

## **FIN 3040 (ECO 3205) - 01, 02**

### **Homework 1**

Due Date: September 26 (11:59 PM)

#### **About:**

Homework 1 tests the relationship between per capita consumption and access to financial institutions using district-level data from India. The dataset combines data from the Basic Statistical Returns and three rounds of the National Sample Surveys. The former is published annually by the Reserve Bank of India and provides district data on bank branches by bank ownership. The latter is based on household surveys conducted every five years. Surveys from the 1999-00, 2004-05 and the 2011-12 rounds are selected. The data is aggregated to districts based on 1991 Census boundaries and covers 361 districts from 19 major Indian states. The primary variables of interest are total, public and private bank branches per capita (per million population) and monthly per capita consumption expenditure (MPCE). Prior to starting the homework, it would be a good idea to spend some time to study the variable labels and gain familiarity with the data.

#### **Rules of Engagement:**

Students will work in groups of 2 with each group submitting one homework. You are required to submit both the homework solutions as a word document or PDF, as well as the Stata log file. Homeworks submitted without the Stata log file will result in an automatic halving of the points scored on the homework. Students are encouraged to collaborate across groups, and also use online resources for assistance with their Stata coding. Students can email the instructor and the teaching fellows with questions on the homework, or use the office hours for the same. There will be an additional office hour on Thursday, September 17 (from 9 to 10 AM) exclusively for the homework. However, the instructor and the teaching fellows will stop responding to queries on the homework after September 24, 11:59 PM (48 hours prior to the homework being due). Except for documented medical exigencies, no extensions will be granted for the homework.

#### **Dataset:**

*hw1data.dta*

# 1 Homework Assignment

## 1.1 Basic Summaries

1. Create a summary statistics table for the four variables of interest – MPCE, total bank branches per capita, public bank branches per capita and private bank branches per capita. [0.25]

1a. What fraction of districts lack private bank branches in each of the years 1999, 2004 and 2011? [0.25]

1b. List the districts which do not have any private bank branches in 2011, and prepare a summary statistics table showing their MPCE, educational statistics (literacy rate, fraction of adults with secondary or higher education), demographic characteristics (fraction of SC/ST and fraction of females), labour force characteristics (LFP, unemployment rate, fraction employed in agriculture, fraction employed in manufacturing, fraction employed as wage workers, fraction employed as casual workers). On the basis of these above characteristics, how different are the districts which don't have a private bank in 2011, from the ones which do? [0.5]

## 1.2 Raw Correlations

2a. Present a scatter-plot documenting the correlation between log MPCE and total bank branches per capita (MPCE on the y-axis). Using the correlation coefficient, describe the strength of this relationship. [0.5]

2b. Present scatter-plots to document the correlation between log MPCE and public and private bank branches per capita (MPCE on the y-axis). Compare the correlation coefficients across public and private banks with MPCE and comment on the same. Is the strength of the relationship between MPCE and branch density stronger for public or private banks? [0.5]

## 1.3 Basic Regressions

3a. Estimate the following regression specification:

$$\ln(MPCE_{dt}) = \alpha + \beta BranchPC_{dt} + \epsilon_{dt} \quad (1)$$

where *BranchPC* denotes per capita bank branches, *d* the district and *t* the year. Interpret the  $\beta$  coefficient. [0.5]

3b. Replicate specification (1) for public and private bank branches. Is the consumption-branch density relationship stronger for public banks? [0.5]

3c. Clustering errors: replicate specification (1) for total, public and private bank branches but cluster the standard errors by district (see command “cluster” and use the variable *dist91*.) Does the coefficient change relative to 3a) and 3b)? Does the standard error change relative to 3a) and 3b)? [1]

## 1.4 Adding Covariates

Let’s think about adding some covariates now! Remember to cluster the errors in all the specifications from hereon by *dist91*.

4a. Let’s start with the fraction of rural households (or urbanization). *A priori*, how do you think including the fraction of rural households as a covariate will affect the relationship between bank branches per capita and MPCE. Write a sentence about your prediction. [0.5]

4b. Now estimate the relationship between MPCE, and bank branches per capita after including the fraction of rural households as a covariate. Run separate regressions to estimate the coefficient on total bank branches per capita, public bank branches per capita and private bank branches per capita. How does the addition of the covariate affect the coefficient in each specification. Does this imply the coefficient estimated in 3c) was biased upwards or downwards? [1.5]

## 1.5 Adding More Covariates

For each question from 5a) to 7h), estimate 3 separate regressions and report 3 separate coefficients – one of total bank branches per capita, one for public bank branches per capita, and one for private bank branches per capita. For all the regressions, include the fraction of rural households as a covariate.

5a. Add in the following demographic covariates: a) log population b) fraction of SC/STs c) fraction of OBCs d) fraction of Muslims e) fraction female. Comment on how the coefficient

changes with the addition of these set of covariates (magnitude and significance)? [1]

5b. Add in the following labour force covariates (include the covariates mentioned in 5a) :  
a) fraction employed in manufacturing b) fraction of salaried workers c) fraction employed in the formal sector. Again, *a priori*, how do you think the addition of these covariates would affect the coefficient on bank branches per capita estimated 5a) (magnitude and significance)? Are the results consistent with your expectation? [1]

5c. Include the covariates mentioned in 5a) and 5b) and also add the following 3 covariates – a) LFP b) fraction unemployed and c) fraction employed in agriculture. Now let's think about educational characteristics – a) fraction of literates and b) fraction of adults educated secondary and higher. How do the coefficients change relative to 5b) (magnitude and significance)? Is this in line with your expectations? [1]

## 1.6 Adding Fixed Effects

Include all the covariates listed in 5c) for all regressions and cluster the errors using the variable *dist91*.<sup>1</sup>

6a. Include year fixed effects (see the command *areg*). How do the coefficients change in each instance? [0.5]

6b. Include district fixed effects (see the command *areg*). How do the coefficients change? Were the coefficients (relative to 5c) biased upwards or downwards? Why do you think the coefficients might be changing even after including a number of covariates? What unobservable factors do you think we might have omitted, which are now included through the district fixed effects? [1.5]

---

<sup>1</sup> Remember, the covariates are: a) fraction of rural households b) log population c) fraction of SC/STs d) fraction of OBCs e) fraction of Muslims f) fraction female g) fraction employed in manufacturing h) fraction of salaried workers i) fraction employed in the formal sector j) LFP k) fraction unemployed l) fraction employed in agriculture m) fraction of literates and n) fraction of adults educated secondary.

## 1.7 Heterogeneity by District Characteristics

We shall now investigate whether the relationship between access to finance and per capita consumption varies by district characteristics. This can assist us in identifying the channels through which access to finance might be affecting consumption. For all regressions, include both year and district fixed effects and all the covariates used in 6a) and 6b) and cluster the errors using the variable *dist91*.<sup>2</sup> We shall estimate the following specification:

$$\ln(MPCE_{dt}) = \alpha_d + \delta_t + \gamma_1 D_{dt} + \beta_1 BranchPC_{dt} + \beta_2 D_{dt} \times BranchPC_{dt} + \phi X_{dt} + \epsilon_{dt} \quad (2)$$

$D$  is a dummy variable which takes the value 1 if a particular district characteristic is satisfied.

7a. Assume  $D$  is a dummy equaling 1 if district  $d$  has a relatively high literacy rate in year  $t$ . What is the interpretation of the coefficients  $\beta_1$  and  $\beta_2$ . What is the interpretation of  $\beta_2$  if it is not statistically significant? What does  $\beta_1 + \beta_2$  tell us? [3]

7b. Now estimate whether the relationship between branch per capita and finance varies by fraction of marginalized populations (share of SC/STs) in a district using specification (2). Use the variables *high\_scst* for  $D$  and *high\_scst\_branch* for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are *high\_scst\_psb* and *high\_scst\_pvt*. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with a high share of marginalized groups for either bank type? [1]

7c. Repeat 7b) but use the fraction of urban population as the heterogeneity of interest. Use the variables *high\_urb* for  $D$  and *high\_urb\_branch* for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are *high\_urb\_psb* and *high\_urb\_pvt*. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with high urbanization for either bank type? [1]

---

<sup>2</sup> Remember, the covariates are: a) fraction of rural households b) log population c) fraction of SC/STs d) fraction of OBCs e) fraction of Muslims f) fraction female g) fraction employed in manufacturing h) fraction of salaried workers i) fraction employed in the formal sector j) LFP k) fraction unemployed l) fraction employed in agriculture m) fraction of literates and n) fraction of adults educated secondary.

7d. Repeat 7c) but use the fraction of population employed in manufacturing as the heterogeneity of interest. Use the variables `high_mfg` for  $D$  and `high_mfg_branch` for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are `high_mfg_psb` and `high_mfg_pvt`. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with higher manufacturing for either bank type? [1]

7e. Repeat 7c) but use the fraction of adults with secondary education as the heterogeneity of interest. Use the variables `high_edu` for  $D$  and `high_edu_branch` for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are `high_edu_psb` and `high_edu_pvt`. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with higher education for either bank type? [1]

7f. Repeat 7c) but use the fraction of workers in agriculture as the heterogeneity of interest. Use the variables `high_ag` for  $D$  and `high_ag_branch` for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are `high_ag_psb` and `high_ag_pvt`. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with a high share of workers in agriculture for either bank type? [1]

7g. Repeat 7c) but use districts with relatively high per capita consumption as the heterogeneity of interest. This in fact tests whether the finance and consumption relationship weakens or strengthens at relatively high levels of consumption. Use the variables `high_mpce` for  $D$  and `high_mpce_branch` for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are `high_mpce_psb` and `high_mpce_pvt`. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with higher MPCE for either bank type? [1]

7h. Repeat 7c) but use districts with relatively low per capita consumption as the heterogeneity of interest. This in fact tests whether the finance and consumption relationship weakens or strengthens at relatively low levels of consumption. Use the variables `low_mpce` for  $D$  and `low_mpce_branch` for  $D_{dt} \times BranchPC_{dt}$ . The corresponding variables for public and private bank branches are `low_mpce_psb` and `low_mpce_pvt`. Report and interpret the coefficients  $\beta_1$  and  $\beta_2$  for each of the 3 regressions. Is the relationship between branch density and consumption stronger in districts with lower MPCE for either bank type? [1]

## 1.8 A “Naive” DiD Exercise!

We will finish off with a “naive” difference-in-difference exercise. Recall that some districts in 1999 did not have any private bank branches, while most of them in 2011 had private bank branches. Let’s assume (incorrectly!) that the expansion of branches into these districts was uncorrelated with observable characteristics. The treatment intervention therefore is the introduction of private bank branches in districts and we shall compare consumption outcomes across districts before and after the introduction of private bank branches.

8a. What is the average log MPCE across districts which had 0 private bank branches in 1999? These are the “treatment” districts, as they later received private bank branches (at least most of them did!). [0.25]

8b. What is the average log MPCE across districts which had at least 1 private bank branch in 1999? These are the “control” districts, as they already had private bank branches. [0.25]

8c. What is the average log MPCE across “treatment” districts in 2011, conditional on them having at least 1 private bank branch? What is the average log MPCE across “control” districts in 2011, conditional on them having at least 1 private bank branch? [0.25]

8d. Subtract the difference between the mean log MPCE for treatment districts between 2011 and 1999 and report the difference (this is the first difference). Call this difference  $\mu_t$ . Do the same for control districts and call this  $\mu_c$ . [0.25]

8e. Now compute  $\mu_t - \mu_c$ . This forms the DiD coefficient. Interpret this coefficient – how does the introduction of private bank branches affect consumption? [1]

8f. Let’s now do this with covariates and fixed effects. Create a dummy variable  $NoPvt_d$  which equals 1 if district  $d$  had no bank branches in the year 1999 (consider exploring a command called *egen*). Define the dummy  $Post1999_t$  to equal 1 for the years 2004 and 2011. Now estimate the following specification:

$$\ln(MPCE_{dt}) = \alpha_d + \delta_t + \beta NoPvt_d * Post1999_t + \phi X_{dt} + \epsilon_{dt} \quad (3)$$

Include district and year fixed effects as well as covariates discussed before and cluster the

errors using *dist91*. Report the coefficient  $\beta$ . How different it is from the DiD coefficient calculated in 8e? Is it biased upwards or downwards? [2]

## 1.9 Wrapping Up

Based on your empirical findings, write a short paragraph describing the branch density consumption relationship between 1999 and 2011. Mention how the relationship varies across bank ownership, and also district characteristics. Which districts appear to benefit most from an expansion in bank branches? What does this say about the role of finance in development in the Indian context? [5]