README File for:

"The Market for High-Quality Medicine: Retail Chain Entry and Drug Quality in India"

Daniel Bennett and Wesley Yin

February 2018

1. Explanation: our analysis is based on primary data that we collected. To reproduce our analysis, the reader should save the input data files and **replication.do** into the same directory and run the do file. We used Stata version 14 for our analysis. You may need to install the command **psacalc** to compute the proportional selection delta's that appear in some tables. Table 4 requires the command **xtgraph**. The file **replication_bootstrap.do** computes Cameron, Gelbach, and Miller (2008) wild cluster bootstrap standard errors for all relevant estimates in the manuscript and the appendix.
2. Input Data Files: per the journal's instructions, we have saved these files in csv format.
    a. **Final**: this file is the primary regression dataset. It includes observations by audit-pharmacy-round on drug quality and prices. We have merged relevant pharmacy characteristics into the dataset. Data are available for Rounds 1 and 2.
    b. **Consumer**: this file contains responses to the consumer survey, which we use to measure consumer socioeconomic characteristics, perceptions of drug quality, and pharmacy preferences. As the text describes, we sampled "shoppers" who had just exited a sample pharmacy and "non-shoppers" who were in the area but had not just visited a pharmacy. The file contains both types of observations, and the variable **int_loc** distinguishes between these subsamples. Data are available for Rounds 1, 2, and 3.
    c. **Traffic**: this file counts the number of shoppers who visited each pharmacy during two pre-defined one-hour intervals. We have combined the counts from these intervals and computed the natural log to create the variable **lntr**.
    d. **Hyd_weather**: this file contains the daily weather data that are used in Appendix Figure 2.
3. Data Dictionary
    a. **Final**
        1. mkt_id: Market identifier
        2. cst_m: Percent of non-shopping consumers who belong to a scheduled caste/tribe by market round.
        3. edu_m: Mean years of schooling for non-shopping consumers by market-round.
        4. deg_m: Percent of non-shopping consumers who have a college degree by market-round.
        5. tfw_m: Percent of non-shopping consumers who own a vehicle by market-round.
        6. inc_m: Mean monthly income of non-shopping consumers by market-round.
        7. fevr_m: For non-shopping consumers, percent of all household members with a fever in the past week, by market-round.
        8. cold_m: For non-shopping consumers, percent of all household members with a cough or cold in the past week, by market-round.
        9. diar_m: For non-shopping consumers, percent of all household members with diarrhea in the past week, by market-round.
        10. injy_m: For non-shopping consumers, percent of all household members with an injury in the past week, by market-round.
        11. dosage: the dosage per tablet obtained in the pharmacy audit: either 250mg or 500mg.
        12. national: an indicator that the audit drug is a national brand.
        13. nn: an indicator that the audit drug is not a national brand.
        14. ph_clean: the cleanliness of the pharmacy on an increasing 5-point Likert scale.
        15. ph_workers_tot: total number of employees at the firm.
        16. ph_shipments: the frequency of shipments that the firm receives from distributors.
        17. ph_ac: the pharmacy has air conditioning.
        18. l_api: lab report of the active pharmaceutical ingredient in the audit sample.
        19. l_api_pct: lab report of the active pharmaceutical ingredient in the audit sample (absolute deviation from 100 percent).
        20. l_api_pct_cont: lab report of the active pharmaceutical ingredient in the audit sample (percent of the indicated dosage).
        21. l_uniform_absmax: lab report of the uniformity parameter.
        22. l_dissol_min: lab report of the dissolution parameter.
        23. tr_all: total number of observed customers at the pharmacy.
        24. distw: walking distance from the pharmacy to the market center.
        25. ce_mfirms: number of pharmacies in the market based on the pharmacy census.
        26. medplus: an indicator identifying Medplus pharmacies.
        27. disttrt: the distance from control markets to the nearest treatment market.
        28. pass: indicator that the audit sample complies with Indian Pharmacopeia.

29. ip_api_pass: indicator that the audit sample complies with the active pharmaceutical requirement of Indian Pharmacopeia.
30. ip_uniform_pass: indicator that the audit sample complies with the uniformity requirement of Indian Pharmacopeia.
31. ip_dissol_pass: indicator that the audit sample complies with the dissolution requirement of Indian Pharmacopeia.
32. tte: days from the audit until the sample's expiration date.
33. he: indicator that the auditor was high-SES.
34. pe: the average years of schooling of shoppers at the sample pharmacy.
35. retpctb: the percent of the shop dedicated to medicine at baseline.
36. signsb: the shop's number of signs at baseline.
37. mpaspct: the percent of samples that are IP compliant by brand at baseline.
38. qualm: average perceived quality of nearby pharmacies, by market.

b. Traffic

1. mkt_id: Market identifier
2. cst_m: Percent of non-shopping consumers who belong to a scheduled caste/tribe by market round.
3. edu_m: Mean years of schooling for non-shopping consumers by market-round.
4. deg_m: Percent of non-shopping consumers who have a college degree by market-round.
5. tfw_m: Percent of non-shopping consumers who own a vehicle by market-round.
6. inc_m: Mean monthly income of non-shopping consumers by market-round.
7. fevr_m: For non-shopping consumers, percent of all household members with a fever in the past week, by market-round.
8. cold_m: For non-shopping consumers, percent of all household members with a cough or cold in the past week, by market-round.
9. diar_m: For non-shopping consumers, percent of all household members with diarrhea in the past week, by market-round.
10. injy_m: For non-shopping consumers, percent of all household members with an injury in the past week, by market-round.
11. lntr: log customer traffic by pharmacy-round. This variable is total number of customers who visited the pharmacy within two pre-defined one-hour windows.

c. Census

1. mkt_id: Market identifiers
2. present: An indicator that the firm still exists in the market.
3. cst_m: Percent of non-shopping consumers who belong to a scheduled caste/tribe by market round.
4. edu_m: Mean years of schooling for non-shopping consumers by market-round.
5. deg_m: Percent of non-shopping consumers who have a college degree by market-round.
6. tfw_m: Percent of non-shopping consumers who own a vehicle by market-round.
7. inc_m: Mean monthly income of non-shopping consumers by market-round.
8. fevr_m: For non-shopping consumers, percent of all household members with a fever in the past week, by market-round.
9. cold_m: For non-shopping consumers, percent of all household members with a cough or cold in the past week, by market-round.
10. diar_m: For non-shopping consumers, percent of all household members with diarrhea in the past week, by market-round.
11. injy_m: For non-shopping consumers, percent of all household members with an injury in the past week, by market-round.

d. Consumer

1. mkt_id: Market identifiers
2. hsize: Household size
3. int_loc: Interview location. 1 indicates that the person just exited a sample pharmacy; 2 indicates that the person was in nearby but did not exit a sample pharmacy.
4. edu_years: Consumer's years of schooling.
5. tfw: an indicator that the consumer owns a two-wheel or four-wheel vehicle.
6. ln_incr_usd: Log income in 2010 USD.
7. scst: an indicator that the consumer belongs to a scheduled caste or scheduled tribe.
8. qual_nearpharms: Consumer's perception of the quality of medicine sold by nearby pharmacies on a four-point Likert scale.
9. qual_natdrugs: Consumer's perception of the quality of national-brand drugs on a four-point Likert scale.

10. qual_locdrugs: Consumer's perception of the quality of local-brand drugs on a four-point Likert scale.
11. round: Survey round. 1 indicates the baseline, 2 indicates the first follow-up (one year later) and 3 indicates the second follow-up (two years later).
12. mp: an indicator that the consumer visited MedPlus (the chain pharmacy).
13. qs: an indicator that the consumer believes that drug quality is important.
14. cs: an indicator that the consumer believes that pharmacy convenience is important.
15. fs: an indicator that the consumer believes that pharmacy familiarity is important.
16. cst_m: Percent of non-shopping consumers who belong to a scheduled caste/tribe by market round.
17. edu_m: Mean years of schooling for non-shopping consumers by market-round.
18. deg_m: Percent of non-shopping consumers who have a college degree by market-round.
19. tfw_m: Percent of non-shopping consumers who own a vehicle by market-round.
20. inc_m: Mean monthly income of non-shopping consumers by market-round.
21. fevr_m: For non-shopping consumers, percent of all household members with a fever in the past week, by market-round.
22. cold_m: For non-shopping consumers, percent of all household members with a cough or cold in the past week, by market-round.
23. diar_m: For non-shopping consumers, percent of all household members with diarrhea in the past week, by market-round.
24. injy_m: For non-shopping consumers, percent of all household members with an injury in the past week, by market-round.

e. **Hyd_weather**

1. QCP Temp: temperature in Celsius.
2. RHx: Relative humidity.