Impacts of an innovative audit strategy
Pre-Analysis Plan

Nada Eissa, Francois Gerard, and Andrew Zeitlin

May 1, 2019

1 Study overview

A central question for a country's economic development is how to improve the state's ability to raise revenues domestically in the presence of widespread tax evasion (Besley and Persson, 2009). In this context, the Value Added Tax (VAT) has become one of the most important instruments of revenue mobilization in the developing world. In principle, a VAT has several advantages in terms of tax compliance compared to other tax instruments, as it incentivizes accurate reporting of sales and purchases. Firms are essentially taxed on the difference between their sales and the cost of their inputs, and they must report both to the tax authority and keep a paper trail of the transactions (i.e., receipts) in their books. To minimize tax liability, firms would like to under-report sales, but to over-report inputs. This asymmetry should limit the room for "collusive evasion" in the case of business-to-business transactions, i.e., collusion between firms in misreporting transaction values. The existence of a paper trail should also deter "unilateral evasion," i.e., firms unilaterally misreporting transaction values, as the tax authority could crosscheck the values reported by a firm with records from third parties (i.e., suppliers and clients).

Pomeranz (2015) provided empirical evidence showing the importance of these properties of the VAT for tax compliance. Yet, in practice, limited enforcement capacity may prevent tax authorities from efficiently cross-checking what firms report against records from third parties, and thus limit unilateral evasion. To address this concern, several countries around the world, including Rwanda, have mandated the use of Electronic Billing Machines (EBMs), which generate a unique "hard-to-tamper-with" receipt for each transaction and send information on each receipt to the tax authority on a regular basis (e.g., in real time or daily). This technology limits the ability of firms to unilaterally misreport a transaction once a receipt for that transaction has been issued, which can be particularly effective for business-to-business transactions as firms have an incentive to ask for receipts for their input transactions (to decrease their tax liability). Eissa and Zeitlin (2015) show the roll-out of EBMs improved compliance in Rwanda.

However, final consumers do not have similar incentives to ask for receipts, so firms can more easily under-report sales to final consumers, an issue that can persist with EBMs if firms do not actually issue receipts to final consumers. Eissa and Zeitlin (2015) also document that firms failed to issue receipts in a large share of a sample of retail transactions made after the EBM roll-out in Rwanda. One potential policy response is to incentivize consumers to ask for receipts, e.g., with lotteries and tax rebates based on consumers receipts. Naritomi (AER forthcoming) provides evidence that such policies can effectively reduce the under-reporting of sales to final consumers. However, these policies can be quite expensive, as incentives have to be paid to all consumers.

Alternatively, governments can address this "last-mile" problem of the VAT with audits and related enforcement strategies at the retail level. In this light, Electronic Billing Machines can be seen as a technology that changes the cost of such enforcement strategies: by requiring firms to keep electronic accounts, they make it readily visible whether a firm is planning to declare revenue associated with a given transaction. This suggests that tax authorities can combat tax evasion by auditing firms' use of EBMs through unannounced visits, in which staff of the tax authority stop consumers leaving a store, check whether they were issued a valid EBM receipt, and apply penalties in case of non-compliance.

This study investigates the impact of such an enforcement strategy through a randomized control trial conducted in collaboration with the Rwandan Revenue Authority (RRA) with the aim of answering three questions:

1. What are the impacts of such an enforcement strategy on retailers' issuance of receipts, reporting of sales, and tax liability?

2. What is the impact of the associated increase in retailers' tax compliance on prices, i.e., what is the incidence on retailers vs. consumers?

3. Do firms' responses to audit risk depend on the audit risk of their competitors, i.e., does the impact of the treatment vary with the rate of treatment *saturation* among competitors?

The answer to the first question will allow us to evaluate the effectiveness of a policy to address the last-mile problem that is complementary to the EBM technology. The second question is important conceptually to understand the effects of enforcement policies and the possible public support for these policies. There is very little evidence on the incidence of tax evasion (Kopczuk et al., 2016). To the extent that the evasion rents are broadly shared with consumers (through lower retail prices), policies aiming to reduce tax evasion may face little support in the public as the cost of such policies (higher retail prices) may be more salient than their possible benefits (possible gains from increased government revenue). The answer to the third question is important to shed light on the mechanisms behind our answer to the second question. The ability of retailers to pass the costs associated with stronger tax enforcement onto prices likely depends on the competitive structure of the market in which they operate (e.g., monopoly, monopolistic competition, perfect competition) and on whether their competitors are also subject to the same costs. This is important because experimental enforcement interventions, such as the one we study, often focus on a subset of firms (e.g., often treat competing firms differently) but hope to inform the effect that such interventions could have when applied more broadly (e.g., treating all competing firms similarly).

In practice, this project evaluates the impact of an RRA initiative to send auditors to make surprise visits to retail firms in Kigali over a 10-month period and to observe whether ordinary consumers of those retailers are receiving receipts for their transaction. The evaluation combines innovations in measurement – we combine VAT declarations with both real-time data on the universe of transactions reported through EBMs and receipt outcomes from natural transactions conducted by "mystery shoppers" sent to control and treatment firms throughout the study period for research purposes only – with two dimensions of experimental variation: at the firm level, firms are assigned to (zero or) variable frequencies of audit, and across firms, we vary the saturation of the treatment among economic competitors to study strategic complementarities.

2 Data

The study makes use of four types of data.

1. *Enumerator baseline visits to shops.* At the outset of the study, firms meeting enrollment criteria on the basis of administrative data (see Section 3.1 for details) were visited by enumerators from the survey firm. Enumerators collected survey data from firm operators, collected information about items available for sale in the store, and categorized the firm in terms of four-digit ISIC categories.

2. *VAT declarations.* The study makes use of VAT declarations from the first quarter of 2016 (more than a year prior to the start of the intervention) through the fourth quarter of 2018 (one quarter after completion of the intervention). Firms in the study sample comprise a mix of those who file on a monthly basis and those who file on a quarterly basis. All VAT declaration information is aggregated to the quarterly level for purposes of comparability.

3. *Mystery-shopper visits.* To study the effects of study interventions on retailer behavior as experienced by consumers, the study employed a team of *mystery shoppers*, who conducted natural transactions in study firms. Two "baseline" transactions were undertaken prior to intervention, over the period from April–May of 2017, and eight transactions were undertaken after intervention, over the period of December 2017–August 2018.

   These transactions involved baskets of goods (mostly non-perishable items) that ranged in (sticker-price) value but were typically valued at between RWF 6,000 and 15,000. For any given firm-transaction, mystery shoppers were provided with a "shopping list" that detailed one or more items that they were requested to purchase, as well as a ranked list of backup purchase goods, in case the requested good(s) could not be found. In subsequent analysis, we use only the *intended* attributes of the basket—i.e., attributes of the originally requested items, but not of replacements—to guard against the possibility that the likelihood of finding the first-priority item(s) may itself be affected by the treatments under study.

   As described in Section 3.3 below, in post-intervention rounds 3–8 of the mystery-shopper visits we randomly varied the value of these baskets in order to test for purchase-value effects, and their interactions with the intervention, on the likelihood that an EBM receipt was issued. In addition to the intended basket of goods (the 'shopping list') and the content of the purchased basket, data collected about mystery-shopper visits include (i) an indicator for whether an EBM receipt was issued to the customer; (ii) the price paid for the basket of goods; and (iii) the price appearing on the EBM receipt, if one was issued.

4. *Electronic Billing Machine data.* The study further makes use of data on the universe of transactions for which an EBM receipt was issued over the period from January 2016 to December 2018. Meta-data from these transactions are used to produce measures of EBM-reported transaction counts and values at the transaction and daily level. In addition, attributes of products sold at the line-item level are used to define measures of the degree of economic 'connectedness' among firms (see Section 3.2), and may be used in exploratory analysis to examine impacts of audit risk on prices charged conditional on issuing an EBM receipt.

3 Experimental design

3.1 Enrollment criteria

631 VAT-registered firms were identified for inclusion in the study. The threshold for VAT eligibility is an annual turnover in excess of RWF 20 million per year, or RWF 5 million per quarter. At the time of the study, all VAT-registered firms had been issued with EBMs.

Starting from the administrative universe of VAT-registered firms, inclusion criteria were as follows:

1. Firms with VAT filings dating back to at least the first quarter of 2016, to avoid potentially high business failure rates among new, small businesses that might create high rates of attrition;

2. Single-location firms based in the city of Kigali, to avoid the risk that firms might shift business across locations in response to an intervention;

3. Firms in a relevant economic sector, which excludes wholesale, repair, and cleaning activities, as well as a small number of miscellaneous or unidentifiable types;

4. Firms with at least 5 percent of EBM-receipted transactions under RWF 6,000 in value, since this was the intended value of our own 'mystery shopper' measurement activities; and

5. Firms that could be located by a team of enumerators, guided by local ('zone'-level) RRA officers.

3.2 Assignment to treatment arms

Randomization was done following the baseline mystery shopper visits. The process of assigning firms to treatment arms was designed to generate the same treatment probabilities for all firms, while inducing patterns of correlation across the network of firms that would facilitate an analysis of spillovers, as described in details below.

The compound lottery that this assignment mechanism produced assigned firms to 'treatment' with probability 0.6 and to 'control' with probability 0.4.

- 375 firms were randomly assigned to the audit treatment. These were then assigned (by simple randomization) to two groups, characterized by varying frequency of audits:
  - 284 firms were assigned to receive an expected number of 2 visits per month;
  - 91 firms were assigned to receive an expected number of 4 visits per month;
- 256 firms were randomly assigned to the control group. These were then assigned (again by simple randomization) to two groups:
  - 182 received a pre-intervention visit by the RRA; and
  - 74 firms did not receive any pre-intervention visit at all.

Treatment and control firms were visited by an RRA auditor prior to the start of the intervention, between September 2017 and November 2017, and handed a formal letter from the RRA. The letter reminded firms about their rights and responsibilities as VAT-paying firms, especially regarding the use of EBMs, and usual enforcement strategies. The 375 treated firms were also informed about the new enforcement strategy and the expected frequency of visits per month that they would be subject to in subsequent months. Among control firms, 74 firms did not receive any pre-intervention visit by the RRA (and thus did not receive any letter). This will allow us to separate the effect of the RRA visit/letter and the effect of the enforcement intervention (EBM use audits) itself.

In practice, the assignment mechanism involved a two-stage saturation design that takes advantage of knowledge of firms' connections in product and geographic space and anticipates an Athey et al. (2018) approach to inference on a connected network of firms. Our assignment strategy is related to recent work on saturation designs in settings where interference can be assumed not to occur across distinct clusters in a study (Baird et al., 2018). Athey et al. (2018) provide one solution to the problem of inference in settings where that assumption cannot be maintained; we describe how we plan to deploy that approach to estimation and inference in Section 4.2.7. Our design anticipates use of the inferential approach of Athey et al. (2018) and considers designs that will be well powered under that approach to testing interference hypotheses.

Our assignment mechanism involved the following steps:

1. We partition the firms in our sample into those that, for testing interference hypotheses, will be *focal* firms (Athey et al., 2018), and those that, for testing interference hypotheses, will be *variant* firms. This partition is chosen to maximize the number of focal nodes subject to the constraint that no two focal nodes are adjacent to one another in the network.

2. Focal nodes are assigned to treatments at random, with probabilities that match the proportions in the study design.

3. Focal nodes are assigned saturation rates that are either *high* (in which case variant nodes that are their direct neighbors are treated with probability 0.95) or *low* (in which case variant nodes that are their neighbors are treated with probability 0.19). This yields the same probability of treatment among variant nodes as among the focal nodes.

4. Variant nodes are assigned to either treatment or control with probabilities that are the average of the saturation rates of all focal nodes to which they are directly connected.

5. Finally, the high- and low-frequency treatment statuses are assigned to treated variant nodes by simple randomization, and the letter and no-letter statuses are assigned to control variant nodes by simple randomization.

The above process yields the same assignment probabilities for both focal and variant nodes, while increasing the variance of saturation rates surrounding focal nodes for the testing of interference hypotheses.

3.3 Experimental variation in purchase amounts

Outcomes measured through mystery-shopper transactions are necessarily constrained to represent only transactions of a value that the study measurement budget could allow. But if patterns of EBM utilization and VAT compliance are different for larger transactions, then study findings may be relevant only locally, for transactions in the studied range. To address this concern, we cross-randomized experimental variation in the value of goods assigned for purchase in specific mystery-shopper visits. In particular, a random subset of mystery-shopper visits were assigned to 'high-value' baskets, which on average cost more than 2.5 times as much as the standard baskets. This randomization was done at the firm-by-visit-round level, for visits in post-intervention rounds 3 through 8 only. In each of those rounds, 25 percent of firms were assigned to high-value purchases. This implies that all firms have a mixture of both high- and low-value baskets across the 8 rounds of post-intervention mystery-shopper visits.

4 Analytical specifications

4.1 Primary hypotheses

4.1.1 Audit treatment impacts on EBM utilization in mystery-shopper visits.

This outcome is defined as a binary indicator, taking on a value of one if an EBM receipt was issued, and zero if it was not, for all mystery-shopper visits in which goods were successfully purchased.

We analyze these data with a linear probability model of the form

$$receipt_{it} = \tau_1 T_i + \tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \rho \cdot receipt_{i0} + \phi_t + e_{it} \tag{1}$$

where T_i is an indicator for whether firm i is assigned to the audit treatment (pooling both audit frequencies; henceforth we refer to this as the 'pooled audit treatment').

As described in Section 3.2, our assignment mechanism may induce correlations in treatment status among firms that are connected to one another on the basis of their geographic and product-market proximity. We define the adjacency matrix G such that its entry ij is an indicator taking a value of one if firms i and j are connected, and zero otherwise; we adopt the convention that $G_{ii} = 0$. Then $\sum_j G_{ij}$ is the 'degree' of firm i, and $\sum_j T_j G_{ij}$ represents the number of treated firms to which firm i is connected. In the presence of spillovers, the coefficient τ_1 would suffer from omitted variable bias in a restricted model that omits τ_2, so we include that term throughout. In our secondary analysis of interference (Section 4.2.7 below), we will be interested in testing hypotheses about interference directly; for now, we note that in this framework, estimates of τ_1 derived from equation (1) represent the treatment effect of audits at average levels of saturation.

Regarding the additional controls in this specification, ϕ_t is a round-specific intercept and $receipt_{i0}$ is a vector that comprises the values for each baseline mystery shopper visit of (i) an indicator for whether this variable is non-missing (i.e., if the shoppers could not conduct a transaction at that firm in that round), and (ii) an indicator for whether a receipt was issued in that transaction, which is set to zero if no transaction occurred.

Our test of the hypothesis will be conducted by randomization inference, using the set of feasible randomizations and the *studentized* regression coefficient τ_1, following Chung and Romano (2013) and DiCiccio and Romano (2017).

4.1.2 Audit treatment impacts on prices paid for baskets of goods in EBM visits.

Extending the analysis of equation (1), we model the log price paid for a given basket of goods, purchased in round t at firm i as a function of (pooled) treatment, local treatment saturation, log prices paid in baseline mystery shopper visits, round indicators, and a vector of basket characteristics:

$$\ln price_{it} = \tau_1 T_i + +\tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_{it} + \rho_0 \ln price_{i0} + \phi_t + e_{it}. \tag{2}$$

In this specification, baseline prices paid $\ln price_{i0}$ are defined analogously to baseline receipts in equation (1): this is a vector including an indicator for whether a transaction occurred in each baseline mystery shopper visit, and the log price of the transaction in each baseline visit. We set this control for log price in baseline mystery shopper visits to an arbitrary value (zero) for visits in which no transaction occurred to avoid missing post-intervention outcomes.

The set of basket attributes, x_{it}, will be chosen as those attributes with non-zero coefficients in a lasso regression run on control firms only, in the spirit of Belloni et al. (2014). The set of characteristics includes indicators for the product, brand, and quantity of goods assigned for purchase. The lasso regularization parameter, λ, will be chosen to minimize the out-of-sample mean squared error as in Chernozhukov et al. (2018); Jones et al. (2018). Note that the resulting attribute vector, x_{it}, is time varying because baskets assigned to purchases differ over time even within a given firm. We use attributes of the *assigned* rather than the realized items purchased to avoid potential endogeneity of baskets, as described in Section 3.3.

As in the preceding hypothesis, we use the randomization distribution of the studentized coefficient τ_1 as our primary test of this hypothesis.

4.1.3 Audit treatment impacts on total value of sales for goods subject to VAT for which EBM receipts were issued (using EBM data)

Our primary test for changes in the distribution of EBM quarterly totals is a Kolmogorov-Smirnov (henceforth KS) test statistic for the difference in distributions across the (pooled) treatment and (pooled) control populations.

This KS statistic is defined as

$$KS = \sup_{\ddot{y}} \left| \hat{F}_1(\ddot{y}) - \hat{F}_0(\ddot{y}) \right|, \tag{3}$$

where $\hat{F}_T(\ddot{y})$ represents the empirical cumulative distribution function for (residual) daily total EBM-receipted revenues, $\ddot{y}$, under pooled control and treatment statuses $T \in \{0, 1\}$ respectively.

To guard against chance imbalance on observables and the possibility that local correlations in treatment assignment create a source of omitted variable bias, we do not use the *raw* daily total EBM revenues in equation (3). Rather, we use residuals, $\ddot{y}$ from a regression of the form

$$y_{it} = \tau_1 T_i + \tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_i + \rho_0 y_{i0} + \phi_t + e_{it}. \tag{4}$$

As in previous specifications, the baseline value y_{i0} is a vector comprising (i) an indicator for whether receipt totals are nonzero and (ii) values of receipt totals, in pre-intervention periods. Covariates x_i are firm-level variables chosen from the baseline survey, pre-trial EBM receipt submissions, and pre-trial VAT declarations selected via post-double lasso with regularization parameter chosen to minimize cross-validated out-of-sample mean squared error, as in Chernozhukov et al. (2018).

As an input into the KS test, we then construct residuals $\ddot{y}$ for post-intervention periods as

$$\ddot{y}_{it} = y_{it} - \left(\tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_i + \rho_0 y_{i0} + \phi_t \right). \tag{5}$$

Note that we omit the predicted impact of the treatment for firm i, $\tau_1 T_i$, from the construction of these residuals.

For this and all subsequent outcomes in which the primary test of the 'sharp null' of no treatment effects is based on a KS statistic, we will also estimate secondary, regression-based tests using equation (4) directly. We estimate this secondary specification for the following functional forms for the outcome, y_{it}:

- A binary indicator of whether the outcome is strictly greater than zero;
- Median value of receipts (implemented as a quantile regression analog of equation (4))
- Mean value of receipts, in levels.

In each case, we use the randomization distribution of T_i to conduct inference with regard to equation (4). Our test statistic of interest is the studentized coefficient τ_1.

4.1.4 Audit treatment impacts on total value of sales subject to VAT, as reported in VAT declarations

Our primary test for impacts on total VAT-liable sales in quarterly VAT declarations uses a KS statistic, as in equation (3) to test for changes in distributions between the (pooled) treatment and (control) populations. We again residualize these outcomes using the same procedure employed in the preceding section. While the KS statistic provides our omnibus test against the sharp null of no treatment effect for any unit, we follow the analysis of outcome 4.1.3 to provide a parametric test for impacts on whether a firm declares strictly positive VAT-liable sales; median VAT-liable sales declared; and mean VAT-liable sales declared.

4.2 Secondary hypotheses

4.2.1 Audit treatment impacts on total costs subject to VAT, total VAT liabilities, total sales not subject to VAT, and total sales, as reported in VAT declarations

Here, we follow the structure of analysis for outcome 4.1.3 to test for impacts of the (pooled) treatment on total costs subject to VAT, total VAT liabilities, total sales not subject to VAT, and total sales, as reported in quarterly VAT declarations. Each of those analyses provides a KS test using the residualized outcome as its main specification, followed by tests for impacts on a binary indicator of strictly positive values, mean values in levels, and median values.

4.2.2 Audit treatment impacts on total number of transactions with EBM receipts, total number of transactions with EBM receipt that include at least one VAT-liable item, and total value of sales for which EBM receipts were issued (including VAT exempt items).

Here, we test for impacts of the (pooled) treatment on total number of transactions with EBM receipt, total number of transactions with EBM receipt that include at least one VAT-liable item, and total value of sales for which EBM receipts were issued (including VAT exempt items), as measured in the quarterly EBM data. For each outcome, we follow the analysis in Section 4.1.3 by using a KS test as our primary test for differences in distributions, followed by regression-based tests for differences in a binary indicator for values of the outcome strictly greater than zero, impacts on the mean of the outcome, and impacts on the (conditional) median of the outcome.

4.2.3 In a model where treatment effects are not constant, firms' response to the pooled treatment are non-zero in at least one period

To allow for the possibility that firms' responses to the audit treatment are not constant over time – and in particular they may be non-zero in only a subset of periods – we modify the linear models used in specifications for receipt probabilities (equation 1), log prices (equation 2), and average effects on EBM transaction values and total value of sales subject to VAT (equation 4) to allow for time-varying treatment effects, as in

$$y_{it} = \tau_1^t T_i + \tau_2^t \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_{it} + \rho y_{i0} + \phi_t + e_{it}, \tag{6}$$

where the coefficient τ_t now takes on different values for each time period t. We define time periods over which treatment effects may vary at the quarterly level for all outcomes. Covariates x_{it} and ANCOVA controls y_{i0} are defined as in the corresponding constant-effects specifications above. We use the randomization inference distribution of the *minimum* of the studentized coefficients τ_1^t to test the null of zero effects in all periods.

This specification also allows us to construct an analogous test for whether the spillover effect, τ_2^t, is non-zero in any period, and to test for pooling in the direct and spillover effects.

4.2.4 Responses to the audit treatment are sensitive to audit intensity

To evaluate the impact of audit intensity, we augment models for EBM utilization (equation 1), log prices paid (equation 2), and average effects on EBM transaction values and total value of sales subject to VAT (equation 4) to include an indicator for whether the firm was in the high-intensity audit treatment T_i^H. These specifications now take the general form:

$$y_{it} = \tau_1 T_i + \tau_1^H T_i^H + \tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_{it} + \rho y_{i0} + \phi_t + e_{it}, \tag{7}$$

such that the parameter τ^H denotes the *differential* impact of the high-intensity audit arm. Covariates x_{it} and ANCOVA controls y_{i0} are defined as in the corresponding specifications in equations (1), (2), and (4). Randomization inference is undertaken with respect to the set of possible draws that hold constant the pooled audit indicator, T_i, and permute only the set of firms allocated to high-intensity audits. The test statistic of interest in this randomization inference exercise is the studentized coefficient τ^H.

4.2.5 Control firms respond to contact letters from the RRA

We test whether control firms' likelihood of issuing EBM receipts is impacted by receiving a letter informing them of their obligations under the EBM program (but with no audit risk described) by extending the specification in equation (1) to include an indicator, L_i, for the 'letter' treatment.

We therefore estimate:

$$receipt_{it} = \tau_1 T_i + \lambda L_i + +\tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} \rho receipt_{i0} + \phi_t + e_{it}. \tag{8}$$

Within this framework, we test (a) whether the letter has an impact relative to 'pure' control, and (b) what is the incremental effect of the audit treatment, over and above the letter. In each case, we use studentized regression coefficients as the test statistic of interest and conduct randomization inference with respect only to permutations of the pairs of arms being compared.

4.2.6 Estimates of treatment effects are affected by the value of the basket of goods in mystery shopper visits

One concern for the generalizability of our estimates of the treatment effect on EBM utilization rates and prices in mystery shopper visits is that they are limited to the types of transactions that our mystery shoppers' budgets allow. For instance, if patterns of EBM utilization are different for larger transactions, then effects found in our study will be relevant only locally, for transactions in the studied range. To evaluate this concern, we exploit experimental variation in the value of the baskets of goods as described in section 3.3.

We test for a potentially different impact of higher-value baskets on EBM utilization rates and prices with a modified version of the corresponding specifications in equations (1) and (2). To do so, we define an indicator for high-cost baskets, C_{it}, and estimate

$$y_{it} = \tau_1 T_i + \gamma_0 C_{it} + \gamma_1 C_{it} T_i + \tau_2 \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_{it} + \rho y_{i0} + \phi_t + e_{it}. \tag{9}$$

for post-intervention outcomes y_{it} defined variously as an indicator for whether a receipt was issued or the log price of the transaction. Randomization inference with respect to assignment to high-cost baskets (C_{it}) allows testing for the effect of basket value on the likelihood that an EBM receipt is issued in general, and on the effectiveness of the intervention in particular.

4.2.7 Treatment interference

We are interested in an empirical model that allows for interference between units, i.e., we are interested in allowing differential spillovers from treatment onto both untreated and treated units. While spillovers onto untreated units have clear implications for the returns to this intervention, in a cost-benefit sense, interference that affects other firms' *response* to treatment can have further implications for the optimal targeting of interventions and for equilibrium effects of audit policies.

To study interference, we define the graph G whose entry (i, j) denotes a connection between firm i and j, and by convention we define $G_{i,i} = 0$. Our measure of the connectedness of any pair of firms is given by the application of two criteria: we consider two firms to be connected if (i) they are located in the same geographic 'sector' of Kigali; and (ii) their pre-experimental EBM receipts reveal that they sell products in common with one another.

We focus on testing for interference on two outcomes in particular: the likelihood of issuing an EBM receipt in mystery-shopper visits, and the log price paid in those visits. To do so, we estimate the following model for outcomes of firm i at time t:

$$y_{it} = \tau_1 T_i + \tau_2 \sum_j T_j G_{ij} + \tau_3 T_i \sum_j T_j G_{ij} + \delta \sum_j G_{ij} + \beta x_{it} + \rho_1 y_{i0} + \phi_t + u_{it}, \tag{10}$$

where covariates x_{it} and ANCOVA controls y_{i0} are defined as above for each outcome. Notice that although the degree of a firm, $d_i \equiv \sum_j G_{ij}$, is potentially correlated with unobserved factors u_{it}, random assignment of peer firms to treatments T_j gives us that $\mathrm{E}[u_{it}|d_i, \sum_j T_j G_{ij}] = \mathrm{E}[u_{it}|d_i]$. In other words, the count of treated neighbors is uncorrelated with the error term, conditional on the number of neighbors.[1] Here, parameter τ_2 can be interpreted as the impact of treated neighbors on the behavior of untreated firms, while τ_3 characterizes the moderating effect of treated neighbors on the treatment *response* of the firm in question. Parameter τ_1 can be interpreted as the effect of treatment on a firm with no treated neighbors.

Inference about the 'sharp null' of no interference, which implies $\tau_2, \tau_3 = 0$, should be conducted *without* imposing the additional restriction that $\tau_1 = 0$. To do so we follow the strategy proposed by Athey et al. (2018): we restrict attention to a subset of 'focal nodes', and we consider permutations of the treatment assignment vector that were allowable under the original randomization strategy and that *do not alter* the treatment status of these focal nodes themselves, but only vary the saturation rates of their neighborhoods. The choice of focal nodes was undertaken at the time of randomization to balance goals of (a) a reasonably large sample of focal nodes, and (b) a selection of focal nodes on the graph G that allows us to generate high amount of variance in the saturation rates of their variant-node neighbors.

[1] This issue would not arise for a linear-in-means model, which down-weighted the influence of treatment for any firm that is highly connected in our network. Because our sample graph, G, is only a sample of the true competitive network that surrounds any given firm, however, this specification may introduce measurement error, and we prefer the linear-in-totals specification set forward here.

4.2.8 Audit treatment impacts on prices reported on EBM receipts for similar baskets of goods as those purchased in mystery shopper visits.

The overall treatment effect on the price paid in mystery shopper visits can come from three effects: an effect on the price paid when no EBM receipt is issued, an effect on the price paid when an EBM receipt is issued, and a compliance effect, i.e., a change from the price paid when no EBM receipt is issued to the price paid when an EBM receipt is issued. The overall effect will be estimated through the specification in equation (2). The compliance effect will be estimated through the specification in equation (1). The mystery shopper visits may not record prices for both transactions with and without an EBM receipt issued. However, we will use the line-item information in the universe of the EBM transaction data to construct for each firm a variable capturing the price of the baskets of goods purchased in mystery shopper visits when an EBM receipt is issued. Concretely, we will flag for each firm any EBM receipt involving some of the goods included in the baskets of goods purchased in mystery shopper visits in that firm and extract the price of these goods on the receipt. It is not clear yet whether this will be feasible in practice,[2] but it would allow us to estimate an effect on the price paid when an EBM receipt is issued using a specification akin to the one in equation (2). Combining estimates of the overall price effect, of the effect on the price paid when an EBM receipt is issued, and of the compliance effect, we would then be able to back out an estimate of the effect on the price paid when no EBM receipt is issued.

Note that we could follow a similar strategy and study changes in the price of any good that we can follow over time in EBM receipts, irrespective of whether or not it is part of the baskets of goods purchased in mystery shopper visits.

4.2.9 Subgroup analyses

We plan for three types of subgroup analyses.

1. A subgroup analysis that describes heterogeneity in treatment effects of the primary treatment by firm's sector of economic activity, defined as the two-digit ISIC code, coded at baseline.
2. A subgroup analysis that allows the effect of audits to depend on the number of businesses in the same (geographic) sector and in the same two-digit ISIC code–thereby allowing the degree of competition from non-VAT-liable firms to moderate the effect of treatment; and
3. A subgroup analysis that uses the approach of Athey and Imbens (2016); Wager and Athey (2018) to estimate subgroup heterogeneity and characterize optimal targeting.

[2] It requires some customers to be issued an EBM receipt for transactions including those goods on a regular basis; it also requires us to be able to match the description of the goods on EBM receipt to the goods included in the mystery shopper baskets.

References

Athey, Susan and Guido Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 2016, *113* (27), 7353–7360.

_ **, Dean Eckles, and Guido W. Imbens**, "Exact p-Values for Network Interference," *Journal of the American Statistical Association*, 2018, *113* (521), 230–240.

Baird, Sarah, J Aislinn Bohren, Craig McIntosh, and Berk Özler, "Optimal design of experiments in the presence of interference," *Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "High-dimensional methods and inference on structural and treatment effects," *The Journal of Economic Perspectives*, 2014, *28* (2), 29–50.

Besley, Timothy and Torsten Persson, "The origins of state capacity: Property rights, taxation, and politics," *American Economic Review*, 2009, *99* (4), 1218–1244.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 2018, *21* (1), C1–C68.

DiCiccio, Cyrus J and Joseph P Romano, "Robust permutation tests for correlation and regression coefficients," *Journal of the American Statistical Association*, 2017, *112* (519), 1211–1220.

Eissa, Nada and Andrew Zeitlin, "The incidence and impact of Electronic Billing Machines for VAT in Rwanda," International Growth Centre, Report submitted to the Rwanda Revenue Authority 2015.

EunYi Chung and Joseph P Romano, "Exact and asymptotically robust permutation tests," *The Annals of Statistics*, 2013, *41* (2), 488–507.

Jones, Damon, David Molitor, and Julian Reif, "What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study," NBER Working Paper No. 24229 2018.

Kopczuk, Wojciech, Justin Marion, Erich Muehlegger, and Joel Slemrod, "Does tax-collection invariance hold? Evasion and the pass-through of state diesel taxes," *American Economic Review: Economic Policy*, 2016, *8* (2).

Naritomi, Joana, "Consumers as tax auditors," *American Economic Review*, forthcoming.

Pomeranz, Dina, "No taxation without information: Deterrece and self-enforcement in the Value Added Tax," *American Economic Review*, 2015, *105* (8), 2539–2569.

Wager, Stefan and Susan Athey, "Estimation and Inference of Heterogenous Treatment Effects using Random Forests," *Journal of American Statistical Association*, 2018, *113* (523).