	213	620	2,082
Output	166	9,853	22,353	176	10,113	20,766	258	663	2,233	213	1,128	3,622
Productivity	166	1.54	1.53	176	1.53	1.61	258	1.63	1.88	213	1.71	1.57
Wheat												
Area	437	60,088	81,807	349	64,550	81,240	923	49,803	65,451	689	52,869	67,471
Output	437	204,344	353,065	349	225,183	353,261	923	126,363	200,516	689	147,671	224,604
Productivity	437	2.21	1.25	349	2.38	1.27	923	1.78	0.97	689	1.93	1.02

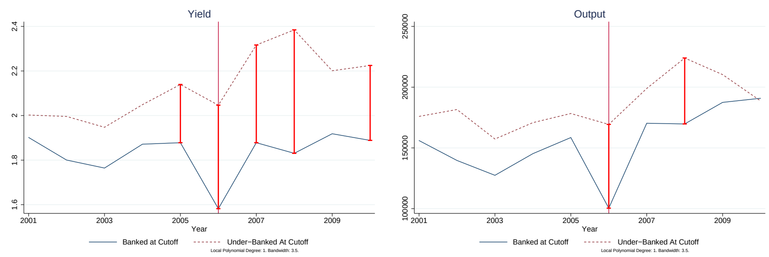
Source: Rainfall data from TRMM satellite, crop data from State Agricultural Reports. Sample includes years 2000-2010 for districts falling within 5 thousand persons per branch of the national average. Observations are crop-years; the number of districts varies by crop as not every crop is grown in all districts. 302 of 572 districts are eligible for sample. Area is reported in Hectares square, output in tonnes, and productivity is output divided by area. Cotton reported in bales instead of tonnes.

Figure 8: Discontinuity from Reduced Form: Individual Crops

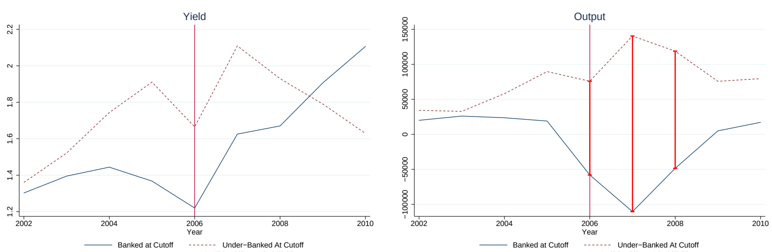
Rice Statistics



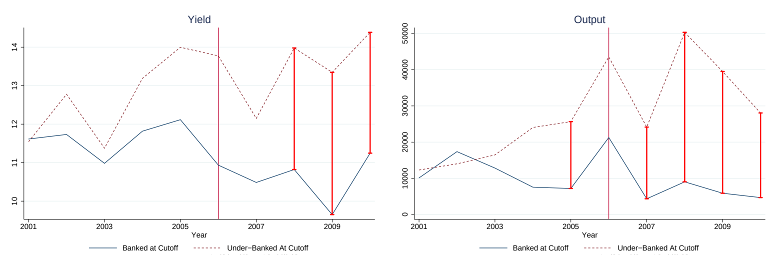
Wheat Statistics



Cotton Statistics



Onion Statistics



Note: Yield [Tonnes/Hectare] (Left), Output [Tonnes] (Right). Cotton output measured in bales of cotton lint rather than tonnes. Estimated using local linear regressions with controls for district population and its square, the percent of workers in agriculture, a control for monsoon rainfall and the pre-randomization 2001 value of the dependent variable. Bandwidths are set 3.5 thousand persons per branch and estimated using a uniform kernel. See notes from figure 6 for graph description.

4 Additional Data Appendix

4.1 Districts

The majority of analysis in this paper is conducted at the administrative district level in India. Districts constitute the administrative level directly below the state government (and union territory). Data sets at the district level rarely provide numerical identifiers. When available, these identifiers typically do not easily map to other data sets. Further, the anglicized spelling of district names is often inconsistent across and even within data sets. Renaming and redistricting also occur relatively frequently in India. As such, each data set required the assignment of a numerical identifier before conducting analysis. To ensure consistent measures in the data across time, I adjust all data to their 2001 district boundaries from the Population Census. I first assign each district its 2001 state and district numerical codes from the 2001 Population Census, or an auxiliary district code if the district was formed post 2001. Then using the atlas provided in the 2011 Population Census, I map new districts back to their source districts in 2001. Although super-districts, created when newly formed districts drew land from more than one source district, are identified, they are dropped from the analysis.[6] District websites, newspapers and other internet based resources were used to help map alternative spellings to numerical codes.

4.2 Banking

Branches and Licenses Data on the number of operating branches and licenses are from the Master Office File (MOF) accessed from the RBI website in spring 2012. Opening and closing dates (when applicable) are provided for each bank, as well as information regarding branch location and type of business conducted at the branch (e.g. General Banking, Specialized Banking, ATM). "Brick and mortar" branches are used in the analysis, meaning branches classified as being general banking or specialized banking. Not Administratively Independent Offices such as extension counters and ATMs are excluded from analysis. The number of operating branches for each year is calculated as the number of branches with an opening date prior to January 1st of that year and a closing date afterward or missing. Operating branches by subsets of bank group classification are calculated similarly. Licenses are considered to be

[6] New districts since 2001 that claimed territory from more than one source district are dropped along with the source districts due to issues with the aggregation. In addition, Thane and Pune districts in Maharashtra are dropped. These districts are located close to Mumbai but are not technically classified as belonging to the greater Mumbai area. They constitute outliers as they achieve rapid growth more likely to be attributable to their proximity to Mumbai. Thane is on the under banked list while Pune is not, though the RBI amended the policy to 2008 to make centers within 100km of Mumbai ineligible for under banked status, effectively removing Thane's status. Varanasi district in Uttar Pradesh is also dropped after 2002 due to the 2002 merger of the private sector Banaras State Bank with Bank of Baroda which is a nationalised bank. Banaras State Bank primarily operated in Uttar Pradesh with the bulk of its branches in districts designated as under banked. However, 20 branches operated in Varanasi which happens to be located right at the cutoff on the banked side. The vast majority of branches affected by the merger belonged to districts designated as under banked. However, the reclassification of 20 branches to public sector bank status just on the banked side of the cutoff results in a sudden drop in the banked intercept in 2003 for private banks. Since most of these branches continued to operate under the public sector, the drop-off creates an exaggerated representation of the policy effect, which does not accurately represent the change to the banking environment. While these branches could be "added back" using the detailed data from the MOF, the same cannot be done for the aggregated data on credit.

operating if issued before January 1st of a given year with a branch close date afterward or missing. Thus, licenses can be in "operation" even if branch opening occurs at a later date. After the December 2009 reform granting blanket permissions to open in low population centers, the incidence of unreported license dates for branches in such centers increased. The assumption is made that these constitute branches exercising the blanket permission, such that the effective license date is taken to be the date of branch opening.[7]

Credit The Basic Statistical Returns 1 (BSR1) provides information on credit accounts, credit limits and credit outstanding by scheduled commercial banks including RRBs (last accessed spring 2014). The data are reported annually by banks with values as of March 31st for that year. Credit captured by BSR1 relates to gross bank credit such as term loans, cash credit, overdrafts, etc. Detailed descriptions are provided by the RBI. The financial year 200X-200Y is reported as 200Y in the paper and is reported with consistent notation across analyzed data. Values are delineated by bank group and population group at the district level (e.g. number of credit accounts with Nationalised Banks, by semi-urban areas in Rangareddy). Locations, such as semi-urban Rangareddy, represent the area of credit utilization for loans exceeding 2 lakh Rs. for which detailed account information is collected. Loans of lesser amounts are reported with less information, and the RBI assumes they are utilized in the same area as which the loan was sanctioned. Credit amounts are further delineated by utilization purpose, coined "occupation," and include : agriculture, industry, professional and other services, personal, trade, transport operators, finance and all other. These are broken down further for agriculture into "direct" and "indirect," for industry by "construction" "mining" "manufacturing and processing" and "electricity, gas and water" and trade by "retail" and "wholesale." Personal loans are also presented disaggregated, but the delineation between subgroups appears to be inconsistent through time so are always treated as aggregated personal loans in the analysis. A reclassification of loans to make occupations consistent with a 2004 update of industrial codes occurred in 2008. The reclassification should not have affected aggregate measures of account and amounts, though caution should be taken when attempting to draw comparisons at the occupation level before and after 2008.[8]

The BSR2 provides analogous information for deposits and is structured similarly (last accessed spring 2014). Values are reported for the number of deposit accounts and deposit amounts.

The BSR7 provides quarterly data on credit, deposits and reporting branches. Analysis on BSR7 is not included in this paper.

All credit and deposit limits and amounts are adjusted using the Consumer Price Index for Industrial Workers provided by India's Labour Bureau. I adjust all values to 2011, fourth quarter prices. Amounts are reported in Rupees.

[7] A similar pattern for license dates from branches in urban centers in the Northeast region that had a special exception for blanket permissions for urban centers, and only in that region, provides additional support for this assumption.

[8] Two districts exhibit measures of credit accounts and amounts that appear to reflect coding errors in the data. Mallapuram, Kerala is dropped in 2004 due to an unexplainable jump in the magnitude of credit unmatched in the district in any other years. Ghaziabad, Gujarat in 2008 displays even more erratic values for certain credit measures. These values are set to missing as the remaining appear unaffected. In both instances, private sector banks with a presence in the concerned district were acquired by the public sector. The reclassification of the bank to the public sector may have created underlying issues in the data reported in those places for those years.

Population Groups The RBI follows a specific assignment procedure for population groups. Based on the Population Census, locations with populations less than ten thousand are designated rural; 10,000 - 100,000 semi-urban; 100,000 - 1 million urban; and greater than 1 million metropolitan. Prior to 2005 locations were assigned status based on their 1991 Population Census values. The switch to the 2001 Population Census for reports in 2006 and later make strict comparisons between the sets of years complicated at the disaggregated population group level. The problem appears to be greater for the metropolitan and urban population groups, as fewer centers exist in these categories. The scope for problems appears smaller for rural and semi-urban classifications due to the high volume of centers in these categories. Still, the caveat should be kept in mind for analysis at the disaggregated level.

4.3 Agriculture

Crop output and area The data on crop output and area are reported in the Annual Crop Yields at District Level from the Crop Production Statistics. The production output in tonnes and area cultivated in square hectares are reported by crop at the district level either annually or by season, depending on the crop and state. Reported crops vary across districts, and the detail of information on variety and growing season also varies across states and years. I develop the data from a file made available from the Government of India for years 1998-1999 to 2010-2011 (years reported July-June). Extensive cleaning of district and crop names, as well as accounting for redistricting, is required to analyze the data as a panel. I match each district reported to their 2001 Population Census identification number or to a 2011 ID number constructed for this analysis when dealing with new districts since 2001. Analysis is restricted to years 2001-2010 which exhibit lower frequencies of missing data. Missings values after 2010 are reported to be due to unfiled state reports. Districts never reporting positive statistics for a crop over the sample period are dropped from analysis for that individual crop. Missing values for crop statistic may be due either to null values being reported, or changes to the set of crops surveyed across years in a state. Missing values are left as missing.

Crop prices The data on crop prices are from the Farm Harvest Prices of Principle Crops. States are responsible for reporting crop prices for a set of prominent crops each year. The prices are supposed to be collected during the peak harvest times of each crop and account for variations in quality. States vary in their reporting of crop prices by season and detail on variety. Further, states vary in reporting price for some crops by product (e.g. some report prices for sugarcane while others only report prices for raw sugar, cotton lint or whole cotton, etc.) Technical conversion factors for raw crops to agricultural outputs provided by the Statistic Division of the FAO are used where applicable to match prices to corresponding crop outputs. Prices are reported in Rupees per Quintal (an Indian quintal is 100 kg) and must be converted to Rupees per tonne for consistent units with the output data. I have developed the data from pdf reports available in separate sets by state for 2001-2002 to 2003-2004, 2004-2005, 2005-2006, 2006-2007 to 2007-2008 and 2008-2009 to 2009-2010. Extensive cleaning of district names, accounting for redistricting, and assignment to identification numbers was similarly required.

4.4 Industry

Annual Survey of Industries The Annual Survey of Industries (ASI) is a detailed survey of registered manufacturing firms in India conducted by the Central Statistical Organisation. The ASI is used extensively in economic research (Hsieh and Klenow, 2009; Bollard et al., 2013) to name just a few). I use fiscal years 2001-2010 in my analysis. In these years, all firms with greater than 100 workers were enumerated, as were all firms operating in the five less developed states/UTs (Manipur, Meghalaya, Nagaland, Tripura and Andaman & Nicobar Islands). The remainder of registered firms (those with greater than 10 workers, assuming compliance) were surveyed from samples representative at the State by NIC-2004 4 digit industry code. In addition to the values reported directly in the ASI, I construct the capital labor ratio as the average of the opening and closing values of assets net of depreciation divided by the sum of the firm's wage bill plus benefits, as in Hsieh and Klenow (2009). Due to the joint census-sampling methodology, I conduct my analysis at the state level in order to apply proper weighting for a representative sample of all registered firms. A thorough discussion of the ASI data can be found in Bollard et al. (2013).

4.5 Remote Sensing

DMSP-OLS Nightlights The Defense Meteorological Satellite Program (DMSP) maintains data sets of night lights data, constituting a yearly average of the amount of light emitted into space at night for a roughly 1km square grid. Using satellite images, algorithms to control for reflection, cloud cover and other confounding factors assign a digital number between 0 and 63 for each cell that may be downloaded as a finely pixelated map of the Earth. Using the boundary outline of India's administrative districts in 2001, I construct the district average of the digital numbers in each district. I then calculate the percentage change of this average as the log of the district mean value minus the log district mean from 2004. Analyzing changes in growth across districts, as opposed to levels is important due to measurement error introduced through machine learning and the algorithms applied to eliminate glare and light bleed. I have processed data from satellites F16 and F18, that cover calendar years 2004-2012. A thorough discussion of the nightlights data is included in Henderson et al. (2012).

TRMM Rainfall Data Rainfall strongly affects agricultural productivity. To the extent that rainfall varies annually across districts, conditioning on it will improve my precision for estimates related to agriculture. I use the publicly available data collected by the Tropical Rainfall Measuring Mission (TRMM) satellite jointly maintained by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA). Fetzer (2014) gives a detailed description of these data and their verification processes. These data are collected from a satellite orbiting approximately 250 miles above the Earth's surface that completes an orbit several times a day and is able to detect rainfall falling as lightly as 0.7 millimeters per hour. Daily rainfall measures are available from 1998-2012 on a 0.25 by 0.25 degree grid, making it the finest available spatial resolution for India to the best of my knowledge.

These data are likely favorable to those generated using ground rainfall gauges as the latter require local monitoring and maintenance, the quality of which may vary systematically with the prosperity of districts. Further, the spatial diffusion of gauges is not uniform across India, requiring different levels of interpolation between rain gauges that can introduce measurement error that may be difficult to account for and change in less transparent ways as the number and location of gauges vary across time.

5 Theoretical Framework Appendix

This section sketches out the theoretical framework for anticipated competition in the second period leading to increased levels of credit at the time of policy announcement in the first. Details of intermediate steps are omitted in the interest of space. The participation constraint for borrower i in each period is given by,

$$\begin{cases} E[\pi_i^1] = P_s(R_s^A)[R_s^A(1-\theta_i) - (1+r^1)] - s > \mu & Period\,1 \\ E[\pi_i^2] = \{P_s(R_s^A)[R_s^A(1-\theta_i) - (1+r^2)] - s1(First\,Time\,Borrower) > \mu & Period\,2 \end{cases} \tag{1}$$

where $\theta_i \sim uniform[0,1]$ is a privately known cost to the borrower that is constant across periods (as is being a safe type), r^t denotes the interest rate in period 1 or 2, and $1(First\,Time\,Borrower)$ is the indicator function for the borrower's first period of borrowing from the specific bank. Consider the borrower participation constraint from period t. The indifferent borrower with type θ_i facing interest rate r^t will satisfy

$$P_s(R_s^A)R_s^A(1-\theta_i) = P_s(R_s^A)(1+r^t)] + s1(First\,Time\,Borrower) + \mu \tag{2}$$

Rearranging terms, the indifferent borrower may be expressed as a function of the interest rate r^t,

$$\hat{\theta}_i(r^t) = 1 - \frac{P_s(R_s^A)(1+r^t) + s1(First\,Time\,Borrower) + \mu}{P_s(R_s^A)R_s^A} \tag{3}$$

such that all borrowers with $\theta_i < \hat{\theta}_i(r^t)$ will demand a loan with interest rate r^t. For a market of size M, total demand for loans at interest rate r^t will be $M\beta\hat{\theta}_i(r^t)$.

Assume banks are profit maximizers, face an exogenous marginal cost of funds plus administrative costs of lending equal to $(1+\rho)$, and cannot discriminate in the interest rate offered to repeat and first time borrowers. The bank's participation constraint from each period is,

$$E[\pi_B^t] = P_s(R_s^A)(1+r^t)\theta(r^t) > (1+\rho)\theta(r^t) \quad for\,t = 1,2 \tag{4}$$

In deciding the interest rate for each period, the incumbent bank will anticipate its outcome in the second period if facing entry and take that into consideration in setting its first period interest rate. Specifically, if a new bank enters the market in the second period, the incumbent will expect to compete in interest rates such that the entrant must offer his zero-profit condition interest rate and the incumbent will offer the interest rate making his first period borrowers that do not pay the screening fee if they stay indifferent between borrowing from him and the incumbent.

Sketch of Proof: *If the incumbent offers an interest rate higher than that making first period borrowers indifferent between borrowing from the incumbent while avoiding switching costs and borrowing from the entrant while incurring the switching costs, then the incumbent loses the entire market to the entrant. If the incumbent offered an interest rate lower than that value, then he loses profits from the locked in first period borrowers but gains no new borrowers since new borrowers must pay the screening fee regardless and the entrant's interest rate is strictly lower. If instead the entrant offered a price above the zero profit condition interest rate, then the incumbent would increase his rate to earn higher profits off of his first period borrowers. This, however, creates incentive for the entrant to lower his interest rate a small amount and capture the entire market. If the entrant instead lowers his interest rate he will serve the entire market at a loss.*

Taking the second period equilibrium into consideration, the incumbent knows his second period interest rate when facing entry will be $1 + r_2^I = \frac{1+\rho+s}{P_s(R_s^A)}$ by equating demand for the zero profit interest rate and demand for an interest rate when the switching cost need not be incurred. The incumbent will then maximize first period interest taking the second period predetermined interest rate into consideration as the first period interest rate will determine the demand faced in both periods. Thus, the incumbent's maximization problem is

$$\max_{r_1^I, r_2^I} P_s(R_s^A)(1 + r_1^I)\theta(r_1^I) + \delta P_s(R_s^A)(1 + r_2^I)\theta(r_2^I) - [(1 + \rho)\theta(r_1^I) + \delta(1 + \rho)\theta(r_2^I)] \quad (5)$$

Substituting in the value for r_2^I and setting demand equal in both periods reduces the problem to

$$\max_{r_1^I} P_s(R_s^A)(1 + r_1^I)\theta(r_1^I) + \delta P_s(R_s^A)(\frac{1 + \rho + s}{P_s(R_s^A)})\theta(r_1^I) - (1 + \delta)(1 + \rho)\theta(r_1^I) \quad (6)$$

Taking the first order condition with respect to r_1^I, setting it equal to zero and solving for the optimal first period interest rate for the incumbent yields,

$$1 + r_1^{I*Entry} = \frac{1}{2P_s(R_s^A)}\{P_s(R_s^A)R_s^A - (1 + \delta)s - \mu + (1 + \rho)\} \quad (7)$$

Intuitively, the incumbent increases the interest rate with the expected payoff of the project to capture additional surplus as well as the cost of lending the funds and lowers the interest rate with the borrower's reservation utility. The incumbent lowers the interest rate as the switching cost increases, as this relaxes the constraint on the interest rate he offers in the second period, allowing for higher profits from each continuing first period borrower.

To determine the effect of anticipated competition on first period lending, consider an incumbent that does not expect entry in the second period. He will find it optimal to set interest rates so as to maximize total profit from both periods, increasing the interest rate in the second period to extract the additional surplus from the repeat borrowers no longer paying the screening cost. Since no other changes occur to the environment, the incumbent will maximize profits by serving the same set of borrowers in both periods, setting the second period interest rate so as to make the marginal borrower indifferent between accepting the loan and not. The maximization for the incumbent not expecting entry may be expressed as,

$$\max_{r_1^I} P_s(R_s^A)(1+r_1^I)\theta(r_1^I) + \delta P_s(R_s^A)(1+r_1^I+s)\theta(r_1^I) - (1+\delta)(1+\rho)\theta(r_1^I) \tag{8}$$

Taking the first order condition with respect to r_1^I, setting it equal to zero and solving for the optimal first period interest rate for the incumbent yields,

$$1 + r_1^{I*NoEntry} = \frac{1}{2P_s(R_s^A)}\{P_s(R_s^A)R_s^A - (1 + \frac{\delta P(R_s^A)}{(1+\delta)})s - \mu + (1+\rho)\} \tag{9}$$

Finally, since the interest rate determines the first period quantity of credit, anticipated competition will lead to an expansion of credit if $1+r_1^{I*Entry} < 1+r_1^{I*NoEntry}$. This inequality reduces to the simple expression, $\frac{P_s(R_s^A)}{1+\delta} < 1$ that must always be true. Hence, introducing the potential of future competition leads to an expansion of credit at the time announcement.

Online Appendix References

References

(2004). Report of the advisory committee on flow of credit to agriculture and related activities from the banking system. Technical report, Vyas Committee, Reserve Bank of India.

Bollard, A., P. J. Klenow, and G. Sharma (2013). IndiaÊŒs mysterious manufacturing miracle. *Review of Economic Dynamics 16*(1), 59–85.

Fetzer, T. (2014). Can workfare programs moderate violence? Evidence from India. Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Klonner, S. and C. Oldiges (2014). Safety net for India's poor or waste of public funds? Poverty and welfare in the wake of the world's largest job guarantee program. Technical report, University of Heidelberg, Department of Economics.

Mohan, R. (2002). Transforming indian banking: In search of a better tomorrow. Technical report, RBI Bulletin.

Zimmermann, L. (2012). Labor market impacts of a large-scale public works program: Evidence from the Indian employment guarantee scheme.