Mobile Money in Tanzania

Online Appendix

Nicholas Economides[*]

Przemyslaw Jeziorski[†]

July 28, 2016

1 Data

This section of the appendix contains supplementary details of the data assembly process.

1.1 GPS location of the receivers

We are able to match 75% of the cellular IDs into fine-coded GPS positions. The remaining 25% of location IDs are inferred in the following way. For each sender, we obtain the previous active location (within a few hours), and match that location ID to the GPS data. The previous transactions can include any activity excluding being a receiver of the P2P transfer, for which we do not directly observe the location data. In this way, we are successful in matching 24% of the data. The remaining 1% is approximated by using the first few numbers of the location ID, which defines the location of sender with an error of a few miles. For our objective to distinguish short-, medium- and long-distance transfers, the matching procedure is reliable. Since we do not observe the location ID of the receiver, we have to approximate it. Using a procedure similar to one we used for the unmatched senders' IDs, we match the most recent transaction of the receiver. In the vast majority of cases, we are able to match

[*]Stern School of Business, NYU; NET Institute; economides@stern.nyu.edu

[†]Haas School of Business, UC Berkeley; przemekj@haas.berkeley.edu

the receiver within a few hours of receiving the transfer, usually cashing-out or checking the balance of the account.

2 First-stage estimation

We postulate the following semi-parametric model for the cash-out fees

$$c_t^m = \theta_0^m + \theta_1 b_t^m + \theta_2 (b_t^m)^2 + \tilde{\epsilon}_t^m, \tag{1}$$

where c_t^m is an incurred cash-out fee at day t by user m. The above equation can be classified as a censored regression model because cash-out fees may be zero if no cash-out actions are executed on day t, that is, c_t^m is censored at 0. In addition to censoring from below, the model is censored from above by the maximum cash-out fee of $5,000$. Both bounds are accounted for in the estimation. The censored regression specification embeds a semi-parametric model of cash-out frequency or $Prob(c_t^m > 0 | b_t^m)$, since it does not assume the distribution $\tilde{\epsilon}_t^m$. In addition, conditional on executing a cash-out, the model predicts cash-out size in a similar way as a linear regression model.

The term θ_0^m represents the baseline propensity to cash-out of user m. If θ_0^m is high, the user prefers cash over m-money and cashes out frequently. If θ_0^m is low, the user tends to keep his money in the mobile wallet. In the data, the cash-out patterns vary from user to user, therefore it is necessary to model heterogeneity in the marginal cash-out propensity θ_0^m.

The term θ_1 represents the marginal effect of the account pre-cash-out balance on the cash-out propensity and intensity. Additionally, the term θ_2 models a second-order effect of the balance on the cash-out. We stipulate that $\theta_1 > 0$. Thus, if $\theta_2 = 0$, high balances would be associated with incurring proportionally higher average cash-out fees. However, because the cash-out pricing schedule is concave and carrying higher balances may lead to greater usage of the platform in general, users with higher balances may incur lower cash-out fees per Shilling. For this reason, the effect of the balance on the cash-out fees is likely to be non-linear, and we predict that $\theta_2 < 0$.

By construction, θ_0^m is correlated with the current balance b_t^m because clients with high θ_0^m cash-out frequently and tend to carry low balances. We allow for this correlation by modeling θ_0^m as fixed effects. This specification, however, introduces an incidental parameters issue because the model becomes non-linear after accounting for censoring.[1] To overcome this problem, we use a GMM estimator developed by Honoré (1992) and extended by Alan et al. (2014) for two-sided truncation.

Next we use equation (1) to obtain estimates of the cash-out fee imposed on receiver m by sender n

$$e_t^n(a_f, b_t^m) = \mathbb{E}[\text{daily cash-out fees}|b_t^m + a_f] - \mathbb{E}[\text{daily cash-out fees}|b_t^m], \tag{2}$$

We employ a result from Honoré (2008) that establishes a way to compute marginal effects in the censored regression model with fixed effects. We are interested in the marginal effect $\frac{\partial E[c_t^m|b_t^m]}{\partial b_t^m}$, where the conditional expectation is taken with respect to ϵ keeping constant the fixed effect θ_0^m. We compute the marginal effect of the transfer, which determines the cash-out externality in the following way:[2]

$$e_t^n(a_f, b_t^m) \approx (\theta_1 + 2\theta_2 b_t^m) a_f P(c_t^m > 0). \tag{3}$$

The marginal effect of transferring one Shilling is composed of two parts: (i) a first-order effect $\theta_1 a_f$, capturing the effect of a higher account balance, and (ii) a second order effect $2\theta_2 b_t^m a_f$, capturing lower fees per Shilling (economies of scale) when cashing-out large transfers.

3 Larger discontinuity window

This section of the appendix presents results with a larger discontinuity window. We estimated the P2P transfer model without differentiating the location of the origin over a window spanning 2 weeks before and 2 weeks after the price change. The estimates of the P2P transfer model are presented in Table 1. We observe negligible change in coefficients compared to a model estimated on the shorter window. In particular, the transfer price coefficient is estimated to be -0.71 in the longer window, compared to an average price coefficient of $-1.17 + 0.56 \times 0.4 = -0.95$ estimated using a model in the shorter window.

[1] We would encounter a similar issue if we modeled cash-out propensity using, for example, a logit model with fixed effects.

[2] Note that we use $P(c_t^m > 0)$ instead of $P(c_t^m > 0|b_t^m)$, because for small b_t^m the probability of cashing-out is close to zero which would produce a distortion in the estimates of the externality.

Price coef. - mean ($\bar{\alpha}$)	-0.71 (0.025)
Cash out coef. - mean ($\bar{\beta}$)	-43.33 (1.0598)
Price coef. - std. dev. (σ_α)	0.54 (0.031)
Cash out coef. - std. dev. (σ_β)	0.0216 (0.0010)
Transfer size coef. - std. dev. (σ_κ)	0.2174 (0.0026)
Price coef./Cash out coef. - Pearson correlation ($\rho_{\alpha\beta}$)	0.46 (0.060)
Price coef./Transfer size coef. - Pearson correlation ($\rho_{\alpha\kappa}$)	-0.58 (0.022)
Cash out coef./Transfer size coef. - Pearson correlation ($\rho_{\beta\kappa}$)	-0.57 (0.033)
Covar. of ϵ for 2 adjacent transf. (ρ_ϵ)	0.06 (0.003)

Table 1: Estimates of the P2P transfer model without location origin on the longer time window.

4 Supplementary graphs and tables

This section of the appendix contains supplementary graphs.

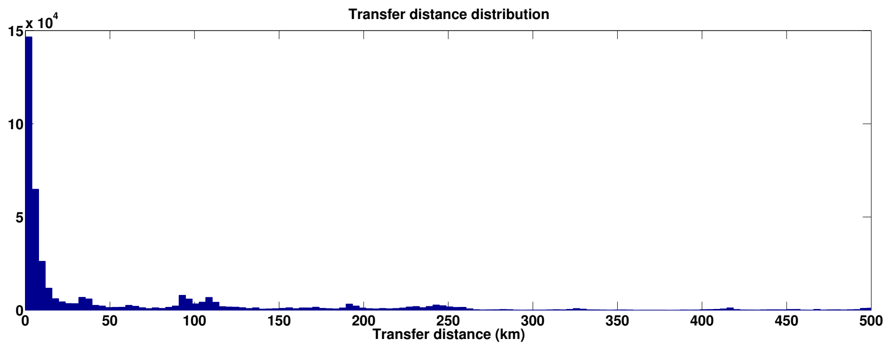

Figure 1: Distribution of the geographical distance of transfers.

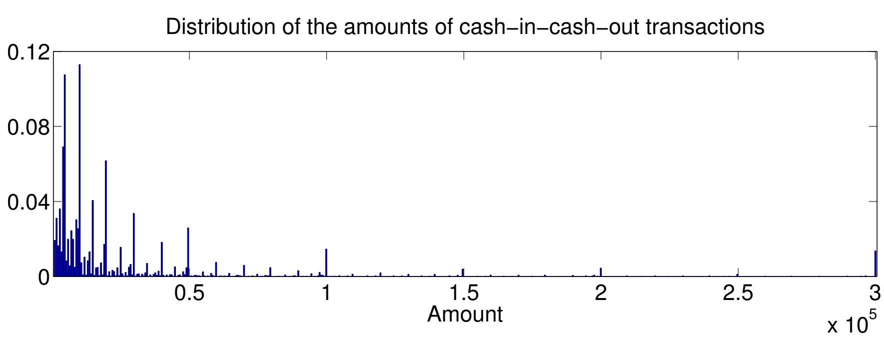

Figure 2: Empirical distribution of the cash-in-cash-out transaction sizes.

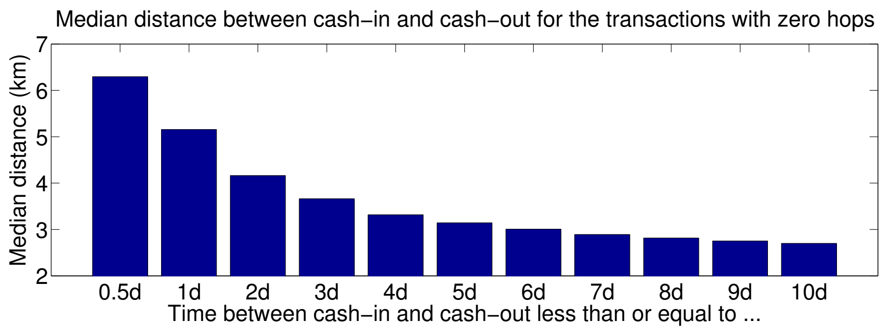

Figure 3: Median distance between cash-in and cash-out for transactions with zero hops as a function of the lifetime that the money stayed in the network, for long lifetimes.

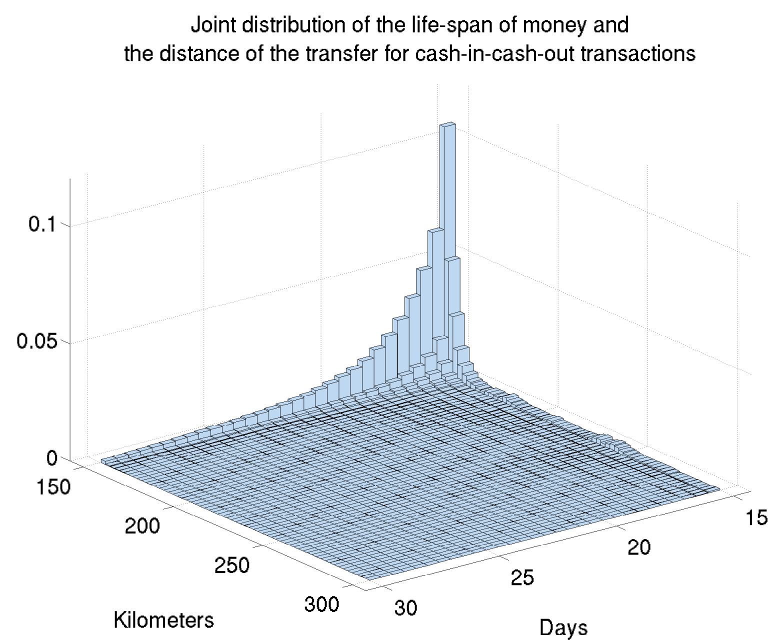

Figure 4: Joint distribution of the time and distance between cash-in and cash-out for the transactions with zero hops.

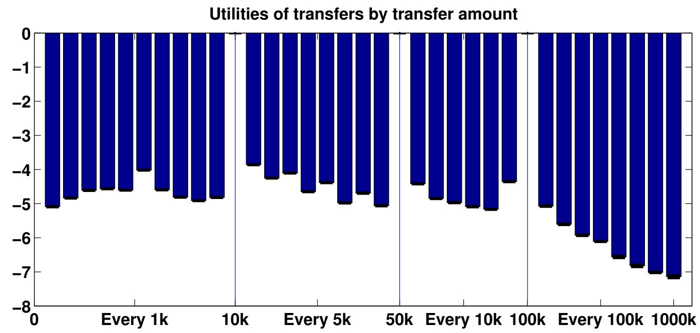

Figure 5: Transaction size fixed effects in the utility function for transfers.

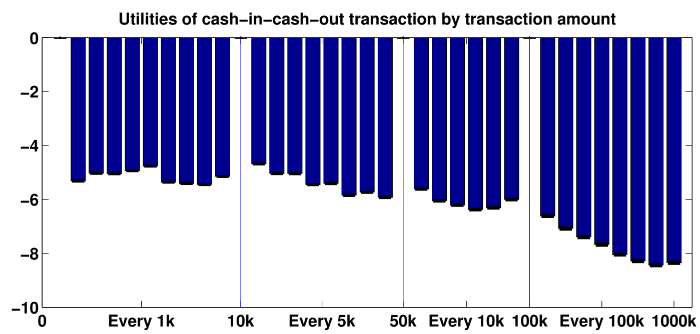

Figure 6: Transaction size fixed effects in the utility function for transportation and storage.

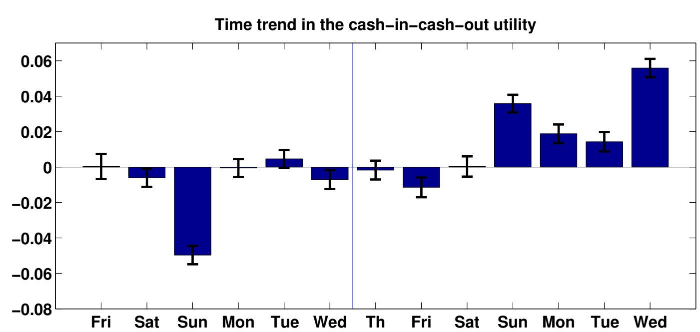

Figure 7: Trend in the utility function for transportation and storage.

References

Alan, S., B. E. Honoré, L. Hu, and S. Leth-Petersen (2014): "Estimation of panel data regression models with two-sided censoring or truncation," *Journal of Econometric Methods*, 3, 1–20.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533–65.

——— (2008): "On marginal effects in semiparametric censored regression models," *Available at SSRN 1394384*.