Pre-Analysis Plan: Measuring the demand for public transport in Lagos, Nigeria

Daniel Björkegren* Alice Duhaut† Geetika Nagpal‡ Nick Tsivanidis§

May 28, 2023

Abstract

This project will measure demand for public transport in Lagos. It has two components. The first study consists of a field experiment to measure participants value of wait time. The experiment hinges around an SMS-based app (playable on all cellphones) in which participants arrive at a bus stop and are offered a payment amount to wait for a number of minutes before boarding their bus. Enumerators stationed at the bus stops have a tablet which displays a secret code that changes each minute; participants text these codes to our SMS service when they arrive to register their arrival time and receive their offer. If they accept the offer, they send the code displayed after the specified amount of time has elapsed to verify they waited. We randomize the offers of payments and waits; participants' decisions to accept or reject the offers reveal their value of wait time. The second study consists of a field experiment that measures the price elasticity of demand for transport services. In collaboration with the transit regulator in Lagos, we have developed a way in which price subsidies can be administered via the swipe cards used to the pay for the formal bus system in Lagos.

1 Background

Cities in low income countries play a crucial role in structural transformation, and already account for a majority of non-agricultural GDP. But their contribution to growth is being limited by traffic congestion. Congestion is a spatial friction, which distorts interactions between firms and workers. Moreover, transportation is one of the most prominent factors driving air pollution and greenhouse gas emissions in large cities. As more people in low- and middle-income countries move into urban areas, access to efficient public transportation will play a key role both in limiting carbon emissions and reducing the amount of pollution people living in cities are exposed to. However, in many cities in sub-Saharan Africa, more people commute using personal vehicles or informal transportation services than using formal public transit.

*Brown University. Email: dan@bjorkegren.com
†World Bank. Email: aduhaut@worldbank.org
‡Brown University. Email: geetika_nagpal@brown.edu.
§University of California at Berkeley. Email: ntsivanidis@berkeley.edu.

Cities are considering a menu of investments to ease these frictions. What are the impacts from and returns to government investments in new transit systems in Africa? Could similar impacts be obtained by cheaper options, such as regulation of the informal sector? A first part of this project will address the first question by conducting a quasi-experimental impact evaluation of the rollout of 820 new buses across 28 lines in Lagos state combined with three RCTs aimed to causally identify the demand and supply sides of the market. The second part of the project will address the second question by combining these RCTs with a new census of the informal transit network in Lagos and a structural model that characterizes optimal policy for its regulation. Our results will provide evidence-based insights into the relative returns to alternatives for reducing spatial frictions in low income countries.

The remainder of this analysis plan details the two RCTs to measure price and time elasticities of demand, to be extended later for the supply side RCT before it is run.

2 Price RCT

Researchers are partnering with the Lagos Metropolitan Area Transit Authority on a randomized evaluation to study the impacts of introducing a formal public bus system (BRI) on commuters.

Researchers will conduct a household recruitment of 2800 commuters from areas near transit to participate in the commuter intervention, 2000 of which will participate in the "Price Experiment" and 800 of which will participate in the "Wait Time Experiment".

In the Price Experiment, commuters will be randomly assigned to one of 5 groups:

1. 50% ticket subsidies for 10 weeks
2. 50% ticket subsidies for 20 weeks
3. 75% ticket subsidies for 10 weeks
4. 75% ticket subsidies for 20 weeks
5. No subsidies

Commuters will be surveyed at baseline and endline to understand the previous day's travel patterns, as well as through a midline SMS survey. We also have access to the smartcard backend to measure each trip on the formal system. The SMS survey will be used to verify which trip was taken from a set of options, to try and identify card sharing.

There will be two main empirical approaches we will follow. One is a simple reduced form diff-in-diff regression where we regress the change in outcomes on a dummy for whether the individual was in a treatment group. The core outcomes will include whether they took the BRI, the number of trips taken. Additional outcomes will include the distance of the main trip, number of trips taken using car and informal minibus, as well as other "real" outcomes such as income and hours worked. We will include control variables to increase the precision of our estimates, including respondent demographics (age, gender, education, income, number of children) and locational attributes (distance to closest BRI stop, local government area fixed effects). We will also conduct a heterogeneity analysis interacting the treatment variable with (i) gender and (ii) income.

The second empirical approach will be structural. The simplest model of transport demand we have in mind is a discrete choice model

$$U_i(\omega) = \alpha - \gamma p - \eta t + \epsilon(\omega)$$

where $i \in \{Bus, NoBus\}$ and mean utility for the no bus option is normalized to zero. Here α is an amenity from the bus, γ is the marginal utility of income (or disutility from paying fares) and η is the value of time. p is the bus fare, and $t = t^V + t^W$ is total travel time (wait time plus in-vehicle time). If ϵ is drawn from a T1EV with unit shape, then

$$P(Bus) = \frac{\exp(\alpha - \gamma p - \eta t)}{1 + \exp(\alpha - \gamma p - \eta t)}.$$

The RCT provides exogenous variation in p that allows us to estimate $dP(Bus)/dp$ (i.e. γ) using maximum likelihood.

3 Wait Time RCT

The second RCT aims to measure η from the model above, i.e. individuals' sensitivity to time.

The experiment works as follows. Our team has developed an SMS-based app which allows us to measure participants' value of time at bus stops. After being recruited at their home on weekends (which is important, so as not to select participants with low value of time willing to speak with enumerators at bus stops during peak times), participants show up to their registered bus stop and find the enumerator waiting there. The enumerator will hold up a tablet, which displays a random code which changes every minute. The participant texts this code to our SMS-service, which then sends them back an offer to wait X minutes for a payment of Y Naira. To accept the offer, a participant will wait at the bus stop and then send back the new code that the enumerator's tablet displays after X minutes, allowing us to verify they had waited the requisite amount of time. If the participant does not accept the offer, they can continue with their day.

By observing the combinations of wait time and payments that are accepted and rejected, we are able to identify participants' value of time. Suppose a commuter is at a bus stop and we offer them to pay them s to wait for t^W minutes. The commuter decides whether to accept or reject the offer. She will accept the offer if

$$\alpha - \gamma(p - s) - \eta(t + t^W) + \epsilon(\omega) > \alpha - \gamma p - \eta t + \epsilon(\omega)$$
$$\Leftrightarrow s/t^W > \eta/\gamma$$

Our idea is to offer individuals different combinations of (s, t^W) in order to identify η/γ. The price experiment separately identifies γ, allowing us to recover η.

Our power calculations estimated this through maximum likelihood: To allow for different people will make different choices given the same bundles, we add a noise/idiosyncratic term so that on day t, the individual has a utility $\nu_{it} \sim N(0, \sigma)$ from taking the quicker option. Then

$$D_i = \mathbb{I}\left\{\gamma s - \eta t^W > \nu_{it}\right\}$$
$$\Rightarrow E\left[D_i\right] = \Pr\left(\frac{\gamma s - \eta t^W}{\sigma}\right).$$

The likelihood is as follows

$$\mathcal{L}(\gamma, \eta | D_{it}, s_{it}, t_{it}^W, \sigma) = \prod_{it}\left[\Pr\left(\frac{\gamma s_{it} - \eta t_{it}^W}{\sigma}\right)\right]^{D_{it}}\left[1 - \Pr\left(\frac{\gamma s_{it} - \eta t_{it}^W}{\sigma}\right)\right]^{1-D_{it}}$$

$$l(\gamma, \eta | D_{it}, s_{it}, t_{it}^W, \sigma) = \sum_{it} D_{it} \log \Pr\left(\frac{\gamma s_{it} - \eta t_{it}^W}{\sigma}\right) + (1 - D_{it}) \log\left[1 - \Pr\left(\frac{\gamma s_{it} - \eta t_{it}^W}{\sigma}\right)\right]$$

We will conduct extensions of this simple model:

1. Heterogeneity: allowing this to vary with income, either by adding in group- or income-specific coefficients (i.e. essentially re-estimating the same model for different groups) or by dividing the fare by worker income, which delivers a value of time proportional to income.

2. Individual-level Coefficients: We have included a sub-sample of 80 participants (out of the 400 we expect to play) which will play for 5 weeks (rather than 3 for everyone else). Our power calculations suggest this should be enough to uncover individual-level estimates of γ/η, which would allow us to recover the distribution non-parameterically. Otherwise we can estimate a random coefficients model a la BLP where we parameterize the distribution of these as normal and allow the parameters to vary across groups such as gender and income/education.

3. Non-Linear Value of Time: It is possible that people really dislike long waits, so we will estimate an extension which allows for a non-linear value of time.

Our power calculations suggest that power is maximized when the fraction of offers accepted is 0.5. This is intuitive: if all offers are rejected, for example, this only puts a bound on the value of time and does not allow estimation of a point estimate. While we have conducted pilots to get some information on users value of time, we will update the distribution of offers up to 3 times during the RCT to aim for a fraction of accepted offers closer to 0.5. We do this at the bus-stop level, and use other bus stops in the same “zone” (4 large geographic areas of the city) to compute users implied value of time (we index our distributions by average value of time (in terms of naira per minute), so if we see a fraction of offers accepted > 0.5 we will adjust to a lower naira per minute distribution). We use other similar bus stops to avoid previous bus stop participants’ choices affecting the future distribution of offers, but can conduct robustness where we compute standard errors via boostrap to account for this dependence.

4 Further Analysis

The first part of the project will provide reduced form evidence on the impact of the BRI on the informal network, and estimate a model of demand and supply for transit using the RCTs to measure the total impact of the system on consumer surplus and its distributional effects accounting for impacts on drivers. A key aspect of this approach is it will capture indirect effects of formal transit provision (since users of the informal system may be affected if prices and supply responds) and impacts on emissions (since changes in supply will impact emissions if danfo are particularly polluting, or if users near the margin of driving are induced to use the new formal system).

The second project will use the same RCTs to estimate a model of a decentralized transport network and characterize optimal policies to regulate prices and entry in the industry. This is an important question because the solutions of formal mass rapid transit like Bus Rapid Transit and Light Rail are very expensive and not a medium-run solution for improving public transit in Sub-Saharan Africa. The backbone is a dynamic queueing model which captures the various external effects present in the transit market (increasing returns to scale or "wait time" externality; business stealing; network effects through commuters combining multiple routes via transfers). The researchers will use the model to characterize how governments should regulate prices and entry to maximize societal welfare, inclusive of both consumer surplus and emissions.