Vallée, Geneviève

Working Paper

How long does it take you to pay? A duration study of canadian retail transaction payment times

Bank of Canada Staff Working Paper, No. 2018-46

Provided in Cooperation with:
Bank of Canada, Ottawa

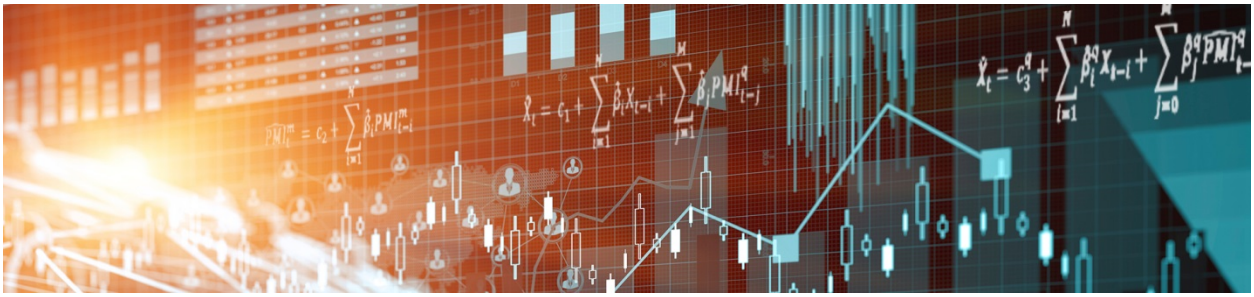
Suggested Citation: Vallée, Geneviève (2018) : How long does it take you to pay? A duration study of canadian retail transaction payment times, Bank of Canada Staff Working Paper, No. 2018-46, Bank of Canada, Ottawa, https://doi.org/10.34989/swp-2018-46

This Version is available at:
https://hdl.handle.net/10419/197899

BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2018-46

How Long Does It Take You to Pay? A Duration Study of Canadian Retail Transaction Payment Times



by Geneviève Vallée

Bank of Canada Staff Working Paper 2018-46

September 2018

How Long Does It Take You to Pay? A Duration Study of Canadian Retail Transaction Payment Times

by

Geneviève Vallée

Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
valleeg@econ.queensu.ca

ISSN 1701-9397

Acknowledgements

I'd like to thank Kim P. Huynh and Marcel Voia for co-supervising this research and providing valuable support and guidance. Furthermore, I'd like to thank visiting scholars and speakers to the Bank for their useful comments: Victor Aguirregabiria, David M. Drukker, Thomas Lemieux, Oz Shy, and Jeffery Wooldridge. I'd also like to thank my colleagues for their valuable insights: Ben Fung, Alex Shcherbakov, Gradon Nicholls, Chiyoung Ahn and Kerry Nield. Finally, thank you to Casey Jones for providing valuable documentation and preliminary analysis and cleaning of the 2014 Transaction Duration Study data. This research was completed as part of the MA internship program at the Bank of Canada.

Abstract

Using an exclusive data set of payment times for retail transactions made in Canada, I show that cash is the most time-efficient method of payment (MOP) when compared with payments by debit and credit cards. I model payment efficiency using Cox proportional hazard models, accounting for consumer choice of MOP. I propose two instruments to identify and estimate the causal relationship between MOP and payment time: (1) the value of the transaction, and (2) the duration of the payment preceding the one under observation. Discounting consumer selection underestimates the efficiency of cash relative to cards. Overall, the efficiency of MOPs is an important component of the private and social costs of making and accepting payments. The efficiency of cash helps explain its continued use in Canada, which is motivated by its low cost in terms of payment time for consumers and merchants.

Bank topics: Bank notes; Econometric and statistical methods; Payment clearing and settlement systems
JEL codes: C25, C36, C41, D23, E41, E42

Résumé

À l'aide d'un ensemble de données sur la rapidité de paiement en magasin au Canada, je montre qu'il est plus rapide de payer en argent comptant que par carte de crédit ou de débit. Je me sers de modèles à risques proportionnels de Cox pour modéliser l'efficacité des paiements, compte tenu du mode de paiement choisi par le consommateur. Je suggère deux variables pour dégager et estimer la relation causale entre le mode et le temps de paiement : 1) la valeur de la transaction et 2) le temps de paiement de la transaction précédente. Si le choix du consommateur est ignoré, l'efficacité de l'argent est sous-estimée par rapport à celle des cartes. L'efficacité des modes de paiement est un élément important des coûts sociaux et privés associés à l'exécution et à l'acceptation des transactions. L'efficacité de l'argent permet d'expliquer son utilisation continue au Canada : son emploi n'entraîne, pour les consommateurs et les marchands, qu'un faible coût en temps de paiement.

Sujets : Billets de banque; Méthodes économétriques et statistiques; Systèmes de compensation et de règlement des paiements
Codes JEL : C25, C36, C41, D23, E41, E42

Non-Technical Summary

I make use of the 2014 Transaction Duration Study data, a collection of information on payment times for retail transactions in Canada, to study the payment efficiency of cash. This data set was collected as part of the larger 2015 Retailer Survey on the Cost of Payment Methods and provides purchase characteristics for a set of retail transactions, such as consumer and clerk gender or total number of cash registers in store. All transactions under study were completed in "brick and mortar" stores, with online payments and bill payments excluded.

With these data, I study payment efficiency with the use of a duration model. Typically, these models are used to study the efficacy of a drug at reducing mortality in a treatment group, or the effect of job-training programs on the length of unemployment spells. Thus, applying a duration model to data on payment times allows me to evaluate the time efficiency of cash for processing payments against the combined alternative of debit cards and credit cards. I control for observable transaction characteristics that could influence the time taken for payment.

However, consumers make decisions on which method of payment to use based on factors that remain unobservable at the point of sale. These unobservable factors that affect payment choice range from reward programs, to habitual use of a payment instrument, or even preferences. Given that these are unobserved, their impact on the payment processing time cannot be controlled for. Therefore, I augment the analysis by considering the selection mechanism that influences the consumer's method of payment choice. To achieve identification, I propose to use the value of the transaction and the duration of the preceding payment as exogenous variations that affect the consumer's choice of method of payment, while having no direct impact on the payment time.

I find that cash is more efficient in terms of time than cards, allowing for the processing of more payments. However, ignoring consumer selection regarding method of payment underestimates this payment duration efficiency. This ability to process more payments has implications on how researchers define the transaction-variable cost of cash, as well as how consumers and merchants make decisions regarding which methods to adopt and accept, respectively. In part, it helps explain the continued use of cash to pay for low-value retail purchases, despite the increase in payment innovations.

1 Introduction

In the study of payments, making a purchase involves costs other than those associated with the price of the goods or services being transacted. At first glance, these costs appear trivial; however, they are non-negligible and have daily implications for how we choose to make payments. This has led several central banks to study the different costs of transacting associated with the use of different methods of payment (MOP) (European Commission, 2015; Norges Bank, 2014; Stewart et al., 2014; Jonker, 2013; Danmarks Nationalbank, 2012; Hayashi and Keeton, 2012; Segendorf and Jansson, 2012; Brits and Winder, 2005). In 2015, the Retailer Survey on the Cost of Payment Methods (RSCPM) was undertaken to calculate the costs of making and accepting cash, debit card, and credit card payments to Canadian stakeholders, namely, consumers, merchants, financial institutions and infrastructure, the Royal Canadian Mint, and the Bank of Canada (Kosse et al., 2017). This study found that the costs associated with the total resources required to make payments accounted for 0.78% of gross domestic product (GDP) (Kosse et al., 2017). Included among these resources was the time spent on payments at the point of sale.

The time taken to make a payment with a particular MOP affects the overall number of transactions that are processed, in turn affecting merchants' front-office costs and consumers' preference for certain MOPs. Payment instruments that process a higher number of transactions within a given interval are more efficient in terms of time. Studies around the world, including in Canada, have found that, on average, cash is the fastest MOP (Kosse et al., 2017; Polasik et al., 2012; Brits and Winder, 2005). For consumers, Klee (2006) shows that payment speed affects choice of MOP, and Arango et al. (2015b) find that the faster payment time of cash increases the likelihood of using it. Likewise, merchants utilize the payment speed of cash as an incentive for consumers to switch away from MOPs that are costly to accept, such as credit cards (Welte, 2016).

Yet, it's unclear how much more efficient in terms of time cash is relative to debit and credit cards at processing payments at the point of sale. In other words, how many more transactions can be processed using cash relative to cards for a given interval of time? With the help of the 2014 Transaction Duration Study data, which collected payment times of over 5,000 retail transactions, I estimate the rate of retail payments completed by cash relative to cards using a duration model and controlling for transaction characteristics. This allows me to show that cash is more efficient than debit cards and credit cards, processing more payments for a given interval of time.

As Polasik et al. (2012) note, time spent at the point of sale is made up of both queuing and time spent making the payment. In this paper, I focus exclusively on the time spent making payments at the point of sale and exclude time spent queuing. While MOPs with faster processing time should lead to a reduction in time spent queuing, other factors might influence this wait time. Given that I do not have data on queuing and that MOPs' processing efficiency at the point of sale is what I analyze, I exclude the time spent queuing.

In the face of competing MOPs, consumers optimize their choice by using the MOP that provides the greatest utility. Given that the data set used in this study doesn't originate from a controlled laboratory experiment, where consumers are randomly assigned to make a

payment with cash or cards, it's crucial that I account for the selection mechanism that dictates consumers' choice of MOP. However, this poses a challenge, since some important factors that determine MOP, such as preferences, are unobserved at the point of sale when payment is processed.

Therefore, to deal with the omitted variable bias present when estimating the impact of MOP on payment time, I model consumer selection and propose the use of two excluded variables: (1) the transaction value and (2) the time spent paying for the preceding transaction. These instruments allow for the correction of the endogeneity that arises from the correlation between MOP choice and consumer unobservables.

In Canada the usage of cash is strongest for transactions less than $25, with cards being more often used to pay for larger purchase transactions (Henry et al., 2015). Thus, I exploit this influence of transaction value on choice of MOP, since the value of the purchase will indirectly impact the payment time by influencing the consumer to select a particular MOP to pay for the purchase. Additionally, within the data set exists a detailed ordering of the transactions, allowing me to build a previous duration variable. Thus, when the consumer ahead takes a long time to complete the payment, the following consumer may use that time to switch to a perceived faster payment instrument to make up for the time lost due to waiting. By performing a two-stage analysis, where I identify the MOP selection mechanism in a reduced-form probit model, I show that the efficiency of cash is underestimated in the baseline model, which does not account for consumer selection.

As stakeholders of the payment system, central banks conduct research on topics such as the cost of making and accepting payments. Within this context, the time it takes payments to be processed and the corresponding relative time efficiency of the MOPs are components that influence how the costs of respective MOPs are defined. In turn, these costs have implications not only on researching the private and social costs of payments to societies, but also on whether merchants choose to accept certain payment instruments, and the daily decisions consumers make when they transact.

Moreover, central banks study the trends and factors that influence consumer adoption and usage of MOPs, as well as merchant acceptance. In fact, several studies show that time spent paying at the point of sale is a significant factor in determining consumers' choice of MOP (Polasik et al., 2012; Klee, 2006, 2008; Jonker, 2007), with speed of payment a MOP characteristic that is highly ranked by consumers, second only to security (Arango and Welte, 2012). Therefore, analyzing the time efficiency of MOPs not only allows for an estimation of the number of transactions faster MOPs can process, but also provides more information on a significant factor that influences consumer usage of MOPs, and inevitably demand.

To the best of my knowledge, no one has attempted to estimate the payment processing time efficiency of MOPs, while accounting for transaction characteristics and consumer selection. Thus, the aim of this paper is to fill this gap by analyzing the payment times of the two most popular payment options used in Canadian retail purchases: cash and cards.

The rest of the paper is structured as follows: Section 2 presents the 2014 Transaction Duration Study data set used in this paper, while the identification strategy is outlined in Section 3, and the results showcased in Section 4. Finally, I conclude in Section 5.

2 2014 Transaction Duration Study

To estimate the impact of MOP on duration, I use the 2014 Transaction Duration Study data from the Bank of Canada, which provide information on payment methods used at the point of sale and payment times measured in seconds. This study was conducted between October and December 2014 and is part of the larger 2015 RSCPM conducted by the Bank of Canada to estimate the cost of MOPs to retailers (Kosse et al., 2017).

The 2014 Transaction Duration Study data focus on purchases from retailers and exclude online purchases, online transfers, utility payments, and payments made by businesses and government institutions. I use data from 27 retail store locations,[1] which I categorize into the following seven types of business: *Grocery stores*, *Gas stations & convenience stores*, *Alcohol*, *General stores* (which constitutes large national retailers), *Coffee shops*, *Pharmacies*, and *Hardware stores*, from three cities in Ontario and Quebec, namely, Toronto, Ottawa, and Montreal. Each observation in the data set includes information on the day of the week; intervals for time of day; whether the transaction was completed at a regular checkout aisle, an express aisle, or in the back of the store; and the gender and estimated age of both the clerk and consumer. While studies show that card acceptance influence consumers' choice of MOP at the point of sale (Arango et al., 2015a; Huynh et al., 2014), all merchants included in the data set accepted all MOPs under study, that is, cash, debit cards, and credit cards. Therefore, the effect of MOP acceptance by merchants is controlled for here.

I drop two observations that present excessively long durations for reasons not related to the payment method, i.e., the consumer requesting a price correction and membership adjustment.

2.1 Payment duration

During the study, 12 observers were assigned to a specific store on a given day for the three-month period, and they recorded transaction characteristics for shifts of six hours. Payment duration was recorded in seconds, using a stopwatch application on a smartphone, from the moment the transaction amount was made known to the consumer to the moment the consumer received their receipt and any change back. If waiting time occurred during the purchase, then the observer stopped the timer until the waiting time was over and the transaction was resumed. Waiting time was defined as activities that did not pertain to accepting payments, such as packaging, social interactions with colleagues or the consumer, answering the phone, price checking of items, and so on. Given that only the time to make the payment is under observation, it's important to note that the payment duration observed here constitutes the time spent making the payment for the consumer, and the time spent receiving the payment for the merchant. This time is different from the total consumer time, which includes both queuing and time spent on making payments. Alternatively, this can be viewed as the merchant time which Garcia-Swartz et al. (2006a,b) refer to as "tender time" and constitutes a merchant cost in the form of a front-office cost.

Upon closer investigation of the data, I note the presence of measurement error in the pay-

[1] Data were collected in 29 retail locations. However, two locations were excluded due to measurement error in the payment time introduced by the observers assigned to these stores.

ment duration. Two of the 12 observers collected duration to the nearest tenth of a second, whereas the others rounded to the nearest second. The decimal observations account for approximately 10% of observations, presented in Table 1. Despite the higher precision of the decimal durations, I exclude these from the analysis, since they are restricted to only two types of businesses: *Grocery stores* and *Hardware stores*. Furthermore, I do not attempt to round these observations, in order not to introduce additional bias.

While there is no information in the data set that allows me to know what type of queuing configuration was used (pooled or multiple queues),[2] data on transactions are ordered chronologically, following the transaction order. This allows me to build an instrumental variable (IV) based on the previous transaction's payment time. Within the data, there exist 168 transaction groups. I define a *transaction group* as the set of transactions observed by an observer, stationed in a particular store, at a random cash register, for a given shift. These transaction groups reflect the continuous flow of consumers making payments at a given cash register. The first transaction to occur in a group is considered *first* and is given a previous transaction duration of zero.

To attenuate the bias in the IV arising from the measurement error of transaction duration, if transaction $i-1$ happens to have a decimal duration, then the previous transaction duration for transaction i is set to be missing.

Figure 1 illustrates the distribution of payments times, which is leftward skewed. Because payment durations are so short, ranging from 1 to 180 seconds, there is no right censoring within the data set. For this analysis, I treat payments as non-repeatable events, since a completed purchase is considered as a single occurrence.

2.2 Descriptive statistics

For the 5,107 transactions under study in the full model, Table 2 presents the mean and median durations for cash, debit cards, credit cards, and the pool group of cards. I frame the analysis with a focus on cash versus the combined main alternatives, debit cards and credit cards. Even though consumers choose to use debit cards or credit cards for different reasons, at the point of sale both use the same payment-processing technology. Regardless of the type of card chosen, the clerk inputs the transaction value into a card terminal, which causes a visible delay in the card payment times when compared with cash, as shown in Figure 1. Ultimately, it's this technology that determines the payment time when using a card, making debit and credit cards comparable in the context of payment time.

Within the pool of payment cards there are three card subtypes available to consumers, presented in Table 2, which affect payment times using cards. Firstly, swipe cards represent the oldest type of card payment technology and are the least used. Users of these cards are required to swipe the magnetic stripe of their card through the terminal and provide a signature for a credit card or a personal identification number (PIN) for a debit card to clear the payment. However, when comparing swipe debit and credit cards in Table 2, we notice a marked difference in the payment time. As Polasik et al. (2012) show, from a merchant or "pure" payment time

[2] Approximately 4.5% of transactions were conducted in a store with a single cash register, which necessarily implies a single or pooled queue; the remaining 95.5% were in a store with multiple cash registers.

perspective, cards that require a signature rather than a PIN are faster, despite being slower from a consumer perspective. When considering exclusively the payment time, swipe and sign cards are faster than swipe and PIN, since for the former, a consumer can sign and grab the receipt, while for the latter, the PIN must be authorized before the receipt is printed.

Second, chip and PIN cards utilize a chip that is inserted into the card terminal instead of swiping. Following this, the consumer provides a PIN, regardless of whether the card is a debit card or a credit card, to authorize the payment. At the time of this study, these types of cards were the most widely used.

Third, contactless cards are the latest technology, requiring consumers to press the card to the terminal without the need for a PIN or signature, causing an even further reduction in payment time. This differentiates contactless cards from swipe and chip and PIN, since both the processing of the card as well as the authorization of the payment require only one action from the consumer. However, when the value of the transaction exceeds a certain amount, most contactless card users are required to use the chip and PIN method as an added security feature. At the time of this study, contactless technology was relatively new; thus, too few observations were sampled to compare cash exclusively against contactless cards.[3]

For the purpose of this analysis, each of these subtypes are included, since they represent the full range of debit and credit cards available to consumers. Moreover, newer card subtypes, such as contactless and chip and PIN, are backwards compatible. That is, they can still utilize older technologies. Contactless cards can make chip and PIN or swipe payments in the event the contactless card is not accepted. Similarly, chip and PIN can make swipe payments. Including contactless cards in the analysis provides cards with an advantage when they are compared with cash.

2.3 Transaction value

To illustrate the sample composition of transactions, I segment the data into four transaction value groups: (1) transactions with a value of $15 or less; (2) those with a value between $15 and $25; (3) those with a value between $25 and $50; and (4) those with a value of $50 or more. Figure 3 presents the number of observations in each of the transaction value groups, and shows that cash is most often used for the lowest value transactions. This is consistent with Henry et al. (2015), who show that, for low-value transactions, cash dominates debit cards and credit cards by total number of transactions. Similarly to Henry et al. (2015), the number of card payments overtakes cash payments for the higher-transaction groups.

Figure 4 contrasts the impact of transaction value on payment time for cash and cards. From here, we note that the payment time is affected differently by the use of cash or cards, depending on the value of the transaction. Cards possess the same median processing speed, regardless of value. However, cash's median speed is positively correlated with the value of the transaction.

When taken together, Figures 3 and 4 show that a correlation exists between the payment time and the payment instrument chosen. As cash becomes slower for processing payments,

[3]When running an estimation that excludes contactless cards, the processing efficiency of cash against cards is only slightly reduced.

consumers appear to move away from it towards cards. However, many factors, most of which are unobservable at the point of sale, influence the choice of MOP as the value of the transaction increases.

3 Identification strategy

3.1 Model for payment efficiency

I model the payment time of cash and card transactions using a duration model. In the context of payments, duration analysis allows me to estimate the relative efficiency of cash-based payments against card-based ones with a hazard function. Let $T \geq 0$ be the set of payment durations, measured in seconds, and t be a specific value of T. The realization, defined as the completion of the payment, is represented by the following density function,

$$f(t) = \Pr(t \leq T < t + dt),$$

and a cumulative density function (CDF),

$$F(t) = \int_0^t f(s)ds = \Pr(T \leq t), \quad t \geq 0.$$

Using the CDF, we can obtain the survivor function $S(t)$, which represents the probability of the payment still not being complete at time t,

$$S(t) = 1 - F(t) = \Pr(T > t).$$

Combining this information, we get the hazard function, which represents the instantaneous rate of payment completion, conditional on the payment still not being completed at time t,

$$h(t) = \Pr(t \leq T < t + dt | T \geq t) = \frac{f(t)}{S(t)}.$$

Thus, as in the context of drug trials on patient mortality, I can use a duration model to evaluate the impact of MOP on payment time, while conditioning on transaction characteristics represented by covariates X. Given the short duration of payment times, I model this payment efficiency with the use of a Cox proportional hazard model (CPHM), which consists of time-invariant covariates, with parameter of interest γ:

$$h(t|MOP_i, X_i, v_i) = h_0(t)v_i \exp\{\gamma MOP_i + X_i'\beta\}. \tag{1}$$

The choice of a CPHM to model payment efficiency is in part motivated by the payment time, which is measured in seconds. While the set of payment times is technically discrete, the short duration of the observations allows me to view these as part of a continuous set. Estimating the hazard ratios allows me to measure the different rates at which payments are completed on an interval of time between cash and cards, conditioning on the transactions still not completed at the beginning of that interval.

Another advantage of using a CPHM is its semi-parametric specification. The baseline hazard is left as non-parametric, since the CPHM makes no assumption about its functional form. This allows for some additional flexibility in estimating the impact of MOP on payment time without tying the model to any parametric baseline hazard function.

I include a vector of exogenous covariates X, presented in Table 3, which represent transaction characteristics of the payment under observation, as well as determinants in MOP choice.

I augment the duration model with an unobserved heterogeneity (UH) term v_i, also called *frailty*, that captures the latent differences between transactions completed in the seven types of business, using gamma-distributed latent random effects. As seen in Figure 5, each type of business exhibits heterogeneous payment patterns due to their unique environments. This leads to clerks and consumers adopting different behaviours at the point of sale, which can impact the time it takes to complete the payment. In essence, a large purchase at a grocery store is viewed differently than paying for a daily coffee at a café. Failing to account for UH causes misspecification, since transactions with equal covariates are considered identical. On the other hand, including UH allows for within-group correlation of the transactions completed in a specific store, which leads to more efficient standard errors. In this case, the hazard conditional on the UH is now the instantaneous probability that the payment is completed before time t, given that a transaction is completed in a type of business with a specific *frailty* (Wooldridge, 2010). The choice of introducing a UH term that is gamma distributed is justified in both theory and analytical tractability (Huynh et al., 2010).

3.2 Accounting for consumers' MOP choice

Under a controlled setting, where consumers are randomly assigned to use either cash or cards, identifying the impact of MOP on payment efficiency in equation (1) is straightforward. However, consumers exercise free will in making decisions on which MOP to use, based on factors that affect their overall preferences. Arango et al. (2015b) show that the consumer's choice of instrument is determined by the attributes it possesses, for example fee, rewards, speed and security. Additionally, Klee (2006) demonstrates that choice of MOP is affected by the payment speed of the instrument, showing that the payment time of debit cards causes consumers to favour them over cheques. As Garen (1984) explains, the decision-making process of agents is endogenously determined from an optimal action taken. In the context of payment efficiency, consumers choose the MOP they believe to be the most efficient (or fastest at completing payments) by maximizing a utility function that considers the payment instrument's attributes, yet remains latent at the point of sale. Thus, failing to account for selection of MOP leads to inconsistent and biased parameter estimates in the model for payment efficiency outlined in equation (1).

Following Terza (1998), I consider the endogenous selection by modelling the likelihood that consumer i chooses to use cash with the following latent utility function, where X_i is a vector of observable transaction characteristics affecting consumer choice, Z_i a vector of excluded variables, and ε_i a random error term:

$$MOP_i = \begin{cases} 1 & \text{iff} \quad X_i'\alpha + Z_i'\delta + \varepsilon_i \geq 0, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where cash is selected by consumer i when $MOP_i = 1$.

To handle the consumer selection, I propose two IVs to correct the endogeneity arising from the choice of MOP. Firstly, I use transaction value as a determinant of payment choice. Transaction value can be viewed as exogenous to the transaction itself, since at the point of sale, consumers are usually aware of the value of the purchase and take it as a given. However, despite being exogenous, the value impacts the payment method chosen. Indeed, Arango et al. (2015b) outline this strong relationship between choice of MOP and transaction value as one of the "key-stylized facts in retail payment literature." Many reasons can partially explain this, such as the desire to reduce the amount of coins held (Chen et al., 2017) or because consumers value convenience and speed (Henry et al., 2015; Arango and Welte, 2012).

Evidence that transaction value has an impact only on payment time through the choice of MOP can be seen in Figure 4. When consumers choose to use cards, the transaction value has no impact on the payment time. However, when cash is chosen, the median payment time increases with the value of the purchase. This difference in payment time for cash across transaction values arises from the counting process required to make a payment. On the side of the payment, consumers need to count the amount of cash to hand to the clerk. More expensive transactions require that more denominations be counted, thereby increasing the payment time. We also see evidence of this counting from the clerk side. Transactions that have non-whole values, that is \$10.15 rather than \$10, require more counting time, since clerks need to count the change to be given back.

Secondly, I use the payment time of the preceding transaction as an instrumental variable, since this time can influence the consumer's choice of MOP. Imagine an individual standing in a long line, waiting to pay, and that the consumer ahead is struggling to find their MOP. If the individual happens to be an impatient person, then perhaps they prepare their MOP ahead of time, or switch to a faster one. Ultimately, this choice to prepare ahead of time or change MOP impacts the payment time.

Like Klee (2006), I assume clerks don't change their behaviour to compensate for long payment times; thus, each payment time remains a mutually independent event. Therefore, the time it takes to pay affects only a consumer's payment time through the selection of the preferred MOP. The transaction value, the time of day, day of the week, province, or any other controls are uncorrelated to the previous payment duration, since the amount of time the consumer in front takes to pay is exogenously determined.

3.3 Two-stage model

Following Terza (1998) and Wooldridge (2014), I let ε_i have a standard normal distribution, and augment the payment efficiency model with a first-stage probit to capture the consumer selection of MOP:

$$\Pr(MOP_i = 1|X_i, Z_i) = \Phi(X_i'\beta + Z_i'\delta), \tag{3}$$

$$h(t|MOP_i, X_i, v_i, \hat{gr}_i) = h_0(t)v_i \exp\{\gamma MOP_i + X_i'\beta + \theta \hat{gr}_i\}. \tag{4}$$

Using the control-function (CF) method, I augment the second stage with the generalized residuals $\hat{gr}_i$ from the first by including them as an additional covariate (Terza, 1998; Wooldridge, 2014). Conditioning payment efficiency on $\hat{gr}_i$ captures the variation in MOP choice from the utility of usage consumers receive, essentially making the estimation on a non-random sample similar to a randomized one. In essence, inclusion of $\hat{gr}_i$ in the duration model removes the selection due to unobserved differences in consumer preferences.

Additionally, treating the selection with the use of the CF method with $\hat{gr}_i$ provides the added advantage that exogeneity of MOP can be easily tested using a t-test with the null hypothesis that $\theta = 0$ (Wooldridge, 2014).

I estimate two main specifications of the duration models using the CF method and compare them with the baseline CPHM (Table 4, column 1), as well as a model estimating consumer selection using a linear probability model (LPM) (Table 4, column 7). The first CF model (CPHM-CF I) includes an additional regressor $\hat{gr}$ generated from the probit with the two instrumental variables (Table 4, column 3). For the second model (CPHM-CF II), I include two additional squared transformations of the IVs in the probit and generate the regressor $\hat{gr}$ from this first stage (Table 4, column 5) (Dong, 2010; Escanciano et al., 2016).

3.4 Estimation method

The estimation follows Wrenn et al. (2017), such that the goal is to instrument a duration model using the CF method. However, the first departure from Wrenn et al. (2017) arises from the endogenous regressor in the duration model being binary. Secondly, while Wrenn et al. (2017) estimate a discrete duration model with the use of a probit, I estimate a proportional hazard model.

Thus, I estimate the first-stage probit using maximum likelihood estimation and, following the CF method, add the generalized residuals as a generated regressor in the duration model as seen in equation (4). The generalized residuals are computed using the inverse Mills ratio, $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$, such that $\hat{gr}_i = MOP_i\lambda(X_i'\hat{\alpha} + Z_i'\hat{\delta}) - (1 - MOP_i)\lambda(-X_i'\hat{\alpha} - Z_i'\hat{\delta})$.

I estimate the second-stage parameters $\{\gamma, \beta_1, \beta_2, ..., \beta_{13}, \theta, v_i\}$ using partial likelihood, following Cox (1975). Unlike full likelihood estimation, partial likelihood simplifies the estimation by eliminating the baseline hazard $\lambda_0(t)$, which is a nuisance parameter.[4]

Given that I deal with a large data set of retail transactions, it's not surprising that several unique transactions have the same payment time. Unfortunately, in Cox's original model, failure times are continuous and therefore unique to each observation. Thus, to account for the payment durations, I use Breslow's method for tied times (Breslow, 1974). While other methods exist to deal with the tied times, such as the exact-marginal calculation, the exact-partial calculation,

[4]With a large enough sample, partial likelihood estimates are asymptotically normal (Tsiatis, 1981).

and the Efron approximation, each of these are much more computationally intensive than Breslow's approximation.

3.5 Reporting results

Due to the difficulty of interpreting and comparing the estimated coefficients of choice models, I present in Table 5 the marginal effects of a change in the probability of using cash. The marginal change in the kth covariate for a categorical variable x_k, such as North American Industry Classification System (NAICS) code or province, is

$$\Pr(MOP = 1|x_k = 1; Z) - \Pr(MOP = 1|x_k = 0; Z).$$

For continuous variables, such as the number of registers in the store, the marginal effect of unit change in the value of the continuous variable leads to a change in probability of using cash, which is measured as

$$\frac{d\Pr(MOP = 1|X, Z)}{dx_k} = \phi(x_k'\beta)\beta,$$

where ϕ represents the CDF of the standard normal distribution.

For the second-stage regression output, I present the exponentiated coefficients or hazard ratios in Table 4. These values should not be confused with the estimated coefficients, which are calculated by taking the exponential of the hazard ratios estimated by the model.

Unlike the estimated coefficients in a duration model, the hazard ratios are non-negative. For binary variables, like MOP, hazard ratios represent the likelihood that the cash transaction will be completed over the benchmark choice of cards, at any point in time. A hazard ratio greater than 1 would signify that cash transactions are more likely to be completed than card-bases ones, whereas a hazard ratio less than 1 would imply that cards are more likely. For continuous variables, hazard ratios represent the likelihood of the event occurring, in this case the transaction being completed, following a unit change in that variable.

4 Results

4.1 First-stage results

The results of the reduced-form model for consumer selection of cash are compiled in Table 5. I present several specifications. Columns (1) and (3) present the probit results for the models with two and four instruments, respectively. For comparison, I showcase the estimates from an LPM in column (5). Finally, columns (2), (4), and (6) show the results when no control variables are included as a robustness check.

I conduct Wald tests to determine the validity of the instruments and list the p-value of the tests in Table 5 for all specifications. For each instrumental variable, I test the null hypothesis that the coefficient is zero.

I reject the null that transaction value has no impact on the choice of MOP for all models. This is consistent with other studies that find transaction value to be a significant determinant

of MOP choice (Fujiki and Tanaka, 2017; Arango et al., 2015b; Klee, 2008; Bounie and François, 2006).

However, I cannot reject the null hypothesis that previous payment duration doesn't affect MOP selection. This inability to reject the null hypothesis could be due to the construction of the instrument. While the detailed ordering in the data set allows me to know how long the previous consumer took to pay, it does not provide information on the inter-transaction time. That is, time elapsed between transactions is not available; therefore, I do not know the wait time experienced by the consumer while queuing. The previous duration instrument is a compelling exclusion restriction, yet it may be the case that for some transactions in the data set too much time passed between transactions for the previous duration to have any meaningful impact.

Turning to the estimated parameters of the model, the coefficient for transaction value is a significant variable when determining the choice of MOP, with consumers moving towards card-based MOPs when the transaction values increase. However, the marginal effect of the change in transaction value is relatively small, suggesting that consumers exhibit a preference for a type of MOP, and may be induced to switch only once the transaction value becomes sufficiently large. This is consistent with the findings of Wakamori and Welte (2017) that preferences and habit drive usage of cash, specifically for low-value transactions. On the other hand, several reasons can explain why consumers prefer using cards for high-value transactions. For instance, more expensive purchases make reward programs on cards more appealing, providing consumers with an incentive to utilize them. Additionally, given that coins are burdensome (Chen et al., 2017), using cards eliminates the amount of change received and limits the need to carry coins.

4.2 Second-stage results

Table 4 shows that accounting for consumer selection increases the efficiency of cash relative to cards. Column (1) presents the rate of transactions completed by cash in the baseline CPHM model. When considering consumers' choice of MOP, shown in columns (3) and (5), this rate increases. While this jump in the rate of transactions completed with cash is large, I justify it by performing a t-test on the coefficient of the generalized residual term $\hat{gr}$, testing whether the model is exogenously determined. To achieve this, I test the null hypothesis that the coefficient on the generated residual parameter in the duration model is equal to zero. That is, I test whether the generated regressor's coefficient is zero, which would imply that there is no endogeneity present in the MOP variable (Wooldridge, 2010). I reject this at the highest level of significance, indicating that there exists endogeneity, which leads to a downward bias in the estimates. Thus, accounting for consumer selection increases the estimated efficiency of cash.

Once again, as a comparison I present the results of the CPHM model using an LPM to estimate the probability of a consumer selecting cash in column (7) of Table 4. In this case, the fitted residuals of the LPM are included as a generated regressor in the duration model.

Figure 6 illustrates the smooth hazard functions for both cash and cards when estimating the conditional probability that either MOP completes a retail payment over the time interval under study. The shape of these two functions provides a few insights into how payments are completed for cash and cards. First, the hazard rate is highest for cash for each of the three

models estimated.

Second, the rates of completion appear to be highest within the first minute of the payment for both cash and cards, as shown in Figure 6, signifying that most transactions are completed within this interval. This isn't surprising, given that 98% of retail transactions in the data are completed within that time.

Third, hazard rates for cash and cards are much lower for transactions that remain incomplete after a minute. Thus, the probability of completing the payment as each second passes is reduced relative to those completed within the first minute. The hazard rates remain relatively constant for card-based transactions, while hazard rates of cash-based transactions decrease after approximately 130 seconds spent paying. This suggests that transactions being paid for in cash that remain incomplete after 130 seconds become less probable of being completed as each second passes. This could be explained by the few transactions that are paid for in cash at this point, which drives down the hazard rate. In fact, of the eight transactions that have yet to be paid after 130 seconds of payment time, only one is cash-based. In other words, transactions with long payment times are more likely to be card-based than cash-based, since most cash-based payments have already been completed by this time.

Several types of residuals can be used to assess the fit of a Cox model. The plot of the Cox-Snell residuals, presented in Figure 7, are used to assess the overall fit. From Figure 7, the data appear to fit well, but deviate from the reference line on right-hand tail of the distribution, which could be explained by the smaller sample size of transactions with long payment times.

In Figure 8 and Figure 9, the martingale and deviance residuals are presented, respectively. These are used to identify potential outliers. Finally, the Schoenfeld residuals, shown in Figure 10, are used to test the hypothesis of proportionality.

4.2.1 Proportionality assumption

The Cox model relies on the assumption of proportional hazards, which implies hazard ratios are time invariant. When this assumption is satisfied, coefficients can be interpreted as a relative effect on the instantaneous hazard. Thus, if the assumption holds, the hazard ratios estimated in Table 4 have the same effect at all points in time. However, I present the results testing the null hypothesis of proportional hazards using the Schoenfeld residual in Table 6 and show that this assumption is violated. Thus, the hazard ratios can be viewed as an average rate over the interval of time under study.

The presence of endogeneity in the model provides an explanation for the violation of proportionality. In Figure 4, the median payment time for cards remains constant over all purchase values, while cash transactions vary.

4.2.2 Split sample test

To determine if there exists unobservable heterogeneity in the data, which could bias the estimates, and motivate the use of the frailty term v_i, I conduct a split sample test. I follow the algorithm presented by Huynh et al. (2010),[5] which is based on the methodology advanced

[5]For more detail on the steps required to perform this test, consult the Appendix in Huynh et al. (2010).

by Dufour and Jasiak (2001). I employ this empirical test due to the lack of a formal test for misspecified correlation between observables and unobservables.

For practical purposes this test is conducted on an accelerate failure time (AFT) model, which can be viewed as a linearization of the Cox model. This test allows for the confirmation of the presence of UH in the data by:

1. Estimating a first AFT model on half of the data;
2. Using the estimated coefficient to generate predicted values of duration on the second half of the data;
3. Subtracting the predicted duration from the actual duration for the second half;
4. Regressing the differenced duration against the same covariates in the first step;
5. Testing the null hypothesis of no bias in the parameter of interest ($H_0 : MOP = 0$) using a $\chi^2(1)$ distribution.

When performing a split sample test on the MOP coefficient, I cannot reject the null hypothesis of no bias. However, the inconclusive result does not exclude the possible presence of UH, which could bias the estimates.

In this paper, frailty represents the overall time efficiency that a particular store has at processing payments. Given how different stores might induce different types of behaviour, which could impact payment times, I base the frailty term on the seven types of business surveyed in the Transaction Duration Study and included in this analysis.

4.3 Discussion

The result that cash is more efficient than cards in terms of payment time for Canadian retail transactions has implications for consumers and merchants.

First for consumers, the efficiency of cash provides the reasoning behind its continued dominance for low-value retail purchases. Evidence shows that consumers rate cash as easier to use compared with debit and credit cards (Fung et al., 2015; Arango et al., 2015b), with speed and convenience being significant determinants in the decision regarding payment choice (Polasik et al., 2012; Klee, 2006, 2008; Jonker, 2007). Thus, for low-value transactions, consumers perceive that cash is the fastest MOP and actively select it to minimize the time spent paying, which constitutes an opportunity cost for them. To some extent this motivates the assumptions of the shopping time model, which claims money demand is driven by consumers minimizing transactions times. Thus, showing that cash is the most efficient instrument in terms of time at processing payments suggests it minimizes payment time for consumers, leading to its continued demand. This provides a basis on which to challenge the idea that Canada is moving towards a cashless economy.

Second, the time spent accepting payments represents a front-office cost to merchants that varies with the number of transactions processed. Payment instruments efficient at processing payments lead to a reduction in the total time spent accepting payments. Thus, by showing that cash can process a higher number of retail transactions than cards, I demonstrate that

the transaction-variable cost of cash is less than that of cards. In fact, many retailers will exploit this processing-speed feature of cash by offering non-pecuniary incentives to consumers, such as cash-only checkouts, to induce switching away from credit cards, since these carry fees that eat away at merchants' profit margins (Welte, 2016). Other merchants even resort to unprofitable methods, for example, discounting, in an attempt to steer consumers towards payment instruments that are cheaper to accept than credit cards (Welte, 2016). Given that merchant acceptance fuels consumer usage of MOPs (Huynh et al., 2014), understanding the cost structure of cash and cards gives insight into how these MOPs proliferate.

Yet, some merchants still choose to accept these costly MOPs (typically credit cards), despite high costs, to remain competitive when attracting consumers (Bounie et al., 2016). In Canada, approximately two-thirds of small and medium businesses accept credit cards (Fung et al., 2017). This has led researchers and policy-makers to question whether acquirer fees imposed on merchants are too high. By applying the merchant indifference test (MIT), Fung et al. (2018) compare cash and credit card acceptance in Canadian data to determine the level of fees at which merchants would be indifferent to accepting credit cards over cash. They show that the results are sensitive to the underlying assumptions of how the costs of cash are allocated. Thus, by showing that a relative difference between the tender time of cash and card payment exists, I demonstrate that a component of costs must be left to vary with number of transactions processed.

Despite showing that cash is more time efficient at processing payments, I conduct this analysis solely within the context of retail purchases. Demographic characteristics help explain some of the preferences and habits of consumers for certain payment options. However, even for a single consumer, these preferences may vary based on the type of purchase. The wide variations that exist between different types of purchases will inevitably dictate the behaviour consumers adopt towards the payment. For instance, speed of payment may be more important when paying for groceries, than major appliances. What's more, I explore only the payment times associated with purchases completed in "brick and mortar" stores, as opposed to online purchases. Not only are payment options different when shopping online, but the medium through which the purchases occur (physical versus online) may change consumers' perceptions towards payment times.

One of the major challenges of this paper resides with the data set. Ambiguous instructions provided to the observers on how to collect and round the duration data introduced some measurement error. Furthermore, the 2014 Transaction Duration Study focused only on retail transactions from a set of stores that are not representative of Canadian businesses. Thus, the non-random sample of transactions, coupled with the lack of data on the purchase of durable goods, could lead to an over-representation of low-value transactions, causing a bias in the estimated measure of efficiency of cash.

Finally, I fail to show that the previous payment duration instrument I propose is valid. However, previous payment duration remains a compelling exclusion restriction in the model for payment efficiency when dealing with non-randomized MOP usage. Unfortunately, within the context of this analysis, the inter-transaction wait time was not available. In the future, collecting data on total queuing time could provide an opportunity to test the validity of the

previous payment duration instrument on MOP choice.

5 Conclusion

Using data collected from retailers in Canada, I find that cash allows for more retail payments to be completed than cards on average, indicating that it is more efficient in terms of payment time. Furthermore, I show that the efficiency of cash is underestimated if consumer selection of MOP is not accounted for. To do this, I model and estimate a reduced-form probit for consumer selection of cash and, using the CF method, estimate a duration model for payment time. I provide two exclusion restrictions, namely, the value of the transaction and the time of the preceding payment.

The ability of cash to process a higher volume of payments than cards has implications for its usage by consumers and acceptance by merchants. This supports evidence that shows the speed of cash is used by retailers to induce consumers to switch away from high-cost MOPs such as credit cards (Welte, 2016).

Given its lower volume cost relative to cards for merchants, and that acceptance fuels usage of MOPs (Huynh et al., 2014), the demand for cash is likely to continue if the time efficiency of cash relative to current and new payment instruments is maintained. This is particularly true for low-value retail purchases, since the efficiency of cash is greatest for these transactions.

Finally, if new and more recent data could be acquired, it isn't clear that these results could be replicated. Innovations in MOPs have been expansive since 2014, with contactless cards becoming more widely implemented and new forms of payment technology emerging, such as payments completed with smartphones. Given the transaction amount limits imposed on contactless card payments by issuers, studying the time efficiency of these payment instruments compared with cash is particularly relevant. As these alternative payment options become more widely accepted by merchants, demand for cash could be affected (Huynh et al., 2014). However, with payment times constituting a non-trivial cost to consumers and retailers, tracking the evolution of payment durations in the face of new or improved MOPs can allow researchers and central banks to understand how and when cash continues to be the most competitive MOP in terms of time. More specifically, they can recognize how payment times impacts money demand.

Bibliography

Arango C, Hogg D, Lee A. 2015a. Why Is Cash (Still) so Entrenched? Insights from Canadian Shopping Diaries. *Contemporary Economic Policy* **33**: 141–158.

Arango C, Huynh KP, Sabetti L. 2015b. Consumer Payment Choice: Merchant Card Acceptance Versus Pricing Incentives. *Journal of Banking & Finance* **55**: 130–141.

Arango C, Welte A. 2012. The Bank of Canada's 2009 Methods-of-Payment Survey: Methodology and Key Results. Staff Discussion Paper 2012-6, Bank of Canada.

Bounie D, François A. 2006. Cash, Check or Bank Card? The Effects of Transaction Characteristics on the Use of Payment Instruments. Working Paper No. ESS-06-05, Telecom Paris Economics and Social Sciences.

Bounie D, François A, Van Hove L. 2016. Merchant Acceptance of Payment Cards: "Must Take" or "Wanna Take"? *Review of Network Economics* **15**: 117–146.

Breslow N. 1974. Covariance Analysis of Censored Survival Data. *Biometrics* **30**: 89–99.

Brits H, Winder C. 2005. Payments Are No Free Lunch. DNB Occasional Studies 302, Netherlands Central Bank, Research Department.

Chen H, Huynh K, Shy O, et al. 2017. Cash Versus Card: Payment Discontinuities and the Burden of Holding Coins. Staff Working Paper 2017-47, Bank of Canada.

Cox DR. 1975. Partial Likelihood. *Biometrika* **62**: 269–276.

Danmarks Nationalbank. 2012. Costs of Payments in Denmark. Danmarks Nationalbank Papers.
URL `http://www.nationalbanken.dk/en/publications/Documents/2012/04/betaling_engelsk_samlet_web.pdf`

Dong Y. 2010. Endogenous Regressor Binary Choice Models Without Instruments, With an Application to Migration. *Economics Letters* **107**: 33–35.

Dufour JM, Jasiak J. 2001. Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review* **42**: 815–844.

Escanciano JC, Jacho-Chávez D, Lewbel A. 2016. Identification and Estimation of Semiparametric Two-Step Models. *Quantitative Economics* **7**: 561–589.

European Commission. 2015. Survey on Merchants' Costs of Processing Cash and Card Payments, Final Results. European Commission, Directorate-General for Competition.
URL `http://ec.europa.eu/competition/sectors/financial_services/dgcomp_final_report_en.pdf`

Fujiki H, Tanaka M. 2017. Choice of Payment Instrument for Low-Value Transactions in Japan. International Cash Conference 2017 – War on Cash: Is there a Future for Cash? 162909, Deutsche Bundesbank.

Fung B, Huynh K, Kosse A, et al. 2017. Acceptance and Use of Payments at the Point of Sale in Canada. *Bank of Canada Review* : 14–26.

Fung B, Huynh KP, Nield K, Welte A. 2018. Merchant Acceptance of Cash and Credit Cards at the Point of Sale. Staff Analytical Note 2018-1, Bank of Canada.

Fung B, Huynh KP, Stuber G, et al. 2015. The Use of Cash in Canada. *Bank of Canada Review* : 45–56.

Garcia-Swartz DD, Hahn RW, Layne-Farrar A. 2006a. The Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics. *Review of Network Economics* **5**.

Garcia-Swartz DD, Hahn RW, Layne-Farrar A. 2006b. The Move Toward a Cashless Society: Calculating the Costs and Benefits. *Review of Network Economics* **5**.

Garen J. 1984. The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica* **52**: 1199–1218.

Hayashi F, Keeton WR. 2012. Measuring the Costs of Retail Payment Methods. Economic Review Q II, Federal Reserve Bank of Kansas City.

Henry CS, Huynh KP, Shen QR. 2015. 2013 Methods-of-Payment Survey Results. Staff Discussion Paper 2015-4, Bank of Canada.

Huynh K, Schmidt-Dengler P, Stix H. 2014. The Role of Card Acceptance in the Transaction Demand for Money. Staff Working Paper 2014-44, Bank of Canada.

Huynh KP, Petrunia RJ, Voia M. 2010. The Impact of Initial Financial State on Firm Duration Across Entry Cohorts. *The Journal of Industrial Economics* **58**: 661–689.

Jonker N. 2007. Payment Instruments as Perceived by Consumers - Results from a Household Survey. *De Economist* **155**: 271–303.

Jonker N. 2013. Social Costs of POS Payments in the Netherlands 2002-2012: Efficiency Gains from Increased Debit Card Usage. Technical report, Netherlands Central Bank, Research Department.

Klee E. 2006. Paper or Plastic? The Effect of Time on Check and Debit Card Use at Grocery Stores. FEDS Working Paper No. 2006-02, Board of Governors of the Federal Reserve System.

Klee E. 2008. How people pay: Evidence from Grocery Store Data. *Journal of Monetary Economics* **55**: 526–541.

Kosse A, Chen H, Felt MH, Jiongo VD, Nield K, Welte A. 2017. The Cost of Point-of-Sale Payments in Canada. Staff Discussion Paper No. 2017-4, Bank of Canada.

Norges Bank. 2014. Costs in the Norwegian Payment System. Norges Bank Papers. No. 5.

Polasik M, Górka J, Wilczewski G, Kunkowski J, Przenajkowska K, Tetkowska N. 2012. Time Efficiency of Point-of-Sale Payment Methods: Empirical Results for Cash, Cards and Mobile Payments. In *International Conference on Enterprise Information Systems*. Springer, 306–320.

Segendorf BL, Jansson T. 2012. The Cost of Consumer Payments in Sweden. Sveriges Riksbank Research Paper Series. No. 262.

Stewart C, Chan I, Ossolinski C, Halperin D, Ryan P. 2014. The Evolution of Payment Costs in Australia. Reserve Bank of Australia Research Discussion Papers. No. 2014-14.

Terza JV. 1998. Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects. *Journal of Econometrics* **84**: 129–154.

Tsiatis AA. 1981. A Large Sample Study of Cox's Regression Model. *The Annals of Statistics* **9**: 93–108.

Wakamori N, Welte A. 2017. Why do Shoppers Use Cash? Evidence from Shopping Diary Data. *Journal of Money, Credit and Banking* **49**: 115–169.

Welte A. 2016. Wait a Minute: The Efficacy of Discounting Versus Non-Pecuniary Payment Steering. *Journal of Financial Market Infrastructures* **4**: 17–25.

Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge JM. 2014. Quasi-Maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables. *Journal of Econometrics* **182**: 226–234.

Wrenn DH, Klaiber HA, Newburn DA. 2017. Confronting Price Endogeneity in a Duration Model of Residential Subdivision Development. *Journal of Applied Econometrics* **32**: 661–682.

6 Appendix

Table 1 – Number of Observations by Decimal and Unit Measurement of Payment Time

	Decimals	Unit
Grocery stores	271	1,288
Alcohol	1	429
Gas & convenience	4	802
General store	2	921
Coffee shops	2	772
Pharmacies	0	120
Hardware stores	238	775
Total	518	5,107

Note: Breakdown of observations that have payment times to the nearest tenth of a second (decimals) and those that are rounded to the nearest second. Decimal observations are excluded from the analysis rather than rounded, since this would introduce measurement error.

Table 2 – Descriptive Statistics

	Obs	Mean	Median	Std Deviation	Min	Max
Cash	2,407	15	11	14	1	151
Cards	2,700	24	23	14	1	180
Debit	1,582	25	24	14	1	180
Swipe & PIN	18	30	25	20	10	89
Chip & PIN	1,485	25	24	14	1	180
Contactless	79	24	16	20	4	101
Credit	1,118	23	21	14	3	150
Swipe & sign	81	19	13	19	3	150
Chip & PIN	919	24	22	13	7	143
Contactless	118	16	14	12	3	87
Total	5,107	20	18	15	1	180

Note: Descriptive statistics for the sample of payment times under study in the 2014 Transaction Duration Study data.

Table 3 – Variable Names and Description

Variable	Description
Variable of interest	
MOP	1 if cash is used to pay for transaction, 0 if card.
Instruments (Z)	
PPD	continuous variable for previous payment duration.
value	continuous variable for value of the transaction.
Controls (X)	
consumer_age	categorical variable (3) for estimated age of the consumer.
clerk_age	categorical variable (3) for estimated age of the clerk.
consumer_gender	1 if consumer is female.
clerk_gender	1 if clerk is female.
couple	1 if more than one consumer paid for the purchase.
express	1 if the transaction was completed in an express checkout lane.
totalnum_cashreg	number of cash register in the store of business.
business	categorical variable (7) for business type.
naics	three-digit North American Industry Classification System codes.
day	categorical variable (7) for the day of the week.
time	categorical variable (4) for time of day.
prov	1 if Quebec (0 if Ontario).
observer	categorical variable (12) for each observer assigned to record data.

Table 4 – Hazard Ratios of Cash Over Cards Payment for Point-of-Sale Retail Purchases

	(1) CPHM	(2) CPHM	(3) CPHM-CF I	(4) CPHM-CF I	(5) CPHM-CF II	(6) CPHM-CF II	(7) CPHM-CF III	(8) CPHM-CF III
MOP	2.459***	2.268***	15.537***	11.158***	14.814***	11.023***	20.914***	18.050
	(0.390)	(0.384)	(3.652)	(6.974)	(2.918)	(5.436)	(18.531)	(29.045)
Frailty (v)	0.103	0.241	0.131	0.260	0.132	0.262	0.115	0.241
	(0.058)	(0.126)	(0.072)	(0.135)	(0.072)	(0.136)	(0.064)	(0.126)
Generated regressor			0.318***	0.364***	0.327***	0.366***	0.115**	0.121
			(0.039)	(0.138)	(0.033)	(0.109)	(0.099)	(0.193)
Controls	✓		✓		✓		✓	
No. of IVs	None	None	2	2	4	4	2	2
First-stage			Probit	Probit	Probit	Probit	LPM	LPM
Observations	5,107	5,107	5,107	5,107	5,107	5,107	5,107	5,107
AIC	75,462.97	76,438.42	75,220.36	76,176.66	75,206.74	76,160.52	75,365.87	76,275.33
BIC	75,554.51	76,444.96	75,318.44	76,189.75	75,304.82	76,173.61	75,463.95	76,288.41
T-test			0.000	0.000	0.000	0.000	0.000	0.117

Note: This table compares the different Cox Proportional Hazard models (CPHMs) estimated using the 2014 Transaction Duration Study data. Columns (1) and (2) present the model that does not account for the consumer selection (CPHM). Columns (3) and (4) present the model accounting for consumer selection, using a probit for choosing cash with two instruments and using the control-function method in the Cox model (CPHM-CF I). Columns (5) and (6) present the model accounting for consumer selection, using a probit with four instruments (squared transformation) and using the control-function method in the Cox model (CPHM-CF II). Columns (7) and (8) present the model using a linear probability model (LPM) for choosing cash with two instruments and using the control-function method in the Cox model (CPHM-CF III). Instruments are transaction value (value), previous payment duration (PPD), squared transaction value (value^2), and squared previous payment duration (PPD^2). Controls are presented and listed in Table 3. Standard errors are bootstrapped with 1,000 replications and presented in parentheses. Significance is * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Results of the T-test of exogeneity are presented in the table, with null hypothesis that the coefficient on the generated regressor is equal to 0.

Table 5 – Marginal Effects of the Reduced Form for Consumer Probability of Choosing Cash for Point-of-Sale Retail Purchases

	(1) Probit 1	(2) Probit 1	(3) Probit 2	(4) Probit 2	(5) LPM	(6) LPM
Value	-0.003***	-0.003***	-0.003***	-0.004***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
PPD	-0.001	-0.001*	-0.001	-0.002	-0.000	-0.001*
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Value^2			0.000	0.000		
			(0.000)	(0.000)		
PPD^2			0.000	0.000		
			(0.000)	(0.000)		
Controls	✓		✓		✓	
Observations	5,107	5,107	5,107	5,107	5,107	5,107
Pseudo R^2	0.100	0.078	0.105	0.084	0.101	0.063
Wald test on:						
Value	0.000	0.000	0.001	0.000	0.000	0.000
PPD	0.261	0.045	0.220	0.043	0.284	0.016
Value^2			0.713	0.685		
PPD^2			0.382	0.202		

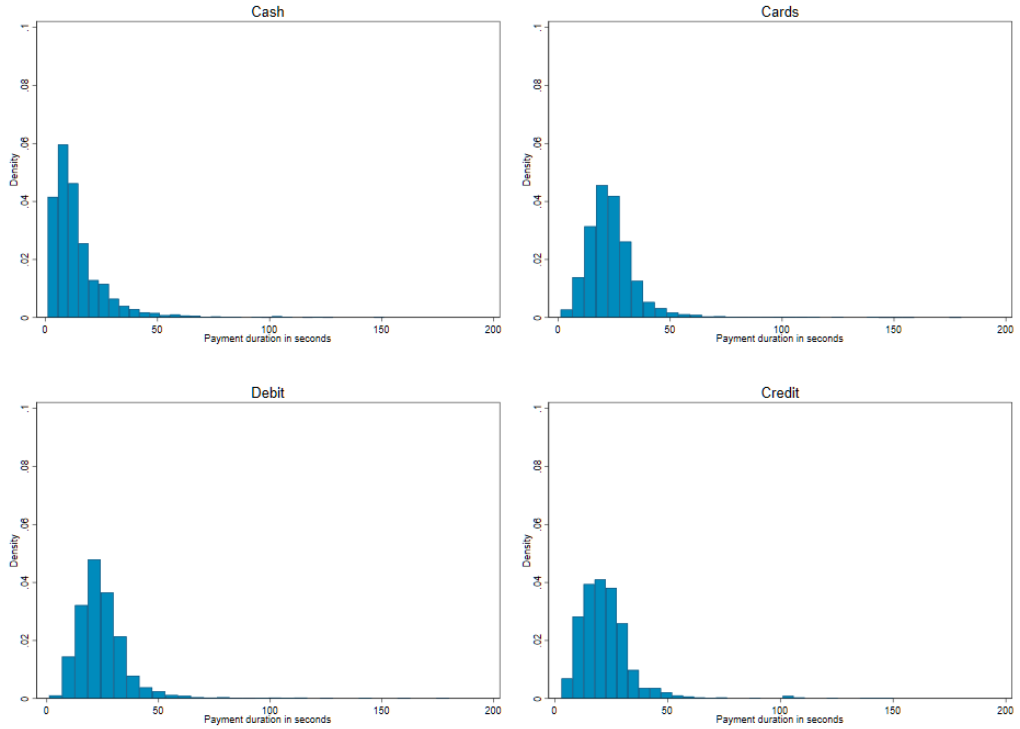
Note: This table compares the results of the models for consumer choice of cash using the 2014 Transaction Duration Study data. Columns (1) and (2) present the result of the probit with two instruments (Probit 1). Columns (3) and (4) present the results of the probit with four instruments, adding the squared transformation of the original two instruments (Probit 2). Columns (5) and (6) present the results of the linear probability model for consumer choice of cash with two instruments (LPM). Instruments are transaction value (value), previous payment duration (PPD), squared transaction value (value^2), and squared previous payment duration (PPD^2). Controls are presented and listed in Table 3. Standard errors are bootstrapped with 1,000 replications and presented in parentheses. Significance is * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The results of the individual test of validity of instruments are completed with a Wald test and presented in the table, with the null hypothesis that the coefficient on the instrumental variable is equal 0.

Table 6 – Test of Proportionality Assumption

	Critical value	P-value
CPHM	35,984.85	0.000
MOP	27.28	0.000
CPHM-CF I	33,000.06	0.000
MOP	56.08	0.000
CPHM-CF II	34,854.09	0.000
MOP	71.70	0.000
CPHM-CF III	34,227.50	0.000
MOP	9.87	0.002

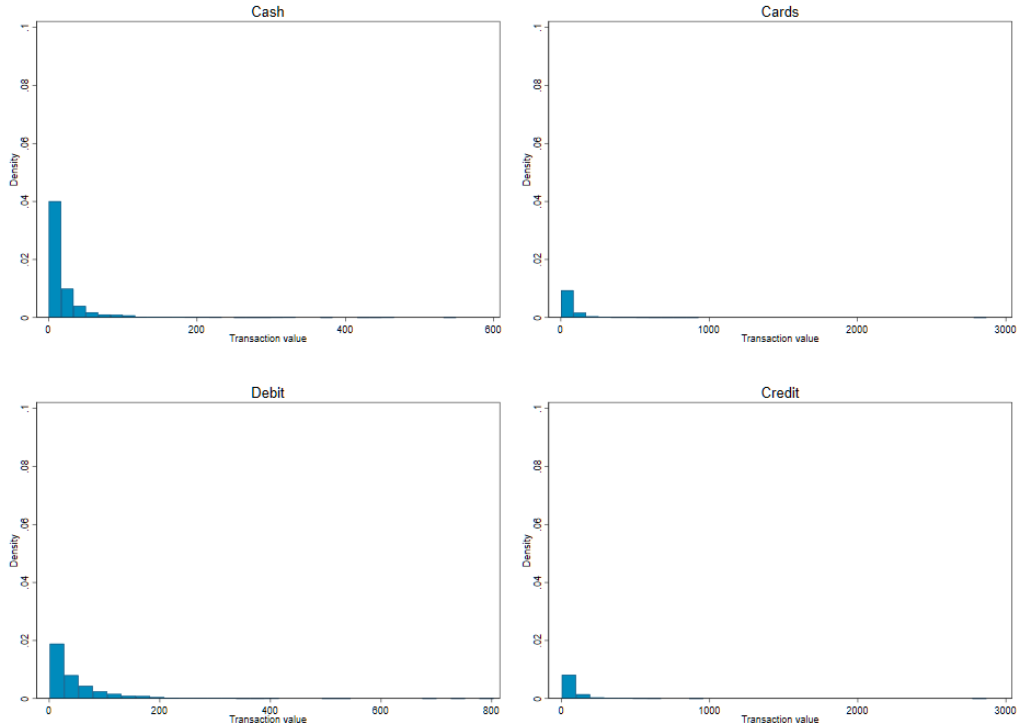
Note: The null hypothesis of the test tests whether proportionality holds using the Schoenfeld residuals. The table presents the results of the global test and the test on the MOP variable for each of the four models. All reject the null indicating proportionality is violated.

Figure 1 – Distribution of Payment Times by Method of Payment



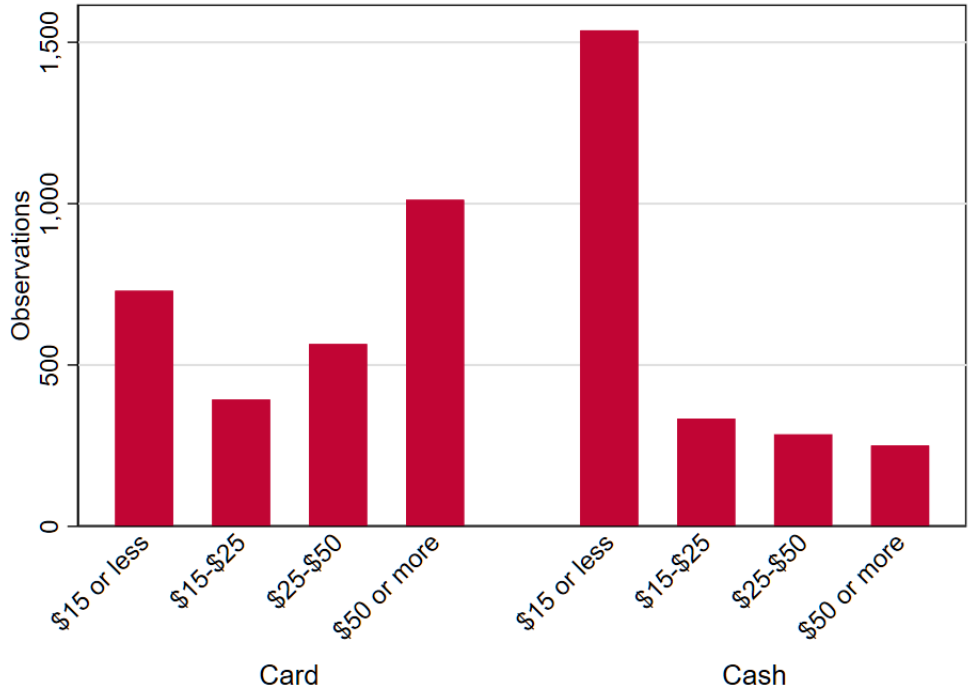
Note: These graphs present the distribution of payment times for the transactions studied from the 2014 Transaction Duration Study data for cash, cards (debit and credit combined), debit, and credit.

Figure 2 – Distribution of Transaction Values by Method of Payment



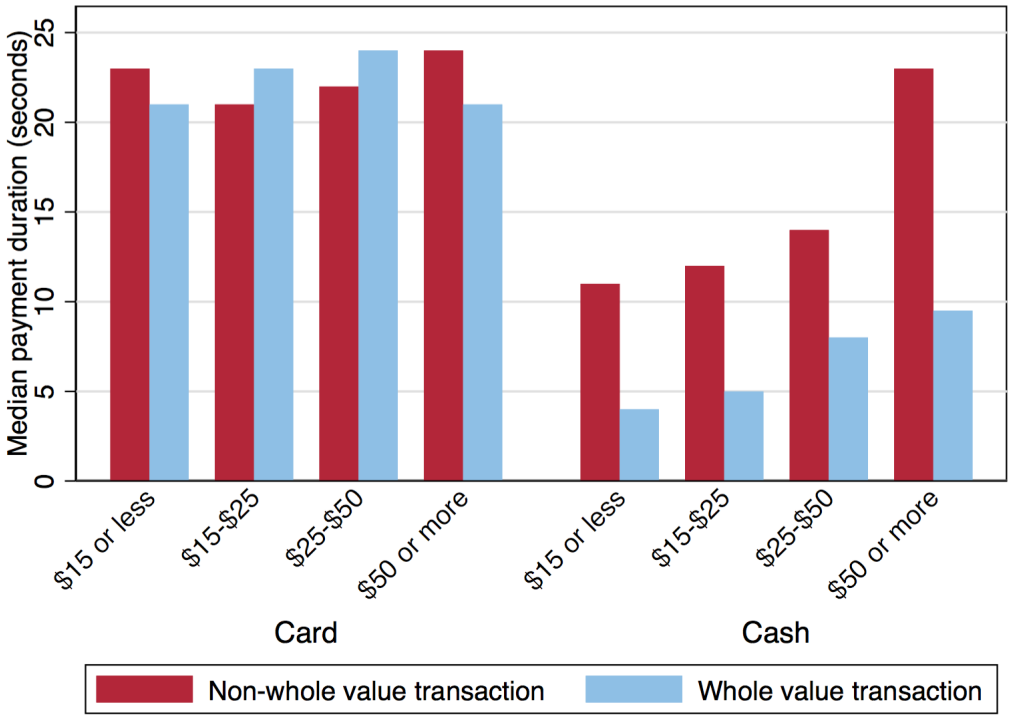
Note: These graphs present the distribution of transaction values for the transactions studied from the 2014 Transaction Duration Study data for cash, cards (debit and credit combined), debit, and credit.

Figure 3 – Number of Payments by Transaction Value



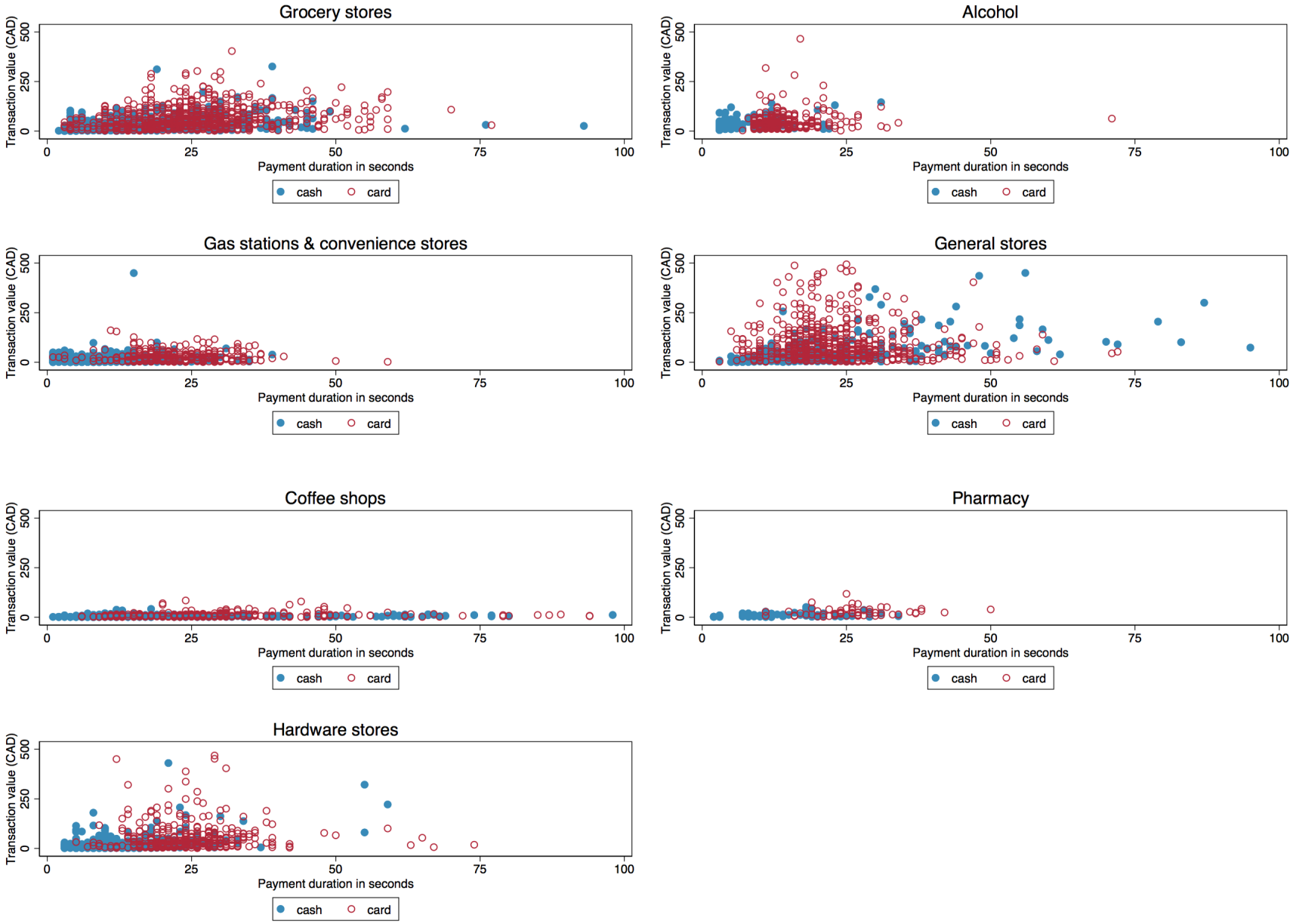
Note: Number of transactions completed using cash or cards by transaction value for the 2014 Transaction Duration Study data.

Figure 4 – Median Payment Times by Transaction Value



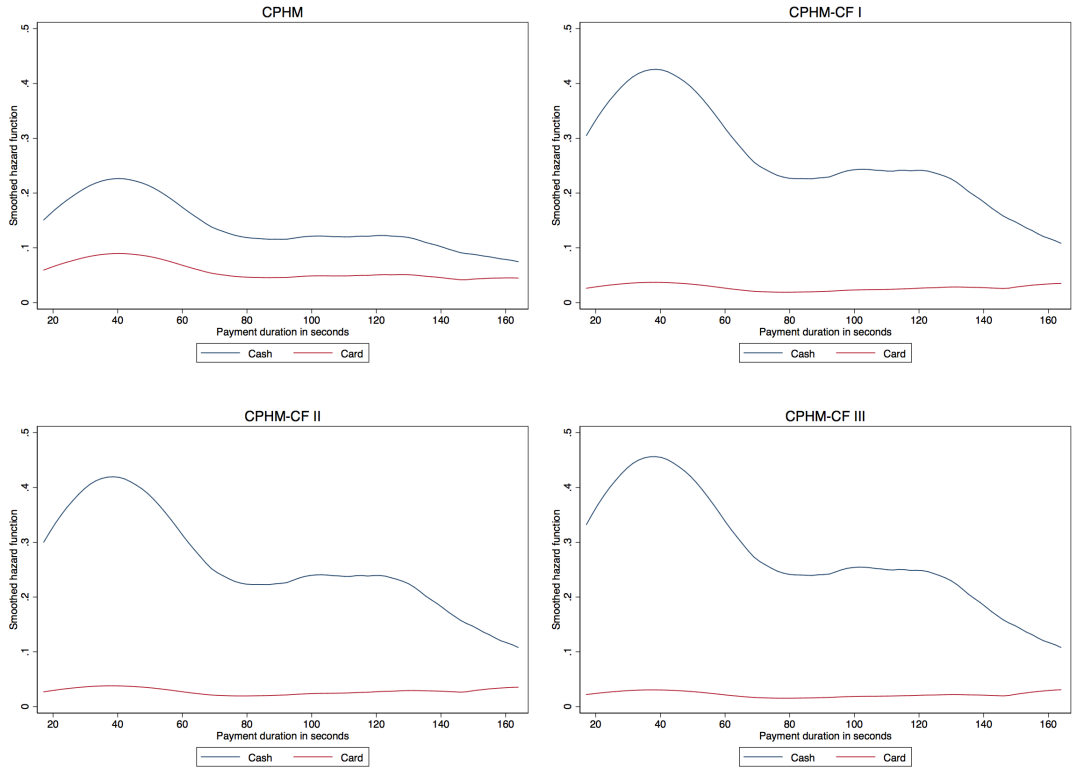
Note: Median payment times for cash and card payments by transaction value bins using the 2014 Transaction Study data. Whole transaction values represent values that are rounded to the nearest dollar, whereas non-whole transaction values will include cents.

Figure 5 – Payment Times Grouped by Business Type



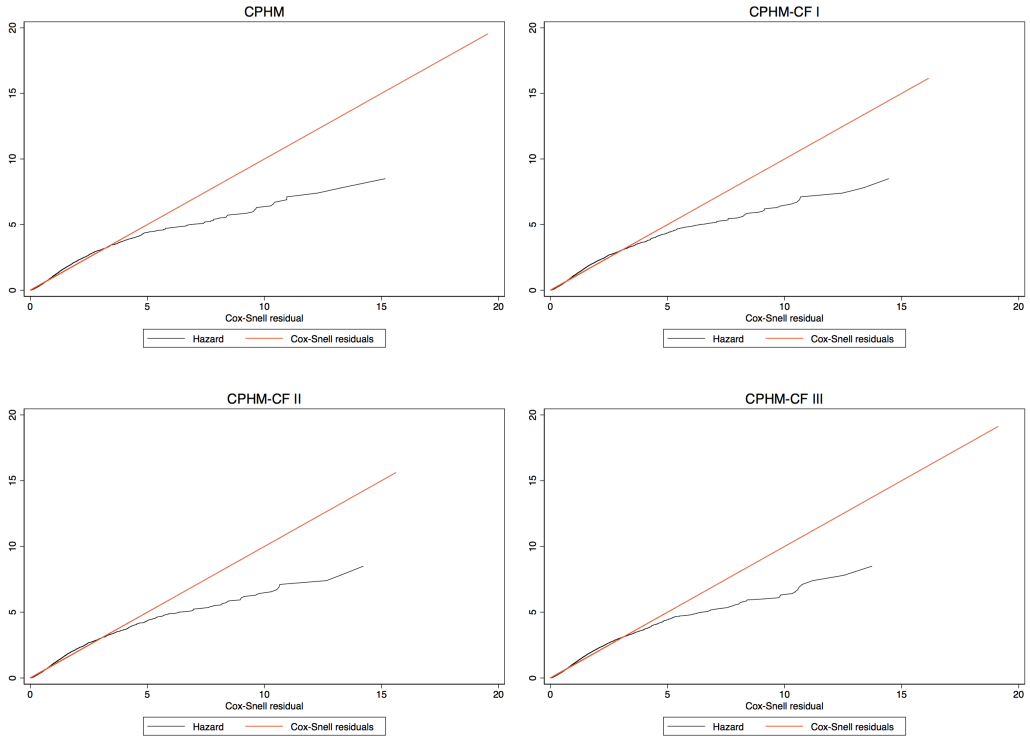
Note: These plots illustrate the various patterns of payment times for each of the seven types of businesses included in the 2014 Transaction Duration Study data. Presented are a subset of payment times for cash-based and card-based transactions that were completed in less than 100 seconds and were no more than $500 in value.

Figure 6 – Smoothed Hazard Functions



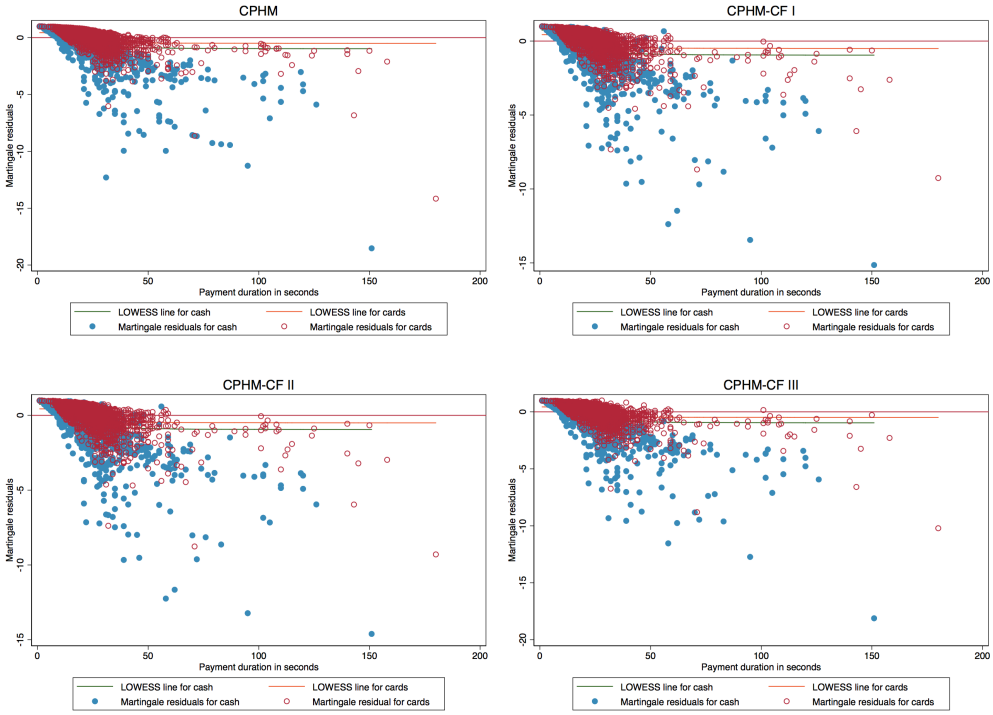
Note: These graphs present the smoothed hazard functions and compare the CPHM with the three control-function models. All plots are evaluated for a frailty v_i equal to 1.

Figure 7 – Cox-Snell Residuals



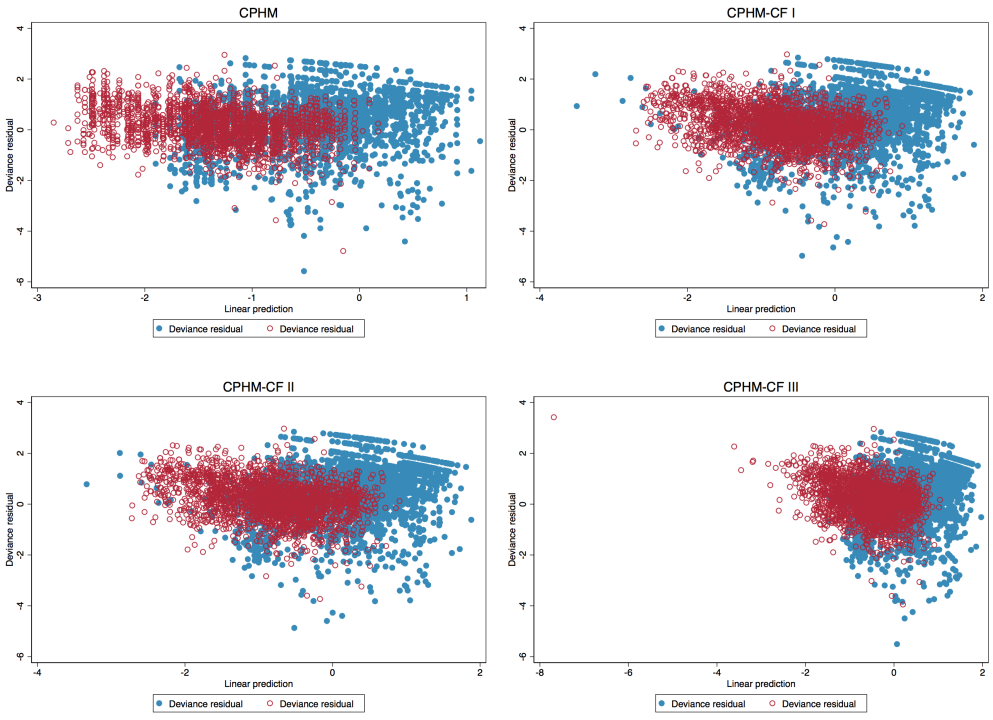
Note: These graphs present a comparison of the Cox-Snell residuals for all models estimated. These residuals are used to assess fit of the model by comparing how closely the hazard line follows the 45° line.

Figure 8 – Martingale Residuals



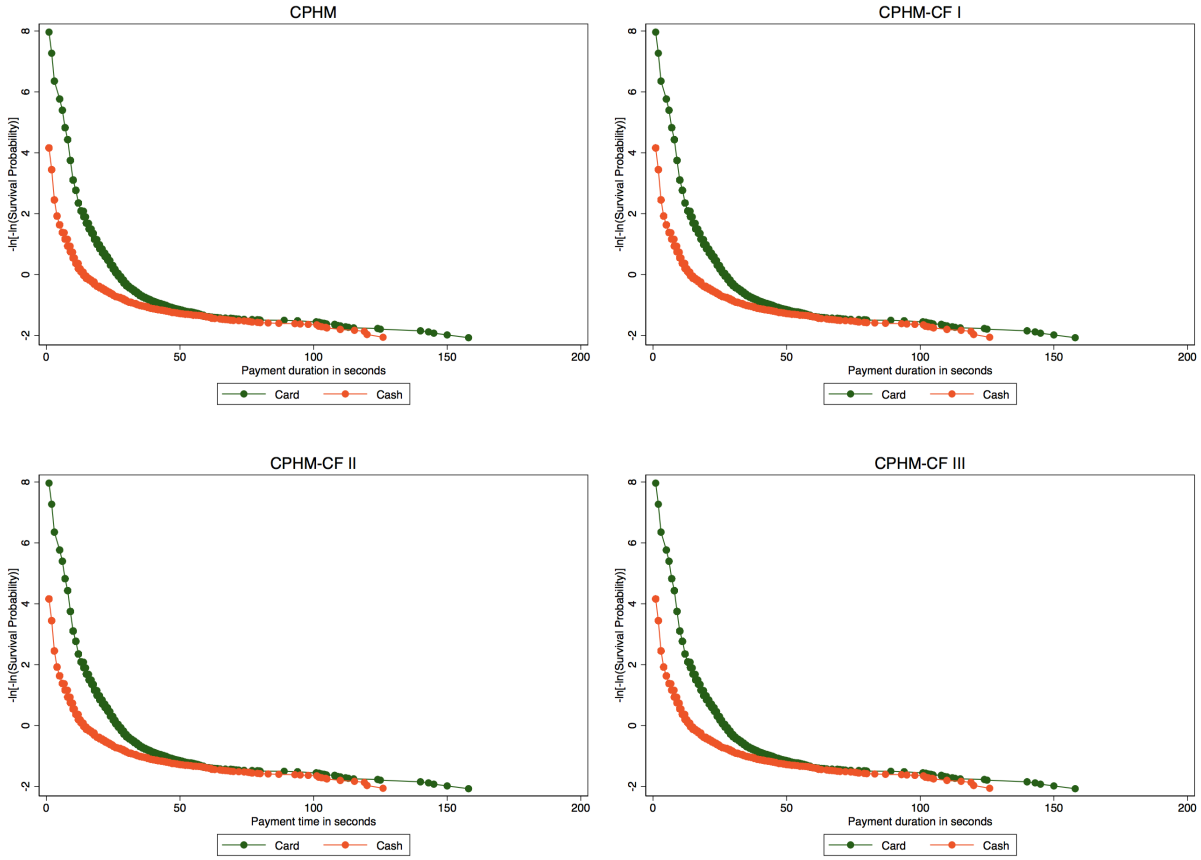
Note: These graphs present the scatter plot of the martingale residuals with the running-mean smoothing LOWESS (locally weighted scatterplot smoothing) line for both cash and card payments. Martingale residuals are used to determine the functional forms of the covariates by assessing how close the LOWESS line is to 0.

Figure 9 – Deviance Residuals



Note: These graphs present a comparison of the deviance residuals for all models estimated. These residuals are a transformation of the martingale residuals such that the residuals are centered on 0. These are often used to identify potential outliers.

Figure 10 – Proportionality of the MOP Variable



Note: The proportionality assumption of the MOP variable is tested using the Schoenfeld residuals. The null hypothesis of the test tests whether the groups are proportional. Referring back to Table 6, the assumption of proportionality is violated for the MOP variable. This can be visually observed from the above figure, since the survival probabilities of cash and cards converge as payment time increases.