Relational Frictions Along the Supply Chain: Evidence from Social Commerce Among Senegalese Traders

Pre-Analysis Plan

Deivy Houeix & Edward Wiles

February 2024

1 Introduction

1.1 Purpose of this PAP

This PAP serves as a verifiable record of our intentions for this research project before conducting data analysis. The main goals for this PAP are to register the following four items according to our *ex ante* thinking:

1. The primary hypotheses;

2. The primary outcomes and how we intend to measure and define them;

3. The primary dimensions along which we plan to conduct heterogeneity analysis;

4. The primary method of analysis.

   In line with Duflo, Banerjee, Finkelstein, Katz, Olken, and Sautmann (2020), we view the eventual research paper as a distinct entity from a document that would populate exactly the analysis described in this PAP. We expect to learn many things over the course of data collection and analysis: in the interests of scientific advancement, we intend for our eventual research paper(s) to take these into account.

1.2 Study Background and Goals

With the rise of smartphone ownership, social media, and mobile money, there has been a surge in "social commerce". This growing phenomenon is especially pertinent for small and medium enterprises (SMEs) seeking foreign market access, historically hindered by search and trust frictions. We study how these recent digital advancements might reduce frictions in searching and contracting along international supply chains. In particular, we design an RCT that leverages key features of social media platforms to vary such frictions in relationships between merchants in Senegal and exporters in Turkey. We designed the experiment with a view to connecting the results to a quantitative trade model to quantify the frictions and evaluate counterfactuals.

1.3 IRB Approval

This study received IRB approval from the MIT Committee on the Use of Humans as Experimental Subjects (COUHES) with protocol #2305000998.

2 Sample

The study consists of two main types of participants: merchants and suppliers. All randomised interventions are randomised across merchants.

2.1 Merchants

The primary participants in the study are merchants in the ready-to-wear garments industry in Senegal. These merchants are of two main types: "offline" and "online" merchants. The first are merchants with physical establishments in Dakar where the second are merchants without physical establishments that sell exclusively through social media. The inclusion criteria for both types of merchants are that the merchant must: (1) be responsible for input purchases, (2) have made at least 5 sales of ready-to-wear garments in the past week, (3) have used WhatsApp at least once for both sales and input purchases in the past 30 days. For offline merchants only, we additionally required that the merchant must have sold garments made in Turkey at least once in the past 6 months.[1]

Offline merchants were recruited in markets in Dakar by a team of surveyors. We identified the largest markets in Dakar and conducted a census market-by-market until the quota was achieved.

Online merchants were recruited in two ways. First, we used Facebook advertisements that targeted individuals in Dakar between the ages of 18 and 40. Second, we used snowball sampling: at the end of the baseline survey, we asked if merchants would like to recommend any other online merchants for us to contact.

Our target sample size was 1500-1800, with approximately one-third offline merchants and two-thirds online merchants. We chose these shares because we hypothesised that the intervention would be most relevant for online merchants, but we wanted to include a sufficiently large share of offline merchants in order to test this hypothesis.

2.2 Suppliers

The study also consists of suppliers in Istanbul. We chose Turkey as the exporting country for three reasons. First, pilot surveys showed that it was a large source of ready-to-wear garments in Senegal (the second largest after China). Second, there was already a well-established export industry taking place over social media. Third, pilot surveys showed that garments made in Turkey were generally viewed as higher quality on average than garments made in China, which is well-suited to studying contracting concerns about quality. The inclusion criteria were that a supplier must (1) sell ready-to-wear garments manufactured in Turkey, (2) live in Istanbul, (3) visit Senegal at most once per year on average, (4) report making at least 5 sales per week via WhatsApp on average, (5) be willing to make a WhatsApp group to advertise their products for the purpose of the study.

---

[1]We wanted to ensure that merchants in the study had a serious intention to sell Turkish goods. Our piloting activities found that–conditional on opting in to the study–this was typically not a problem for online merchants, but often was among offline merchants. We thus introduced this additional criterion for offline merchants to screen out merchants with no experience selling Turkish goods.

In addition to these inclusion criteria, in order to participate in the study, suppliers were asked to agree to several conditions governing the use of the WhatsApp group. These were straightforward and primarily designed to ensure that suppliers engaged appropriately with the merchants in the study. In particular, the conditions were that they post (1) regularly in the group (at least 20 posts per week, on average), (2) items of varying quality levels, (3) only items within their chosen sector,[2] (4) with prices included.

In particular, suppliers were recruited in three stages. First, two surveyors in Istanbul conducted a listing of suppliers in Aksaray and Laleli, quarters with a large presence of exporters to Africa. The listing was a short survey that asked the questions necessary to assess the inclusion criteria. Suppliers with physical stores were listed by a census, and suppliers without physical stores were listed by being approached when they visited cargo stores (businesses that manage transportation logistics for exporting). Second, among listed suppliers who met the inclusion criteria, we conducted a mystery shopping exercise several months later in which we hired and trained a small team of merchants in Senegal to order 2 dresses and 2 polo shirts from each supplier. Based on this ordering experience, we identified suitable suppliers according to a range of criteria, whom we then invited to participate in the study, subject to agreeing to respect the conditions discussed above. Our target sample size for the number of suppliers was 15-30.

3 Treatments

The full design tree is shown in Figure 1. We designed treatments to test for the importance of search and trust frictions in the context of social media commerce. There are three treatments. We use the notation T1 to indicate that a merchant is treated with Treatment 1 and C1 to indicate that a merchant is in the control group for Treatment 1, with analogous notation for treatments 2 and 3. Note that all randomisation is across merchants.

3.1 T1: Search

The purpose of this treatment is to generate exogenous variation in search costs of accessing Turkish suppliers and Turkish goods among merchants in Senegal. A randomly selected 80% of merchants are treated with T1. Merchants are added to the WhatsApp groups of 3 different suppliers. The suppliers to match with are selected at random, subject to being a match to the merchant's chosen sector.

Supplier WhatsApp groups involve one supplier advertising their products to a large number of customers using high quality photos and videos. They are a one-way conversation where only the supplier is allowed to post. If a customer sees a good that they are interested in, they typically follow up with the supplier via private message. Suppliers sell more than just what they post, meaning that customised orders or requests are also possible.

We communicate to the control group that unfortunately we cannot add them to any supplier groups at this time, but that we might do so at the conclusion of the study.

3.2 T2: Hidden Types "Social Feedback"/ Reputation Building

The purpose of this treatment is to generate exogenous variation in merchants' beliefs about the trustworthiness of the suppliers. A random 50% of those in T1–that is, those treated with the Search treatment–are treated with T2. The treatment consists of two things. First, merchants

---

[2]The three sectors were (1) women's clothing, (2) men's clothing, (3) women's shoes and women's bags.
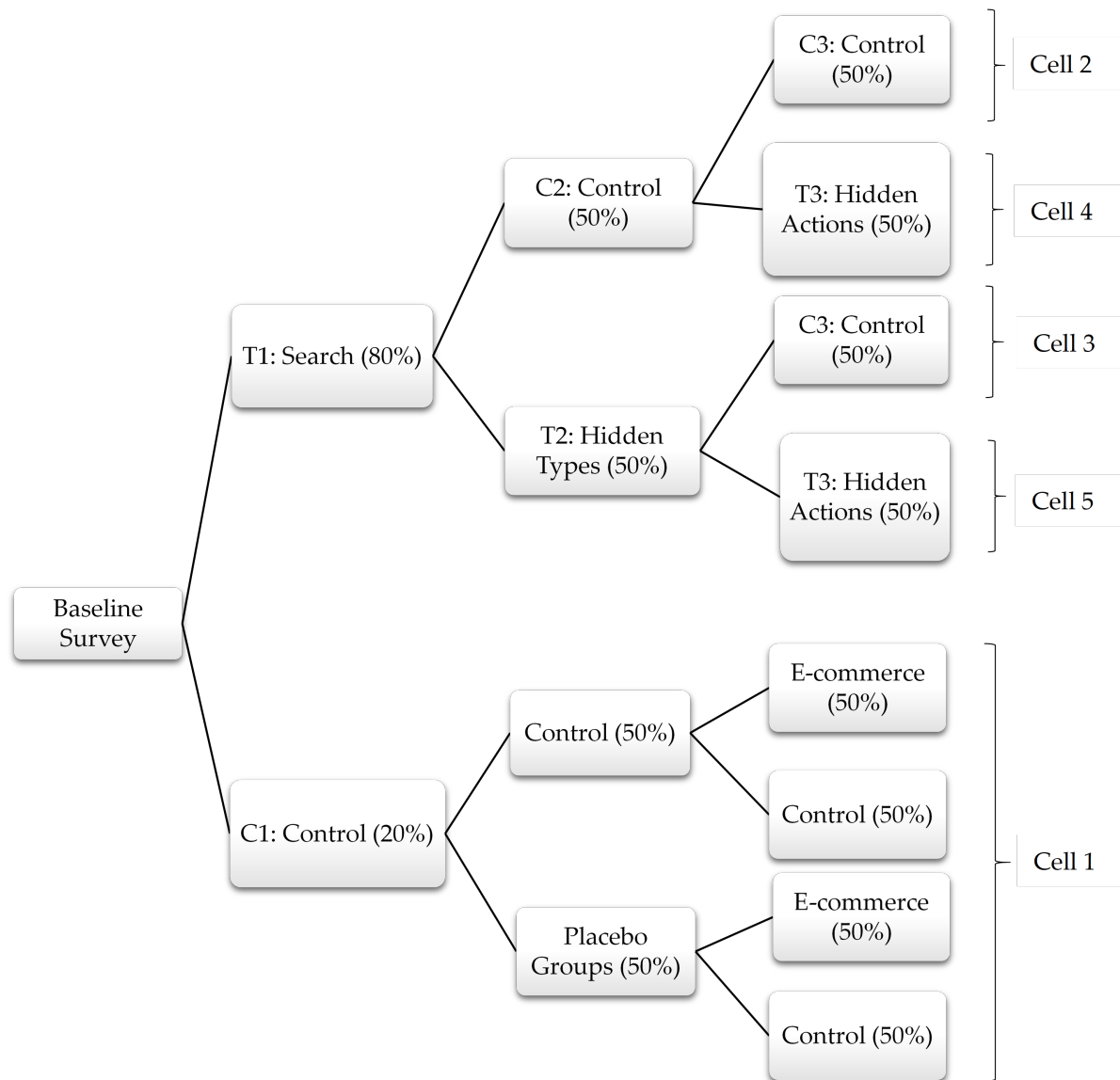
Figure 1: Design Tree

are added to a fourth WhatsApp group–this group contains other merchants in the study that were matched with the same suppliers, but does not contain any suppliers. We explain to such merchants that the purpose of this group is to discuss the suppliers and share experiences. A member of the survey team moderates discussion and enquires whether anyone has experiences to share from time to time. Second, merchants receive a phone call 2-3 days after recruitment from a recommender, who is one of a team of merchants hired and trained before the study to order standard products from a large number of suppliers. In this call, the recommender describes their experience ordering from one of the suppliers that the merchant was matched with and sends a photo of the item as ordered as well as the item that was delivered, but does not reveal that they are part of the study team. All merchants not in this treatment group instead receive a "placebo" phone call from a surveyor (see Section 4.2 below for more detail).

3.3 T3: Hidden Actions "Social Review" / Moral Hazard

The purpose of this treatment is to generate exogenous variation in merchants' beliefs about the incentives of the suppliers to honour contracts. A randomly selected 50% of merchants in T1 are treated with T3. In this treatment, merchants are told the following at the end of the baseline survey:

> I have one last piece of information to give you. As you know, you have been added to WhatsApp groups of Senegalese suppliers in Turkey.

> We work with many suppliers in our study. We want to assure you that they are motivated.

> We would like to collect feedback on these suppliers so that we can recommend the best ones in the future. To do this, we will ask merchants to rate your experience with the suppliers we have presented to you on a scale of 1 to 5 on product arrival and quality. These reviews help identify the best suppliers, which is beneficial to them and allows us to continue recommending them to others. They are therefore motivated.

> If a supplier gets a bad rating, we will investigate and remove them from the study if they did not do a good job. They will therefore lose access to around 150 merchants if they do not do a good job.

> I will give you a phone number that you can use to give your rating or report a problem.

> Lastly, I want to emphasise that the suppliers are aware that they are being rated and that, if they receive bad ratings, they will be removed from the study. We can thus assure you that they are motivated.

After delivering this message, the surveyor gives a business card to the merchant. The business card has a phone number to call, and prominently highlights that this number should be used to rate the suppliers and/or to signal any problems. Untreated merchants receive a similar card, but without any mention of ratings–instead saying that the phone number is for questions. Both cards can be seen in Figure A1. For merchants for whom the baseline survey was conducted over the phone, the surveyor sends a high-quality image of the business card to the merchant via WhatsApp. All suppliers are told a similar message about how the ratings will work.

3.4 Social Media

We intentionally implemented T1-T3 through social media as much as possible. We did so because social media platforms are rapidly changing how business is conducted across Africa, and key features of such platforms have the potential to provide low-cost solutions to the frictions that we study in this project. We thus wanted to measure the extent to which these features can and do alleviate such frictions, which is why we isolated the different frictions by activating or shutting down key features of the widely used platform, WhatsApp.[3]

3.5 Additional Treatments

We also implemented a few additional treatments.

Placebo Groups: Within C1, a random half of merchants are added to a WhatsApp group with other merchants in the same treatment condition and the same sector. We explain to merchants in this treatment that the purpose of this group is to share information amongst merchants about suppliers. A member of the survey team moderates discussion and prompts the group from time to time to ask if anyone has questions about suppliers or if anyone has suppliers of Turkish goods that they would recommend. The purpose of this treatment is to serve as a placebo for T2: that is, to account for whether simply grouping merchants who otherwise do not know each other to share information in general has an impact in and of itself. We expect the effect of this to be zero on most outcomes, so, to maximise power to detect main effects, in our main analyses we intend to pool it with the rest of C1. We may explore results from this treatment in more detail in subsequent analyses.

E-commerce: Within C1, a random half of merchants receive a short training on the use of Alibaba, one of the few e-commerce platforms available in Senegal. The training is conducted at the end of the baseline survey and covers (1) what the platform does and how it can help merchants, (2) how to install and set up the app, (3) how to use the app to search for products and find suppliers. The purpose of this treatment is to test whether limited usage of formal e-commerce platforms at baseline is due to lack of knowledge and understanding about how to use these platforms. Our prior is that the effect of this treatment is zero on most outcomes, and thus, to maximise power to detect main effects, in our main analyses we intend to pool it with the rest of C1. We may explore results from this treatment in subsequent analyses.

Experimenter Demand Treatment: To measure any experimenter demand effects on questions about merchants' views of the study suppliers, we implemented a version of the "weak demand treatment" from de Quidt, Haushofer, and Roth (2018) in both Call 2 and the Followup Survey. In particular, the surveyor reads the following message to a random 20% of merchants in T1 at the beginning of the interview: "Just to inform you, we hope that the merchants in the study trust our suppliers and know that the suppliers are motivated." The purpose of this is to compare any treatment effects from our main treatments to this treatment to assess the scope of experimenter demand.

Mystery Shopping Purchase: When doing mystery shopping, among merchants that get to the end of the protocol (i.e., they find a suitable good with an acceptable price, etc.), we only purchase from a randomly selected 80%. The rationale for this is that our budget did not allow us to purchase from all merchants that reached the end of the protocol, so we selected those we did purchase from at random. We may use this treatment to test whether the act of making an order has an impact on propensity to establish a relationship, but we do not expect to be well powered

---

[3]WhatsApp is ubiquitous in this setting: in a separate, representative survey of 138 merchants in a few markets in Dakar that we conducted in early 2023, we found that 97% of merchants had a WhatsApp account.

Figure 2: Timeline

to study such an effect. We thus do not expect to include it in our main analysis. Note that this treatment is distinct from our use of mystery shoppers as an outcome measure.

3.6 Treatment Assignment

For the main three treatments, Placebo Groups, and E-commerce, we stratified by online vs offline, product, and a dummy for whether the merchant has ever imported directly from abroad. We conducted this randomisation in Stata using the command *randtreat*.[4] We implemented this stratification by generating strata-specific lists of unique identifiers linked to treatment conditions, with proportional sizes based on pilot summary statistics. Upon a merchant satisfying the inclusion criteria and completing questions sufficient to identify their stratum, the surveyor entered the next available identifier on the appropriate stratum's list into their tablet, thereby determining a merchant's treatment condition. If a list was complete, so there is no next available identifier, the merchant was unconditionally randomised into a treatment condition (implemented through the surveyor typing in an ID from an additional list–not linked to any stratum–that was generated in Stata).

4 Timeline

See Figure 2 for a timeline.

4.1 Recruitment and Baseline Survey

Upon recruitment, merchants answer a baseline survey lasting approximately 30 minutes. All merchants receive the same survey, and any treatment conditions are explained at the end of the survey.

Merchants recruited by Facebook advertisements or snowball sampling had their baseline survey conducted over the phone; otherwise, the baseline survey was conducted in person.

4.2 Call 1

2-3 days after recruitment, merchants receive a short phone call lasting a few minutes. If the merchant is in T2 and T1, this is the recommendation call described in Section 3.2. If the merchant is in T1 and C2, this is a "placebo" call that asks a few questions about what the merchant thinks of a named supplier's WhatsApp group. If the merchant is in C1, this is a "placebo" call that asks a few questions about supplier WhatsApp groups in general.

---

[4]We first randomised merchants evenly across the 5 cells shown in Figure 1. Then, among merchants in C1 (Cell 1), we did a second evenly-proportioned randomisation across the four sub-groups. Both randomisations used the option "strata" to deal with misfits.

The purpose of this call is twofold. First, it is to deliver the treatment for merchants in T2. Second, it is to ensure that merchants in T1 were correctly added to the groups and nudge them to look and think about them.

4.3 Call 2

2-3 days after receiving Call 1, merchants receive a second phone call lasting a few minutes. This call asks a few survey questions to measure their beliefs about access to Turkish goods, trust in suppliers, and supplier motivation. After asking these questions, the surveyor reminds merchants of their treatment condition. Finally, as described above, a random 20% of merchants receive the experimenter demand treatment at the start of this call.

The purpose of this call is twofold. First, it is to measure perceptions and beliefs about the supplier groups after merchants have had enough time to form an opinion about the groups, but before merchants are likely to have ordered. Second, it is to remind merchants of their treatment condition.

4.4 Mystery Shopper

8-9 days after recruitment, all merchants are contacted over WhatsApp by a mystery shopper, played by a trained surveyor. Merchants are expecting this contact, as we explain to them at the end of the baseline survey (and remind them at the end of Call 2) that we will put them in touch with customers who are looking to buy high-quality goods from Turkey. The mystery shopper explains that they would like to purchase a certain high quality product from Turkey for an event. We leave a detailed description of the mystery shopper protocol to the research paper. In short, each product has a total of 5 horizontal criteria, and the mystery shopper proceeds with the purchase–including asking about price and delivery–if the merchant can find a good with at least 3 criteria. If the mystery shopper successfully reaches the end of the protocol, they buy the good in a random 80% of cases and we then measure its quality (see Section 4.5). Whether they buy the good is only revealed to the surveyor conducting the mystery shopping after having reached the end of the protocol so as not to bias their behaviour.

In addition, after having concluded the protocol, but before revealing whether they will purchase the good, the mystery shopper also does what we call a "Rapid Product" exercise: they send a photo of a product recently posted in a supplier group that the merchant was added to, and ask if the merchant has this product (and at what price). This is designed to be fast: the merchant either finds the exact product or not (there are no criteria), and the surveyor waits a maximum of 24 hours to see if the merchant finds it.[5] If the merchant finds it, we purchase the rapid product in a random 5% of cases.

The purpose of the mystery shopper is twofold. First, it is to measure a set of revealed preference outcomes concerning whether the treatments affect merchants' ability to supply a customer demanding high quality goods from Turkey. Second, it is to indirectly subsidise merchants to experiment with the study suppliers.

4.5 Quality Measurement

When the mystery shopper purchases the product, and it eventually arrives, two expert tailors independently assess the quality according to a 50-point scorecard, shown in Figure A2. The

---

[5]The rationale for this exercise is to test a second product where search is shut down by making the good as easy to find as possible. We only allowed a maximum of 24 hours in order to not disrupt the main mystery shopping exercise.

tailors also opine as to the where the product was likely manufactured, both by assessing whether the sewing or design can identify where it was made, and by looking at any tags or labels that indicate the origin.

4.6 Followup Survey

Around 2 months after recruitment, merchants receive a survey lasting approximately 30 minutes conducted over the phone.

5 Hypotheses and Outcomes

Below, we describe different sets of outcomes. Within each set of outcomes, we describe both primary and secondary outcomes, as well as our hypotheses.

5.1 Mystery Shopper Outcomes

5.1.1 Primary Outcomes

1. Horizontal Product Fit:

    (a) Extensive Margin: Whether the merchant provided any product quote. We hypothesise that all three treatments have ambiguous results on this outcome. On the one hand, treatment should make it easier for merchants to access Turkish goods. On the other hand, we expect there to be a non-trivial share of merchants who have limited prior experience with Turkish goods, but conclude–after learning about the typical price and quality of Turkish goods due to treatment–that they would prefer to focus on cheaper, lower-quality goods. In such cases, we expect treatment to cause them to be less likely to provide a quote for Turkish goods to the mystery shopper. We thus expect that heterogeneity analysis will be key.

    (b) Intensive Margin: Conditional on providing any quote, whether the merchant found a suitable product. We intend to define suitable as matching 3 or more of the 5 criteria, for two reasons. First, it is in line with our purchasing rule. Second, it is robust to whether one or two criteria on some products were particularly rare. If we find that 3 out of 5 generates too little variation to be meaningful, we may adjust this. We hypothesise that all three treatments should have a positive impact on this outcome.

    (c) Overall: Whether the merchant found a suitable good unconditionally–that is, coding those who do not provide a quote as 0. We hypothesise that all treatments should have an ambiguous impact on this outcome, as it combines the ambiguous effect on the Extensive Margin with the positive effect on the Intensive Margin.

2. Vertical Product Fit. We intend to define this in two ways. First, we will define a binary measure of "high" vs "low" quality. A good will be defined as high quality if it is above the median quality score. We may also define low quality as a good receiving a score of half or less in at least one section of the scorecard.[6] Second, we will use the score on our 50-point

---

[6]The rationale is that, when we piloted the quality measure, we observed that we often found products where most aspects were high quality but the fact that one or two aspects were poor was sufficient for it to be unreasonable to call the product high quality. For example, in some cases the fabric was of high quality, but the sewing was so poor that the product would clearly fall apart after wearing it a few times. In such cases, the good can score quite highly on a linear scale, even though a merchant would be unable to command a quality premium for it.

scale.[7] We hypothesise that all treatments should have a positive impact on this outcome, as Turkish goods are generally viewed as higher quality and our treatments aims to reduce barriers to accessing Turkish goods. As this outcome is only measurable when the good is purchased, we expect to be less powered for it than other outcomes.

3. Customer Contract: Did the merchant accept payment on arrival? We view this as a measure of trust in the supplier. We hypothesise that the impact on this variable is negative for T1, and ambiguous for T2 and T3. It is negative for T1 because we expect T1 to make merchants more likely to purchase directly from Turkey, where suppliers typically insist on payment in advance. This channel exists for T2 and T3 as well, but if T2 and T3 also increase trust, then–conditional on ordering from Turkey–we expect a positive effect, thus making their overall effects ambiguous.

5.1.2 Secondary Outcomes

1. Price: Conditional on finding a suitable product, the unit price of the good (delivery included). We hypothesise that this is ambiguous a priori. On the one hand, if treatment causes merchants to be more likely to sell Turkish goods (as opposed to, for example, pretending their current stock is Turkish), then–because Turkish goods are higher quality–we expect price to go up. On the other hand, if the study suppliers offer lower prices than merchants' previous way of accessing Turkish goods, then we expect price to decrease.

2. Product Origin: For goods purchased, was the product made in Turkey? We hypothesise that our treatments should have positive effects on this outcome. While the mystery shopper only purchases the good if the merchant confirms it was made in Turkey, we expect variation here because some merchants will either lie or be unsure about the origin of products from their regular suppliers. We intend to measure this using a combination of the expert tailors' assessments and any tags on the product.

3. For merchants in T1: Did the merchant contact any of the suppliers in the study to search for the good? We hypothesise that T2 and T3 should have a positive effect.

4. For merchants in T1: Rapid Search: Did the merchant find the rapid product? We intend to restrict this analysis to only merchants who engaged with the mystery shopper, as merchants who made no effort at any point will trivially not find the product for reasons unrelated to trust. We hypothesise that T2 and T3 should have a positive effect.

5.2 Supplier Relationship Outcomes

We hypothesise all treatments to have a positive effect on outcomes relating to ordering from Turkey and supplying Turkish goods.

We are agnostic as to whether treatments should have a positive or negative effect on outcomes relating to orders from and relationships with current suppliers. If the study suppliers are substitutes for existing suppliers, we expect a negative effect; if the study suppliers are complements for existing suppliers (for example, if the study causes merchants to learn about Turkish goods and then inquire about these to their existing suppliers), we expect a positive effect. However,

---

[7]We will make minor adjustments for goods where some criteria are not applicable: for example, one of the criteria for polo shirts concerns whether the buttons are securely attached. If a merchant supplies a polo shirt with a zipper instead of buttons, this criterion is not applicable.

we expect any such effects to be small–particularly on the short timeline at which we conduct the Followup Survey–and thus do not expect to be well-powered for these outcomes.

We intend to measure outcomes related to orders with questions in the Followup Survey, but we are also–at the time of writing–seeking access to a mobile money dataset that could be used to measure orders at transaction level. If we obtain access to this dataset, we plan to use it to measure further outcomes described below.

5.2.1 Primary Outcomes

1. Orders (to all groups, including pure control).

    (a) New relationships: Did the merchant develop a new regular supplier relationship in the past 3 months? "Regular supplier" is defined as having made at least 2 orders in the past 3 months.

    (b) Previous relationships:

        i. Are there any suppliers that you previously had that you no longer intend to order from?

        ii. Total value of purchases from already-existing suppliers in the past 3 months.

    (c) Total inputs: Total value of purchases from all suppliers in the past 3 months.

2. Orders (to merchants in T1).

    (a) Did the merchant contact any of the suppliers?

    (b) Did the merchant make any orders?

    (c) If the merchant made orders:

        i. Total number of orders

        ii. Total value of orders

        iii. Modal contract on the order–share of orders where the merchant paid in advance. There may be too little variation in this variable to meaningfully analyse, as the vast majority of first-time purchases are entirely paid in advance.

3. Orders as measured by mobile money transactions.[8] We will use this to validate the survey-based measures. Depending on what the eventual dataset looks like in terms of features including data quality, coverage, and which variables are available, we may use it instead of survey-based measures for certain outcomes (e.g., value of orders), as well as to measure ordering outcomes over a longer time horizon (e.g., transaction frequency, size).

5.2.2 Secondary Outcomes

1. Sector:

    (a) Has the merchant entered a new sector in the past 3 months? We hypothesise that treatments should have a positive effect on this outcome.

2. Supplier Bargaining:

    (a) In the last 3 months, has the merchant forwarded something they saw from a supplier WhatsApp group or Status to ask if their current suppliers could provide this good?

---

[8]This is conditional on obtaining access to the mobile money dataset.

(b) In the last 3 months, has the merchant used information from a supplier WhatsApp group to ask their current suppliers to reduce prices?

5.3 Sales Outcomes

5.3.1 Primary Outcomes

1. Sales of Turkish goods (we hypothesise a positive effect on this outcome):

    (a) Did the merchant sell any goods made in Turkey in the past 3 months?

    (b) Total value of Turkish good sales in the past 3 months.

    (c) Share of sales made up of goods from Turkey.

2. In the past 3 months, share of sales that the merchant describes as high quality. We hypothesise a positive effect on this outcome.

3. In the past 3 months, how many clients has the merchant lost owing to the fact that they didn't have the product they were looking for? We hypothesise a negative effect of treatment on this outcome, although we expect that the mean value in control may already be quite low and the outcome noisy, meaning that we might not be well-powered to detect treatment effects.

5.3.2 Secondary Outcomes

We hypothesise a positive effect of our treatments on revenue and profit. We also potentially expect a change in the probability to travel abroad to buy products. However, since the Followup Survey is conducted after only 1-2 months, we might not be well powered to detect such effects (which we expect to shift over a longer horizon). This is why we list them as Secondary Outcomes in this PAP.

1. Total revenue

2. Total number of customers

3. Total profit

4. Propensity to travel outside of Senegal to purchase goods

5.4 Social Media Outcomes

We hypothesise positive effects of our treatments on merchants' propensity to use social media in various ways as part of their business. We have two primary channels in mind: (1) merchants may directly advertise goods from the suppliers on their social media; (2) merchants may learn from, copy, and adapt the suppliers' social media practices.

5.4.1 Primary Outcomes

1. Social media advertising:

    (a) In the past 7 days, has the merchant advertised using WhatsApp Status?

    (b) Does the merchant have a WhatsApp group with their customers?

    (c) Does the merchant use other social media platforms to learn about suppliers?

    (d) T1 only: In the past 7 days, has the merchant advertised goods from the study suppliers using WhatsApp status?

2. Social media sales:

    (a) In the past 7 days, has the merchant made a sale using WhatsApp status?

    (b) In the past 7 days, out of 10 sales, how many originated through social media?

5.5 Other Outcomes

5.5.1 Call 2 Intermediate Outcomes

These are not outcomes of interest in and of themselves, but–as they are measured after less than a week–are intended to capture how the treatments change initial perceptions of the supplier groups and the suppliers. All three questions are on a scale of 1-5, where 1 is the most positive answer. We intend to use a dummy capturing whether the merchant answered 1 or 2; if this generates too little variation, we may amend it.

1. "In your opinion, is it easy or difficult for you to find good that interest you from Turkey?" We hypothesise that all three treatments should positively impact this outcome.

2. "According to you, we simply want your opinion, imagine that tomorrow you see a product from the supplier [supplier name] in the group and you want to place an order for 100,000 FCFA. Would you trust this supplier [supplier name] in terms of the arrival and quality of the product?" We hypothesise that T2 should positively impact this outcome.

3. "According to you, we simply want your opinion, what is the level of motivation of the study suppliers to do a good job with the merchants?" We hypothesise that T3 should positively impact this outcome.

5.5.2 Beliefs

1. Supplier Trust: We will conduct an exercise to measure beliefs about the propensity of the study suppliers to honour a contract. In particular, we will elicit WTA to purchase a good from a study supplier compared to an existing supplier, in a game where the goods are identical but there is a large reward for the merchant for successful delivery.

2. Varieties. The below are all measured on a scale of 1-5, where 1 indicates "very easy" and 5 indicates "very hard". We intend to use a dummy capturing whether the merchant answered 1 or 2; if this generates too little variation, we may amend it.

    (a) Is it easy or hard to have access to new varieties?

    (b) Is it easy or hard to find a specific variety when the merchant is looking for it?

    (c) Is it easy or hard to have access to high quality items?

5.5.3 Alibaba/e-commerce usage

1. Has the merchant used Alibaba or e-commerce platform to search for goods *or* to make an order *or* to compare prices/quality from Alibaba to your current suppliers?

5.5.4   International trade knowledge

1. Does the merchant know the shipping cost from Turkey to Senegal?

2. Does the merchant feel that they know how international trade works?

6   Heterogeneous Treatment Effects

We specifically plan to explore heterogeneous treatment effects along the following dimensions:

- Whether the merchant is online vs offline. This is important because these types of merchants have very different business practices that affect the way in which they interact with suppliers (e.g., offline merchants likely prefer to sell their existing stock, meaning that we expect quite different response patterns to the mystery shoppers). As we expected this to be a key margin of heterogeneity, *we stratified by this variable*.

- Whether the merchant has previous direct import experience. This is important because importing directly may be too large of an experiment for a non-trivial share of merchants, particularly for one mystery shopper. This means that merchants with experience may have different response patterns to the mystery shopper. As we expected this to be a key margin of heterogeneity, *we stratified by this variable*.

- Whether the merchant's activities have a substantial wholesale component. This is important because merchants for whom wholesale represents a larger share of their business likely interact with the study suppliers quite differently–for example, making smaller and more infrequent orders, or visiting the supplier in person from time to time.

- Whether the merchant already sells a large share of Turkish products. This is important because we expect that a non-trivial share of merchants that do not currently sell Turkish products may decide, after learning that Turkish products are typically more expensive, may prefer not to sell Turkish products. The treatments could thus have *negative* effects for this group if they cause merchants to learn about Turkish goods and update their beliefs away from Turkish goods.

- Baseline digital literacy. This is important because some level of digital literacy is likely needed, as the study takes place over social media. Merchants with very limited digital literacy may find it too challenging to interact with the suppliers, the mystery shopper, and other merchants in T2. For instance, we will use smartphone adoption year as one proxy for digital literacy.

- Recruitment method. Merchants were recruited using three different methodology: snowball listing, in markets, and through social media advertising.

We also plan to use machine learning methods, such as those presented in Wager and Athey (2018), to detect and explore key dimensions of heterogeneous treatment effects. We intend to use primarily baseline covariates, although we may include covariates collected after treatment if they are plausibly unaffected by treatment.

7 Methodology

7.1 Treatment Cells

As noted in Section 3.5, we intend to pool together all sub-groups within C1 in our main analyses. This creates a total of 5 *ex ante* equally-sized cells, shown in Figure 1. We refer to the cell numbers in this figure throughout the remainder of this section.

7.2 Regression Specification

7.2.1 Main Analysis

We intend to run the following "long" regression as our baseline specification.

$$y_{it} = \alpha + \gamma_s + \sum_{j=2}^{5} \beta_j T_{ij} + \delta y_{i0} + \sum_{k=1}^{K_1} \rho_k X_{i0,k} + \varepsilon_{it} \tag{1}$$

- T_{ij} is a dummy for whether merchant i is in treatment cell j. Note that Cell 1, pure control, is the omitted category.

- γ_s are strata fixed effects.

- y_{i0} is baseline value of y_{it}, which we include where this is available or where a very similar variable is available (for example, the same survey question asked over a 12 month horizon rather than a 3 month horizon).

- $X_{i0,k}$ is one of K_1 merchant-level covariates, discussed later.

We will report estimates of the coefficients $\{\beta_j\}_2^5$ to directly measure the effect of the different treatment combinations. Our primary estimands of interest are linear combinations of these coefficients corresponding to the three frictions:

- Search (T1): $\beta_2 - \beta_1$

- Hidden Actions (T2): $0.5 (\beta_3 - \beta_2) + 0.5 (\beta_5 - \beta_4)$

- Hidden Types (T3): $0.5 (\beta_4 - \beta_2) + 0.5 (\beta_5 - \beta_3)$

We intend to report the estimates of these linear combinations, including p-values, alongside the raw coefficients. We choose this approach of focusing on the long regression and explicitly aggregating coefficients for two reasons. First, it allows us to derive all of the key estimates from one regression. Second, it is explicit about the fact that our estimands are (convex) weighted averages of treatment effects, some of which have other treatment conditions active.

For outcomes that are only relevant for merchants in T1, we intend to drop the merchants in C1 but otherwise use an analogous method.

We also intend to run additional regressions to study the impact of the additional treatments.

7.2.2 Heterogeneous Treatment Effects

As discussed in Section 6, we intend to analyse heterogeneous treatment effects. In such regressions, we intend to use the following regression specification.

$$y_{it} = \alpha + \gamma_s + \sum_{j=2}^{5} \beta_{1j} T_{ij} + \sum_{j=2}^{5} \sum_{h=1}^{H} \beta_{2j,h} T_{ij} X_{i0,h} + \delta y_{i0} + \sum_{k=1}^{K_2} \rho_k X_{i0,k} + \varepsilon_{it}, \tag{2}$$

where $H \leq K_2$ is a subset of the included covariates. We may de-mean some covariates to facilitate interpretation. As in Section 7.2.1, we intend to aggregate coefficients to isolate the three frictions.

7.2.3 Quantile Treatment Effects

We intend to compute quantile treatment effects using quantile regression.

7.3 Covariates

Aside from strata indicators, we may use machine learning techniques to select appropriate covariates to include in regressions. We intend to only use variables that are measured at baseline or are invariant over the study period.

7.4 Inference

We plan to use randomisation inference as our primary method of computing p-values and conducting inference, although for computational convenience we may use standard sampling-based approaches for inference from time to time in various presentations and similar settings.

7.5 Outliers

We intend to winsorise variables with extreme outliers at either the 1% or 5% level, depending on the extent of outliers.

7.6 Robustness to Contamination Bias

At the time of writing, a recent working paper, Goldsmith-Pinkham, Hull, and Kolesar (2022), highlights that treatment coefficients in regressions of the form of (1) may not estimate convex-weighted averages of heterogeneous treatment effects. While we prefer (1) as our primary regression specification for ease of interpretation, we will compute the decomposition and interacted estimator in Goldsmith-Pinkham, Hull, and Kolesar (2022) to assess robustness to this issue.

7.7 Indexes

We intend to combine variables corresponding to the same outcome construct into indexes.

8 Model

We designed the experiment and treatments with a view to estimating a model of importing under search, hidden types, and hidden action frictions. Specifically, the three treatments are designed

to map to these three theory-based frictions. The goals of the model are to (1) quantify the relative and absolute sizes of the frictions, (2) compute gains from trade based on primitives (i.e., in absence of these frictions), (3) consider counterfactuals.

9 References

Banerjee, A., Duflo, E., Finkelstein, A., Katz, L.F., Olken, B.A. and Sautmann, A., 2020. In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics (No. w26993). National Bureau of Economic Research.

De Quidt, J., Haushofer, J. and Roth, C., 2018. Measuring and bounding experimenter demand. American Economic Review, 108(11), pp.3266-3302.

Goldsmith-Pinkham, P., Hull, P. and Kolesar, M., 2022. Contamination bias in linear regressions (No. w30108). National Bureau of Economic Research.

Wager, S. and Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), pp.1228-1242.

(a) Treatment
(b) Control

Figure A1: Business Cards for T3

		More points = better quality	
		SCORE POLO	SCORE DRESS
Fabric	Is the fabric soft?	3	3
	Are there pills?	2	2
	Is it transparent under light?	4	4
	Are the threads consistent?	3	3
Sewing	Are the seams straight and neat?	3	3
	Do the seams come apart upon gentle pulling?	3	3
	Are there stray threads or loose stitches?	2	2
	Do the seams lie flat without puckering?	2	2
	Is the hem seam reinforced?	2	2
Shoulders	Do the shoulders fall correctly?	3	3
Collar and Buttons	Is the collar well-attached and flat?	3	
	Are the buttons securely attached?	3	
Darts	Are there darts on the bust and waist?		3
	Are the darts neatly sewn?		3
Colourfastness	Does it lose colour upon soaking?	7	7
Pattern	Does the pattern line up with the seams?	2	2
Hem	Is the hem well done?	4	4
Sleeves	Are the sleeves the same length and width?	4	4
Absence of Other Defects	Are there any other defects not covered above?	-10	-10
		50	50

Figure A2: Quality Scorecard