

## Appendix A: Sample Selection and Randomization

We use data from 50 service area pairs<sup>1</sup> across three districts. The average service area of a branch spanned 3-5 km from the branch office and encompassed an average of 10 villages and 2,400 households. Villages serviced by a single branch were typically well-connected by roads and bus routes. Below we first describe randomization and then surveying.

The selection of potential branch sites and randomization across them proceeded as follows: In conjunction with the bank, potential location sites were identified using a global position system (GPS)-based population survey which determined relevant political, administrative and social boundaries. Once all feasible branch locations in the district had been designated, we used Edmond's algorithm for minimum distance matching to construct pairs of service areas. This matching for treatment and control allows the study to overcome issues in seasonality and geographic correlation in outcomes by minimizing differences between paired branches. It also improves balance across treatment and control villages on observed and unobserved factors, and provides a strong service-area-level control variable. For several 2001 census village outcomes (including caste composition, number of primary schools, water facilities and proportion of irrigated land), we find that controlling for pair fixed effects explains roughly 70% of the variance. One service area in each pair was then randomly selected to receive a bank branch first (treatment area). Expansion in the other area (control) was delayed for 36 months. Bank employees were not informed about the study or whether their branch is a study branch or not. Treatment and control areas of the same pair will be surveyed simultaneously. Surveyors do not know the treatment status of villages and are rotated across treatment and control.

The opening of bank branches happened in three rounds due to operational constraints following the Indian microfinance crisis in late 2010. We will account for this fact in the empirical analysis and may, at times, restrict the analysis to certain rounds only, for example when looking at the influence of the Pradhan Mantri Jan Dhan Yojana (PMJDY) scheme which started in 2014.

---

<sup>1</sup> 101 service areas are covered, due to one triplet with two control areas.