informs
http://pubsonline.informs.org/journal/mksc/

MARKETING SCIENCE
Articles in Advance, pp. 1–23
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Mobile Money in Tanzania

Nicholas Economides,[a,b] Przemyslaw Jeziorski[c]

[a]Stern School of Business, New York University, New York, New York 10012; [b]Networks, Electronic Commerce and Telecommunications Institute, New York, New York 10012; [c]Haas School of Business, University of California, Berkeley, Berkeley, California 94720

Contact: economides@stern.nyu.edu (NE); przemekj@haas.berkeley.edu (PJ)

Received: July 31, 2015
Revised: April 22, 2016; July 28, 2016
Accepted: September 19, 2016
Published Online in Articles in Advance:

https://doi.org/10.1287/mksc.2017.1027

Abstract. In developing countries, mobile telecom networks have emerged as major providers of financial services, bypassing the sparse retail networks of traditional banks. We analyze a large individual-level data set of mobile money transactions in Tanzania to provide evidence of the impact of mobile money on alleviating financial exclusion in developing countries. We identify three types of transactions: (i) money transfers to others, (ii) short-distance money self-transportation, and (iii) money storage for short to medium periods of time. We utilize a natural experiment of an unanticipated increase in transaction fees to identify the demand for these transactions. Using the demand estimates, we find that the willingness to pay to avoid walking with cash an extra kilometer (short-distance self-transportation) and to avoid storing money at home (money storage) for an extra day are 1.25% and 0.8% of an average transaction, respectively, which demonstrates that mobile money ameliorates significant amounts of crime-related risk. We explore the implications of these estimates for pricing and demonstrate the profitability of incentive-compatible price discrimination based on type of service, consumer location, and distance between transaction origin and destination. We show that differential pricing based on the features of a transaction delivers a Pareto improvement.

History: K. Sudhir served as the editor-in-chief and Juanjuan Zhang served as associate editor for this article.
Funding: The authors thank the Bill and Melinda Gates Foundation for financial support and the University of Chicago, Using Mobile Phones as a Platform for Banking [Award 035250-002].
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2017.1027.

Keywords: mobile money network • financial exclusion • transaction costs • Tanzania • banking • social network • price discrimination • location-based price discrimination • crime

1. Introduction

Developing economies are often characterized by *institutional voids* (see Khanna and Palepu 1997, Narasimhan et al. 2015, Sudhir et al. 2015), that is, deficiencies in their institutional and regulatory environments compared to developed economies. These deficiencies often result in insufficient infrastructure and underdevelopment of core industrial sectors that are crucial for economic growth. Recent technological innovations created new markets and institutions (see Aker 2010, Jack and Suri 2014) and revealed opportunities to mitigate institutional voids. Despite the potential significance of alleviating institutional voids by the new technology-enabled markets, there is only scarce economic and marketing literature studying the efficacy of this process. Moreover, little is known about which regulatory and firm policies are effective in facilitating this transition.[1]

This paper studies the impact of a new banking technology on a particular type of institutional void, that is, *financial exclusion*—the dearth of banking services in developing regions caused by significant deficiencies in financial infrastructure. Financial exclusion is seen as an important impediment to growth in developing countries (see Schumpeter 1912, King and Levine 1993, Levine 1997). Filling the infrastructure gap, the rapidly expanding mobile phone networks introduced mobile money (m-money) wallets. These mobile wallets are attached to the cell phone numbers of customers and provide many functions of traditional bank accounts. Mobile money allows consumers to perform financial transactions in a relatively inexpensive and reliable way, potentially augmenting money liquidity and ameliorating crime-related risk.

We provide evidence on the impact of m-money on alleviating financial exclusion using a unique data set of millions of m-money transactions between December 2012 and January 2013. Transaction data were procured from Tigo, the second largest telecommunication network in Tanzania—a market well suited to study financial exclusion, since banking infrastructure is particularly deficient in sub-Saharan Africa (see Demirgüç-Kunt and Klapper 2012). We analyze the transaction patterns and classify the uses of m-money into those that improve financial liquidity and those that alleviate the risk of exposure to different types of crime, such as street crime and burglaries. We explore the implications of a natural experiment created by an exogenous and unanticipated increase in the transaction fees. We make two contributions.

First, we identify and estimate willingness to pay (WTP) for three modal types of transactions executed using the mobile money network. In particular, consumers use the network for peer-to-peer transfers, secure money self-transportation, and secure money storage.[2] We obtain the estimates of the WTP for the above transactions using a structural model of supply and demand. The identification of causal effects within our model is obtained using a discontinuity approach, examining transaction propensity within a short time horizon before and after an exogenous fee increase.[3] We use the estimates of the WTP to quantify which dimensions of financial exclusion and risk amelioration are affected by the advent of mobile money. In particular, the WTP for peer-to-peer transfers informs us on the extent of economic harm caused by low financial liquidity and on the extent of its alleviation. Similarly, the WTP for secure money transportation and storage informs us about the economic harm generated by the high level of crime and the provision of risk amelioration thereof.

Second, we show that consumers have heterogeneous WTP and price sensitivity for different types of transactions. The demand curves vary depending on the location of transaction origin as well as on the physical and temporal distance between transaction origin and destination. Using a model of supply combined with detailed data on marginal cost, we generate insights into the profitability of several novel pricing strategies. We develop incentive-compatible pricing strategies that enable differential pricing across the three types of provided services depending on the location of the transaction origin and the distance between transaction origin and destination. We show that our pricing strategy delivers higher profits to the network. Moreover, we show that such differential pricing delivers a Pareto improvement; that is, it increases both the network's profit and consumer surplus.

1.1. Description of the m-Money Network

A mobile money account can be best described as a checking account associated with a mobile phone number. Users can cash in and cash out money from the account using a dense network of local agents serving as ATMs. Additionally, users can perform cash-free transactions, such as peer-to-peer transfers, using a mobile phone with mere support of legacy short message service technology. Mobile money networks differ from traditional banks by having significantly lower capital and institutional barriers to their creation and operation because (i) they leverage the existing dense cell phone network and (ii) the only capital requirement to become an agent, who facilitates deposits and withdrawals, is to have a mobile phone, and therefore the networks do not need to invest in costly banking branches. Consequently, mobile money networks can offer a viable alternative to traditional banking. Transfers across users within the same m-money network are relatively inexpensive (the fee is 1.1% on average), while conversion from m-money to cash ("cashing out") is relatively expensive (the fee is 7.3% on average).

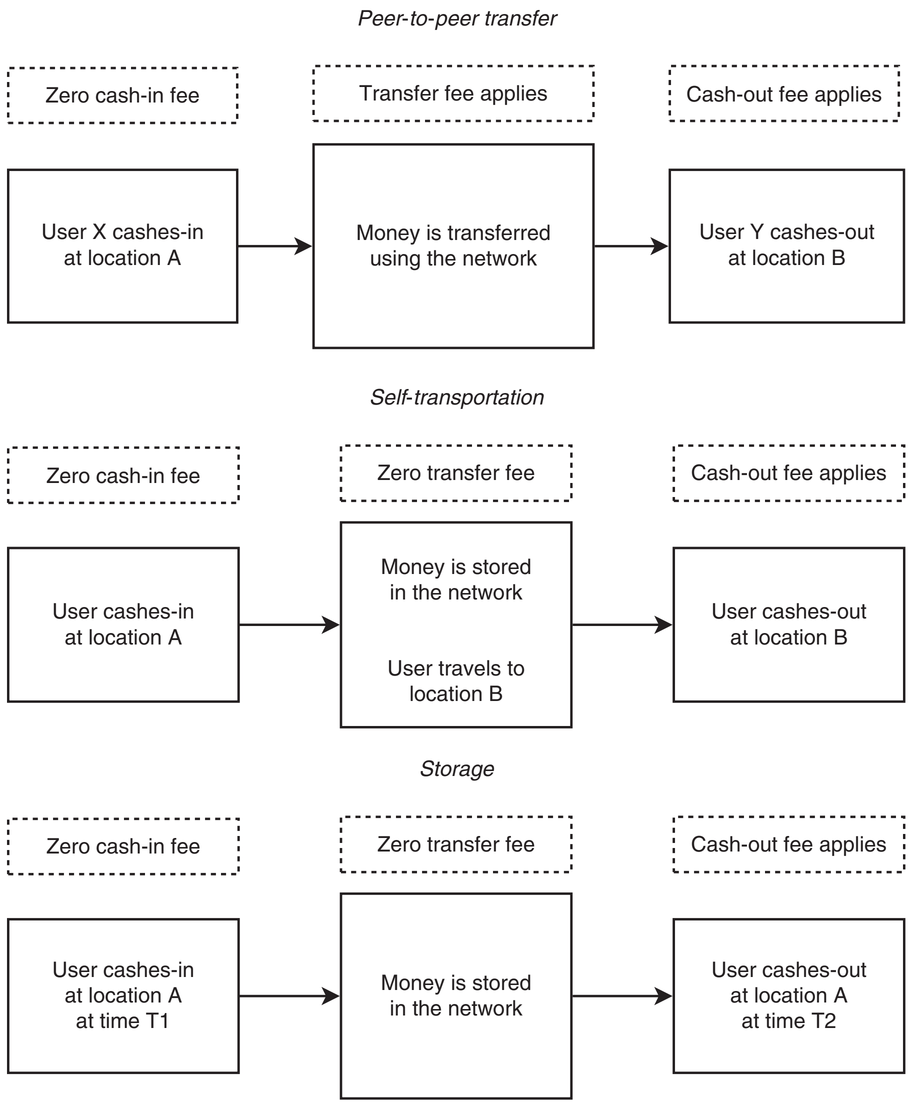
In Tanzania, mobile banking has significant adoption, that is, almost 35% of households have at least one m-money account. Thirty-two percent of the population uses exclusively m-money as a provider of financial services, and only 2% have an active traditional bank account.[4] There are three major m-banking networks: Vodacom (Vodafone), with 53% market share in m-money; Tigo, with 18% market share; and Airtel, with 13% market share. During the span of our data, it was impossible to send money across mobile networks,[5] and making telephone calls across mobile networks was relatively expensive. Thus, many consumers have a different phone appliance or subscriber identification module (SIM) card for each phone network that they substitute in the same phone. Additionally, Tigo advertises phone appliances that take multiple SIMs.[6] Some consumers use m-money for business transactions: 21% of Vodacom M-Pesa users do, as do 12% of users of Tigo and Airtel.

1.2. Summary of Results

As we have noted, there are three uses of m-money networks that resolve different dimensions of financial exclusion (see Figure 1). The first function is the ability to execute instantaneous peer-to-peer (P2P; person-to-person) transfers, compared to the alternatives of transporting money in person, using a bus driver or using Western Union. P2P transfers are predominantly used to transfer remittances from urban areas to rural locations.[7] Approximately 30% of users with a Tigo account make at least one transfer a week.

We find that, on average, the consumers (that is, senders) of P2P transfers are price inelastic and that they are heterogeneous and respond quite differently to changes in fees. In particular, customers who execute large transactions are usually more price inelastic than consumers who execute smaller transactions, possibly due to income effects. Additionally, we find that demand for long-distance transfers is less elastic than that for short-distance transfers. The difference in demand elasticities is consistent with the prices of the traditional alternatives, such as using a bus driver, being higher for long-distance transfers than for short-distance ones.[8] Thus, despite significantly lower transaction fees when using any of the m-money networks, we believe that many marginal customers are choosing between antiquated money transfer means and a particular m-money network (in this case, Tigo), rather than

Figure 1. The Types of Transactions and the Applicable Fees



choosing between two competing m-money networks. Indeed, more than 70% of Tigo users report that they have never used another network (see InterMedia 2013).

The second function of the m-money network is to securely carry money for short distances, typically up to 10 kilometers (km). These transactions involve depositing to the m-money account (cashing in) using a local agent and withdrawing the money at another location (cashing out), without making a P2P transfer. We find that such self-transportation is common in Tanzania, as 13% of transactions do not involve a transfer and are characterized by less than one day between cash-in and cash-out, with a median distance of 8.7 km between the two. These transactions are aimed at minimizing the risk of being robbed while walking with medium and large amounts of cash. To our knowledge, such transactions have not been identified before in the economics literature. Transactions incur a relatively high average fee of 7.3%, which suggests that the risk of walking with cash is substantial and that mobile money mitigates a considerable amount of that risk.

The third function of the network is to store money for short and medium periods of time. Compared to peer-to-peer transfers and transportation transactions, savings transactions are not very prevalent, as 90% of the money leaves the network within five days of being cashed in. Moreover, less than 1% of users keep the money in the network for longer than a month, indicating a low contribution of mobile money transactions to long-run savings rates.[9] We find evidence of heterogeneous use of the above services. In particular, transporting and storing money are positively correlated with sending a transfer, and negatively correlated with receiving a transfer.

Consumers executing transportation and storage transactions are, by contrast to the consumers of P2P transactions, moderately price elastic. This difference may be related to higher urban penetration of transportation and storage transactions, which results in better access to substitutes. We find that consumers are willing to pay up to 1.24% of the transaction amount to avoid walking an extra kilometer carrying cash. Urban customers are willing to pay up to 2.75% per kilometer, while rural customers only up to 0.3%. Similarly, users are willing to pay up to 0.8% to avoid storing cash at home for an extra day (1.25% for urban and 0.25% for rural). Thus, we provide the first monetary estimates of an economic harm, based on revealed preferences from field data, caused by high levels of street crime and burglaries.[10] These estimates of harm suggest a significant loss of welfare arising from poor law enforcement in Tanzania and are consistent with an extremely high crime rate as reported by the United Nations Office on Drugs and Crime (2009) and the Financial Inclusion Tracker Surveys project (InterMedia 2013). High estimated levels of WTP for safe money transportation suggest both that crime risk is high and, more importantly, that m-money ameliorates this risk to a significant extent. In addition, relatively high WTP and the popularity of transportation and storage transactions in urban areas are evidence that these areas are particularly affected by crime risk, or that crime risk amelioration using mobile money is more effective in urban locations, because of denser agent networks in cities than in villages.[11]

For P2P transfers, we find that a sender takes into account, to some extent, the cash-out fee paid by the receiver. This suggests that the incompatibility of m-money with other forms of money, including m-money of other phone companies, as evidenced by high cash-out fees, has a negative effect on the propensity to make P2P transfers. The network would realize a higher revenue from P2P transfers and subsequent cash-outs if it decreased cash-out fees for the receiver and simultaneously increased transfer fees. However, because some users use the network to transport or store money without making a transfer, decreasing the overall cash-out fees would decrease revenue from such transactions. This decrease would lead to an overall reduction in the network's profit, with users who transfer money subsidizing users who do not transfer. We propose a feasible and incentive-compatible price discrimination strategy that solves this problem. In this pricing strategy, the network would charge a zero cash-out fee for withdrawals that do not exceed a recently received transfer amount. For all other cash-outs, the network would charge a positive cash-out fee that is slightly smaller than the transfer fee. This pricing scheme, coupled with a simple fixed mark-up pricing of transfers, delivers a Pareto improvement in which both the network's profit and consumer surplus from transfers, transportation, and storage are higher than under the current pricing.

We find that it would be profitable for the network to charge different prices depending on the location of the transaction origin. We propose to measure the profitability of price discrimination on the margin (for small price changes) in a way that is robust to typical assumptions on pass-through rates and forms of supply-side competition (see Berry et al. 1995, Fabinger and Weyl 2015). Our measure, which we call differential pricing pressure (DiPP), measures the impact on profitability of marginally decreasing prices on the more elastic segment of the demand and simultaneously increasing prices on the less elastic segment of the demand. We find that decreasing the cash-out fee charged to rural-originating transportation/storage transactions by 1% and increasing it in urban-originating transactions by 1% increases profits by 3.1%. We also find that decreasing the transfer fee charged to rural-originating transfer transactions by 1% and increasing it in urban-originating transactions by 1% has a much smaller impact, increasing profits by 0.22%.

Next, we study pricing policies that depend on the consumers' network topology. Since the WTP depends on the distance between a transaction's origin and destination, we find that lowering the fees for short-distance transfers by 1% and increasing the fees for medium-distance transfers increases profits by 0.01% (0.03% when the fee increase is for long-distance transfers). The corresponding profitability increases for transportation/storage transactions are 0.6% after lowering short-distance fees and increasing medium-distance fees (2.58% for long-distance fees). Thus, location-based price discrimination is more profitable for transportation/storage transactions. This can be explained by the fact that the demand for transportation/storage transactions depends on crime rates, which vary by location and are particularly high in urban areas (see Kinabo 2004, Louw et al. 2001).

Our results are related to the literature on heterogeneity of consumer preferences (see Allenby and Rossi 1998) and spatially distributed preferences (see Yang and Allenby 2003, Bronnenberg and Mahajan 2001). The latter literature considers markets in which consumers' preferences are correlated depending on location, which is econometrically accommodated using various random effects specifications. In this work, we leverage the detailed location data of each consumer and propose a fixed effects specification accommodating spacial correlation in demand. After estimating this specification, we measure the implications of spatial heterogeneity of preferences and network topology for pricing.[12]

2. Data

The data set contains all mobile financial transactions among subscribers of a major cellular phone service provider in Tanzania for the months of December 2012 and January and February 2013. Each record contains

Table 1. Change in the Nonlinear Fee Schedule for Transfers

	Transfer bands		
	Tigo		
Transfer fee	Before Jan. 24	After Jan. 24	Vodacom
10	200–999	200–999	
20	1,000–1,999	1,000–1,999	
30	2,000–2,999	2,000–2,999	
40	3,000–3,999	3,000–3,999	
50	**4,000–9,999**	**4,000–4,999**	200–9,999
100		**5,000–9,999**	
200	**10,000–49,999**	**10,000–19,999**	10,000–49,999
250		**20,000–49,999**	
500	50,000–299,999	50,000–299,999	50,000–299,999
1,000	300,000–499,999	300,000–499,999	300,000–499,999
1,500	500,000–1,000,000	500,000–1,000,000	500,000–1,000,000

Notes. All numbers are in Tanzanian shillings. Bold values highlight the price change.

a user ID, balance before and after the transaction, the type of transaction, and the location identifier. There are six types of transactions: peer-to-peer transfer, balance check, cash-in, cash-out, bill payment, and recharge of the phone account. We focus our attention on P2P and cash-out transactions.[13] To distinguish among short-, medium-, and long-distance transfers, we match the mobile transaction data with GPS data on cellular location IDs in Tanzania obtained from locationapi.org (see Section 1.1 of the online appendix).

As mentioned earlier, P2P transfers and cash-out transactions carry fees. According to anecdotal evidence from Tanzanian sources, changes in these fees are infrequent and unanticipated. In the data, we observe one such change on January 24, 2013, and we report the tariffs before and after the change in Tables 1 and 2. We utilize these tariffs' change to identify the demand for m-money transactions. We use the full three-month sample to identify the population of m-money users. We define a user as a person who made at least one transaction between December 2012 and February 2013. To minimize potential price endogeneity issues, we employ a discontinuity strategy and utilize only transactions that occur a week before and a week after the price change. We utilize the fact that the price change acts as a quasi-experimental variation under the assumption that the exact timing of the tariff increase was uncorrelated with other time-varying demand factors. Thus, the period immediately before the price change acts as a control group, and the period immediately after the price change acts as a treatment group. Consequently, the average effect of the price increase informs us about the elasticity of the demand. In addition, we explore the data variation generated by the fact that not all transaction types experienced a price increase. In particular, we are able to identify a common time trend in m-money transactions separately from the effect of the fee increase using a method of moments approach, which explores the same data variation as a difference in difference regression.

Table 3 reports aggregate statistics about the executed P2P transfers. We utilize a subsample of 1,400,644 unique consumers observed over the period of 14 days. We include customers that are registered but made no transactions over the focal 14 days, which minimizes the self-selection problem usually present in data sets that include only executed transactions.[14] The average transfer is approximately 38,000 Tanzania shillings (TSh), or about $24 USD. The transfers are large compared to the 2007 average monthly consumption in Tanzania of 58,000 TSh (see the United Nations Office on Drugs and Crime 2009) and average monthly consumption of m-money users of 130,000 TSh (see InterMedia 2013). Moreover, the transfers are quite dispersed, with a standard deviation of 84,000 TSh. These statistics suggest that a large portion of a user's household income is channeled through a m-money platform. Figure 2 contains the histogram of all Tigo P2P transfers. We note that the amounts are clustered

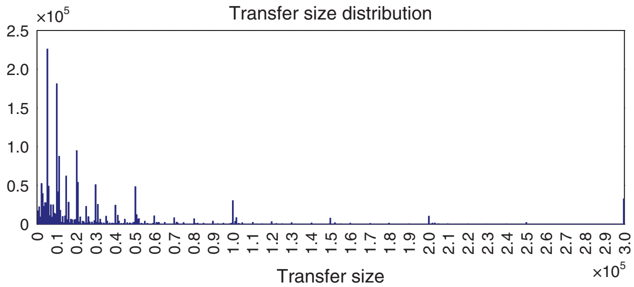
Table 2. Change in the Nonlinear Fee Schedule for Cash-Outs

	Cash-out bands	
Cash-out fee	Before Jan. 24	After Jan. 24
500	1,000–9,999	1,000–4,999
600		5,000–9,999
800	10,000–19,999	
1,000	20,000–49,999	10,000–24,999
1,250		25,000–49,999
1,400	50,000–99,999	
1,500		50,000–99,999
1,800	100,000–199,999	
2,000		100,000–199,999
3,000	200,000–299,999	200,000–299,999
4,000	300,000–399,999	300,000–399,999
5,000	400,000–1,000,000	400,000–1,000,000

Table 3. Aggregate Statistics About P2P Transfers in the Tigo Network

Unique customers	1,400,644
Mean transfer size	38,751
Std. dev. of transfer size	83,954
Percentage of day-senders with no transfer	95.5%
Percentage of day-senders with 1 transfer	3.7%
Percentage of day-senders with 2 transfers	1.1%
Percentage of day-senders with 3+ transfers	0.02%
Percentage of day-receivers with no transfer	95.0%
Percentage of day-receivers with 1 transfer	4.7%
Percentage of day-receivers with 2 transfers	0.39%
Percentage of day-receivers with 3+ transfers	0.01%
Mean transfer distance (km)	69.2
Std. dev. of transfer distance (km)	109.5
Mean transportation/storage size	27,798
Std. dev. of transportation/storage size	67,911

Figure 2. (Color online) Distribution of the Amount Transferred in Tanzanian Shillings



at the edges of the intervals, suggesting that choosing the transfer amount would be more reasonably approximated by a discrete rather than a continuous choice. Furthermore, on an average day, approximately 4% of all people with a Tigo m-money account make one transfer, and approximately 1% make two transfers. The high frequency of transfers coupled with large transfer sizes suggests that part of the household money flow may not be captured by the official consumption figures.

The average geographic distance of a transfer is 70 km, with a standard deviation of 110 km. A high percentage of transfers are short distance; however, we also observe a fat tail of transfers that span hundreds of kilometers. The full distribution of transfer distances can be found in Figure 1 in the online appendix. The left panel of Figure 3 presents a logarithmic frequency of transfer originations by geographic location. The transfer origination points are concentrated in more developed areas, such as Dar es Salaam and Zanzibar. We also observe significant transfer activity along roads and rivers. The right panel of Figure 3 depicts the topology of the first 1,000 transfers in the data. The star network topology confirms that Dar es Salaam and Zanzibar are the main sources of the money flow. We find that 90% of extremely short-distance transfers (less than 10 meters between the sender and receiver) occur at the center of the star in Dar es Salaam or Zanzibar. Moreover, receivers of such transfers carry an average pretransfer balance of 50,000 TSh, whereas the average balance of the receivers that are further than 100 km from the sender is only 25,000 TSh. Thus, we believe that the short-distance transfers are likely related to commercial activity in the urban areas.

In the following section, we present the descriptive analysis of the data.

3. Network Description and Reduced-Form Analysis

Because we observe all financial transactions within the network, we can track an m-money shilling from its birth (cash-in) to its death (exit from the network). For this purpose, we tag an m-money shilling cashed into the network at the beginning date of our data (December 1, 2012) and simulate its path through the network. We find that only about 1% of shillings do not exit the network during the three months of our data. The money leaves the network relatively quickly, with 60% of shillings being cashed out within two to three days of being cashed in (see Figure 4). The money exits the network predominantly in the form of regular cash-outs (about 85% of cases). Other exit ways include topping up of the mobile telecom account (7%), electricity bill payment (6%), and other in-network purchases such as telecom data bandwidth packages (1%).

Figure 5 shows the distribution of the number of P2P transfers (hops) during the lifetime of a shilling in the Tigo network since cash-in. About 35% of the

Figure 3. (Color online) Logarithmic Heat Map of Number of Transfers by Origin and Graph Representation of First 1,000 Transfers in the Data

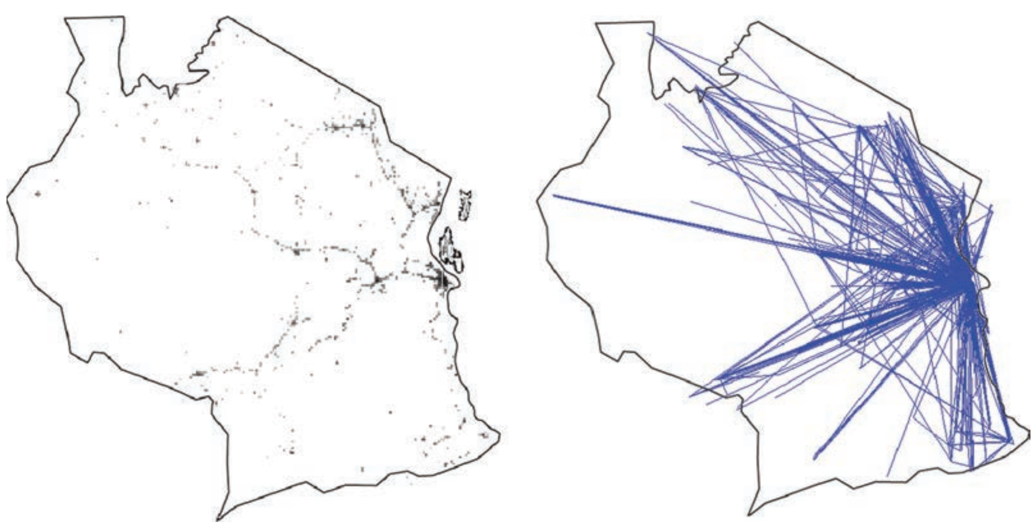
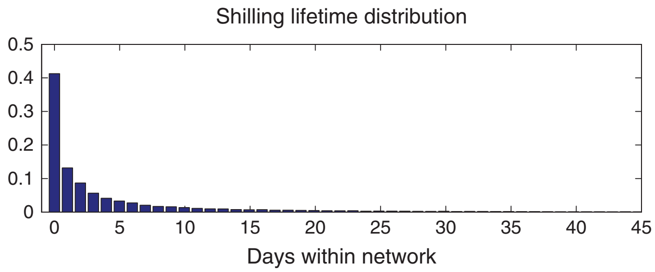


Figure 4. (Color online) Distribution of the Lifetime of a Shilling in the Tigo Network from Cash-In to Exit



money leaves the network without a transfer. Over 50% of the money in the network gets transferred only once and then leaves the network, 10% exits after two hops, and the percentage that stays for three or more hops is miniscule (see Figure 5). As we noted earlier and depict in Figure 4, the lifetime of money in the network is short.[15] This suggests that the m-money network is used predominantly for one-time transfers and not as a full-fledged financial system where the money would circulate extensively from user to user. We come back to this finding when discussing pricing experiments and counterfactuals later in Section 7.

We now investigate the instances in which money leaves the network without ever being transferred (cash-in–cash-out or zero-hop transactions). Approximately 35% of the funds entering the network do not get transferred to anyone, but get cashed out by the original depositor (who has to pay the cash-out fee). The histogram of the transaction amounts for zero-hop transactions is presented in Figure 2 in the online appendix. We find that this distribution is quite similar to that of peer-to-peer transfers, suggesting that the relative sizes of the two types of transactions are comparable. For zero-hop transactions, we find that the median time the money stays in the network (lifetime) is approximately 2.5 days. We present the full histogram of the lifetime of money for these transactions in Figure 6. The money is cashed out within three hours in 17% of the cases, while it stays in the network more than 10 days in 20% of the cases. Thus, we believe that cash-in–cash-out transactions serve two

Figure 5. (Color online) Distribution of Number of P2P Transfers During the Lifetime of a Shilling in the Tigo Network Since Cash-In

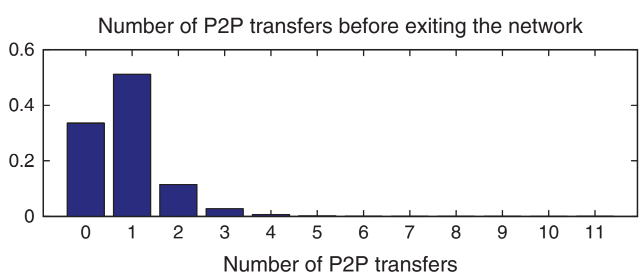
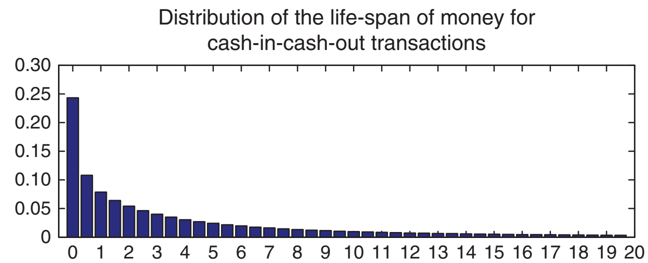


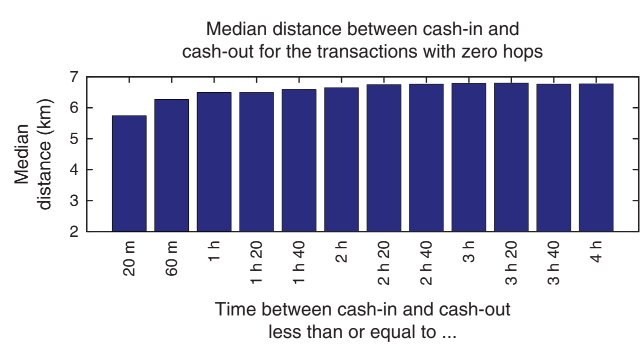
Figure 6. (Color online) Distribution of the Time Between Cash-In and Cash-Out for the Transactions with Zero Hops



distinct purposes: (i) as a short-distance, short-lifetime money transportation vehicle and (ii) as a short- or medium-term savings account. The short-lifetime cash-in–cash-out transactions occur primarily within the capital city of Dar es Salaam. According to the United Nations Office on Drugs and Crime (2009), nearly 40% of the 1,884 surveyed households in Dar es Salaam reported that they were victims of theft or robbery in 2007. These crimes included a large share of car, bicycle, or motorcycle hijackings. Thus, even driving with the monthly paycheck in cash may not be advisable.[16] Using a cash-in–cash-out strategy can serve as viable insurance against money loss, and is facilitated by the high density of the agents' network in the capital city.

To investigate zero-hop transactions further, we compute the median distance between cash-in and cash-out, conditional on transaction lifetime. Median distances for smaller lifetime values reflect predominantly money transportation, while the median distances for longer lifetimes reflect long-term savings. The medians for transactions with a lifetime smaller than a few hours are presented in Figure 7. We observe that the median travel distance between cash-in and cash-out is increasing as a function of lifetime.[17] In particular, the transactions with extremely short lifetimes of less than 20, 40, and 60 minutes have median traveling distances of 5.2, 6, and 6.2 kilometers, respectively. In these cases, the user cashes in, boards a means of transportation,

Figure 7. (Color online) 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 Short Lifetimes

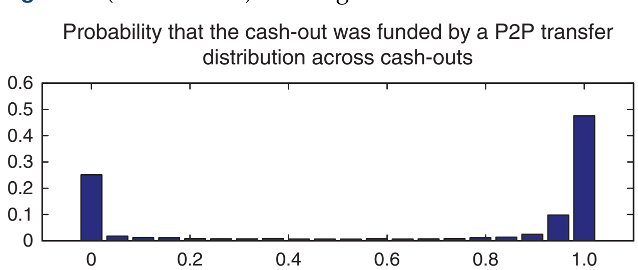


and immediately cashes out on arrival.[18] Considering that the average cash-out fee is approximately 7%, the prevalence of cash-in–cash-out transactions can be rationalized by at least a 7% probability of being robbed, which is modest in light of the aforementioned crime statistics.[19] By contrast, the median traveling distance for zero-hop transactions with longer lifetimes is decreasing in the cash-in–cash-out time interval (see Figure 3 in the online appendix). This reflects the fact that many longer lifetime transactions are not executed to physically transport cash, but rather use the network to keep savings.[20] Examining the joint distribution of distance and lifetime (depicted in Figure 4 of the online appendix) uncovers the existence of transactions with relatively long distance and long lifetime that serve both savings and transportation purposes. The existence of such transactions suggests that savings and transportation should be modeled jointly.

The above discussion suggests that a cash-out preceded by a transfer can be considered as a separate product from a cash-out not preceded by a transfer. To investigate this further, we compute the probability of a cash-out originating from a peer-to-peer transfer rather than from a cash-in by the same customer.[21] As shown in Figure 8, the distribution of funding sources of cash-outs is bimodal; that is, the cashed-out money is funded almost entirely by either cash-ins or peer-to-peer transfers. Exploiting this dichotomy, we label cash-outs funded more than 50% by a cash-in as cash-in–cash-out transactions.[22] In the remainder of this paper, we analyze the demand for two products: first, a transfer product consisting of a cash-in, peer-to-peer transfer, and cash-out, and second, a transportation/savings product consisting of a cash-in followed by a cash-out by the same user. The transfer product and the savings product resemble retail banking products in developed countries; however, the transportation product is unique to Tanzania, and possibly other developing countries. As far as we know, it has not been identified before in the literature.

We find a positive correlation between the number of send transfers and the number of transportation and storage transactions (Pearson correlation coefficient of 0.1 with $p < 0.001$). However, we find a negative correlation between the number of received transfers and the number of transportation and storage transactions (Pearson correlation coefficient of −0.07 with $p < 0.001$). This indicates that people using the network to transport and store money are also initiating transfers, but the people who receive transfers are unlikely to transport and store. These correlations reflect that the senders are usually located in urban areas, where transportation and storage are relatively more popular.

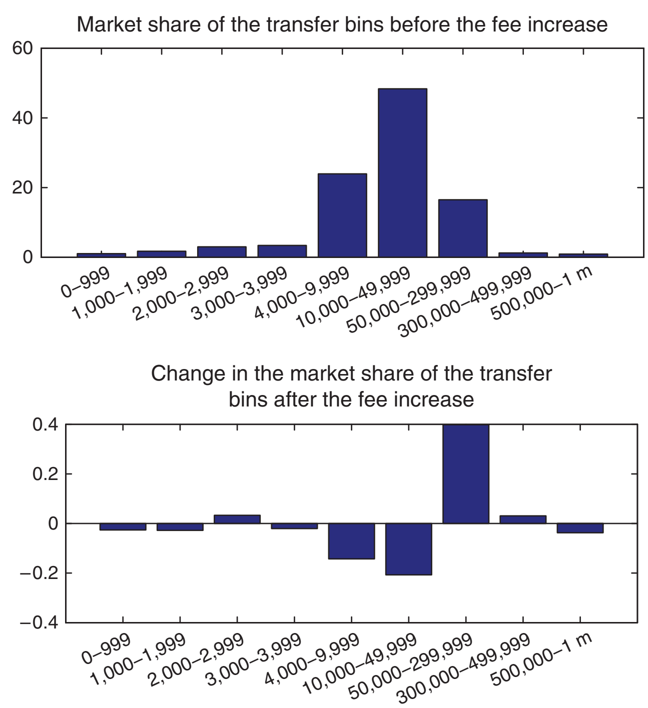
Figure 8. (Color online) Funding Sources of Cash-Outs



We now describe the network pricing specifics. On January 24, 2013, Tigo changed the transfer fees by splitting the transfer amount band 4,000–9,999 into two bands (4,000–4,999 and 5,000–9,999) and doubling the fee for the higher band. It similarly split the band 10,000–49,999 to 10,000–19,999 and 20,000–49,999 and increased the fee on the higher band by 25%. It also increased cash-out fees in some bands (see Tables 1 and 2).[23]

Figure 9 depicts the market share of each transfer bin before the fee increase and the changes in the market shares as a result of the fee increase. As shown, the price increase affected the most popular transfer bins. For this reason, even though most of the bins were unaffected, more than 50% of the transactions executed before the fee increase would have had higher transaction costs after the fee increase. We observe that the lower transfer band experienced a 15% reduction in sales, while the higher transfer band experienced only a 0.2% reduction in sales, as a result of the fee increase. We also find evidence of substitution to adjacent bins, with the 50,000–299,999 bin experiencing a

Figure 9. (Color online) Market Shares of Transfer Bins and Impact of Price Increase



relatively large market share increase. Moreover, we obtain regression discontinuity estimates of price elasticities, which amount to −0.0127 (0.0062) for the interval 5,000–9,999 and −0.0177 (0.0232) for the interval 20,000–49,999.[24]

In the focal transfer bins, that is, 4,000–9,999 and 10,000–49,999 TSh, Vodacom's pricing schedule is the same as Tigo's schedule before the price change. After the price increase, Tigo is 100% more expensive for the 4,000–9,999 TSh band and 25% more expensive for the 10,000–49,999 TSh band. As a result, after Tigo's fee increase, some customers switched to using rival Vodacom's m-money network, and others switched to the antiquated outside option (typically hiring a bus driver to carry cash). The price to use Vodacom does not depend on the transfer distance, while the price of the antiquated outside option does so. Thus, the variation in the distance of the executed transactions before and after the fee increase informs us about the degree of switching to the antiquated outside option as opposed to switching to Vodacom. Particularly, in the world where there is no switching to the traditional outside option, we should observe no change in the average distance of the Tigo transfer after the price increase. In the data, we observe that the transfer distance does change as an effect of the Tigo fee change. Average transfers in the bins where the price has increased span 1% longer distances. This variation suggests that some marginal consumers switch to the traditional outside option and that the demand for short-distance transfers is more elastic.

To analyze the geographical heterogeneity of the demand for various transactions, we classify each transaction as urban or rural by its point of origin. A transaction is marked as urban if it originates within 30 km of the center of Dar es Salaam or Zanzibar, and it is marked as rural otherwise. In the two-week estimation window, 72% of the transfers were urban originated, and similarly, 73% of the transportation/storage transactions were executed in the urban areas.

To classify consumers by location, we track their GPS positions over the full three-month window. We compute a user's average location using any mobile account activity when the GPS location is available (including cash-in, cash-out, balance check, and outbound transfer). This procedure allows us to mark all customers as urban or rural even if they did not execute a transaction within the two-week estimation window. We find that more than 40% of customers are in urban areas more than 80% of the time, while more than 35% of customers are in rural areas more than 80% of the time. We mark a customer as "urban" if he is in urban areas more than 50% of the time, resulting in labeling 56% of customers as urban and 44% as rural.[25]

To investigate the descriptive variation in the demand for transfers, we estimate the following reduced-form demand equation: $q_{zt} = \lambda_0 + \lambda_1 p_{zt} + \epsilon_{zt}$, where $z \in \{1, \ldots, 11\}$ indexes transfer brackets from Table 1, t indexes days within the examined two-week window, q_{zt} is a daily number of transfers, and p_{zt} is the price of the transfer. We estimate λ_1 to be −5.6, which is statistically significant with $p < 0.01$. This coefficient does not change and remains significant after including day dummies.[26] The price coefficient of −5.6 translates to an approximately 0.8% decrease in the demand for transfers due to the observed price increase. Note that such analysis does not account for substitution across transfer bands, treating each transfer band as a separate product, and it does not account for consumer heterogeneity. Nevertheless, the negative price coefficient provides preliminary evidence about the magnitude of the demand elasticity.

4. The Model

In this section, we provide structural models of peer-to-peer transfers and cash-in–cash-out transactions. We start with an m-money network with N subscribers. Each subscriber may transfer money to another person, transport money over some distance, or securely store money as savings.

4.1. Demand for Peer-to-Peer Transfers

Each day t, user n has up to J money transfer needs, indexed by j. The transfers include remittances and mobile payments. Each transfer (t, j, n) has a recipient m. Conditional on wishing to make a transfer, the user may execute it using Tigo or transfer the money using an outside option, which includes other m-money providers. If the transfer is made using Tigo, we model the decision of how much to transfer as a discrete choice, because, as explained in Section 2, the amounts transferred are clustered at the edges of intervals. Consequently, we discretize the size of the transfers into F discrete choices, indexed by f. Each choice is characterized by a transfer amount a_f.

Each transfer is characterized by a distance to the recipient d_{tj}^n as well as the recipient's current account balance b_t^m. Both of these quantities are assumed to be known to the sender. This assumption is close to reality for transfers between family members. If the transfer is a payment for a service, the sender is likely to correctly estimate that the balance of the receiver is high and therefore that recipient's average percentage cash-out fees are small. The sender has beliefs about the average daily cash-out strategy of the recipient as a function of the recipient's account balance, b_t^m. Using these beliefs, the sender can compute the size of a cash-out fee imposed on the recipient as the difference between incurred daily cash-out fees before and after the transfer. Formally, the fee imposed on the recipient when sending a_f Tanzanian shillings is given by

$$e_t^n(a_f, b_t^m) = \mathbb{E}_t[\text{daily cash-out fees} \mid b_t^m + a_f] - \mathbb{E}_t[\text{daily cash-out fees} \mid b_t^m], \quad (1)$$

where both expectations have subscript t to stress that they are conditional on the current cash-out tariff.

The utility of executing a mobile transfer f in the Tigo network is given by

$$u^n_{tjf} = v^n_f + r_t + \alpha^n p_t(a_f) + \beta^n e_t(a_f, b^m_t) + \epsilon^n_{tjf}, \quad (2)$$

where v^n_f is the baseline utility of the transfer f, r_t is the time trend, $p(a_f)$ is a transfer fee for transferring a_f of m-money expressed in Tanzanian shillings, and ϵ^n_{tjf} is a random shock to the utility of the transfer. The parameter α^n represents a disutility of the sender associated with paying a transfer fee p_t. The parameter β^n captures a disutility of the sender associated with imposing a cash-out fee e_t on the receiver.

The term r_t captures a daily seasonality effect and an overall time trend in the utility of mobile transfers. Including this term is necessary because we observe that some days of the week have higher transfer rates than others,[27] and we need to control for the fact that mobile transfers become more popular over time.

The term v^n_f expresses time-persistent preferences for mobile transfers in general and for transfers of particular sizes. For example, some users may intrinsically prefer to use the outside option, possibly because of a lack of an agent network or a preference for another mobile provider. Furthermore, we observe in the data that some users tend to execute only large transfers, while others tend to execute small ones. It is important to capture these features, because they impose restrictions on switching patterns.

We model v^n_f using a flexible parametric specification. Let a_f be the amount transferred with a transfer f expressed in thousands of shillings. We set

$$v^n_f = \bar{v}_f + \bar{v}^U \times \mathbf{1}(n \text{ is urban}) + \bar{\eta}^n + \eta^n_f + \kappa^n \log(a_f). \quad (3)$$

The term $\bar{v}_f$ is a fixed effect capturing the average preferences for a particular transfer size f in the population. The fixed effect $\bar{v}^U$ is active if n is an urban customer and captures the differences in WTP for transfers between urban and rural customers. Random shocks $\bar{\eta}^n$ and η^n_f are assumed to be independent and identically distributed (IID) mean-zero Gaussian random variables with scale parameters $\sigma_{\bar{\eta}}$ and σ_η, respectively. The first random term, $\bar{\eta}^n$, modifies the utility of all Tigo transfers for a particular user and reflects the persistent preferences for the Tigo network, overall. The second random term, η^n_f, modifies the utility of a transfer choice f and reflects intrinsic preferences to send transfers of a particular size. Random coefficient κ^n captures the persistent preferences for sending small versus large transfers. For example, if $\kappa^n < 0$, the user prefers small transfers over large transfers, and vice versa. The introduction of κ^n is motivated by the fact that users who sent large transfers in the past are likely to also do so in the future. The parametric specification for κ^n is described below.

The heterogeneity in the utility sensitivities to transfer and cash-out fees are captured by the variation in the random coefficients α^n and β^n.[28] Mobile transfers involve two fees, the transfer fee and the cash-out fee, thereby adding a modeling dimension. Some consumers may be insensitive to the cash-out price paid by the receivers and maximize their utility separately from the utility of the receiver. The same users may still be highly responsive to the transfer price. Senders with high price sensitivity in general are likely to have high α^n and β^n. Moreover, it proves crucial to allow users who send large transfers to be less price sensitive. Such a possibility is quite natural since the transfer size is likely to be correlated with wealth. To capture these patterns, we model α^n, β^n, and κ^n as jointly normal with an estimated variance–covariance matrix, which enables all coefficients to be different and correlated in a flexible way. We allow the mean of α^n to be different for rural and urban customers to capture the possibility that urban customers may have different price sensitivity than rural customers. We denote the difference between the urban and rural means of α^n as $\bar{\alpha}^U$. We normalize the mean of κ^n to be zero because it is absorbed by the fixed effects $\bar{v}_f$.[29] We discuss further details and goodness of fit in Section 6.

We model the spatial component of preferences using location fixed effects. This differs from the approach of Jank and Kannan (2005), who use spatially correlated random effects. Researchers choosing between the two approaches face the usual fixed effects versus random effects trade-offs. The former are usually more robust to endogeneity, while the latter may be more flexible and have better asymptotic properties. In our case, the endogeneity is less of a concern because we apply a discontinuity approach, nevertheless fixed effects in both level and price intercept of the utility function allow us to estimate an arbitrary correlation structure between the slope and the intercept of the demand functions across different locations.

The term ϵ^n_{tjf} represents idiosyncratic shocks to the utility of making a transfer f. In particular, ϵ^n_{tjf} alters users' average preferences for transfers and rationalizes the fact that users may not make the same transfer even if the prices, account balance, and distance to the receiver are the same. Usually, idiosyncratic shocks are modeled as IID normal random variables (probit) or IID extreme-value random variables (logit). Note that ϵ is a residual idiosyncratic shock after accounting for the user-level random coefficients described above. However, for a particular user, the IID assumption would imply that shocks are independent across transfer sizes. This assumption would be too restrictive for the application to monetary transfers. This is because we observe that, when the price of a particular transfer size increases, users substitute disproportionally into adjacent transfer sizes that did not experience price increases. Part of this effect is already captured by the parameter κ. Nevertheless, in response to the increase in the transfer fee and beyond his persistent preferences expressed by κ, a user may decide to send more money than he initially intended to reach a lower fee bracket and experience economies of scale. To capture this regularity, we allow adjacent ϵ^n_{tjf} to be correlated. In particular, ϵ^n_{tjf} are distributed as mean zero normal random variables with a tridiagonal variance–covariance matrix. We place 1 on the diagonal (normalization) and estimate the off-diagonal term ρ. This is a minimal structure that generates correlated preferences for adjacent transfer amounts and can accommodate disproportional substitutions into neighboring transfer sizes.

The specification for consumer-level coefficients v^n_f, α^n, β^n, and ϵ^n_{tjf} can be described as a panel version of a random-coefficient, nested probit model. An observation is a transfer need/day/user combination. The model is a generalized probit because the adjacent ϵ^n_{tjf}'s are allowed to be correlated. It has two levels of nesting: the outside option is a separate nest as a result of allowing for an aggregate shock $\bar{\eta}^n$, and every transfer size is a separate nest because of the persistent shock η^n_f.

The outside option utility includes using competing m-money networks, wire transfer, cash, and asking a bus driver to transport the money. The cost of using competing m-money networks and executing a wire transfer does not depend on distance d^n_{tj}. Wire transfers are very expensive and thus infrequent. The cost of using cash or a bus driver depends on distance. Consequently, we model the utility of the outside option as

$$u^n_{tj0} = u_0 + \gamma d^n_{tj}.$$

The fees of competing m-money providers are constant in our sample period and thus absorbed in the constant term. As a consequence, we are unable to identify the cross-price elasticity of demand across networks.

Because we observe behavior of each user before and after the change in fees, we are able to use differential price responses of users to identify two selection mechanisms. First, we estimate the unconditional distribution of random effects, and thereby are able to separate which types of users are affected by the price change. The crucial identification assumptions are that the knowledge about the fee change is uncorrelated with the type of user, in particular, that the fee change is unanticipated by all users, and that the knowledge about the fee change diffuses instantly. Second, we estimate an unconditional distribution of the distance between the sender and the receiver. This allows us to assess the selection of executed transfers by distance. We identify this selection using within-user differences in the propensity to execute short- and long-distance transfers before and after the price change.

These two selection mechanisms are important for assessing the welfare implications of the changes in transfer fees. Primarily, some users would be less elastic to the fee change, which would result in them internalizing most of the fee increase. Additionally, the demand for long-distance transfers may be less elastic because the outside option is more expensive for longer distances.

We omit the cash-out fees from the utility specification of the outside option. One could argue that once the money is in the mobile account it has to be cashed out before one can execute the transfer using the outside option. There are two reasons why we think that such considerations are of second order. First, we typically see m-money being cash-out immediately as opposed to accumulated for future transfers. In other words, we rarely see shillings being sent in the network more than once before cash-out. Second, most of the peer-to-peer transfers are executed using freshly cashed-in funds as opposed to accumulated ones.

4.2. Demand for Transporting and Storing Money

The model of transportation and storage/savings is similar to the peer-to-peer transfer model. We concentrate our discussion on the differences between the two models. Each consumer n has J needs to transport or store money, indexed (with a slight abuse of notation) by j. Each transaction is characterized by a distance $d^n_{S,tj}$ and time span (lifetime in the network) between cash-in and cash-out $s^n_{S,tj}$. The transactions with high distance d and low lifetime s are money transportation, whereas the transactions with low d and high s are storage transactions. It is possible to combine storage with transportation when both d and s are high. Formally, we model distance and time span as jointly log-normally distributed with an arbitrary variance–covariance matrix.

The utility of executing a transaction of size f in the Tigo network is given by

$$u^{S,n}_{tjf} = v^n_{S,f} + r_t + \beta^{S,n} p^S(f) + \epsilon^{S,n}_{tjf}, \tag{4}$$

where the superscript S distinguishes the parameters of the transportation/storage model from the parameters of the peer-to-peer transaction model. The price sensitivity parameter $\beta^{S,n}$ multiplies the cash-out fee $p^S(f)$. In comparison, β multiplies the cash-out fee in the transfer model. The transfer price is absent here because no mobile transfer is executed in the transportation/storage transactions. Other aspects of the inside good utility specification are the same as in the peer-to-peer transfer specification.

The outside option in the case of transportation and storage transactions includes moving with money in one's pocket by using public transit or going by foot or private car, storing cash at home, using traditional banking, or using a competing mobile network. The outside option of carrying cash has significant danger. It is likely that the distance traveled carrying cash affects the probability of being robbed. In a similar fashion, the number of days for which cash is stored at home affects the probability of losing it through burglary. For these reasons, we propose the following specification for the utility of the outside option:

$$u_{tj0}^{S,n} = u_{S,0} + \gamma^{S,1}\log(d_{tj}^{S,n}+1) + \gamma^{S,2}\log(s_{tj}^{S,n}+1),$$

where $\gamma^{S,1}$ and $\gamma^{S,2}$ represent the disutilities of walking with cash and storing cash at home, respectively.

5. Estimation

The peer-to-peer transfer model is estimated in two stages. First, we estimate a model of the recipient's m cash-out activity on day t as a function of his account balance b_t^m. We postulate the following semiparametric model for the cash-out fees:

$$c_t^m = \theta_0^m + \theta_1 b_t^m + \theta_2 (b_t^m)^2 + \tilde{\epsilon}_t^m, \tag{5}$$

where c_t^m is an incurred cash-out fee at day t by user m. 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 his 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 $\theta_1 > 0$ and $\theta_2 < 0$, indicating economies of scale when cashing out directly implied by the concave fee schedule. By construction, account balance is endogenous; that is, θ_0^m is correlated with the current balance b_t^m. We control for the endogeneity using fixed effects and apply Honoré (1992), extended by Alan et al. (2014), to deal with the double-sided truncation of c_t^m resulting in the incidental parameter problem. Subsequently, we apply Honoré (2008) to compute average marginal effects yielding the quantity of interest, given by

$$\begin{aligned} e_t^n(a_f, b_t^m) = \mathbb{E}[&\text{daily cash-out fees} \mid b_t^m + a_f] \\ &- \mathbb{E}[\text{daily cash-out fees} \mid b_t^m]. \end{aligned} \tag{6}$$

The details of the first-stage estimation are contained in Section 2 of the online appendix.

In the second stage, we estimate the demand model using the simulated method of moments following Pakes and Pollard (1989). Specifically, we compute 99 micromoments from the data and choose the parameters that generate the closest match of the simulated moments. We utilize the following moments: market share of each transfer bin[30] before and after the price change (66 moments), mean number of transfers each day of the week (13 moments), mean and standard deviation of the distance for the transfers executed before and after the price change (4 moments), mean and standard deviation of receivers' balances before and after the price change (4 moments), probability of sending the same amount of money at least twice, interaction between the amount and the distance, interaction between the amount and the balance of a receiver, joint distribution of executing n transfers before the price change, and m transfers after the price change (9 moments).

The moments are simulated 10 times for each customer.[31] We set the maximum amount of daily transfers that the user can execute, denoted as J, to three. Executing more than three transfers is infrequent and constitutes less than 1% of the data.

We employ an identity weighting matrix to obtain initial estimates of the model. Next, we estimate the optimal weighting matrix, which is used to obtain the final estimates. Standard errors are clustered on the consumer level.[32]

The transport/storage model is estimated in the same way as, but separately from, the second stage of the transfer model. It includes additional moments identifying the joint distribution of transaction distance and lifetime; that is, we additionally match the average lifetime of a transaction before and after the price increase, as well as the average interaction of distance and lifetime for the executed transactions. By construction, the estimation does not have moments pertaining to the receiver.

6. Results

In this section, we report and discuss the estimates of the transfer and transport/storage models.

6.1. Peer-to-Peer Transfer Model

Table 4 contains the estimates of the cash-out model. The model is estimated separately for before and after

Table 4. Effect of the Recipient's Account Balance on an Average Incurred Daily Cash-Out Fee

Account balance (in thousands)—before price change	8.64 (0.059)
Account balance (in thousands) squared—before price change	−0.0024 (0.0001)
Account balance (in thousands)—after price change	8.75 (0.049)
Account balance (in thousands) squared—after price change	−0.0026 (0.0001)

Note. The estimation contains user fixed effects (not reported).

the fee change to account for differences in the conditional distribution of cash-out events caused by the change in the cash-out fees. We find that the coefficient of the recipient's account balance (in thousands) equals 8.64 before and 8.75 after the fee change, respectively. In other words, conditional on a cash-out, an extra 1,000 shillings in the receiver's account increase his expected cash-out fees by approximately 9 shillings. The difference between the coefficients is statistically significant, confirming that the consumers are incurring higher cash-out fees as a function of their account balance after the fee change.

Negative and statistically significant coefficients on the square of the recipient's account balance suggests that the average cash-out fees are a concave function of the balance. This is evidence of economies of scale in cashing out. We find that such economies of scale are higher after the price change.

Figure 10 contains a graph of the cash-out fee for a modal transfer of 10,000 shillings and different pretransfer balances of the receiver before and after the price change. The fee is a decreasing function of the pretransfer balance, confirming our initial predictions that accounts with larger balances experience a smaller cash-out fee for a transfer of a given size. The difference between the cash-out fees imposed on the receiver before and after the fee change is significant. We utilize this variation to separately identify the impact of the changes in transfer and cash-out fees on the propensity to execute a transfer.

Figure 11 reports the daily seasonality in a transfer activity. We observe that the coefficients are small with the exception of the coefficient of the Sunday dummy. Transfer sizes fixed effects are mostly negative indicating that transfers are tail events as measured on a daily basis (see Figure 5 in the online appendix).

The remaining parameters of the demand model for transfers (see Equations (2) and (3)) are presented in the left column of Table 5. We find substantial heterogeneity in the preferences for Tigo transfers compared to the outside option; that is, the standard deviation of the persistent shock to the utility of any Tigo transfer, σ_A, is estimated at 0.35 compared to the standard deviation of the idiosyncratic shocks, which was normalized at 1. Consequently, a high percentage of the population is averse to mobile banking irrespective of pricing. This is consistent with previous qualitative studies that find nonprice barriers to use of mobile banking, such as insufficient trust and the necessity of technical sophistication for both the sender and the receiver (see Mallat 2007). Additionally, we find considerable persistence in the types of executed transfers. Specifically, we find that the transfer size parameter κ has a standard deviation of 0.24. An interpretation of this parameter is presented in Figure 12, which depicts the distribution of executed transfers by size for five types of consumers, such that $\kappa \in \{-3\sigma_\kappa, -\sigma_\kappa, 0, \sigma_\kappa, 3\sigma_\kappa\}$.

We find that the standard deviation σ_η of the persistent transfer level shocks η_{nf} is low. This suggests that

Figure 10. (Color online) Increase in Daily Cash-Out Fees Resulting from Receiving 10,000 as a Function of the Pretransfer Balance of the Receiver—Before and After Price Change

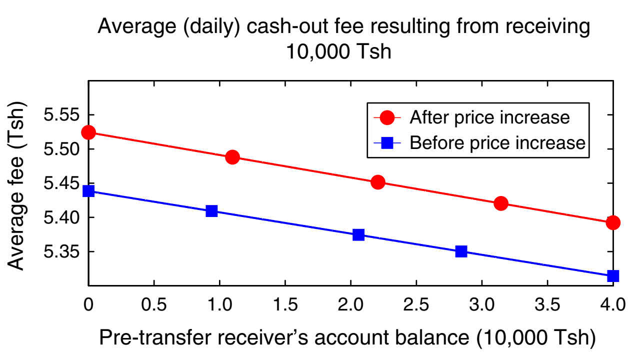


Figure 11. (Color online) Trend in the Utility Function for Transfers

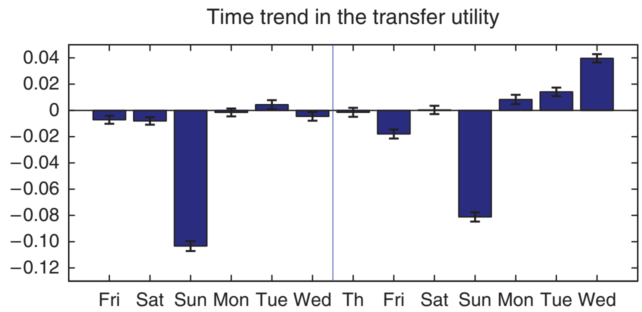
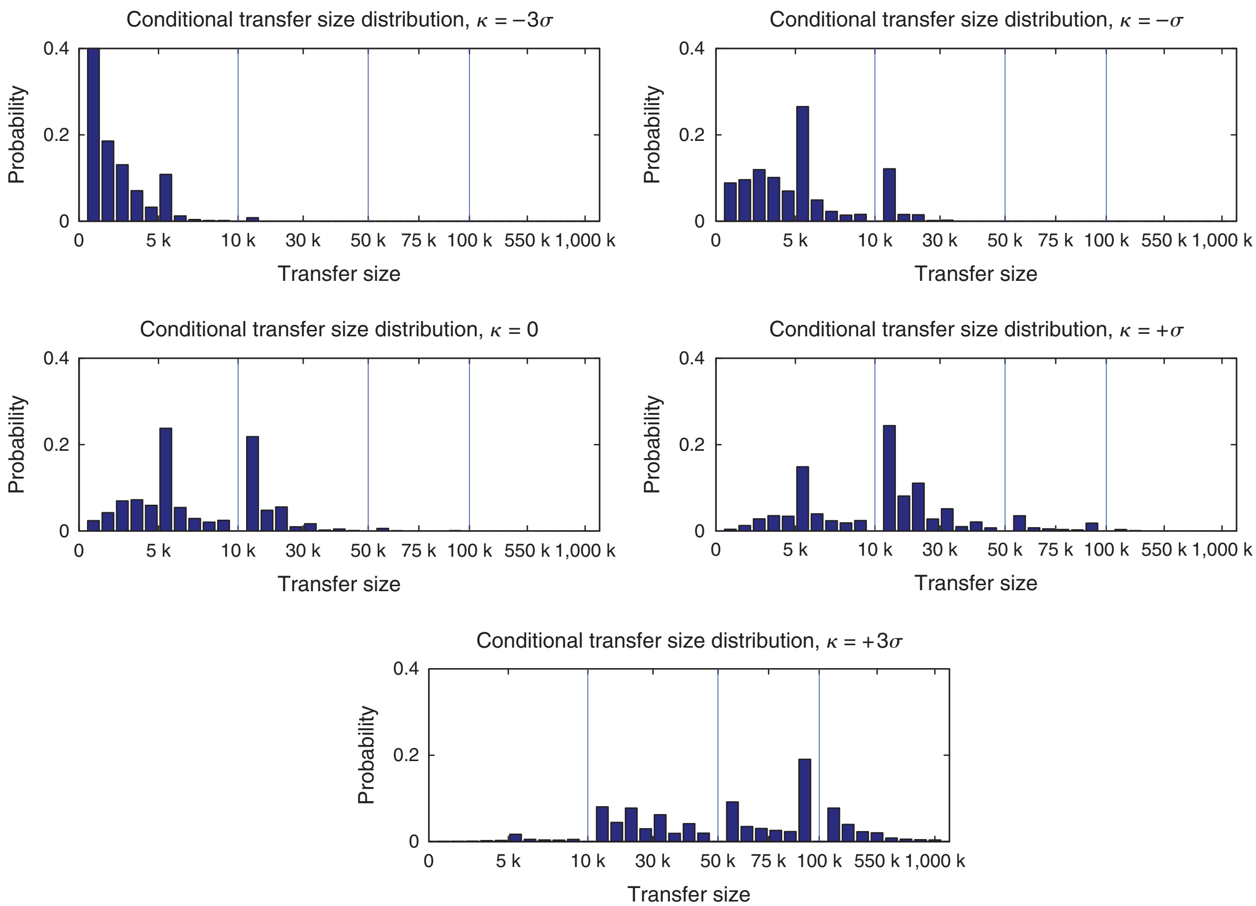


Table 5. Structural Parameters of Demand for Peer-to-Peer Transfers and Transportation/Storage Transactions

	P2P transfers	Transportation/ storage
Urban dummy ($\bar{v}^U$)	0.16 (0.006)	0.11 (0.018)
Price coeff., mean ($\bar{\alpha}$)	−1.17 (0.029)	−1.12 (0.02)
Urban dummy, price coeff. ($\bar{\alpha}^U$)	0.40 (0.036)	0.55 (0.02)
Cash out coeff., mean ($\bar{\beta}$)	−43.66 (0.96)	—
Price coeff., std. dev. (σ_α)	0.41 (0.013)	0.42 (0.01)
Cash out coeff., std. dev. (σ_β)	0.02 (0.001)	—
Transaction size coeff., std. dev. (σ_κ)	0.24 (0.002)	0.21 (0.002)
Price coeff./Cash out coeff., correlation ($\rho_{\alpha\beta}$)	0.57 (0.053)	—
Price coeff./Trans. size coeff., correlation ($\rho_{\alpha\kappa}$)	0.49 (0.01)	0.51 (0.02)
Cash out coeff./Trans. size coeff., correlation ($\rho_{\beta\kappa}$)	0.53 (0.018)	—
Covar. of ϵ for 2 adjacent transf. (ρ_ϵ)	0.06 (0.003)	0.10 (0.002)
Std. dev. of the persistent part of ϵ ($\sigma_{\bar{\epsilon}}$)	0.01 (0.008)	0.01 (0.016)
Std. dev. of the aggregate shock (σ_A)	0.35 (0.01)	0.42 (0.01)
Distance coefficient (γ)	−0.05 (0.003)	−0.11 (0.01)
Time-span coefficient (γ)	—	−0.09 (0.01)

Figure 12. (Color online) Transfer Size Distribution Conditional on Making Transfers for Different Values of κ



most of the user heterogeneity is related to a transfer size and accommodated by κ.

Covariance between idiosyncratic shocks for adjacent transfers is moderate, that is, ρ_ϵ is estimated to be 0.06. Thus, we find evidence that people substitute to the adjacent transfer bins if the price of a particular bin increases. However, this substitution is somewhat limited, which is consistent with the raw data.

We find that coefficients on the transfer fee and on the cash-out fee are negative and statistically significant. This suggests that both fees have an impact on the propensity to transfer. The negative coefficient on the cash-out fee suggests that the sender takes into consideration, to some degree, the externality he imposes on the receiver when sending him m-money. We find that both price coefficients are heterogeneous in the population and highly correlated. In other words, people that are sensitive to cash-out fees are also sensitive to transfer fees. We also find that people transferring larger amounts of money have different price sensitivities. The Pearson correlation coefficient between the transfer price sensitivity α and κ is equal to 0.5. Also, the Pearson correlation coefficient between the cash-out price sensitivity β and κ is equal to 0.53. The conditional means of price coefficients for representative values of κ are presented in Table 6.

6.2. Transport/Storage Model

The estimates of the size fixed effects and time trend in cash-in–cash-out transactions are contained in Figures 6 and 7 in the online appendix. The interpretations of these parameters are qualitatively similar to the interpretation of the corresponding parameters of the transfer model discussed above. We find that the baseline utility of the transport/storage action is lower than that of the transfer action, which is consistent with the fact that the majority of the shillings in the network are transferred at least once.

Table 6. Means of Price Coefficients Conditional of Different Values of κ

	$\kappa=-3\sigma$	$\kappa=-\sigma$	$\kappa=0$	$\kappa=\sigma$	$\kappa=3\sigma$
Transfer fee coefficient, rural, α^R	−1.77	−1.37	−1.17	−0.97	−0.56
Transfer fee coefficient, urban, α^U	−1.37	−0.97	−0.76	−0.56	−0.16
Cash-out fee coefficient, β	−76.89	−54.74	−43.66	−32.58	−10.43

The right-hand column of Table 5 contains the remaining estimates of the structural model of transport and storage. We find that the users are significantly averse to the size of the cash-out fee; that is, we estimate $\bar{\beta}^S$ to be −0.94. We find that users are more price sensitive when executing transport/storage transactions compared to transfer transactions.

We find that the covariance of the shocks ϵ for adjacent transactions is somewhat higher in the model for transport and storage than in the model for peer-to-peer transfers. We suspect that in the case of transport/storage, the consumer can fully control the amount of the transaction, and therefore is able to decrease the transaction amount when the fee increases.[33]

We find that both the distance coefficient and lifetime coefficient are significantly negative. This suggests that walking with money or storing money at home carries significant disutility. While it is hard to interpret these numbers in isolation, they suggest that people are willing to pay extra to not walk with cash or store cash at home. We quantify this in Section 7.

6.3. Elasticity of Demand

We use the estimated price coefficients to compute price elasticities. Because the product mix contains multiple prices, the price elasticity can be computed in multiple ways. Price elasticities that rely on an increase in transfer price of one transfer bin, keeping the transfer prices of other bins constant, are not very informative as to the total demand for Tigo transfers. Instead, we compute elasticities that rely on the impact of uniform percentage change in all of the prices on the total demand for mobile transfers. Formally, we compute

$$\frac{\partial\,\text{demand}(\mathbf{p}+\rho\mathbf{p})}{\partial\rho}\text{demand}^{-1}(\mathbf{p}).$$

We compute the elasticities measuring demand (i) as the probability of a transfer and (ii) as a total transfer shilling amount. In the former case, we obtain −0.1 transfer fee elasticity and −0.36 cash-out fee elasticity. In the latter case, we obtain −0.18 and −0.40, for transfer and cash-out elasticities, respectively.[34]

The estimated elasticities show that the network is pricing on an inelastic part of the demand. Such pricing as well as documented high heterogeneity in the overall value for mobile banking suggests that the network may be engaging in penetration pricing. Another piece of evidence consistent with penetration pricing is a steady upward price trend in both transfer fees and cash-out fees that we observe during the 2010–2014 time period. We also find that the network sets transfer fees on the more inelastic portion of the demand than cash-out fees. In other words, the cash-out fees are set to be relatively higher than transfer fees, which may be related to an attempt to lock-in the users' money inside the network. This possibility is discussed in Section 7.

The coefficient, γ, representing the contribution of the distance of the transfer to the utility of the outside option is negative and statistically significant, which reflects the nature of the outside option. The longer the distance between the sender and the receiver, the more inferior the outside option. The result is not surprising considering that the outside option consists of giving cash in person or sending it using a bus driver. The negative impact of the distance on the outside option implies that the demand for consumers of longer distance transfers is more inelastic. We present the price elasticities in Table 7. Variation of the elasticity by distance in conjunction with the anecdotal evidence that the bus driver option is inferior and costs about 10% of the transfer implies that there is limited substitution to other mobile networks. If substitution to other networks was accounting for the majority of switching, we should have been observing no elasticity heterogeneity across transfer distances. Although this result may seem somewhat counterintuitive, it should be interpreted within the cultural reality of Tanzania. Anecdotally, one of the biggest barriers to adoption of m-money is mistrust in the reliability of the network operator. Such trust issues may be particularly important for a rural population, which is the primary recipient of the long-distance remittances. Once trust with one operator is established, the senders are likely to face switching costs when changing networks, especially in a short time horizon.

The demand for transfers is more elastic for rural origin points than for urban origin points. This variation may be due to an income effect and that many of the urban-originating transactions are remittances used to mitigate income shocks and thus may be more time sensitive and less elastic.

The lower part of Table 7 contains the elasticity of demand for transportation/storage transactions. We note that the firm is pricing on a more elastic part

Table 7. Price Elasticities of Demand

	Overall	Rural	Urban
Peer-to-peer transfers			
Transfer price elasticity	−0.104	−0.152	−0.080
Transfer price elasticity—0 km transfers	−0.11	−0.161	−0.085
Transfer price elasticity—20 km transfers	−0.107	−0.157	−0.080
Transfer price elasticity—100 km transfers	−0.102	−0.151	−0.079
Transfer price elasticity—500 km transfers	−0.101	−0.147	−0.078
Cash-out price elasticity	−0.36	−0.37	−0.36
Cash-in–cash-out transactions			
Cash-out price elasticity	−0.52	−0.94	−0.34
Cash-out price elasticity—1 day, 0 km	−0.54	−0.98	−0.35
Cash-out price elasticity—1 day, 0.5 km	−0.54	−0.98	−0.36
Cash-out price elasticity—1 day, 5 km	−0.53	−0.97	−0.35
Cash-out price elasticity—0 day, 1 km	−0.52	−0.96	−0.38
Cash-out price elasticity—0.5 day, 1 km	−0.49	−0.95	−0.34
Cash-out price elasticity—5 days, 1 km	−0.49	−0.95	−0.36

of the demand curve compared to transfer transactions. This suggests that consumer lock-in may be less important for transportation/storage than for transfers. Indeed, for transfer transactions, both the sender and receiver have to adopt the same network, whereas for transportation/storage transactions, only a single user needs to be in the network. The fact that the firm sets the same cash-out fee for transfers and transportation/storage is another reason that the cash-out fees are set on the more elastic part of demand for transfers, compared to the transfer fees.

Similar to the demand for transfers, we observe that rural customers are more elastic than urban customers. In addition to the previously mentioned income effect, we note that crime-related risk is higher in the city (see Kinabo 2004), which lowers the utility of the outside option (carrying and storing cash) in urban areas.

The last rows of Table 7 contain estimates of the elasticity of demand for cash-in–cash-out transactions with fixed distance d and lifetime s, isolating elasticity of transporting money from elasticity of storing money. We find no significant differences between these two elasticities.

As in the transfer model, we find significant consumer heterogeneity in price sensitivity, utility for transfer size, and the correlation between price sensitivity and transfer size. These results are in general agreement with the earlier estimates of the peer-to-peer transfer model, indicating that both groups of users are similarly heterogeneous.

7. Pricing Counterfactuals

In this section, we use the demand model to conduct pricing counterfactuals. First, we estimate the WTP for money transportation and storage and describe how WTP differs across urban and rural locations. We link these estimates to urban and rural crime data.

Second, we quantify the incentives of the network to price discriminate based on the type of transaction, location of the transaction origin, and distance between the origin and destination.

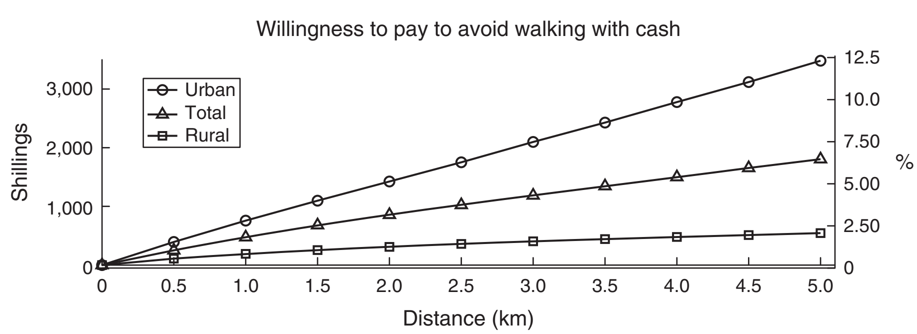
7.1. Willingness to Pay for Money Transportation and Storage

The transport/storage demand model allows us to quantify the consumers' WTP to avoid carrying cash or storing cash at home. We construct a measure of the WTP that is reminiscent of the compensating variation. To estimate the WTP to avoid carrying cash, we compare two transactions: (i) approximately zero kilometers between cash-in and cash-out, within one day and at the current price schedule p, and (ii) x kilometers between cash-in and cash-out, within one day and at the counterfactual price schedule $p + y$. We find the WTP y such that the expected amount of money transacted is the same in (i) and (ii).[35] The results are depicted in Figure 13. We find that consumers are willing to pay, on average, 345 shillings extra not to walk an extra kilometer carrying cash. This amounts to 1.24% of the average transaction size, which translates to a belief that walking an extra kilometer increases the probability of being robbed by 1.24%, if the users are risk neutral. This number is large, but it reflects the numbers cited in the survey by the United Nations Office on Drugs and Crime (2009) for Tanzania.[36]

The WTP for money transportation varies by the location of the transaction origin (Figure 14). Users in rural areas are willing to pay only 91 shillings per kilometer (or 0.3%), whereas users in urban areas are willing to pay over 760 shillings per kilometer (or 2.75%). This variation mirrors high theft rates in urban areas.[37] Although the relationship between theft rates and WTP should be treated as correlational, it is consistent with the theory that users execute cash-in–cash-out transactions to mitigate crime-related risk. This theory also suggests that use of mobile money for transportation and storage may lead to a decrease in crime in the long run.

To quantify the disutility of keeping savings at home, we compare two transactions: (i) 0.5 kilometers between cash-in and cash-out, within approximately zero days and at the current price schedule p, and (ii) 0.5 kilometers between cash-in and cash-out, with

Figure 13. Willingness to Pay to Avoid Walking with Money



Notes. We compare two zero-hop transactions: (i) approximately zero kilometers between cash-in and cash-out, within one day and at price schedule p, and (ii) x kilometers between cash-in and cash-out, within one day and at price schedule $p + y$. We find y such that the expected amount of money transacted is the same as in (i) and (ii). The y-axis is y, and the x-axis is x.

a network lifetime of x days and at the counterfactual price schedule $p + y$.[38] We find the WTP y such that the expected amount of money transacted is the same in (i) and (ii). The results are presented in Table 14. We find that users are willing to pay 217 shillings (approximately 0.8% of the average transfer) for an extra day of secure storage. This number should be interpreted as the difference between the perceived security of m-money and storing the money at home, and again should be regarded as an upper bound. Similarly, the WTP varies by location and amounts to 68 shillings (or 0.25%) per day in the rural areas and 425 (or 1.52%) in the urban areas.

The official criminal records in Tanzania, and similar African countries, are known to be extremely unreliable. For this reason, the usual way to measure the extent of crime is direct surveys, which are likely to be subject to a large survey bias. As far as we know, we provide the first estimates of criminal activity in Africa based on actual consumer behavior rather than based on using official records or surveys. Although the measured level of crime corresponds extensively to other survey measures mentioned earlier in this paper, one should interpret it keeping the risk-neutrality assumption in mind. Our method can be applied to measure cross-sectional and geographical variation in crime.

7.2. Price Discrimination

To quantify the incentives of the firm to price discriminate, we propose a simple supply model. Let $\mathbf{p}^T$ be the vector of the transfer fees. Denote as $\mathbf{p}^C$ and $\mathbf{p}^S$ the cash-out fees charged to a receiver of a transfer and to a user executing a cash-in–cash-out transaction, respectively. Under the current pricing, the two cash-out fees are equal, that is, $\mathbf{p}^C = \mathbf{p}^S$. Per user, daily variable profits from transfers are equal to

$$\pi^T(\mathbf{p}^T, \mathbf{p}^C) = 3\sum_{f=1}^{F} D_f^T(\mathbf{p}^T, \mathbf{p}^C)(Hp_f^T + p_f^C - c_f), \qquad (7)$$

where D_f is the probability of executing a peer-to-peer transfer transaction of size f, as described in the model in Section 4, and H is the number of transfer hops before the money leaves the network. The term c_f is the marginal cost of a transfer transaction, equal to a sum of cash-in and cash-out fees of agents, which we observe in the data.[39]

Per-user variable profits from cash-in–cash-out transactions are

$$\pi^S(\mathbf{p}^S) = 3\sum_{k}\sum_{f=1}^{F} D_f^S(\mathbf{p}^S)(p_f^S - c_f), \qquad (8)$$

where $D_{f,k}^S$ is the probability of executing a cash-in–cash-out transaction of size f. Total daily profits are given by

$$N[\pi^T(\mathbf{p}^T, \mathbf{p}^C) + \pi^S(\mathbf{p}^S)],$$

where N is the total number of users.

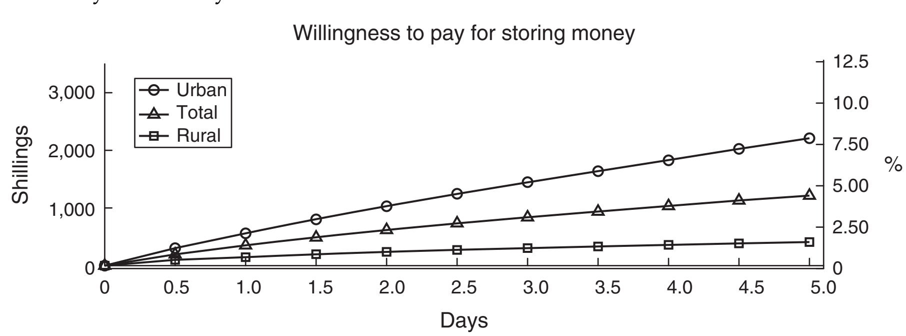
There are two simplifications used in our supply model. Because data limitations prevent us from estimating cross-price elasticities, we keep competitors' prices fixed in the remainder of this section. The estimated effects are conservative if prices are strategic complements because a possible competitive response would increase the profitability of our pricing counterfactuals.[40] We ignore dynamic pricing incentives (see Farrell and Klemperer 2007) and focus on static pricing incentives across different pricing policies.[41]

Using the supply model, we consider two opportunities to price discriminate: (i) depending on the type of service and (ii) based on location and distance.

7.2.1. Price Discrimination Based on Type of Service.

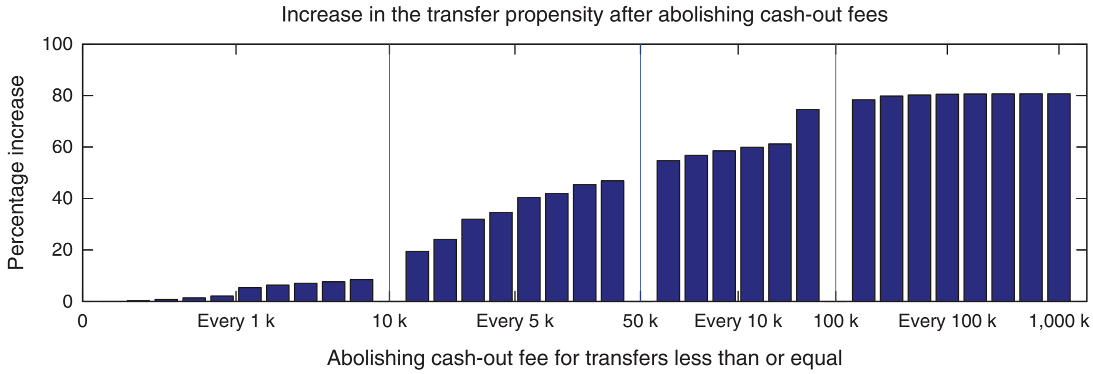
We have identified different demand curves for transfer and transportation/storage services. Presently, the network charges the same cash-out fee schedule to all customers. Since the demand curves for the two types of transactions are different, the network's profits and consumer surplus can potentially be improved through second-degree price discrimination.

Figure 14. Willingness to Pay to Securely Store Cash Over Time



Notes. We compare two zero-hop transactions: (i) 0.5 kilometers between cash-in and cash-out, within approximately zero days and at price schedule p, and (ii) 0.5 kilometers between cash-in and cash-out, within x days and at price schedule $p + y$. We find y such that the expected amount of money transacted is the same as in (i) and (ii). The y-axis is y, and the x-axis is x.

Figure 15. (Color online) Percentage Increase in the Number of Transfers After Abolishing Cash-Out Fee for Users That Make At Least One Transfer



Note. The graph presents an increase in the abolishment of the cash-out fee for transfers smaller than or equal to the values on the *x*-axis.

The limitations of the reduced-form model of cash-out behavior prevent us from evaluating the demand for transfers for arbitrary levels of cash-out fees. The frequency of cash-outs may change in response to a change in cash-out fees, and the parameters of Equation (5) are unlikely to stay the same. This affects the transfer demand because the cash-out parameters enter the demand for transfers through the function $e_t^n(a_f, b_t^m)$. We can, however, evaluate the impact of nullifying the cash-out fees, that is, setting $e_t = 0$. As shown in Figure 15, abolishing the cash-out fees increases the demand for transfers by 81%. Abolishing the cash-out fee for medium-size transfers has the strongest effect on demand since smaller transfers carry negligible nominal cash-out fees, and consumers executing large transfers are inelastic.[42]

Abolishing the cash-out fees can potentially be profitable since it can substantially increase demand for transfers. Nevertheless, unconditionally nullifying the cash-out fee leads to losses that can amount to as much as 500% of current network profits. A significant contributor to the above losses is the 140% expansion of transportation and storage transactions, which generate no revenue but still carry a significant marginal cost. In effect, in such a setup, users who make transfers would be subsidizing users who do not make transfers. A more sophisticated price discrimination schedule is needed.

We start by considering a situation in which the network can price discriminate between users that receive transfers and those that make no transfers. The network can keep $\mathbf{p}^S$ at the current level and set $\mathbf{p}^C = 0$. As part of the same price discrimination scheme, we consider a simple and easy-to-implement fixed percentage markup pricing policy for transfer fees, by setting $\mathbf{p}^T = \delta\mathbf{MC}$, where $\mathbf{MC}$ is the marginal cost. We find that when the firm marks up transfers' marginal cost by approximately 47%, that is, for $\delta = 1.47$, per-user variable profits from transfers do not change, and, of course, variable profits from cash-in–cash-out transactions are constant because we did not alter $\mathbf{p}^S$. We find that, while not altering the network's profits, introduction of this price discrimination scheme doubles the expected amount of transferred money, leading to an increase in consumer welfare.[43] This is a conservative estimate of the possible increase in welfare for the following reasons. First, we consider simple markup policies only; more sophisticated and complex policies would likely achieve better results. Second, the increase in expected consumer surplus would lead to more adoption, and therefore the consumer surplus gain would likely be higher in the long run. Third, more adoption should generate even higher total profits, which would enable the network to use a lower markup δ and maintain the same total profits.

Unfortunately, the above price discrimination policy is hard to implement because setting the cash-out fee for transfers to zero and keeping the cash-out for transportation/storage unchanged is not incentive compatible. Specifically, if the network stops charging a cash-out fee after a transfer, users may execute an unnecessary P2P transfer even if their intent is to transport or store the money. Thus, we propose the following incentive-compatible price discrimination schedule. The network allows users to cash out for free up to an amount received from a transfer, and charges a positive cash-out fee for all other cash-out transactions. To maintain incentive compatibility, the cash-out fee not preceded by a transfer has to be less than the transfer fee. We find that when the network sets $\delta = 1.47$ (in $\mathbf{p}^T = \delta\mathbf{MC}$) and $\mathbf{p}^C = 0.81\mathbf{p}^T$, it obtains the same profit from cash-in–cash-out transactions as before, while the transacted cash-in–cash-out amount increases by approximately 5%. Thus, abolishing cash-out fees for users that conduct peer-to-peer transfers and setting a cash-out fee for the rest of the transactions as described above is a Pareto improvement in comparison to present pricing. We found that the proposed price discrimination schedule increased consumers surplus when profits were kept constant. Clearly, we could also increase profits by increasing fees marginally.

7.2.2. Price Discrimination Based on Location and Distance. Our demand analysis reveals significant heterogeneity in both the intercept and slope and the demand function that the network could exploit through price discrimination. First, urban customers value P2P transfer and storage/transportation transactions more, and they are less elastic on the margin at current prices than rural customers. Second, consumers are willing to pay more for transactions with a longer distance span between cash-in and cash-out.

To identify the incentive for differential pricing across segments, we extend our supply model to handle different segments of consumers. We consider K segments of sizes N_k. As an illustrative example, we consider segments of equal or nearly equal sizes, that is, $N_k \approx N_{k'}$. The fees charged in segment k are given by vector $(\mathbf{p}_k^T, \mathbf{p}_k^C, \mathbf{p}_k^S)$. The corresponding profits per user in segment k are given by π_k^T and π_k^S, and the profit from both types of transactions by π_k.

The first-order condition for static profit maximization is given by $\sum_k (\partial \pi_k / \partial \mathbf{p}_k) = 0$. The current pricing is performed under no price discrimination, or formally, under a series of constraints $\mathbf{p}_k = \mathbf{p}_{k'}$ for all k'. If we introduce profitable price discrimination, we should observe relaxation of these constraints, so that $\partial \pi_k / \partial \mathbf{p}_k > 0$ and $\partial \pi_k / \partial \mathbf{p}_k < 0$ for some k and k'; that is, the network will have incentives to lower the price for some customers and increase the price on others. Without loss of generality, we label $k = 0$, the segment with the lowest derivative $\partial \pi_k / \partial \mathbf{p}_k$. We define a measure of incentives for price discrimination on the margin (for small changes in price), which we call DiPP, as

$$\mathrm{DiPP}_k = (\pi_0 + \pi_k)^{-1} \left[\frac{\partial \pi_k(\mathbf{p}_k + \rho \mathbf{p}_k)}{\partial \rho} - \frac{\partial \pi_0(\mathbf{p}_0 + \rho \mathbf{p}_0)}{\partial \rho} \right].$$

DiPP_k measures the percentage change in joint profitability of segment k and segment 0, if the firm decreases the prices to segment 0 by 1% and increases the prices to segment k by 1%.[44]

DiPP is similar in spirit to the upward pricing pressure measure developed by Farrell and Shapiro (2010) to evaluate the impact of horizontal mergers on the incentives to change prices on the margin. This formulation may not capture the full extent of additional profits generated by price discrimination if the curvature of the profit function or if the pricing responses by competitors are substantial. Despite these limitations, studying incentives on the margin has significant advantages. Our measure does not make strong assumptions about the curvature of the demand out of sample, nor does it assume any form of supply-side equilibrium. An alternative approach, that is, a full supply-side counterfactual, would need to heavily leverage such assumptions.[45]

We use DiPP to quantify the marginal incentives to differentially price urban and rural customers. We consider charging differential transfer fees for transfer transactions and charging differential cash-out fees for cash-out transactions. As mentioned above, we cannot consider arbitrary changes in cash-out fees for transfer transactions, and therefore we keep these fees unchanged. This contrasts with the analysis of Section 7.2.1, where we considered a lumpy change in cash-out fees for transfers.

Table 8 contains measures of DiPP for both types of transactions across different locations and distances of the transaction. The partial derivatives of profits in DiPP calculations were weighted by the sizes of urban and rural segments in the data. We find that increasing the price of transfers for urban customers and decreasing it for rural customers by 1% leads to a 0.22% increase in profits of the network from transfer transactions. Similarly, the same percentage changes in cash-out fees increase profits from transportation/storage transactions by 3.1%. The network can implement changes in both of these fees. We note that the profit effects of the two changes in fees differ by an order of magnitude, with a stronger incentive to price discriminate by location among transportation/storage consumers than among transfer consumers. This is

Table 8. Differential Pricing Pressure for Each Type of Transaction by Location and Distance

P2P transfer	Urban minus rural				
DiPP for change of transfer fees based on originating location	0.22%				
Transportation/storage	**Urban minus rural**				
DiPP for change of transfer fees based on originating location	3.1%				
P2P transfer	**0–12**	**12–24**	**24–43**	**43–85**	**85–**
DiPP for an increase of transfer fees at distance of x (km) with reduction of fees at baseline	Baseline	0.009%	0.013%	0.019%	0.029%
Transportation/storage	**0–0.2**	**0.2–0.8**	**0.8–2.9**	**2.9–13.5**	**13.5–**
DiPP for an increase of transfer fees at distance of x (km) with reduction of fees at baseline	Baseline	0.26%	0.61%	1.25%	2.58%

consistent with the presumption that the demand for transportation/storage transactions is driven mostly by theft rates, which are higher in cities than in rural areas, while the demand for transfers is driven by other factors that are common across rural and urban areas.

Table 8 also contains DiPP indexes for price discrimination based on distance. We find that the network has a strong incentive to discriminate among transportation/storage transactions but only a negligible incentive to do so based on the distance of transfer transactions. This result is consistent with the fact that the outside option to transportation/storage transactions varies more with distance.

8. Conclusion

We analyzed a mobile money network in Tanzania that is widely used (i) to transfer money across users (P2P transfers), (ii) to transport money without a transfer, and (iii) as a savings account to store money for the short and medium term. We analyzed and quantified the demand for P2P transfers in terms of distance, lifetime in the network, and number of transfers (hops) of a shilling before cash-out. We found that the elasticity of demand for P2P transfers decreases in the distance between cash-in and cash-out. We also found that the senders internalize to a significant extent the cash-out fees that the network imposes on receivers when money is cashed out. Despite high cash-out fees, most users cash out after only one transfer.

We found that a significant percentage of transactions on the network do not involve transfers, but are deposits and withdrawals by the same user despite high cash-out fees. The network is also used for storage, despite a zero interest rate. Surprisingly, the network is also used extensively for very short-term (less than two hours), short-distance transportation of cash because of extremely high crime.

Using the demand estimates, we provided measures of WTP to avoid carrying cash in one's pocket when traveling as well as keeping cash at home. We estimate that consumers are willing to pay up to 1.24% of the transaction amount to avoid carrying it in the form of cash for an extra kilometer. Similarly, they are willing to pay up to 0.8% to avoid keeping money at home for an extra day.

We found that the current pricing is inefficient. We proposed an incentive-compatible and Pareto-superior price discrimination scheme that would set the cash-out fee to zero for money received through a transfer while setting the cash-out fee a bit below the transfer fee for money deposited and withdrawn by the same person. Moreover, we demonstrated that the network can improve its profitability by implementing price discrimination based on the type of service, the location of the transaction origin, and the distance between transaction origin and destination. We analyzed several pricing strategies that illustrate that the network would benefit from differential pricing based on the above characteristics. We showed that price discriminating between rural and urban customers, as well as between short-distance and long-distance transportation/storage transactions, is particularly profitable. These results are linked to high crime rates in urban areas.

Beyond the pricing policies considered in this paper, we believe that the network could implement more complex price discrimination that would utilize more information about the consumer's network topology, dynamics of the customer's location, and timing of transactions. The effectiveness of such policies is a possible area for further research.

Acknowledgments

The authors thank MIC Tanzania Limited for providing the data. The authors also thank Paul David; Ben Hermalin; Ganesh Iyer; Jake Kendall; Jacques Thisse; participants of seminars at NYU Stern, Duke University, the University of Minnesota, and Northwestern University; participants of the 2016 American Economic Association Meetings, 2015 Marketing Science Conference, and 2nd Annual Conference on Network Science and Economics; two anonymous referees; the associate editor; and the editor-in-chief for many insightful comments.

Endnotes

[1] The marketing literature studying developing markets is pretty small; see Miller and Mobarak (2015), Sudhir and Talukdar (2015), Qian et al. (2014), and Zhang (2015) for examples.

[2] Beyond executing the focal types of transactions, there are other features of mobile money that may be useful to consumers, such as convenience of account management and integration with other mobile applications. Such features play a negligible role in Tanzania and similar developing countries, but could be important in more developed economies.

[3] The change involved raising the transfer and cash-out fees for certain sizes of transfers and cash-out transactions affecting more than 50% of transactions. It triggered an immediate substitution effect. In particular, comparing the week before the fee increase to that after the fee increase, the aggregate share of transfer transactions that experienced a fee increase decreased by 0.3%, and the average transfer distance increased by 1%.

[4] See Tanzania—Quicksights Report FII Tracker Survey, http://finclusion.org/wp-content/uploads/2014/07/FII-Tanzania-Wave-One-Survey-QuickSights-Report.pdf (accessed December 2014).

[5] Mobile banking networks are "incompatible" with each other and with the national currency, but there exists a costly "adapter" that provides compatibility at a price of 7.3%. See Economides (1991), Farrell and Saloner (1992), Katz and Shapiro (1994), and Economides (1996) for a discussion of adapters in the compatibility decision. In 2015, Tanzanian operators introduced a form of compatibility across networks.

[6] There is an evidence that SIM cards are relatively inexpensive. See InterMedia (2013).

[7] According to InterMedia (2013), only 14% of all households had made or received a nonremittance payment in the past six months using any type of cash delivery, including using m-money. The most

common types of nonremittance payments are school fees, government fees and taxes, utility bills, and salaries.

[8] This is also consistent with the risk of robbery being higher when transporting cash instead of using a P2P transfer.

[9] This is consistent with data from the Financial Inclusion Tracker Surveys project (InterMedia 2013) indicating that only 20% of the population believe that m-money can be used to save money.

[10] In a related study, Blumenstock et al. (2014) show that extreme levels of violence in Afghanistan, which may disrupt or even destroy the mobile money network, decrease the use of mobile money.

[11] Fafchamps and Lund (2003) show the existence of social insurance through networks of friends and relatives. Thus, an alternative explanation for transportation/storage transactions is to avoid friends and family members taking the money. The dependence of the WTP on the distance is harder to explain by this hypothesis. Storage transactions may also be replacing durable goods as a savings mechanism as identified by Rosenzweig and Wolpin (1993).

[12] Our analysis is related to the literature studying spatial competition in the industries with transportation cost. Hotelling (1929), Salop (1979), Economides (1984), and Thisse and Vives (1988) are related theoretical studies, and Thomadsen (2005), Davis (2006), McManus (2007), Houde (2012), and Miller and Osborne (2014) are related empirical studies. In this literature, the level of consumer utility depends on the distance between the consumer and the firm, which leads to the possibility of price discrimination. By contrast, in the case of the mobile money market, the spacial consumer heterogeneity arises because of the possibility of making the transaction bypassing the m-money network. The outside option involves risk related to carrying the money in person, which varies across locations. Thus, both the level of consumer utility and its price sensitivity depend on the customer's location. The distance of the consumer to the firm is a second-order determinant of demand because the m-money agent network is sufficiently dense. Moreover, the literature on spatial price discrimination usually assumes that a consumer cannot arbitrage across markets circumventing the price discrimination. In our model, arbitrage ability is a part of the outside option, since the outside option includes transporting cash across locations without using m-money.

[13] We drop phone account recharge transactions because they are executed at no charge to customers.

[14] Because of the short time span of our data, we are unable to capture new adoption of Tigo's service that results from price changes. Our estimates of the slope of the demand for a price increase should be unaffected; however, our estimates of the impact of price decrease should be regarded as short-run estimates. We note that short-run demand elasticity is likely to be smaller than long-run demand elasticity. We discuss the consequences of this difference when computing pricing counterfactuals.

[15] The inflation rate in Tanzania is relatively low and averages 6% in the relevant period. Thus, we exclude a high inflation rate as a possible cause of the short lifetime.

[16] According to a data set on crime provided by the Financial Inclusion Tracker Surveys project (InterMedia 2013), 6% of households experienced a theft, robbery, burglary, or assault in the last six months as of 2012.

[17] These traveling distances are based on the distance between the center of the cell at cash-in and the center of the cell at cash-out. For small traveling distances, it is possible that the user moves from an edge of one cell to an edge of another cell, or moves within the cell.

[18] For security reasons, agents are obliged to enforce identification checks during cash-out. Thus, cash-out by a person other than the account owner is prohibited.

[19] We note that the median distances are constant beyond the life span of three hours, indicating that the required traveling distance can be covered within this time.

[20] About 13% of respondents in the United Nations survey (United Nations Office on Drugs and Crime 2009) report being a victim of burglary or attempted burglary, which puts into question the long-term safety of keeping savings at home. This is also indicated by focus group studies by Plyler et al. (2010) in Kenya. There is considerable evidence that the crime rate is higher in Tanzania.

[21] We compute the probability by simulating the backward path of each dollar cashed out from the network between January 17 and February 3, 2013. Using this procedure, we can trace back the origin of about 97% of the cashed-out money.

[22] The actual cutoff has a negligible impact on the results in the remainder of this paper.

[23] The nonlinear nature of the pricing schedule offers a cheaper possibility to execute certain transfers (on the left side of the pricing intervals) by splitting one transfer into multiple transfers. Our data do not contain many instances of instantaneous multiple transfers between two individuals; nevertheless, our structural model captures some of the splitting as a substitution to a lower transfer bin, with a possibility of executing more than one transfer in a given period. We account for the fact that possible splitting generates correlation of the error terms within consumers and across transfers, but we do not account for possible time dependence, which would require a complex dynamic model.

[24] These estimates should be treated as back-of-the-envelope approximations, because they do not take into account switching to the outside option and changes in cash-out fees. Nevertheless, they suggest that there is enough variation in the data to reliably estimate the slope of demand.

[25] This classification is quite natural since the distribution of the average location is bimodal. Our results do not depend on this cutoff.

[26] When using product dummies together with time dummies, we obtain a price coefficient of -2.9; however, the coefficient is not statistically significant because the number of aggregate observations is small relative to the size of the price effects. When estimating the main model, we employ individual-level data and obtain extra power by controlling for consumer heterogeneity.

[27] In particular, Sunday has the smallest transfer propensity, while Wednesday has the highest.

[28] For the importance of the heterogeneity in the price sensitivity in the models with a single price, see Villas-Boas and Winer (1999).

[29] We note that when designing the above setup, we tried to write down the most parsimonious model that matches heterogeneity patterns in the data and produces sensible counterfactuals.

[30] We discretize the transfer amounts into 33 intervals that follow the modes of the empirical distribution of transfers. In particular, we ring-fence the following intervals: 10 intervals every 1,000 TSh between 0 and 10,000 TSh, 8 intervals every 5,000 TSh between 10,000 and 50,000 TSh, 6 intervals every 10,000 TSh between 50,000 and 100,000 TSh, and 8 intervals every 100,000 TSh between 100,000 and 1,000,000 TSh. We round down the transferred amount in each of the intervals (which coincides with the mode of the interval) and treat the data as discrete.

[31] Pakes and Pollard (1989) show that the simulated generalized method of moments is consistent for any fixed number of draws and that the simulation error approaches zero as the number of observations approaches infinity. Because our sample size is large, we experience numerically negligible simulation error with as few as 10 draws per observation.

[32] We report the standard errors of the second-stage estimates without the computationally expensive adjustment for the first-stage estimation error because this error is negligible given the size of our data set. We investigate the movement of second-stage estimates along the interquartile range of the asymptotic distribution of the first-stage estimates and find numerically negligible effects.

[33] An adjustment would require the user to keep some money in his pocket while walking home, or store extra shillings at home.

[34] Note that the structural elasticities are obtained using a different formula than those presented in Section 3. Structural elasticities measure total demand response rather than bracket-by-bracket response, which is more relevant for profit maximization. Moreover, the structural estimates of elasticities account for switching to an outside option and changes in cash-out fees.

[35] We choose transactions with a one-day time span to exclude savings motives from our estimates.

[36] This last calculation should be treated as an upper bound because it assumes that users are risk neutral. If users were risk averse, the rationalizing belief of users about the estimated probability of robbery would be smaller, since risk-averse agents would be willing to pay more than the expected value to avoid the risk of crime.

[37] According to Kinabo (2004), "The most prevalent crime [in Dar es Salaam] is burglary with about 43% of the households reporting being burgled over the last five years period. Simple theft is the second most frequent crime. Theft of livestock and crops is common in the rural areas of the city." Also, Louw et al. (2001) report particularly high levels of burglary and theft in Dar es Salaam.

[38] We chose the baseline distance to be 0.5 kilometers because we found that for longer lifetime transactions, the distance between cash-in and cash-out is close to this number.

[39] In the expression for variable profits from transfers, we make the simplification that all of the money transferred is immediately cashed out after a transfer. As shown in Section 2, under the current cash-out fees, this assumption is approximately satisfied and is likely to remain satisfied under the counterfactual pricing policies considered in this paper. The expression is multiplied by 3 because in Section 4 we assume that there are three transaction needs per day for every user.

[40] In case prices fail to be strategic complements, the results should be interpreted as short-run approximations. Pakes (2010) notes that prices may fail to be strategic complements if a price increase pushes price sensitive consumers out of the market.

[41] In a more general model, both N and H would be functions of prices.

[42] Another interpretation of this counterfactual is as a proxy for compatibility across m-money providers and between mobile providers and cash. Imposing full compatibility would allow costless flow of funds from m-money to cash, which can be proxied by setting the cash-out fees to zero, keeping in mind that conversion from cash to m-money (cash-in) is already free. The significant impact of setting cash-out fees to zero suggests a potential for regulation of cash-out fees and imposition of full compatibility.

[43] In both models of peer-to-peer transfers and cash-in–cash-out transactions, consumers have varying sensitivity, depending on their position on the demand curve. This introduces complications when pooling surpluses of individual customers to compute total welfare, since it is hard to compare utils of customers with low incomes with those of customers with high incomes. Any type of aggregation would imply a particular weighting of welfare of different groups. The usual solution is to rescale the individual surplus using dollar (or shilling) units. This solution is appropriate for smaller industries with relatively homogeneous customer bases, but in our case it drastically overvalues welfare of extremely affluent customers (with near-zero price coefficients). We think that such reweighing is not particularly desirable, and therefore we decided to use a total amount of transacted money as a measure of welfare.

[44] If the segments have unequal sizes, the profits have to be weighted appropriately.

[45] Fabinger and Weyl (2015) discuss the importance of curvature assumptions on pass-through rates, and Froeb et al. (2005) show that the full extent of price effects in response to a change in market structure is dependent on the demand system.

References

Aker JC (2010) Information from markets near and far: Mobile phones and agricultural markets in Niger. *Amer. Econom. J.: Appl. Econom.* 2(3):46–59.

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

Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1):57–78.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Blumenstock J, Callen M, Ghani T (2014) Violence and financial decisions: Evidence from mobile money in Afghanistan. Technical report, University of Washington, Seattle.

Bronnenberg BJ, Mahajan V (2001) Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Sci.* 20(3):284–299.

Davis P (2006) Spatial competition in retail markets: Movie theaters. *RAND J. Econom.* 37(4):964–982.

Demirgüç-Kunt A, Klapper L (2012) Financial inclusion in Africa: An overview. Working paper, World Bank, Washington, DC.

Economides N (1984) The principle of minimum differentiation revisited. *Eur. Econom. Rev.* 24(3):345–368.

Economides N (1991) Compatibility and the creation of shared networks. Guerin-Calvert M, Wildman S, eds. *Electronic Services Networks: A Business and Public Policy Challenge* (Praeger Publishing, New York), 39–55.

Economides N (1996) The economics of networks. *Internat. J. Indust. Organ.* 14(6):673–699.

Fabinger M, Weyl EG (2015) A tractable approach to pass-through patterns. Working paper, University of Tokyo, Tokyo.

Fafchamps M, Lund S (2003) Risk-sharing networks in rural Philippines. *J. Development Econom.* 71(2):261–287.

Farrell J, Klemperer P (2007) Coordination and lock-in: Competition with switching costs and network effects. *Handbook Indust. Organ.* 3:1967–2072.

Farrell J, Saloner G (1992) Converters, compatibility, and the control of interfaces. *J. Indust. Econom.* 40(1):9–35.

Farrell J, Shapiro C (2010) Antitrust evaluation of horizontal mergers: An economic alternative to market definition. *B.E. J. Theoret. Econom.* 10(1):Article 9.

Froeb L, Tschantz S, Werden GJ (2005) Pass-through rates and the price effects of mergers. *Internat. J. Indust. Organ.* 23(9):703–715.

Honoré BE (1992) Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* 60(3):533–565.

Honoré BE (2008) On marginal effects in semiparametric censored regression models. Working paper, Princeton University, Princeton, NJ.

Hotelling H (1929) Stability in competition. *Econom. J.* 39(153):41–57.

Houde JF (2012) Spatial differentiation and vertical mergers in retail markets for gasoline. *Amer. Econom. Rev.* 102(5):2147–2182.

InterMedia (2013) Mobile money in Tanzania: Use, barriers and opportunities. Report, Financial Inclusion Tracker Surveys Project, InterMedia, Washington, DC.

Jack W, Suri T (2014) Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. *Amer. Econom. Rev.* 104(1):183–223.

Jank W, Kannan PK (2005) Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Sci.* 24(4):623–634.

Katz ML, Shapiro C (1994) Systems competition and network effects. *J. Econom. Perspect.* 8(2):93–115.

Khanna T, Palepu K (1997) Why focused strategies may be wrong for emerging markets. *Harvard Bus. Rev.* 75(4):41–48.

Kinabo J (2004) Impact of globalization on food consumption, health and nutrition in urban areas: A case study of Dar es Salaam, United Republic of Tanzania. *Globalization of Food Systems in Developing Countries: Impact on Food Security and Nutrition*, 135–154.

King RG, Levine R (1993) Finance and growth: Schumpeter might be right. *Quart. J. Econom.* 108(3):717–737.

Levine R (1997) Financial development and economic growth: Views and agenda. *J. Econom. Literature* 35(2):688–726.

Louw A, Robertshaw R, Mtani A (2001) Dar es Salaam: Victim surveys as a basis for city safety strategies. *African Security Stud.* 10(1):60–74.

Mallat N (2007) Exploring consumer adoption of mobile payments—A qualitative study. *J. Strategic Inform. Systems* 16(4): 413–432.

McManus B (2007) Nonlinear pricing in an oligopoly market: The case of specialty coffee. *RAND J. Econom.* 38(2):512–532.

Miller G, Mobarak AM (2015) Learning about new technologies through opinion leaders and social networks: Experimental evidence on non-traditional stoves in rural Bangladesh. *Marketing Sci.* 34(4):480–499.

Miller NH, Osborne M (2014) Spatial differentiation and price discrimination in the cement industry: Evidence from a structural model. *RAND J. Econom.* 45(2):221–247.

Narasimhan L, Srinivasan K, Sudhir K (2015) Editorial: Marketing science in emerging markets. *Marketing Sci.* 34(4):473–479.

Pakes A (2010) Upward pricing pressure screens in the new merger guidelines; some pro's and con's. Presentation, DG Competition Authority, May 2011.

Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57(5):1027–1057.

Plyler M, Haas S, Ngarajan G (2010) Community-level economic effects of M-PESA in Kenya: Initial findings. Report, IRIS Center, University of Maryland, College Park.

Qian Y, Gong Q, Chen Y (2014) Untangling searchable and experiential quality responses to counterfeits. *Marketing Sci.* 34(4): 522–538.

Rosenzweig MR, Wolpin KI (1993) Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India. *J. Political Econom.* 101(2):223–244.

Salop SC (1979) Monopolistic competition with outside goods. *Bell J. Econom.* 10(1):141–156.

Schumpeter JA (1912) *The Theory of Economic Development* (Harvard University Press, Cambridge, MA).

Sudhir K, Talukdar D (2015) The "Peter Pan syndrome" in emerging markets: The productivity-transparency trade-off in IT adoption. *Marketing Sci.* 34(4):500–521.

Sudhir K, Priester J, Shum M, Atkin D, Foster A, Iyer G, Jin G, et al. (2015) Research opportunities in emerging markets: An interdisciplinary perspective from marketing, economics, and psychology. *Customer Needs Solutions* 2(4):264–276.

Thisse JF, Vives X (1988) On the strategic choice of spatial price policy. *Amer. Econom. Rev.* 78(1):122–137.

Thomadsen R (2005) The effect of ownership structure on prices in geographically differentiated industries. *RAND J. Econom.* 36(4):908–929.

United Nations Office on Drugs and Crime (2009) Victimization survey in Tanzania. Accessed December 2014, https://www.unodc.org/documents/data-and-analysis/Tanzania-victimization-ExSum.pdf.

Villas-Boas JM, Winer RS (1999) Endogeneity in brand choice models. *Management Sci.* 45(10):1324–1338.

Yang S, Allenby GM (2003) Modeling interdependent consumer preferences. *J. Marketing Res.* 40(3):282–294.

Zhang K (2015) Breaking free of a stereotype: Should a domestic brand pretend to be a foreign one? *Marketing Sci.* 34(4): 539–554.