Online Appendix:
Pay as you go: Pre-paid metering and electricity expenditures

Kelsey Jack (Tufts) and Grant Smith (University of Cape Town)

A.1 Data description

We assemble an 11 year panel (January 2004 to November 2014) of pre-paid electricity transactions using the City of Cape Town's vendor database. The dataset records every purchase made by a customer receiving electricity from the City of Cape Town through a pre-paid meter.

Variable definitions

We describe the primary variables in the administrative dataset, including how they are used to construct new variables for analysis.

- *new_aref*: Panel identifier that identifies a customer over time. The *new_aref* variable is a cleaned version of the identifier in the vending database called *aref*. A customer's *aref* is updated when the customer moves to a new location; as a result, we treat customers that move within our panel as different customers.

- *meter*: The meter number associated with the transaction. Transactions generate an encrypted code unique to a pre-paid electricity meter, so each transaction is identified with a meter, which corresponds uniquely to a *new_aref* at any given time in our sample (see Sample restrictions, below). A *new_aref* may correspond to more than one *meter* over time if a meter is replaced.

- *timestamp*: The date and time that a pre-paid electricity transaction was registered by the vending system. In some cases, a single transaction is broken into multiple lines corresponding to different components of the purchase (such as free basic electricity and paid electricity). A transaction-level dataset is constructed by aggregating within *new_aref* and *timestamp*. The *timestamp* variable is used to construct *year*, *month*, month-year (*monthyear*), day of the week (*weekday*) and day of the month (*day*) variables. Consequently, these variables are missing from the panel for a *new_aref* if no transaction occurs in a given year, month, etc.

- *trans_totalvalue*: The total Rand (ZAR) expenditure for a transaction, including the marginal cost (price/kWh), the fixed cost (service charge), and any debts recovered or refunds disbursed, plus the value added tax. All values are adjusted to reflect real 2012 values, using consumer price index adjustments from the South African Reserve Bank (www.resbank.co.za). The variable *trans_totalvalue* is used to generate customer-level average and total expenditures by *monthyear*, *month, weekday*, and *day*.

- *trans_totalunits*: The number of kilowatt hours (kWh) received for a transaction, including free basic electricity. The variable *trans_totalunits* is used to generate customer-level average and total kWh purchases by *monthyear*, *month, weekday*, and *day*.

- *trans_fbe_units*: The kWh of free basic electricity received for a transaction. The variable *trans_fbe_units* is used to generate a dummy variable (*any_fbe*) that indicates whether a *new_aref* received any free basic electricity in a month.

- *mean_propvalue*: The most recent available property value from the City of Cape Town's administrative records. In most cases, this corresponds to the 2012 value, though in 6 percent of the cases, the values have been updated in the dataset since 2012. Property values are assigned to a *new_aref* using the most recent *meter* associated with the *new_aref* (98.6 percent of matches) and using a location identifier (1.4 percent of matches). Where we cannot match on *meter* and there are multiple meters associated with a location identifier, we take the average value for all units associated with the location identifier and assign the resulting property value to all meters at that location (9 percent of the *new_arefs*). We drop 5,187 locations where one of the units has a value greater than one standard deviation from the mean property value, and we cannot match an individual unit to *meter*, resulting in unreliable property value estimates. Importantly, in the property value dataset, we do not observe whether a unit is rented or owner-occupied. According to Statistics South Africa, 54.2 percent of properties in the City of Cape Town were owner-occupied in 2011.[1]

Sample

We impose a number of sample restrictions in constructing the final dataset for analysis, which we describe in detail here.

- Commerical tariffs: All *new_arefs* with commercial tariffs at any point in the sample period are excluded from the dataset.

- Active customers: The sample is restricted to *new_arefs* that are active in 2014. This corresponds to 521,780 out of a total of 748,854 unique *new_arefs* that appear in the administrative dataset since 2004.

[1] See: http://beta2.statssa.gov.za/?page_id=993&id=city-of-cape-town-municipality (accessed January 11, 2015).

- Single transactions: 1,770 *new_arefs* are associated with only a single transaction in the data. These are likely to be temporary identification numbers or administrative errors and are dropped.

- Non-unique meter-customer pairs: Where *new_aref* corresponds to more than one *meter* within a month, we drop that *new_aref* from the sample. An *aref* has multiple *meter*s if the customer has multiple independently metered dwellings on a property. This restriction drops 4,307 *new_arefs* from the sample.

- Non-residential properties: Before merging property values into the dataset, all non-residential building categories are dropped. Out of 917,453 unique properties, 79,969 are non-residential properties. In addition, we drop blocks of flats if only the building as a whole is assigned a value (1,187 properties).

- Missing property values: After the above restrictions are imposed, 183,835 *new_arefs* fail to merge with the property value data. Most of these are associated with restrictions to the property value dataset (non-residential properties, specifically). These *new_arefs* are not included in the analysis.

- Property value outliers: After mergining property values with the vendor data, we drop 5,463 *new_arefs* with property values of 1,000 ZAR, and the top 0.5 percent of property values (1,615 *new_arefs*).

- 10 percent sample: From a list of all unique *new_aref*s that meet our inclusion criteria (N=321,604), we draw a random sample of 10 percent of the residential customers for inclusion in the transaction-level panel for analysis.

A.2 Summary statistics and robustness checks

Figure A.1 shows the evolution of the sample over time, from 2004 to 2014. The number of customers is plotted on the left axis and 2012 property values in thousand South African Rand (ZAR) on the right axis. The number of customers goes from under 10,000 at the start of the panel to over 30,000 by November 2014. Over time, the median 2012 property value in the sample increases, though the change is modest relative to the variation in the median property value among new customers in each year.

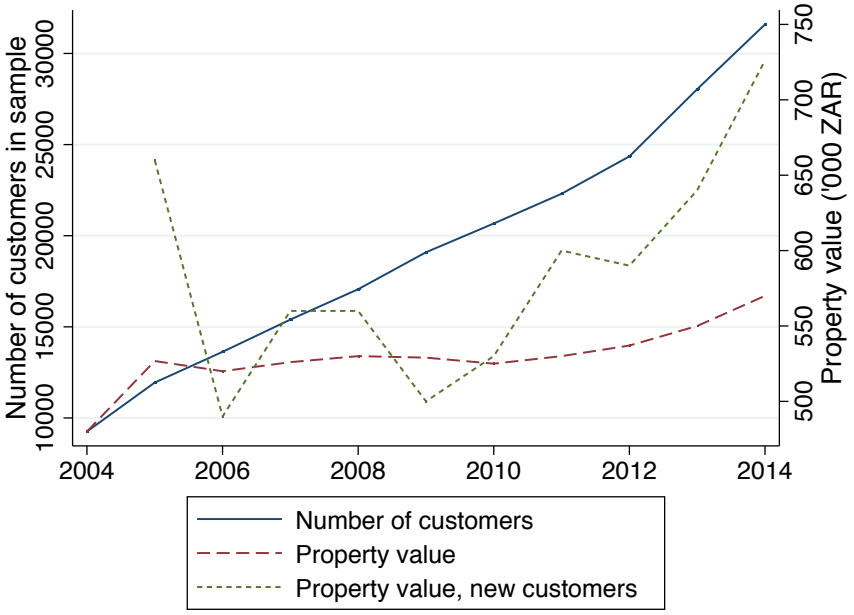

Figure A.1: Sample over time

Note: The solid line plots the number of customers in the sample between 2004 and 2014 (left axis). The right axis scales 2012 property values in thousand ZAR. The dashed line shows the median 2012 property value among customers in the sample. The dotted line shows the median 2012 property value among new customers added to the sample each year.

Table A.1 reports sample statistics by quartile. The table replicates many of the statistics in Table 1 in the main text, broken down by property value quartile. Results can also be compared with the regression analysis reported in Table 2 of the main text. Averages are calculated first for each customer and then averaged across customers within a quartile. The last row reports property value averages for each quartile of the full set of residential properties in the City of Cape Town, after applying our sampling restrictions (See Section A.1), but before merging with the pre-paid transaction dataset. We note some differences between the residential property values in the sample and all residential properties in the city. Specifically, the average property values in the bottom

quartiles are higher in the sampled properties. We can only speculate as to the cause of the difference between the sample and the full dataset. First, the lowest valued residential properties may not be electrified or may not have individually metering. Second, the lowest valued properties may not have property values assigned to individual units, which would exclude them from our sample, and is likely true for public housing.

Table A.1: Summary statistics by quartile

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Months active	79.123	88	75.958	80	66.233	66	65.311	66
Average transactions per month	13.050	10.433	10.773	8.241	6.073	4.471	3.890	3.000
Average ZAR per month	336.827	247.512	463.541	388.347	671.458	585.410	1083.400	861.944
Average kWh per month	364.777	334.264	466.959	436.094	594.136	548.296	927.430	767.370
Fraction of months with FBE	0.723	0.965	0.518	0.533	0.353	0.119	0.240	0.026
Sample property value ('000 ZAR)	160	190	400	400	800	790	2300	1600
All property value ('000 ZAR)	102	92	301	280	702	686	2162	1590

Note: Summary statistics at the customer level (N=31,570). The last row reports property values for all eligible residential properties in the City of Cape Town, prior to merging with the pre-paid transaction dataset. See Section A.1 of the appendix for a description of the restrictions on the property value sample.

Table A.2 replicates Table 2 in the main text, but with a control for free basic electricity (FBE). The variable equals one in each month that the customer receives FBE. Thus, the interpretation of the quartile coefficients changes, since the estimated effect of each quartile omits average differences (which may vary across quartiles) between customers with and without FBE. For all outcomes, the FBE coefficient is negative and significant, implying that customers with FBE have fewer transactions per month, purchase fewer kWh and spend less money. Their average transaction value is also smaller. With the exception of the number of transactions per month, the FBE coefficient takes the opposite sign as the quartile coefficients. This is driven in part by the greater likelihood of receiving FBE in the lower quartiles (see Table A.1). Relative to the results reported in Table 2 in the main text, the quartile coefficients are therefore smaller when FBE is controlled for. The exception to this is the first column, which reports transactions per month. Here the FBE coefficient takes the same sign as the quartile coefficients. At the extreme, customers with FBE may "purchase" only once per month, spend no money and just receive their FBE allowance. This will lower the number of observed transactions for very liquidity constrained customers, who are more likely to be in lower quartiles of the property value distribution. Finally, estimated elasticities are lower when the FBE control is included because it lowers the magnitude of the estimated quartile coefficients for kWh per month.

Table A.2: Expenditure patterns, controlling for free basic electricity

	Transactions per month	kWh per month	ZAR per month	ZAR per transaction	Property value elasticity
Quartile 1					0.168***
					(0.005)
Quartile 2	-2.069***	62.193***	29.843***	7.149***	0.281***
	(0.129)	(2.691)	(2.348)	(0.534)	(0.006)
Quartile 3	-6.740***	137.743***	91.540***	43.901***	0.395***
	(0.120)	(3.621)	(3.097)	(1.670)	(0.006)
Quartile 4	-8.769***	381.308***	291.565***	125.657***	0.468***
	(0.121)	(7.006)	(5.635)	(8.334)	(0.005)
Free basic electricity	-3.516***	-290.270***	-359.241***	-9.093***	
	(0.082)	(2.925)	(2.553)	(0.442)	
Observations	2,262,647	2,262,647	2,262,647	15,667,457	2,262,647
Quartile 1 mean	10.208	340.258	249.384	24.431	

***** p<0.10,****** p<0.05,******* p<0.01

Note: All regressions include month and year fixed effects and cluster standard errors at the customer level (N=31,570). Free basic electricity (FBE) is defined for each customer-month and equals one if the customer received FBE in that month. Columns (1), (2), (3) and (5) are estimated on a monthly panel; column (4) is estimated on a transaction-level panel. Column 5 reports the point elasticity of kWh per month with respect to property value, calculated at the median property value in each quartile.

A.3 Elasticity estimates

We calculate the point elasticities of electricity purchases in kWh with respect to property values at the median of each property value quartile. The results are reported in Table 2 in the main text. We begin by estimating:

$$kWh_{imy} = \delta_1 prop + \delta_2 prop^2 + \gamma_m + \eta_y + \epsilon_{imy} \tag{1}$$

where the outcome is the total kWh purchased by customer i in month m and year y. We regress this on property value and property value squared, where property value is time invariant and measured in ZAR2012. We include month and year fixed effects to account for seaonal and annual trends, and cluster standard errors at the customer level. Each observation is weighted by the number of months the customer appears in the sample.

We use the predicted values from equation 1 to calculate

$$\varepsilon = \frac{\partial kWh}{\partial prop}\frac{prop}{kWh} = (\hat{\delta_1} + 2\hat{\delta_2})\frac{prop}{kWh} \tag{2}$$

at the median of each property value quartile.

Because the relationship between property value and kWh purchases is estimated from the cross-sectional variation in property values, these elasticity estimates should be interpreted as long run elasticities. Though not directly comparable to income elasticity of demand, they offer the closest feasible approximation given the available data.

A.4 Tariffs

Further detail on the tariffs facing customers in the City of Cape Town during the sample period help to interpret the results.[2] Tariffs vary considerably between 2004 and 2014 and tariffs are updated in July of each year. Up to three different tariffs are offered in any given year. Figure A.2 shows the two increasing block tariffs available in 2012. The subsidized tariff includes 50 kWh of free basic electricity (FBE) for customers whose average purchases over the previous 12 months did not exceed 450 kWh per month.

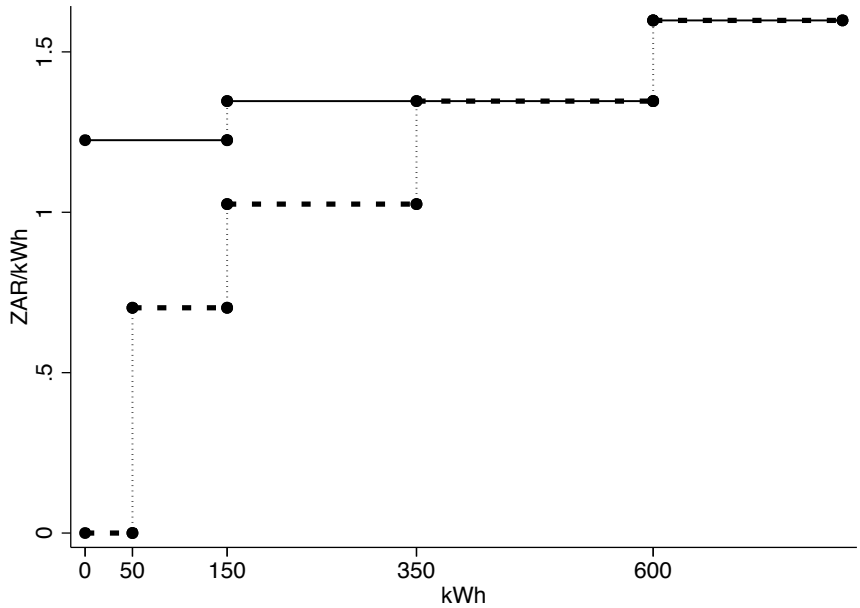

Figure A.2: 2012 tariff schedule

Note: The figure shows the tariff schedule in 2012. Customers are assigned to one of the two tariffs based on a 12 month rolling average of monthly purchases.

For some tariffs in some years, the marginal cost of electricity is constant, but for the majority follow an increasing block schedule based on cumulative expenditures during the month. On the first of each month, the customer returns to the lowest price block. Table A.3 provides the full history of tariffs during the sample period, including both the marginal and fixed costs (service charge) associated with each tariff, inclusive of value added tax. Panel A shows the price schedule for the lowest, subsidized tariff, Panel B shows the next tariff and Panel C, which is relevant only for a subset of years, show the tariffs for the highest monthly electricity purchases. Eligiblity is always based on a 12 month rolling average.

[2]We do not control for electricity prices in our analyses since we wish to explain variation in expenditures that include the variation in tariffs in the sample.

Table A.3: Tariffs 2004-2014

Panel A: Tariff 1

Year[a]	Marginal cost	Fixed cost	Eligibility[b]
2004	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.4422 & Q_{kWh} > 50 \end{cases}$	0	$Q_{AVG} \leq 500kWh$
2005	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.4599 & Q_{kWh} > 50 \end{cases}$	0	$Q_{AVG} \leq 450kWh$
2006	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.4634 & Q_{kWh} > 50 \end{cases}$	0	$Q_{AVG} \leq 450kWh$
	$P = 0.4634$	0	$450kWh < Q_{AVG} \leq 600kWh$
2007	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.4894 & Q_{kWh} > 150 \end{cases}$	0	$Q_{AVG} \leq 450kWh$
	$P = 0.4894$	0	$450kWh < Q_{AVG} \leq 600kWh$
2008	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.5638 & Q_{kWh} > 150 \end{cases}$	0	$Q_{AVG} \leq 400kWh$
2009	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.6144 & Q_{kWh} > 150 \end{cases}$	0	$Q_{AVG} \leq 400kWh$
2010	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.6625 & 50 < Q_{kWh} \leq 150 \\ 0.8034 & 150 < Q_{kWh} < 450 \end{cases}$	0	$Q_{AVG} \leq 400kWh$
2011	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.7022 & 50 < Q_{kWh} \leq 150 \\ 0.9239 & 150 < Q_{kWh} \leq 350 \\ 1.2247 & 350 < Q_{kWh} \leq 600 \\ 1.3459 & Q_{kWh} > 600 \end{cases}$	0	$Q_{AVG} \leq 450kWh$
2012	$P = \begin{cases} 0 & Q_{kWh} \leq 50 \\ 0.7402 & 50 < Q_{kWh} \leq 150 \\ 1.0254 & 150 < Q_{kWh} \leq 350 \\ 1.3465 & 350 < Q_{kWh} \leq 600 \\ 1.5981 & Q_{kWh} > 600 \end{cases}$	0	$Q_{AVG} \leq 450kWh$
2013	$P = \begin{cases} 0 & Q_{kWh} \leq 60 \\ 0.9086 & 60 < Q_{kWh} \leq 350 \\ 2.1090 & Q_{kWh} > 350 \end{cases}$	0	$Q_{AVG} \leq 250kWh$
	$P = \begin{cases} 0 & Q_{kWh} \leq 25 \\ 0.9086 & 25 < Q_{kWh} \leq 350 \\ 2.1090 & Q_{kWh} > 350 \end{cases}$	0	$250kWh < Q_{AVG} \leq 450kWh$
2014	$P = \begin{cases} 0 & Q_{kWh} \leq 60 \\ 0.9611 & 60 < Q_{kWh} \leq 350 \\ 2.3330 & Q_{kWh} > 350 \end{cases}$	0	$Q_{AVG} \leq 250kWh$
	$P = \begin{cases} 0 & Q_{kWh} \leq 25 \\ 0.9611 & 25 < Q_{kWh} \leq 350 \\ 2.3330 & Q_{kWh} > 350 \end{cases}$	0	$250kWh < Q_{AVG} \leq 450kWh$

Panel B: Tariff 2

Year	Marginal cost	Fixed cost	Eligibility
2004	$P = 0.3328$	1.2426 / day	$Q_{AVG} > 500kWh$
2005	$P = 0.3328$	1.4592 / day	$Q_{AVG} > 450kWh$
2006	$P = 0.3477$	2.2686 / day	$Q_{AVG} > 450kWh$
2007	$P = 0.3672$	2.3940 / day	$Q_{AVG} > 600kWh$
2008	$P = 0.6485$	2.8500 / day	$400kWh < Q_{AVG} \leq 800kWh$
2009	$P = 0.8820$	0	$400kWh < Q_{AVG} \leq 800kWh$
2010	$P = 1.0637$	0	$400kWh < Q_{AVG} \leq 1500kWh$
2011	$P = \begin{cases} 1.2247 & 0 < Q_{kWh} \leq 600 \\ 1.3459 & Q_{kWh} > 600 \end{cases}$	0	$Q_{AVG} > 450kWh$
2012	$P = \begin{cases} 1.2905 & 0 < Q_{kWh} \leq 150 \\ 1.3465 & 150 < Q_{kWh} \leq 600 \\ 1.5981 & Q_{kWh} > 600 \end{cases}$	0	$Q_{AVG} > 450kWh$
2013	$P = \begin{cases} 0 & Q_{kWh} \leq 60 \\ 0.7970 & 60 < Q_{kWh} \leq 350 \\ 1.8500 & Q_{kWh} > 350 \end{cases}$	0	$Q_{AVG} > 450kWh$
2014	$P = \begin{cases} 0 & Q_{kWh} \leq 60 \\ 0.8431 & 60 < Q_{kWh} \leq 350 \\ 2.0465 & Q_{kWh} > 350 \end{cases}$	0	$Q_{AVG} < 450kWh$

Panel C: Tariff 3

Year	Marginal cost	Fixed cost	Eligibility
2008	$P = 0.5401$	2.8500 / day	$Q_{AVG} > 800kWh$
2009	$P = 0.7346$	3.8760 / day	$Q_{AVG} > 800kWh$
2010	$P = 0.9117$	0	$Q_{AVG} > 1500kWh$

Note: Up to three tariffs are offered each year. Tariff 1 corresponds to the lowest price, subsidized tariff. All prices are inclusive of VAT.

[a] Tariffs are updated in July of each year.

[b] Eligibility is based on a 12 month rolling average.