THE UNIVERSITY OF SYDNEY

Modelling the Four-Party Billing Payment Scheme: The Case of BPAY

Abraham Akra

A thesis submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy

Business School
University of Sydney
2013

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

____

Abraham Akra

Contents

List of Figures

List of Tables

Chapter 1

Introduction

BPAY is an innovative bill payment platform in Australia established in 1997 and accounts for over 30% of the market, processing over 250,000,000 transactions in 2009. This dissertation involves developing a detailed understanding of the dynamics of a two-sided four-party Bill Payment market utilising proprietary data obtained from BPAY. The operating structure of BPAY is similar to that of other payment providers. BPAY is a four-party payment provider and involves four parties. Each transaction involves a consumer, merchant and their respective banking institutions. This differs from a three-party payment platform, such as Diners Club which includes a merchant, consumer and payment platform. BPAY facilitates the transaction by administrating the information transfer between the respective banking institutions. The other main payment providers in Australia include Visa, American Express and Australia Post. However, Australia Post is not included in this thesis due to data limitations.

The attention that the market for payments has received from the Reserve Bank of Australia (RBA) has brought about a need for greater understanding of the dynamics of the industry. Interchange fees have been the focus of Central Banks and antitrust lawsuits throughout the globe. The interchange fee is the payment made in a four-party payment platform from the merchant's bank to the consumer's bank per transaction. The interchange fee is the key component in formulating the benefits or fees provided to end-users. To date, data availability has been the main restriction in investigating the determinants of demand for bill payments. The data to be employed in this thesis will reveal insights into the key drivers of transactions.

In the past several years in the United States of America, merchants and trade associations filed approximately fifty civil lawsuits against Visa, MasterCard, and several card-issuing banks alleging, among other charges, that interchange fees are too high and that the collective setting of interchange fees by members of the payment card associations constitutes illegal price fixing under antitrust laws. In Australia, the RBA introduced a number of reforms to credit and debit card arrangements that sought to moderate, in its opinion, the excessive use of credit and debit cards. This introduction was aimed at increasing the efficiency of the payments system.

A joint investigation by the RBA and Australian Competition and Consumer Commission (ACCC) in 2000 recommended several changes to the operation of the payments market to increase innovation, competition and social welfare. At the time of the investigation, the cost of a \$100 transaction through EFTPOS was estimated to be approximately \$0.50, once a given amount of fee-free transactions using EFTOS was reached. The effective price for a credit card transaction with a rewards program and interest free period, based on the same transaction value, was estimated to be -\$1.30 by Simon et al. (2009). Hence, consumers therefore receive a rebate equal to \$1.30 on purchases.

Chapter 2 reviews the academic literature relevant to the Bill Payements market. Chapter 3 models the demand for merchant acceptance and consumer usage of a four-party payment scheme in the Bill Payment market. The framework of the model proposed by Rysman (2004) for the Yellow Pages market can be applied to the Bill Payments market as it is an incompatible product market and due to the presence of network effects. For example, the Bill Payments market is an incompatible market as an additional merchant subscribed to the BPAY platform confers no benefit to consumers of an alternative payment platform. Even though Yellow Pages operates as a three-party services provider, from their perspective of the end-users, the demand equations of the end-users will be identical to that of a four-party provider as the consumer and merchant only deal with one other party. Hence, the price and cross price elasticities of the three-party and four-party payment providers are expected to be the same. Within a cointegrating framework, demand equations are estimated using vector error correction models using proprietary data between March 2003 and December 2010. The sampling period is of interest as it is contains a period of rapid change in the market of payments with intervention by the RBA to reduce the high merchant fees of credit card providers.

Results from Chapter 3 illustrate the importance of network effects in determining consumer usage and merchant demand. Additionally, price elasticities suggest the market for payments in Australia is competitive with consumers being price sensitive to other platforms in the Bill Payment market. This is reflected by the magnitude of the cross-price elasticities of Visa and Diners Club. Based on the demand equations estimated of consumer usage and merchant acceptance, it can be deduced that consumers are more vital to the growth of the bill payment platform than merchants. Chapters 4 and 5 seek to develop a deeper understanding of the motivations of consumers to transact with BPAY, with transaction related data and demographics used as explanatory variables. Credit card holding is the main covariate of interest in this thesis. Chapter 4 models the factors that contribute to new consumers on the BPAY platform to cease usage. Whereas, Chapter 5 investigates the influence that credit card membership has on consumer usage.

Empirical studies provide a micro view of the key drivers of transactions. The determinants of a transaction commonly modelled include demographics, the attributes of the payment instrument, and the characteristics of the transaction. Chapter 4 exploits a unique data set that details the demographics and transactions of individuals over a 30 month observational window. Such a data set allows individuals behaviour on the payment platform to be observed over time. Survival analysis techniques are employed to quantify the risks of individuals leaving the platform. Cox (1972) models are estimated in the fixed and mixed framework with the transaction covariate varying in both instances to determine the most appropriate model. Results suggest support for the Hayashi and Klee (2003) finding in the Bill Payment market with individuals having a credit card 10% and 12% less likely to leave the BPAY platform at any point in time after employing a fixed and mixed Cox (1972) model. Also of note is that males are approximately 3% more likely to leave the BPAY platform at any point in time, while there is a geographical influence on whether an individual adopts an innovative bill payments platform, or not.

The motivation of Chapter 5 is to establish whether a link existed between the usage of the BPAY platform by consumers with the adoption of prior payment method technologies, given by credit card holding. Unlike Hayashi and Klee (2003) there is an added layer of complexity as credit cards are another payment instrument individuals can use for bill payments. Hence the impact of credit card holding on the adoption of a new technology in the Bill Payment market is unknown and answered empirically in Chapter 5. Segmenting individuals into usage frequency categories, an ordered generalized logit model is estimated to determine the influence of credit card holding on frequency of usage. The results lend support to the Hayashi and Klee (2003) hypothesis that the adoption of a technology based payment instrument is influenced by the usage of prior technologies. Other findings include a positive relationship between the age of a consumer and being in a high category of usage and gender, with females more likely to be in the higher categories of usage than males.

Chapter 2

Literature Review

The primary factor influencing the effective price of a four-party payment instrument is the interchange fee. By 2007, numerous interventions in the payments market have been made by the RBA. Some of these included the removal of the no-surcharge rule imposed on merchants, a cost-based approach to regulating interchange fees, and the removal of the strict "honour all cards" rule applied by platforms to merchants. Dawson and Hugener (2006) note that rewards account for 44% of the level of interchange fees and in 2006 approximately $\$USD$ 30 billion was collected by Visa and MasterCard card issuing banks from interchange fees. The remaining sources of funding that card issuers use include annual fees, penalty fees and interest paid by card holders. These changes have since brought about several interventions in the payments market by the RBA. This had the effect of decreasing the rewards attached to cards and increasing the average amount required to be spent on a credit card in order to receive reward points.

The average amount required to be spent in order to receive enough points to merit a reward of a \$100 shopping voucher increased from \$12,400 in 2003 to \$16,700 in 2009.

The primary pricing decision faced by four-party payment providers is in setting the level of interchange fees. Interchange fees directly impacts the costs and benefits faced by consumers and merchants alike, upon using a payment instrument. Interchange fees are paid by the merchant's bank (acquirer) to the consumer's bank (issuer) to locate the account holder's name and to transfer funds over the platform (within the network) to the merchant's nominated bank account. Interchange fees are the main component of fees charged to merchants by their banks for each transaction processed and are usually used by the issuing bank to promote the use of a platform to consumers. The benefits offered to consumers include loyalty reward points and having access to credit for an interest free period.

The literature exploring the market for payments is largely theoretical. It examines the two-sided nature of the payments market which focuses on the interchange fee set by the platform. A two-sided market is characterised as a platform providing goods and services to two distinct end-users with prices set for each type of end-user. Examples of a two-sided market include videogame platforms that match game developers and consumers, dating agencies that match partners, and shopping malls that house a variety of stores. Of central importance to two-sided markets are usage and membership (network) externalities. Usage externalities arise because each party in a given transaction evaluates their own costs and benefits associated with a particular payment method, but do not consider the costs and benefits of the other end-user. Network externality refers to the fact that the value of a particular platform to an end-user increases as the number of end-users on the other side of a transaction increases. For example, in relation to the payments market, the more merchants offer a particular payment instrument, the higher the value that consumers place on being part of that platform and induced to join the platform. With more consumers joining the platform, the higher the value merchants place in offering the payment method. This results in an increase in the demand of merchants to join the platform.

2.1 Theoretical Background

Until recently, the set of assumptions assigned to models developed in the literature were limited in their applicability and ability to describe the market for payments. The set of unrealistic assumptions relate to the fixed per-transaction fees imposed on merchants or consumers, and the extent to which consumers and merchants have access to multiple payment instruments. Other assumptions include homogeneity in the benefits received by consumers and merchants from transacting on a given platform; the identical costs faced by platforms in providing payment services; and identical services to consumers and merchants offered by competing payment platforms.

Guthrie and Wright (2007) model the interaction between merchants in a Hotelling world within a game theory framework. The authors analyse the impact of the level of interchange fees set and the structure of fees charged to consumers and merchants through competition between competing four-party card associations. The assumptions of the model developed include no annual fee for consumers, no fixed fee for merchants in joining a platform, as well as both sides having the option of joining both platforms.

The appealing aspect of the model considered by Guthrie and Wright (2007) is the incorporation of the utility of consumers into the benefits that merchants gain by having their preferred payment option. Chapter 3 incorporates this characteristic into the demand equations of consumers and merchants by having the demand of consumers in the merchant demand equation and merchant demand for the platform in the consumer demand equation. This feature is apparent in retail firms where merchants rarely refuse a payment instrument preferred by their consumers due to the benefits derived from a transaction. Additionally, in a competitive market, refusing a payment option will put the firm at a competitive disadvantage and send negative signals to consumers regarding the merchant's quality. The findings suggest that, in a single payment scheme, sellers consider the joint payoff in determining whether to accept the card. Merchants will accept the card if joint utility is greater than zero. Consequently, the resulting interchange fee is the upper bound of the range of possible interchange fees and, assuming the card association will maximise volume, this will result in the over-usage of cards. With two competing card schemes, the equilibrium interchange fee is dependent upon the initial behaviour of consumers and merchants. Assuming consumers always hold both cards whenever it is an equilibrium for them to do so, competing card schemes set interchange fees at the socially optimum level. If merchants always accept both cards whenever it is equilibrium for them to do so, competing card schemes set interchange fees at the same level as in the single card scheme. Thus, unlike other markets, competition does not drive down prices. This is a unique characteristic of two-sided markets.

The "chicken and egg" equilibrium outcome is not isolated to Guthrie and Wright (2007). A series of non-unique equilibria is presented in Gardner and Stone (2009) based on the presumptions of consumers and merchants. One such example includes merchants subscribing to both payment platforms (multi-homing) based on the assumption that consumers will join at most one platform. This would provide an incentive to consumers to subscribe to their preferred platform and thereby, validating the merchant's assumption. The majority of theoretical papers modelling the effects of interchange fees on equilibrium outcomes describe a series of equilibria based on assumptions concerning the initial behaviour of merchants or consumers.

In a variate of Guthrie and Wright (2007), Gardner and Stone (2009) model competition between three-party payment schemes. A three-party payments platform consists of a consumer, merchant and platform. All fees are directly payable to the platform in a three-party payment scheme. The main difference between the three-party and four-party payment model relate to the recruiting of consumers and merchants to join the platform and the extension of credit. Both these functions in a four-party payment scheme are assumed by the respective banks of the consumer and merchant. Whereas, in a three-party scheme, the platform takes responsibility for these functions. Examples of three-party payment platforms include American Express and Diners Club.

The model developed by Gardner and Stone (2009) seeks to replicate the conditions under which consumers and merchants decide to use a payment platform in a competitive market. The key feature of the model developed by Gardner and Stone (2009) relates to the average cost per transaction of the consumer being inversely related to the quantity of transactions placed on the respective platform. The assumption that consumers pay a fixed cost to join a platform and pay no fees per transaction, allows this parameter to vary between consumers. This is in contrast to Guthrie and Wright (2007), where consumers pay no joining fee but are charged per-transaction. The second crucial assumption lacking in prior literature is common in both Guthrie and Wright (2007) and Gardner and Stone (2009). This assumption relates to the ability of consumers and merchants to subscribe to both platforms. Other assumptions include per-transactions fees for merchants, and with platforms facing different fixed and variable costs per-transaction. Fixed costs relate to the costs in signing up the consumer. Any non-uniqueness in equilibrium outcomes are steered via temporary incentives to consumers and merchants to become Pareto efficient outcomes. An example of a non-unique equilibrium is where no consumers join any platform. In this case, it is optimal for merchants to reject any proposal by platform providers to join also. Through temporary incentives to both sides of the market (preferably the value adding side), the development of network externalities will spur greater pick-up of platform registration. Thereby, removing the need for any incentives promoted by the platform once the market reaches maturity.

Gardner and Stone (2009) numerically simulate their model to determine the pricing strategies of platforms for different assumptions relating to the costs involved in operating the platform. The platform costs involved in this simplified model are the cost to process a transaction and the fixed cost associated with registering a consumer or merchant. The results suggest merchants are trying to steer consumers to their preferred payment method by only accepting one platform to carry out transactions. The allocation of the total cost directed to the two-sides of the market has long been argued in the literature to favour the side that joins one platform (single-home). The intuition behind this theory suggests platforms price competitively to the single-homing side to attract them to the platform. In doing so, increasing the network externalities as the scale of the network grows. This result, that the single-homing side attracts a lower cost allocation in relation to the multi-homing side, is based on models that were developed with the assumption that one side cannot multi-home and costs do not vary with transactions on the platform. The simulation results provide evidence to the contrary. The multi-homing side pays less as a proportion of total fees charged than the side that is more likely to single-home. Gardner and Stone (2009) attribute this result to average transaction costs varying between consumers.

A platform's incentive to allocate prices when revenue is fixed per consumer is to price more aggressively to consumers that are not currently subscribed to the platform as the fees extracted as a proportion of the consumer's transactions on the platform will be higher. Whereas, if fees were charged on a transaction basis, then the platform seeks to maximise volume. Hence, the platform will try and induce the single-homing side to join and, since the platform has monopoly access to the single-homing side, are able to extract higher fees from the multi-homing side. This is verified in the model once it is adapted to charge consumer's per-transaction costs.

Gardner and Stone (2009) also alter their parameters to consider the impact of payment instrument choice on the price allocation of total fees. The evidence supports the hypothesis that platforms price less favourably to the side initiating the transaction, commonly the consumer. The favourable pricing to the merchants (the side that doesn't have the final choice at the check-out), may be done by platforms to prevent merchants from steering consumers to their preferred payment option. The bias in price allocation between consumers and merchants decreases as the costs per-transanction incurred by the platform rises.

Theoretical papers provide a regulatory perspective on the pricing motivations of platforms in the market for payments. Gardner and Stone (2009) and Guthrie and Wright (2007) provide insights to the pricing motivations of platforms and different equilibrium outcomes that can eventuate from the preferences of consumers and merchants. Although several hypothesis' can be developed, such as the impact of pricing decisions of platforms on social welfare, empirically testing models developed in the literature and extracting the demand elasticities of consumers and merchants is a difficult task since models are developed with a high level view of the payments market. The different assumptions and platform structures also present challenges in comparing results in the literature.

Rochet and Wright (2010) improve upon the model of Guthrie and Wright (2007) by considering two classifications of credit card consumers and the resultant consumer surplus that eventuates from a monopoly four-party credit card platform setting interchange fees. Credit cards allow consumers to make purchases in the present that they would otherwise delay or avoid, due to the mismatch of cash inflows and outflows. The additional sales that the liquidity functionality of credit cards offer is a key reason merchants accept it as a payment method as a comparison with cheaper forms of payment platforms such as direct debit or cash. Theoretical studies provide the context for Chapter 3. It adds to the literature by examining the demand of platform users, inclusive of the influence of pricing on the demand of consumers and merchants.

2.2 Empirical Background

Humphrey (2010) provides details of the cost to society of transactions from different platforms and shows debit cards are more cost effective than credit cards in processing a transaction. Hayashi and Keeton (2012) notes that the cost to society as a proportion of GDP of different payment platforms varies by country, therefore, the benefits to moving to a electronic payments system are not uniform across different regions. To maximise volume and hence profitability, credit card platforms encourage the over usage of cards via reward programs. In the event that no incremental purchases are made by consumers that need the credit facility that a credit card provides, the welfare of merchants and the efficiency of the payments system is lowered as it is more expensive to process a credit card transaction than their debit card counterpart. The alternative option available to merchants that can remedy the consumers liquidity problem is the provision of store credit. However, this is relatively inefficient in comparison with credit cards as it is costly for every merchant to evaluate the credit risk of a consumer and consumers that don't need store credit receive no benefit.

Rochet and Wright (2010) model consumers having the option to use cash (debit cards), store credit or credit cards at the check-out. Merchants make the decision to accept credit cards if it is more cost effective in providing credit relative to their own capacity. The acquiring banks of the platform are assumed to be perfectly competitive so that merchant discount is the sum of the acquiring cost of processing a transaction and the interchange fee. Whereas, issuing banks are modelled to not face perfect competition so that the fee to consumers is equal to net issuer cost (transaction fee minus interchange fee) plus a profit margin. The motivation for modelling issuing banks as imperfectly competitive is to provide an incentive in the model to maximise joint profit amongst issuing banks. Rochet and Wright (2010) suggest in the past interchange fees were set by member banks to maximise profit between all banks and to do so, volume needs to be maximised. Much like Guthrie and Wright (2007), Rochet and Wright (2010) model merchants competing in a Hotelling framework with merchants operating under a no-surcharge rule. Equilibrium interchange fee arising from the model derived by Rochet and Wright (2010) suggests an unregulated monopoly credit card platform will set the maximum possible interchange fee too high to maximise consumer surplus and regulatory intervention by authorities is beneficial to moving to a more efficient outcome. The extent of the cap that Rochet and Wright (2010) suggest is dependent upon the cost savings merchants gain by accepting credit cards as opposed to providing store credit and the interchange fee with benefits removed, or a cost-based approach.

The finding relating to equilibrium interchange fees being greater than the social optimal level in an unregulated market is not isolated to Rochet and Wright (2010) and Guthrie and Wright (2007). In an alternative market environment of a monopolistic card network and a perfectly competitive environment for acquirers and issuers, Wang (2010) models the demand of end-users to join the platform and the resultant equilibrium interchange fee. Wang (2010) models card networks seeking to maximise the value of card transactions placed through their network. This is in contrast to Rochet and Wright (2010) and Guthrie and Wright (2007) who aim to maximise volume for the card platforms objective function, with free entry and exit of issuing banks, an oligopolistic card market and consumers have an elastic demand function for the card platform. The results of the model applied by Wang (2010) imply interchange fees depend upon the cost advantage of cards relative to alternative payment instruments and consumer demand elasticity. Furthermore, cost efficiencies on the card platform eventuating from economies of scale or advancement of technology, may not translate to a reduction in interchange fees.

Wang (2010) suggests issuers may increase rewards to consumers to further stimulate demand and increase the values of transactions placed on the network. Additionally, the card platform may increase interchange fees to extract the efficiency gains in the market for payments. However, social welfare does not improve as measured by consumer and merchant surplus. In addition, the outcome relating to equilibrium interchange fees being greater than the social optimal level is robust to relaxing the no-surcharge rule. Rochet and Tirole (2003) deduce similar results to Wang (2010) by considering the existence of network effects of the card platform. However, their findings suggest that interchange fees become irrelevant in the event that the no-surcharge rule is removed.

Empirical studies provide a micro view of the key drivers of transactions. The determinants of a transaction commonly modelled include demographics, the attributes of the payment instrument, and the characteristics of the transaction. Empirical studies in the payments market to date have been limited. Restrictive access to propriety data has confined studies to collecting data from surveys and investigating the factors that influence decisions made at checkouts by consumers. In overcoming the restrictive data problem, Rysman (2007) employed a rich data set consisting of two parts. The first part consisted of a panel of households that held multiple payment cards from $1994 - 2004$, while the second part contained the number and amount of transactions by month of every merchant in the Visa network. Rysman (2007) finds consumers maintain cards on multiple networks but tend to use only one network. This result is found to be robust across consumer characteristics such as income, education and spending. This suggests that while consumers have a preference for using a single platform, they recognise that some purchases are valuable enough to warrant using a less preferred network. This type of scenario was reflected in a model developed by Rochet and Tirole (2003) in which consumers have heterogeneous preference for networks. They define a "preference for using a sole payment method" that reflects the level of lost business a merchant faces on abandonment of a network. Additionally, individual consumers may change platforms or networks from transaction to transaction, based on the characteristics of the product and the availability of substitutes. This suggests merchants must accept multiple payment methods, or risk a decrease in sales.

The influence of technology and the adoption of electronic methods of transactions by merchants and consumers on check usage is examined by Schuh and Stavins (2010). The 1995 and 2004 Survey of consumer payment choice in the United States of America notes that the adoption of debit cards, direct debit and internet banking has increased significantly from 18%, 22% and 3% in 1995 to 59%, 47% and 32% in 2004, respectively. On the other hand, the 1995 and 2006 Survey of Consumer Payment Choice shows there has been a decline in the use of checks as a proportion of all noncash payments from 77% in 1995 to 36% in 2006.

Schuh and Stavins (2010) model consumer payment adoption and usage simultaneously using the Heckman (1976) selection model to account for the possible selection bias in payment platform usage. The model includes demographic related covariates, payment instrument characteristics and attitudinal data of individuals. Data is obtained from the 2006 Survey of Consumer Payment choice that includes the adoption and usage of several payment instruments and demographic covariates. The results indicate that 25% of the reduction in check usage as a proportion of total noncash transactions between 2003 and 2006 can be attributed to the number of adopted payment platforms by consumers. Relative to alternative payment methods, the convenience of cheques and the cost transacting with cheques represent 34% and 11% of the total cross-sectional variation in cheque usage. Thus, consumers optimally choose their payment method over time as the relative characteristics of each payment instrument and costs of transacting change. However, in all four models the R-squared value was less than 40%, indicating the model leaves unexplained much of the variation in payment choice at the check-out.

The removal of the no surcharge rule by the RBA locally, as well as countries abroad, allows merchants to communicate to consumers the true cost of different payment instruments and steer consumers to their preferred choice. Rochet and Tirole (2003) show that removing the no surcharge rule enables merchants to completely transfer the cost of a payment instrument to a consumer. Thereby, the allocation of costs between consumers and merchants becomes an irrelevant decision for the platform enabling the transaction. The motivation of central banks to remove the no surcharge rule is to improve the composition of payments between cash, debit cards and credit cards to generate cost savings that eventuate from having an efficient payments system and improve social welfare. Rochet and Tirole (2003) argue social welfare can be improved if there is strong merchant resistance to accept cards in a high interchange fee environment, such that the merchant discount applied by the acquiring bank is not overly high.

Bolt et al. (2010) empirically investigates the behaviour of consumers in choosing payment instruments in the presence of surcharging by merchants, the characteristics of a merchant that applies surcharges, and the impact on the total costs of the payments system in the Netherlands. Without the ability to surcharge, consumers perceive the true cost of all payment instruments to be identical, leading to a socially inefficient outcome.

Data from an EIM (2007) report reveals approximately 84% and 82% of all transactions are completed with cash for transactions below EUR 5 and between EUR $5 - 10$, respectively. However, the percentage of all transactions completed in cash drops to 64% for transaction amounts between EUR $15 - 20$, and the proportion of transactions accounted for cash continues to decrease as the value associated with the transaction rises. The share of debit usage increases with the value of the transaction as cash usage declines.

The dominance of cash for low transaction values implies a surcharge is being placed by merchants for transaction values between $10 - 15$ Euro. Brits and Winder (2005) study the costs associated with the payments system in the Netherlands using data compiled in 2002. By considering the costs bourne by the central bank, retail sector and the banking industry, Brits and Winder (2005) deduce that the socially optimal payment instrument is dependent on the transaction amount. For transactions that are relatively low in value, cash is the most efficient payment instrument, otherwise, debit cards is the preferred payment instrument. Furthermore, Brits and Winder (2005) conclude that for values less than EUR 11.63, it is more cost efficient to pay in cash, otherwise paying by debit card is the most cost efficient payment instrument. Thus, allowing merchants to surcharge is benefiting the payments system in the Netherlands as consumers are directed to the most cost efficient payment instrument for transaction values less than EUR $10-15$.

Bolt et al. (2010) source data from a 2006 DNB Household survey, which observes the behaviour of 2000 households over time that is representative of the Dutch population. The questions include the payment instrument choice of individuals, attitudes relating to merchants applying surcharges and the effect of surcharges on debit cards on payment instrument choice. A separate survey was done in 2006 to obtain merchant data relating to payment instrument acceptance, payment behaviour of customers, debit card surcharges, motivation of surcharging and the effect of surcharging on payment instrument choice.

Overall, 22% of all merchants applied surcharges to debit cards and 80% of all merchants apply surcharges to a transaction that is valued less than EUR 12.50. The average surcharge and transaction value during the study is EUR

0.23 and EUR 10, implying a surcharge of 2.3% by merchants that choose to surcharge. An EC (2006) report notes that the average merchant discount rate on debit cards in the Euro area has been decreasing steadily over time, from 1.7% and 1.2% for a VISA debit card and national debit card scheme in 2000, to 1.3% and 1.1% in 2004, respectively. The further advancement of technology in telecommunication and information technology contributes to the lowering of the costs of card providers to merchants. However, the average surcharge of 2.3% of merchants in the Netherlands implies these savings are not being passed down to consumers. This causes the number of transactions to be settled in cash relative to debit cards to be greater than the optimal amount, resulting in an inefficient outcome for the payments system.

An ordered linear probit model is estimated by Bolt et al. (2010) to quantify the impact of surcharges on the demand for debit card transactions with covariates that controlled for firm specific influences and other external factors. Results indicate merchants that surcharge can expect the share of transactions accounted by debit cards to be less than retailers with no surcharge in place, with merchants successful at steering consumers to their more preferred payment instrument. In an alternative model, the level of the surcharge is included as the key explanatory variable in comparison to a dummy variable to quantify the impact of surcharging. The second model reveals there is a positive relationship between the level of the surcharge and the share of debit transactions of a merchant. Additionally, marginal effects of the model reveal that removing the surcharge causes the share of debit card transactions to rise 8% as a proportion of total transactions for a merchant. Scenario analysis reveals that if the surcharging is removed, the cost savings to flow through the payments system suggests cost savings of up to EUR 50 million. However, the more suitable option would be to allow surcharging but encourage merchants to adjust surcharging to reflect the reduction in costs associated with processing debit card transactions.

The sensitivity of consumers to pricing has implications on the regulation of interchange fees, the strategy issuing firms employ in setting fees and benefits to end-users. Zinman (2009) and Simon et al. (2010) investigate the rate of substitution between payment instruments. The availability of credit, surcharges, fixed fees, per-transaction fees and the accumulation of reward points causes consumers to optimise their choice at the checkout. Empirical studies examining the behaviour of individuals in choosing their preferred payment instrument attempt to extract the behaviour of marginal consumers to a change in pricing to determine the impact of changes in pricing, and consequently, the new composition of payment platforms chosen by consumers.

Zinman (2009) investigates the choice of consumers in choosing between debit and credit cards by modelling the payment instruments as direct substitutes. The primary distinguishing attribute that Zinman (2009) models to explain whether debit or credit is used is the relative pricing between the two payment instruments. Debit and credit cards are close substitutes, given their similarity in relation to acceptance, security, time costs and portability. The debit and credit card differentiate on the basis of pricing, more specifically, access to liquidity, the scale of rewards, interest rate, outstanding balance owing and the credit limit on the credit card.

Data was obtained from the 1995 – 2004 surveys of Consumer Finances, a nationally represented survey from over 4000 American households. Probit models are estimated to quantify the impact of the pricing attributes of credit cards on debit card usage. The results suggest individuals that have binding credit constraints and that are revolvers (that is, individuals paying interest on their amounts owing), are more likely to use debit than credit cards. More specifically, the probability of debit card usage is 9.8% greater when an individual is classified as being a revolver. Additionally, the correlation between revolvers and debit card usage has grown over time, implying the number of individuals viewing debit cards as a direct substitute for credit cards has been steadily increasing with merchant acceptance and fraud in recent years. For example, by only considering the data obtained in 2004, the marginal effect of being a revolver on the probability of debit card usage is 17.3%. The results are robust to several control variables, some of which include proxies for preference for credit cards, consumer spending, internet banking and multi-homing. The impact of holding a credit card on debit usage is expected to be negative and results validate a-priori expectations. An individual holding a credit card is expected to decrease the probability of debit card usage by 6.3%. In addition to the pricing characteristics of credit cards that sought to explain the motivation of individuals in choosing between debit and credit cards, Zinman (2009) found evidence for non-monetary motivations for debit card use, such as the time costs associated with paying a bill at a later date. Using a similar framework to Zinman (2009), Chapters 4 and 5 investigate the underlying themes motivating consumers usage of the BPAY platform.

Simon et al. (2010) study the impact that a reward program and interest free period has on credit card usage. Data collection was outsourced to Roy Morgan Research. Reward programs attached to credit cards enable consumers to observe a negative price in using the payment platform, assuming there is no surcharge attached to the transaction, or the transactional benefit to the consumer is greater than the surcharge imposed by the merchant. The welfare consequences of rewards on the aggregate payments system are unclear. Reward programs can be used as an incentive to shift consumers from inefficient payment methods to more cost effective payment methods, thereby improving the efficiency of the payments system. Additionally, rewards attached to cards can induce consumers to spend more at the checkout, benefiting all stakeholders on the payments platform, which include the platform, merchant, the issuing bank and the acquiring bank. However, rewards may cause the excess usage of credit cards and cause the costs of the payments system to exceed that of an efficient payments market. This imposes unnecessary additional costs on society. In addition, in a no-surcharge environment, merchants that don't observe an increase in spending by consumers that have a rewards program attached to their cards may increase prices to be compensated for the greater costs observed in processing transactions.

The dataset used by Simon et al. (2010) consisted of a sample of consumers whose payment details were recorded in a diary over a two-week period, along with information relating to each transaction. The demographic data relating to these individuals was provided by Roy Morgan Research, along with whether the individual had a rewards program attached to his/her credit card. Simon et al. (2010) initially model the probability that an individual holds a credit card and the probability an individual has a reward program offered on their credit card prior to modelling the effect price incentives have on payment instrument usage. A consumer is assumed to have a interest free period feature on their credit card if the consumer regularly pays off their credit card bill (transactor), otherwise the individual has no access to these interest free funds (revolver).

The authors follow the methodology of Borzekowski et al. (2008) in which a series of probit models are estimated to estimate the effects of various factors, including price incentives on debit card usage. Results indicate income has a large effect on the probability of holding a card. Relative to the base case of a income of $\$40,000 - \$59,999$, there is a statistically significant drop in the likelihood of low income individuals owning a card. Holding a debit card also plays a role in determining the probability of card ownership. Ownership of a scheme debit card decreases the probability of ownership by 13.2%. Income has a significant effect on having a loyalty program attached to card membership. For consumers earning less than $\$20,000$, the probability of being part of the loyalty program is 32.3% less likely, relative to consumers earning between $\$40,000-\$59,999$. This result may be due to the high annual fees or interest rates attached to loyalty programs, where the consumer needs to reach a particular level of expenditure to be compensated. Part-time employment and being retired increase the probability of being part of a loyalty program with a card association, relative to a full-time worker. A retired consumer, will on average have a 22.2% increase in probability of being part of a loyalty program relative to full-time workers. A consumer with a credit card and no loyalty program has a reduced probability of card usage by as much as 22.9% relative to being part of a loyalty program. An individual that is classified as a revolver as opposed to a transactor, reduces the probability of card usage on a given transaction by 15.9%. Furthermore, owning a scheme debit card relative to not owning a debit card, reduces the probability of card use by 14.1%.

Ching and Hayashi (2010) evaluate the contribution of rewards on credit card usage in a similar framework to that of Simon et al. (2010). However, Ching and

Hayashi (2010) study the market for card usage in the United States, which does not allow merchants to surcharge across payment platforms. Such a restriction on merchants may result in the over usage of cards due to negative price that consumers receive via this pricing. Consequently, the contribution of the reward program on credit card usage is expected to be greater than that found in Simon et al. (2010). In addition, Ching and Hayashi (2010) improve upon Simon et al. (2010) by collecting attitudinal data related to payment method choice so as to remove the endogeneity of choosing to have a rewards program attached to the credit card. Consumers may prefer to use a credit card regardless of whether it offers a rewards program. By collecting attitudinal data and using it as a control variable, the direct influence of the rewards program can be observed.

The data set used by Ching and Hayashi (2010) was obtained by consumers completing a survey and it consisted of whether the individual was a member of a card reward program, demographics of the individual, preferences of the individual regarding each payment method, the most frequently used payment method by merchant type and the payment methods the individual perceives is available at different merchant types. Multinomial logit models are used to study the impact of reward programs using four specifications. The four different models depend on whether a homogenous or heterogenous specification of the choice set available to consumers at the checkout is used and whether the attitudinal variable of consumers towards different payment instruments is included as a control variable.

Regression results of Ching and Hayashi (2010) indicate that the coefficients satisfy a-priori expectations that the availability of rewards increases card usage. In addition, the coefficient for both the credit and debit card reward program is statistically significant and robust across all four specifications of the model. The impact of removing reward programs on overall card usage is estimated by determining the marginal change in the probability of using credit cards. It is expected that removing credit card rewards will reduce the volume of transactions of credit cards by 3.3% to 11.4%, the extent of the reduction is dependent upon the industry of the merchant. The effect of removing credit card rewards on usage is most evident in department stores and consumers will typically substitute their rewards credit card with the non-reward card equivalent.

Hayashi and Klee (2003), much like Simon et al. (2010), study the contribution of consumer demographics and transaction characteristics on the choice of payment method chosen in completing a transaction. However, Hayashi and Klee (2003) extend the literature by hypothesising and testing whether consumer technological adoption positively influences the likelihood of adopting an alternative new technology. Huffman and Mercier (1991) find evidence for farmers adopting one type of new computing technology is strongly influenced by the decision to adopt an alternative technology-based product. In relation to the payments market, Hayashi and Klee (2003) test whether technology adoption of alternative electronic services or products strongly influence the propensity to pay bills electronically. Covariates that proxy for technology adoption includes completing the survey via the internet, cell phone, internet shopping and direct deposit. Separate logit models are employed on debit usage, paying bills direct and online bill payments to evaluate the hypothesis. Regression results suggest transaction related variables, such as merchant type and transaction value, influence payment instrument usage. Hayashi and Klee (2003) also find evidence in support of their hypothesis and notice that the significance of demographics and financial related variables become insignificant in explaining payment instrument usage, contradicting the results of Simon et al. (2010), Ching and Hayashi (2010) and Zinman (2009).

The majority of the findings in the literature are fairly represented in the theoretical and empirical papers discussed thus far, with a particular focus on credit cards. The overall focus of these studies is interchange fees. The level of interchange fees set by a payment platform provide price signals to both the consumer and merchant, once costs and benefits are distributed via their respective banking institution (four-party scheme). Thus far, no model has been developed that describes the demand of consumers' and merchants' for transactions to variations in interchange fees. Rysman (2004) proposes a framework to model demand in a two-sided market for the market for Yellow Pages that can be applied to the payments market with network effects and incompatible markets. The market for payments satisfies these conditions. All payment market platforms are incompatible markets as benefits provided to a set of payment platform subscribers, or an increase in platform participants, does not create value for alternative platforms.

Rysman (2004) seeks to model the demand of merchants to place advertisements in Yellow Pages directories and consumers' demand for usage. The Yellow Pages directory is the platform in this two-sided market, with advertisements in the directory (platform) representing the service offered by the platform to merchants. Consumers value the platform to get into contact with a suitable merchant. The network effect in the market for Yellow Pages directories concerns consumer usage and advertising. The more consumers use a Yellow Pages directory, the more advertising merchants wish to place in the directory, which increases the likelihood of the consumer to find worthwhile information that increases usage. Hence, consumer usage and merchant advertising create a network effect.

The data consists of several sources. The data collected to model consumer usage, is obtained from the National Yellow Pages Monitor (NYPM) and consists of the number of consumer references per month in a household in every metropolitan statistical area for the 476 directories in the sample. The number of pages of the directory is used as a proxy for advertising. The Yellow Pages Publishers Association (YPPA) maintains reserves of directories and the Boston

Consulting Group has collected data detailing the number of pages in each directory. Pricing information of advertisements in a directory is obtained by the YPPA, which has prices for advertisements that vary by size and style.

Rysman (2004) formulates the model of merchant demand for advertising by deriving the first-order conditions of maximising profit from advertising in the Yellow Pages directory. The expression for the demand of consumer usage is obtained by the model used in Berry (1994). The first-order condition of maximising profit for the publisher of the directory is also derived to complete the model to calculate equilibrium outcomes for different assumptions relating to social welfare by comparing the outcomes estimated under concentrating the market to a single directory and through competition. The results suggest consumer usage of a directory is positively related to the quantity of advertising in the directory and a merchants willingness to pay increases with consumer usage. Both these results suggest a positive network effect in the market for Yellow Pages directories. Equilibrium outcomes suggest the benefits of competition improve social welfare in comparison with coordinating the two-sides of the market on a single platform (monopoly). Thus, the network effect is not strong enough to compensate consumers and merchants for increased prices.

Milne (2006) develops a model that seeks to describe the differences in adopting technological advancement in the market for payments in Eastern European countries, United States of America and the United Kingdom. From a macro economic perspective, the Milne (2006) model suggests the adoption of innovations in the market for payments is driven by the positive account externality of increasing market share and improving the quality of service offered to bank customers and the degree of concentration in the banking sector. The take-up of a new technology requires banks to coordinate and share costs in establishing the platform. As such, a highly concentrated banking market with coordination amongst the banks and profit to be shared from the efficiencies that arise is predicted by the model.

Chapter 3

Competition in the Bill Payment Market

3.1 Introduction

The Bill Payment market is a two-sided market and the key drivers of transactions are yet to be defined. The literature concerning the market for payments is largely theoretical and have not been confirmed empirically in the Bill Payment market. The focus is on the interchange fee set by the payment platform. A two-sided market is characterised as a platform providing goods and services to two distinct end-users with prices set for each type of end-user.

The motivation of Chapter 3 is to empirically test for the presence of network effects in the market for payments. Network effects refers to the fact that the value of a particular platform to an end-user increases as the number of end-users on the other side of a transaction increases. The literature emphasises the importance of network effects to the number of transactions placed over a platform. There has been marginal progress in establishing empirically the role network effects has on transactions in the Bill Payment market. The key limitation in extending the literature empirically is the unavailability of data. This is reflected in the lack of empirical studies on network effects in the literature.

The data to be used in this chapter will be provided by BPAY and the Reserve Bank of Australia. BPAY is a dominant platform within the Bill Payment market, accounting for approximately 30% of all transactions. Multivariate Error Correction Models will be used to determine whether the network effects hold via a cointegrating framework. Furthermore, impulse response functions are estimated to determine the number of quarters required for equilibrium to be established following an innovation to one of the covariates in the consumer usage and merchant acceptance demand equations.

3.2 Background

The conventional supply and demand theory applied to a good or service cannot be applied to payment instruments. Platform providers offer a joint product to both consumers and merchants. The benefit a consumer receives from using a payments platform to facilitate a transaction, also confers a separate benefit to the merchant. The consumer may receive rewards points, access to credit, while the merchant receives the funds directly into their banking accounts without having to handle cash.

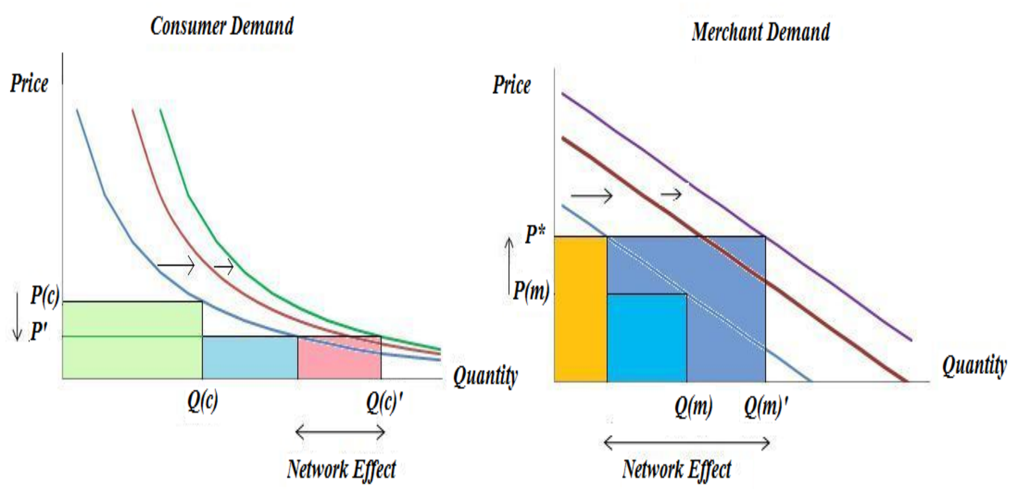
Unlike other markets, the presence of network externalities adds complexity to the pricing decisions faced by platforms to maximise the scale of their networks. The greater the size of the network, that is, the number of merchants and consumers, the greater the value any participant that adopts the network realises from being part of the scheme. Another unique characteristic of the market for payments relates to the party that initiates the transaction not being directly affected by the price signals of the platform. The consumer decides the payment instrument from the platforms to which the consumer and merchant are subscribed to; however, the merchant pays the fee associated with the transaction. Thus the demand of the platform in a given transaction is driven by the consumer; however, merchants pay a fee to be associated with it.

The setting of interchange fees is the main strategic decision faced by platforms, as a means to balance the demand of the service by consumers and the supply of the service by merchants. Prior to the RBA reforms, merchants were not able to convey to consumers the extent of the fees charged by platforms. Subsequently, the lifting of the no surcharge and the no-steering rules has enabled merchants to pass on the costs to the consumer and direct them to preferable platforms.

In traditional markets, prices are expected to be driven downward due to competitive pressures. The presence of network externalities, along with the fact that consumers dictate on which platform a transaction takes place, can lead to prices trending upward over time. Standard economic theory suggests raising prices in a competitive market will result in a decrease in the quantity demanded by consumers and a loss in market share. On the other hand, raising the interchange fee (price) in the payments market, will allow more benefits to be supplied to the consumer. Additionally, a higher fee per transaction will be paid by the merchant, as fees are passed directly from the acquirer to the merchant, thus decreasing merchant demand of the platform. However, the increased consumer demand shifts the demand curve of merchants outward. Thus competition in the market for payments, with network externalities and consumers initiating transactions, may drive prices upward to increase market share. Figure 3.2 presents the outcome of an increase in the interchange fee set by a platform. An increase in the interchange fee, decreases the price consumers face from $P(c)$ to P'. Issuers provide consumers with more incentives, such as access to more credit or reward points, increasing the quantity demanded from $Q(c)$ to $Q(c)'$ as the network effect between consumer usage and merchant acceptance shifts the demand curve outward. Note that, issuers may not pass through all the benefits to the consumer from an increase in the interchange fee and may retain the profits from a price rise.

However, merchants will experience a price rise, the increase in costs by the acquirer are passed to the merchant and consequently, some may leave the platform. The network effects between consumers and merchants will shift the demand curve of merchants and consumers outward repeatedly, as the positive loop created by the network externalities takes effect. Thus, even though there is a price rise for merchants, the increased demand for consumer usage of the platform feeds into the demand function of merchants to transact on the platform simultaneously. The total transactions demanded by merchants thereby rises from $Q(m)$ to $Q(m)'$. The platform consequently becomes more profitable.

Figure *3.2.* An increase in Interchange Fees in a Payments Market.



The payments market, left unregulated will inevitably lead to high interchange fees, over-usage of electronic payment platforms and inefficient outcomes for the economy. Merchants have the option of leaving a platform if fees are excessive; however, this conveys negative signals to consumers and puts them at a competitive disadvantage.

The penetration of cards and BPAY within the Australian market is high. Up to half the value of non-cash retail spending takes place over the card network, while approximately 30% of total transactions are placed over the BPAY platform alone in the bill payments market. To improve social welfare and remove the possibility of entering into a unstable equilibrium of continual interchange price hikes, central banks have intervened in payment markets to improve social outcomes. In addition, a high concentration of transaction volume amongst a few market participants will create artificial barriers to entry, decreasing competition.

The locus of points consisting of platform volume from changes in interchange fees or pricing is the focus of this paper. The circular nature of demand between consumers and merchants requires network externalities to form the basis of demand expressions of their respective demand equations. A platform's goal is to maximise market share through competitive pricing on both sides of the market and thus, maximise network externalities.

3.2.1 Pricing Structure of the Bill Payments Market in Australia

The Australian Bill Payment market is dominated by BPAY, direct debit, Australia Post and credit card providers. The pricing structure to end-users of these platforms is dependent on whether benefits are provided to consumers by transacting on the platform. BPAY, Direct Debit and Australia Post provide no monetary benefit to consumers that choose their platform to pay a bill. However credit cards, such as those belonging to Visa and MasterCard, provide reward points that can be transformed into shopping vouchers or discounts in future purchases. A platform's decision to not allocate benefits or costs to consumers that use the platform suggests growth in volume can only be achieved by maximising merchant acceptance. Choosing whether a platform will allocate fees or benefits to consumers is a strategic decision and it is beneficial to model both scenarios by applying the theory of two-sided markets to the Bill Payment market. This study will concentrate on BPAY and Visa as they best represent the two scenarios of the Billing Payment market in Australia. Figure 3.2 provides a representation of the pricing structure that Visa adopts. The rewards program attached to usage is illustrated by considering the prices of consumers $P(c)$ and P' to be less than zero. The more rewards there are available to consumers for using the Visa platform, the lower the cost in using that payment platform for consumers.

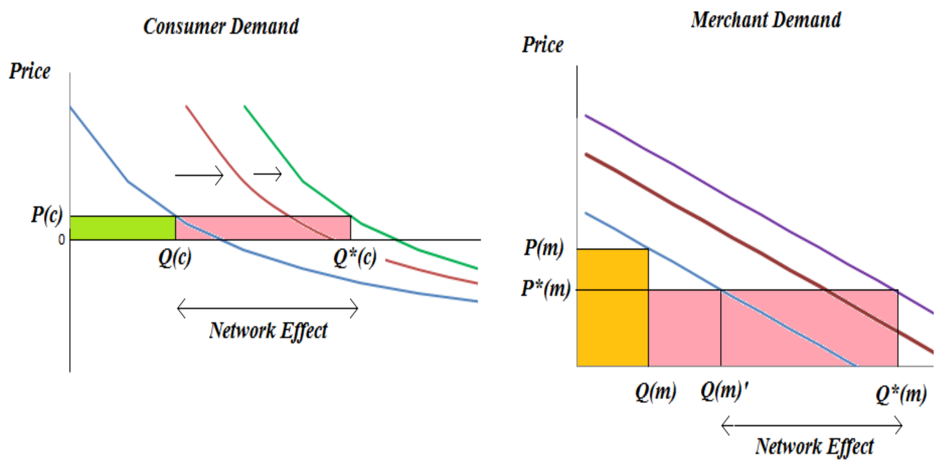
3.2.1.1 BPAY

The BPAY platform has no cost or benefit to consumers that use their platform to pay bills. Figure 3.2.1.1 presents the outcome of a decrease in the interchange fee set by BPAY. The price to the consumer of using BPAY is hypothesised to be a constant whose value is close to zero, and will be modelled as such. This assumption is required in modelling and estimating the demand function of consumers. The positive price, $P(c)$, may be interpreted as the cost to the consumer of going through the process of using BPAY. Some costs may include the disutility of the time and effort required logging into their internet banking portal and inputting the biller and customer reference number. Even though some of these costs are trivial, they are non-zero and must be incorporated into the modelling and estimation process.

To increase consumer usage, the platform must increase retailer acceptance. A decrease in the interchange fee has no financial impact on consumers. Issuers do not provide consumers with any incentives, such as access to credit or reward points. Thereby, there is no increase initially in the demand of consumers to use BPAY.

However, with a decrease in the interchange fee, merchants will experience a price fall. The decrease in costs by the acquirer are passed to the merchant and consequently, more merchants will join the platform. Merchant acceptance increases from $Q(m)$ to $Q(m)'$. The network effects between consumers and merchants will shift the demand curve of merchants' and consumers' outward repeatedly, as the positive loop created by the network externalities takes effect. Thus, even though there is a constant price for consumers, the increased demand in merchant acceptance of the platform feeds into the demand function of consumers to transact on the platform simultaneously. The total transactions demanded by consumers rises from $Q(c)$ to $Q^*(c)$ and merchant acceptance increases from $Q(m)$ to $Q^*(m)$. These respective increases are due to the network effects present in the Bill Payments market.

Figure *3.2.1.1.* A decrease in Interchange Fees by BPAY.



3.3 The Bill Payment Model

To determine the key drivers of transactions in the Bill Payment market, the network effects between consumer usage and merchant acceptance needs to be modelled. The approach of Rysman (2004) will be modified and used to model the interaction between consumers and merchants in order to capture the network effects underlying their respective demand functions.

In a simplified framework, the role of the issuer (consumer's bank) and acquirer (merchant's bank) will not be initially modelled. The network effects within the Billing Payment market can then be described by the following system of equations:

$$U_j^c = a\left(CV_j\right), \tag{3.3.1}$$

$$R_j = b\left(U_j^c, P_j^m, \hat{P}_k^m\right), \tag{3.3.2}$$

$$CV_j = c\left(R_j, P_j^c, \hat{P}_k^c\right), \tag{3.3.3}$$

where: CV_j = Consumer value for platform j.

U_j^c = Consumer usage of platform j.

R_j = Retailer acceptance of platform j.

P_j^m = Merchant fees of accepting platform j.

$\hat{P}_k^m$ = Merchant fees of accepting platform k.

P_j^c = Consumer benefits of transacting over platform j.

$\hat{P}_k^c$ = Consumer benefits of transacting over platform k.

Equation (3.3.1) describes the number of transactions placed over platform j as a function of the value (CV_j) a consumer places on that platform. The value placed on a platform may be dependent upon the ease of use, probability of fraud, benefits provided to the consumer via the platform and merchant acceptance. However, from the perspective of modelling the network effects between consumer usage and merchant acceptance, the primary variable of interest is consumer value. Other factors will also be considered in the estimation process in order to formulate a more robust expression for consumer usage. The impact of the network effect is captured in equation (3.3.2), with merchant acceptance (R_j) being a function of consumer usage $\left(U_j^c\right)$. Lastly, equation (3.3.3) completes the network effect between consumers and merchants. The value placed by consumers (CV_j) on using the platform $\left(U_j^c\right)$ is a function of merchant acceptance (R_j), as well as the fees or benefits of transacting on the platform and that of a competing platform k. This framework can easily be extended to consider multiple payment platforms by modelling $\hat{P}_k^m$ and $\hat{P}_k^c$ as vectors, with each element representing the prices of alternative platforms.

The three first-order conditions that are consistent with network effects in the Billing Payment market are as follows:

$$\frac{\partial U_j^c}{\partial CV_j} > 0, \tag{3.3.4}$$

$$\frac{\partial R_j}{\partial U_j^c} > 0, \tag{3.3.5}$$

$$\frac{\partial CV_j}{\partial R_j} > 0 \tag{3.3.6}$$

These three first-order conditions detail the network effects within the Billing Payment market and complete the network between merchant acceptance and consumer usage. All three equations describe a-priori expectations in modelling the demand functions of consumers and merchants. Equation (3.3.4) expresses the positive relationship between consumer value and consumer usage, while equation (3.3.5) dictates how merchant acceptance increases with consumer usage. Similarly, equation (3.3.6) details how the consumer value placed on a platform increases with retailer acceptance. The work of Rysman (2004) finds that equation (3.3.4) and (3.3.6) are sufficient for identification of the network effects.

Estimating equations (3.3.1) to (3.3.3) is problematic as consumer value (CV_j) is difficult to estimate empirically. It is hypothesised some of the factors that influence CV_j include merchant acceptance, time using the platform, wealth, occupation, access to alternative platforms, fees and rewards. Such a data set will be time consuming to obtain and it does not hinder modelling network effects. Consumer usage will be used as a proxy for CV_j. In the above framework, this amounts to estimating the following equations:

$$U_j^c = a\left(c\left(R_j, P_j^c, \hat{P}_k^c\right)\right), \tag{3.3.7}$$

$$R_j = b\left(U_j^c, P_j^m, \hat{P}_k^m\right), \tag{3.3.8}$$

Network effects are thereby identifiable by expressions (3.3.7) and (3.3.8). Therefore, the modified first-order conditions that are consistent with network effects in the Billing Payment market are as follows:

$$\frac{\partial U_j^c}{\partial R_j} > 0, \tag{3.3.9}$$

$$\frac{\partial R_j}{\partial U_j^c} > 0, \tag{3.3.10}$$

3.3.1 Models of Consumer Usage and Merchant Acceptance

Estimation of equation (3.3.7) and (3.3.8) is of central importance in testing for network effects. In empirical macroeconomics, the Cobb-Douglas production function is used to govern the functional forms of a set of outputs to inputs. The framework for the equations that follow was established in section 3.3. Network effects within the Bill Payments market considers the total usage of both end-users. This macroeconomic perspective of the Billing Payment market provides an opportunity to model equation (3.3.7) and (3.3.8) using Cobb-Douglas functions.

$$V\left(M, P_c, \hat{P}_c, X\right) = aM^{\alpha} P_c^{\tau} \hat{P}_c^{\vartheta} \tilde{P}_c^{\pi} X^{\gamma} \tag{3.3.11}$$

$$M\left(V, P_m, \hat{P}_m\right) = bV^{\beta} P_m^{\delta} \hat{P}_m^{\lambda} \tilde{P}_m^{\phi} \tag{3.3.12}$$

where: V = Volume of transactions.

M = Number of Merchants offering the BPAY platform.

X = Vector of deterministic variables.

P_c = Usage fees or benefits of the BPAY platform.

$\hat{P}_c$ = Usage benefits of transacting on the Visa platform.

$\tilde{P}_c$ = Usage benefits of transacting on the Diners Club platform.

P_m = Merchant fees of the BPAY platform.

$\hat{P}_m$ = Merchant fees of the Visa platform.

$\tilde{P}_m$ =Merchant fees of the Diners Club platform.

Equation (3.3.11) and (3.3.12) form the basis of estimation by applying the Cobb-Douglas production function to the Bill Payments market. Taking logarithms of equation (3.3.11) and (3.3.12) has the attractive property of transforming parameters into elasticities. The parameters of interest in this Chapter are α and β, where a statistically significant value supports the existence of a network effect in the Bill Payments market. Taking logarithms, enables the parameters of equation (3.3.11) and (3.3.12) to be linear. This transforms equation (3.3.11) and (3.3.12), as provided below:

$$\begin{aligned} v\left(m,\hat{p},x\right) &= \bar{a}+\alpha m+\tau p_c+\theta\hat{p}_c+\pi\tilde{p}_c+\gamma x \qquad (3.3.13)\\ v\left(m,\hat{p},x\right) &= \mathring{a}+\alpha m+\theta\hat{p}_c+\pi\tilde{p}_c+\gamma x\\ \mathring{a} &= \bar{a}+\tau p_c \end{aligned}$$

$$m\left(v,p,\hat{p}\right)=\bar{b}+\beta v+\delta p_m+\lambda\hat{p}_m+\phi\tilde{p}_m \qquad (3.3.14)$$

where: $v=\log(V)$

$m=\log(M)$

$p_c=\log(P_c)$

$\hat{p}_c=\log(\hat{P}_c)$

$\tilde{p}_c=\log\left(\tilde{P}_c\right)$

$x=\log(X)$

$p_m=\log(P_m)$

$\hat{p}_m=\log(\hat{P}_m)$

$\tilde{p}_m = \log(\tilde{P}_m)$

BPAY and the consumers' bank do not charge consumers to transact on BPAY's platform. Therefore p_c will be modelled as a constant with a value close to zero. The resulting model for consumer usage of BPAY is given by expression (3.3.13), with $\mathring{a} = \bar{a} + \tau p_c$. Australia post is a key firm in the Bill Payment market, however, data relating to the merchant fees is not available and is thereby omitted by the model.

Of additional interest is the price elasticity of merchants to costs imposed by the platform. It has commonly been hypothesised in the literature that $\alpha < \beta$ in the development of theoretical models in the market for payments as, in practice, the price allocation by the platform has been in the consumers favour as they are less likely to multihome. For example, there is typically no cost in using the BPAY platform for consumers, however, merchants are typically charged a fee per-transaction. Empirically estimating the values of α and β provides an opportunity to evaluate theory with practice.

The price variables in expression (3.3.14) allows the platform to gauge its market power in the Bill Payments market. In absolute terms, the lower the value of δ, the greater the market power of the platform. Similarly, the greater the value of λ and ϕ, the lower the market power of the platform. Theoretically, there should be a sole platform that exists in the long-run in any two-sided market. It is in the best interest of end-users to use the platform with the largest scale or greatest network effect.

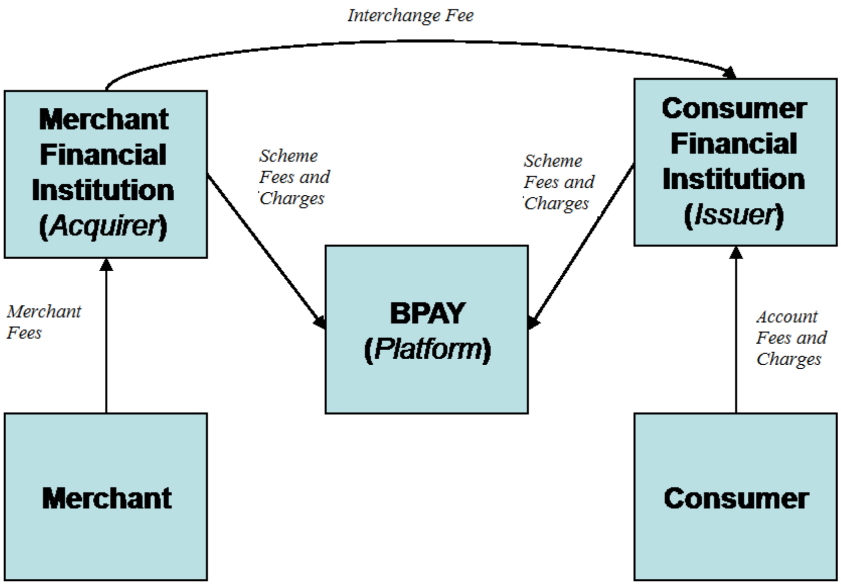
Consider a repetitive game in the Billing Payment's market, with n consumers, m merchants, k platforms and t time periods. Let all consumers and merchants have the ability of subscribing to multiple platforms and let the subscription fees and per-transaction cost of joining a platform be identical. There is no restriction on the allocation of costs to merchants and consumers. This assumption is realistic as all platforms provide identical services. At every time period of the game a joining fee is charged by all platforms to all end-users. In the first stage of the game, the consumers and merchants' subscriptions to the k platforms are governed by expectations. Utility is maximised for consumers when they can transact with the greatest number of merchants, and vice verca. Due to the fees involved in joining a platform, both the consumer and merchant have the ability to choose platforms they wish to join and neglect. In the next period of the game, the utility of consumers is maximised when they join the platform with the highest market share of merchants subscribing to the platform in the previous period, and vice verca. Hence, platforms with the highest market share will increase their market share in the next period. As this game is repeated for T time periods, the Nash equilibrium of all consumers and merchants will be to subscribe to a single platform.

3.3.2 Data

The lack of access to proprietary data in the market for payments has been the primary cause of limited empirical studies of network effects. BPAY has made available quarterly data between March 2003 and December 2010 relating to consumer usage, merchant acceptance and interchange fees. Quarterly statistics from the Reserve Bank of Australia complete the data set. There are two limitations associated with this data set. Prior to outlining the data that are used to estimate the demand equations of merchants and consumers, it is beneficial to view the structure of the BPAY platform.

The structure of the BPAY four-party scheme is similar to that of Visa and MasterCard and is illustrated in Figure 3.3.2. A fee is paid by both the acquirer and issuer to BPAY per-transaction, along with a payment (interchange fee) made by the acquirer to the issuer for services performed in transmitting the information to BPAY. These fees are set by the platform. The merchant is charged a merchant fee per-transaction by their financial institution which is commonly a fixed percentage of the value of the transaction. The cost imposed on the consumer for using BPAY is normally zero and is part of a bundle of services provided by their bank. As such, it is difficult to isolate the pricing of such a service. There is no contact made by either the consumer or merchant with BPAY. It acts as a platform that enables information to be sent between the banks.

Figure 3.3.2. Four-Party Payment Scheme



To estimate the demand equations of merchants and consumers, the merchant fees and the account fees and charges of consumers that use BPAY are required. However, these costs are not publicly available. It has been observed in practice that there is a high correlation between the interchange fee set by the platform and the merchant fee the acquiring bank sets per transaction. Therefore, the interchange fee for BPAY is used as a proxy for merchant fees for this platform. The RBA publishes quarterly the average merchant fee of Visa/MasterCard and Diners Club, thus no proxy is required for $\hat{p}_m$ and $\tilde{p}_m$ in equation (3.3.14).

The second data shortcoming relates to the ability to estimate the benefits of using an alternative platform to BPAY. The literature in the market for payments describes a positive relationship between interchange fees, merchant fees and benefits supplied by the platform or the issuing bank, to the consumer. Effectively, merchants subsidise the cost for consumers to use the platform. The greater the interchange fee, the greater the merchant fee that the acquiring bank charges and therefore, the greater the revenue for the issuing bank. The revenue that is raised by the issuing bank is used to provide benefits to consumers to use the platform. There is no financial benefit or cost in transacting over the BPAY platform. However, there are rewards attached to using a credit card. Data relating to the benefits of using a Visa/MasterCard is only available annually. Average benefits and fees of the credit card providers will be considered as each issuing bank has on offer several cards available with differing reward programs. To increase the number of degrees of freedom in estimation, a proxy is required. The correlation between the average merchant fees and benefits of Visa/MasterCard is 98.5%. Thus, merchant fees will be used as a proxy for benefits in equation (3.3.13). The list of the variables that is required to estimate the parameters of interest are provided succinctly below:

- v : Quarterly transaction volume.
- m : Number of active biller codes.
- $\hat{p}_c$: Quarterly average merchant fees of Visa/MasterCard. Expressed as a percentage of the value of a transaction.
- $\tilde{p}_c$: Quarterly average merchant fees of Diners Club. Expressed as a percentage of the value of a transaction.

- x : Quarterly dummy variables.

- p_m : Quarterly merchant fee of firms accepting BPAY, expressed as a percentage of the value of a transaction. This is estimated by dividing the interchange fee by the average value of a transaction in that respective quarter.

- $\hat{p}_m$: Quarterly average merchant fees of Visa/MasterCard. Expressed as a percentage of the value of a transaction.

- $\tilde{p}_m$: Quarterly average merchant fees of Diners Club. Expressed as a percentage of the value of a transaction.

3.4 Methodology

The estimation of merchant and volume elasticities in equation (3.3.13) and (3.3.14) is obtained within Johansen's Maximum Likelihood cointegration framework. The Vector Autoregressive (VAR) model is advocated by Sims (1980) as a way of estimating dynamic autoregressive relationships between endogenous variables without presupposing exogeneity restrictions on some variables.

In testing for network effects, the time series properties of the data set must be considered in the estimation of the parameters of interest, namely α and β. Assuming stationarity in all the variables enables equation (3.3.13) and (3.3.14) to be estimated via Ordinary Least Square (OLS). The parameters are then evaluated using conventional t-tests, assuming the assumptions underlying (OLS) are met. However, the presence of unit roots complicates estimating and evaluating the elasticities representing network effects. This typically occurs when the dependent and explanatory variables are trending consistently over time, which implies falsely that there is a valid relationship. However, in the event of an cointegrating relationship between the non-stationary variables, OLS is a valid estimator.

The use of univariate error correction models (UECM) to determine the estimates of the network effects elasticities requires the assumptions of one cointegrating vector, as well as the explanatory variables of the model being strongly exogenous to the dependent variable. Such restrictions are not realistic due to the network effects in the Bill Payments market. The usage of the platform by consumers and merchants are influenced by network effects, as shown by equation (3.3.13) and (3.3.14). Consequently, consumser usage and merchant acceptance are classified as endogenous variables. Thereby, imposing the strong exogeneity assumption of merchant acceptance in equation (3.3.13) is an incorrect assumption. Similarly, assuming consumer usage in equation (3.3.14) is strongly exogenous is also inappropriate. Thus, by employing a UECM, the parameter estimates will be biased.

Prior to the implementation of Johansen's Maximum Likelihood Cointegration procedure to estimate the elasticities of interest, it is necessary to determine the time series properties of the variables in the model and establish possible cointegration. The sections that follow describe the panel unit root tests to be implemented, followed by the formulation of a Vector Error Correction model (VECM), Johansen's Maximum Likelihood Cointegration procedure and Impulse Response Functions.

3.4.1 Univarite Unit Root Tests

To evaluate the time series properties of the variables in the consumer usage and merchant acceptance demand equations, the Dickey and Fuller (1979) and Kwiatkowski et al. (1992) tests are implemented. The problem of testing for unit roots using a Dickey and Fuller (1979) test in finite samples is the tendency to not reject the null hypothesis when the alternative is true. A further problem is the low statistical power to distinguish between unit and near unit roots and between trend and drift. Therefore, it has become the norm to test for stationarity in the data series using the Augmented Dickey Fuller Test (ADF) and Kwiatkowski et al. (1992) (KPSS) tests jointly. The ADF test evaluates the null hypothesis that the series contains a unit root. This is in contrast to the null hypothesis of stationarity in the KPSS framework.

However given the relatively short time dimension of the sample, the ADF test and KPSS test may have low power. Thereby, the conclusions reached have the potential to be misleading. By employing a panel data approach to the testing of the time series properties of the variables, more power can be gained.

The Dickey-Fuller (DF) approach tests the null hypothesis that a series contains a unit root, $H_0 : \rho_a = 1$, against the alternative of stationarity, $H_1 : \rho_a < 1$. The DF test amounts to estimating equation(3.4.1) :

$$\begin{aligned} y_t &= \rho_a y_{t-1} + u_t, \qquad (3.4.1) \\ \Delta y_t &= (\rho_a - 1) y_{t-1} + u_t, \\ u_t &\sim IID\left(0, \sigma^2\right), \end{aligned}$$

Under the null hypothesis of a unit root, the test statistic follows a non-normal distribution. Critical values are computed using Monte Carlo simulations, and are commonly referred to as DF statistics.

The Augmented Dickey Fuller Test (ADF) is comparable to the simple DF test, but it involves adding an unknown number of lagged first differences of the dependent variable to capture autocorrelated omitted variables that would otherwise, by default, enter into the error term, u_t. Choosing the appropriate lag length is also important. Over-specifiying the model (too many lags) will reduce the power of the test as additional parameters are estimated, decreasing the degrees of freedom; too few lags may result in over-rejecting the null when it is true, decreasing the size of the test. It is important that the model has little autocorrelation, as too few lags may result in the autocorrelation in the residual term and the inapplicability of applying critical values.

The KPSS test assumes stationarity of the series as the null hypothesis.

As the consequences of non-stationarity are important in relation to spurious regressions and cointegration, it is best to take a conservative approach with non-stationarity being the null hypothesis. However, it is useful to test using both alternatives of the null.

3.4.2 Panel Unit Root Tests

To augment the information in trying to evaluate the time series properties of the variables in the consumer and merchant acceptance demand models, several panel unit root tests were conducted. The common panel unit root test is based on the following regression:

$$\Delta y_{it} = \rho_i y_{it-1} + \kappa z_{it} + u_{it}, \tag{3.4.2}$$

where $i = 1, 2, \ldots, N$ represents each individual, $t = 1, 2, \ldots, T$ the time series observations, z_{it} the deterministic components such as individual specific intercept or time trend and u_{it} is the random error term.

Panel unit root tests can be divided by the degree of homogeneity imposed in the alternative hypothesis. The null and alternative hypothesis adopted by panel unit root tests are:

$$
\begin{aligned}
H_o &: \rho_i = 0 \; for \; all \; i = 1, 2, ..., N \\
H_{1a} &: \rho_i = \rho < 0 \; for \;\; i = 1, 2, ..N, \\
& \quad or \\
H_{1b} &: \rho_i < 0 \; for \;\; i = 1, 2, ..N_1 \text{ where } N_1 < N
\end{aligned} \tag{3.4.3}
$$

The null hypothesis provided in equation (3.4.3) is a test for a panel unit root, with the alternative hypothesis differing by whether the rejection of the null hypothesis is supported for the panel to be stationary as a whole or whether there is a mixture of stationary and non-stationary individuals in the panel as denoted by the alternative hypothesis of H_{1b}. In other words, the null hypothesis is rejected for $N_1 < N$ cross-sections of the panel such that, for consistency as $N \rightarrow \infty$, $\frac{N_1}{N} \rightarrow 0 < \delta \leq 1$. The Levin et al. (2002) (LLC) and Hadri (2000) test for a panel unit root share the common characteristic of imposing a large degree of homogeneity in the null and alternative hypothesis. The main disadvantage with such a test is the degree of homogeneity imposed on the autoregressive coefficients which results in low power, or the failure to reject the null hypothesis when it is false. The Hadri (2000) test is unique in the sense it is a test for stationarity of the panel as opposed to the LLC and Baltagi (2000) tests of a panel unit root. It has also been found by Harris and Tzavalis (1999), that the LLC test suffers from a dramatic loss of power when individual trends are included and it is sensitive to the inclusion of deterministic trends.

In contrast Im et al. (2003) (IPS) and Fisher's ADF share a common null hypothesis of a panel unit root and alternative hypothesis of at least some cross-sections being stationary as denoted by H_{1b}. Simulation results from IPS show in the case of no serial correlation in the error term, the LLC test has a tendency to over-reject the null hypothesis as N increases. For samples with a small T dimension, the IPS test has more power relative to the LLC test, but the LLC test has better size. The tests that impose the restriction of a common unit root process across all cross-sections may be preferred in the case where it is logical to assume that the autoregressive parameter ρ_i, is the same across all cross-sections. Thereby, pooling the data may be advantageous.

The panel unit root tests rely heavily on the assumption of the independence of the errors across cross-sections. However, when the errors are cross-sectionally dependent, all tests suffer from size distortion. The Monte Carlo simulation performed by Maddala and Wu (1999) provides evidence that this problem is less severe with Fisher Type tests in comparison with IPS. In addition, only T needs to approach infinity for the test statistics to approach its limiting distribution. Thus, more weight will be attributed to the Fisher Tests in comparison with the other proposed panel unit root tests in evaluating the time series properties of the data.

However, Pesaran (2007) develops a panel unit root test that does not rely on the assumption of the errors being independent across cross-sections. Pesaran

(2007) proposes estimating the standard ADF regression with the lagged cross-sectional mean and its first difference to capture the cross-sectional dependence that arises through a single factor model.

3.4.2.1 Panel Unit Root Test Allowing for Cross-Sectional Dependence of the Error Terms

All panel unit root tests[1] discussed rely heavily on the assumption of the errors not being cross-sectionally correlated. This assumption is required for the test statistics to approach their limiting distributions. Given that the data involved are economic variables and the sample consists of firms competing fiercley in the Bill Payments market to attract both consumers and merchants, it is likely that the variables of different cross-sections are correlated. For instance, pressure by the RBA and merchant groups to decrease the interchange fees has resulted in the uniform decrease of merchant fees in the Bill Payments market across all platforms. Even though Diners Club has no interchange fee, the threat of losing merchants has forced them to mirror the actions of Visa and decrease merchant fees directly.

Pesaran (2007) Pesaran (2007) develops a modification to the IPS test that accounts for the cross-sectional dependence of the errors while sharing the same null and alternative hypothesis. Additionally, the new test can be modified for

[1] Refer to Appendix A for a detailed explanation of the panel unit root tests employed.

serially correlated errors. The following data generating process is assumed:

$$\begin{aligned} y_{it} &= (1-\rho_i)\,\mu_i + \delta_i y_{it-1} + u_{it}, \\ u_{it} &= \gamma_i f_t + \varepsilon_{it}, \end{aligned} \tag{3.4.4}$$

where μ_i is the deterministic component and it is assumed the initial values (y_{io}) are known and the errors follow a one-factor structure. Pesaran (2007) also states the following assumptions:

- The idiosyncratic shocks, ε_{it}, $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ are independently distributed both across i and t, have mean zero, variance $\sigma^2_{\varepsilon i}$, and finite fourth order moment.

- The unobserved common factor, f_t, is serially uncorrelated with mean zero and a constant variance σ^2_f , and finite fourth moment. Without loss of generality σ^2_f can be set to unity.

- ε_{it}, f_t and γ_i are independently distributed for all i.

- $\bar{\gamma} = N^{-1}\sum_{j=1}^{N}\gamma_j \neq 0$ for a fixed N or $N \rightarrow \infty$

Pesaran (2007) extends the IPS test by including in the ADF test, the cross-sectional mean of the data series, $\bar{y}_{\cdot t} = N^{-1}\sum_{i=1}^{N} y_{it}$, and its lagged values. The unobserved common factor that is causing the errors to be cross-sectionally dependent is essentially proxied by the cross-sectional mean of the data series and its lagged values. The cross-sectionally augmented Dickey Fuller (CADF) regression is expressed as follows:

$$\Delta y_{it} = a_i + b_i y_{it-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + e_{it}, \tag{3.4.5}$$

The test statistic is the average of the N individual t-statistics of the CADF tests. Pesaran (2007) shows the N individual t-statistics of the CADF tests are asymptotically similar due to their dependence on the sole common factor and are independent of the factor loadings. The test statistic thereby has no standard distribution and its critical values are estimated using Monte Carlo methods.

As stated previously, the assumption of the errors being cross-sectionally independent is necessary for the test statistics of LLC, Baltagi (2000), IPS, Hadri (2000) and Fisher Tests to achieve their limiting distribution and a violation of such an assumption will lead to severe size distortion. To evaluate the assumption of cross-sectional dependence of the error terms, the Breusch and Pagan (1980) Lagrange Multiplier Test will be utilised. The characteristics of the sample, with $T = 32$ and $N = 3$ suggest that a test that is designed for fixed N and $T \rightarrow \infty$ is the most suitable for this study. Hence the Breusch and Pagan (1980) Lagrange Multiplier Test is the most suitable test given the dimensions of the sample to test whether the errors are cross-sectionally independent.

The null and alternative hypothesis evaluated are given by:

$$H_o \ : \ \hat{\rho}_{ij} = \hat{\rho}_{ji} = 0 \ \ for \ all \ i \neq j, \tag{3.4.6}$$

$$H_1 \ : \ \hat{\rho}_{ij} = \rho_{ji} \neq 0 \ \ for \ some \ i \neq j,$$

where ρ_{ij} is the ijth residual correlation coefficient and is estimated by:

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{\frac{1}{2}} \left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{\frac{1}{2}}}, \tag{3.4.7}$$

the rejection of the null hypothesis implies the assumption of the errors being cross-sectionally independent being violated.

The Breusch and Pagan (1980) Lagrange Multiplier Test The Breusch and Pagan (1980) Lagrange Multiplier (LM) Test evaluates the hypothesis that the residual correlation matrix, computed over all common observations, is an identity matrix of order N_c, where N_c is the number of cross-sections in the sample. The LM test statistic is given by:

$$\curlywedge_{LM} = T \sum_{i=2}^{N_c} \sum_{j=1}^{i-1} \hat{\rho}_{ij}^2, \tag{3.4.8}$$

The Breusch and Pagan (1980) LM test statistic is distributed as $\chi^2(d)$, where $d = N_c(N_c - 1)/2$, under the null hypothesis.

3.4.3 Cointegration

3.4.3.1 The Engle-Granger (EG) Approach

As a rule, if there are two time series that are $I(d)$, any linear combination of this set of time series will also be $I(d)$. In the event that there exists a vector β, for which the resulting error term is of a lower order of integration $I(d-b)$, then Engle and Granger (1987) define the series y_t and x_t to be cointegrated of order (d, b). Thus, assuming v and m of equations (3.3.13) and (3.3.14) were both $I(1)$, and the associated error term was stationary ($\varepsilon_t \sim I(0)$), then it can be said the two series are cointegrated of order $(1, 1)$.

Assuming cointegration, estimating equations (3.3.13) and (3.3.14) using ordinary least squares (OLS) achieves super-consistent parameters. The parameters converge to their true value at a faster rate as T increases, relative to applying OLS with variables that are $I(0)$. Hence, first differencing the variables in equation (3.3.13) and (3.3.14) to find estimates of the parameters is not efficient. All dynamics and endogeneity issues are invalid asymptocially due to this super-consistent property. Further, in small samples, it has been shown bias is a problem. Consequently, there is an inability to draw inferences about the significance of the parameters of the static long-run model.

3.4.3.2 Problems with the Single Equation Approach

It is convenient to assume that there is only one cointegrating vector in estimating a economic model. However, if there are more than two variables, there is a possibility of more than one cointegrating vector. That is, the variables in a model may feature as part of several equilibrium relationships governing the joint evolution of the variables. This is reflected in the model for merchant acceptance in equation (3.3.14), which includes four variables.

Assuming there is only one cointegrating vector when there is more, leads to inefficiency in the sense that we can only obtain a linear combination of these vectors when estimating a single equation model. Statistically, estimating the cointegrating vector using a single equation is not efficient, that is, there are alternative estimators that obtain a smaller variance. This is due to information being lost, unless each endogenous variable appears on the left hand side of the equation. The exception is when all the variables in the cointegrating relationship are weakly exogenous. However, due to the circular nature of network effects in the Bill Payments market, such an assumption is incorrect.

The problems with estimating an ECM using a single equation with more than two variables can be shown in the example below are given in Appendix B. In the event $y_{2,t}$ and x_t are weakly exogenous to $y_{1,t}$, $\alpha_{21} = \alpha_{31} = 0$. The cointegrating vector from equation $(B.5)$ can be efficiently estimated by OLS. The cointegrating relationship does not enter into the expressions of $\Delta y_{2,t}$ and Δx_t

and hence, there is no loss of information by estimating the model using a single equation. Hence, when all variables in a cointegrating relationship are weakly exogenous to the left hand side variable, the OLS estimator for the ECM is efficient. Johansen (1992*a*) provides support for the multivariate approach and shows the standard error associated with estimating the cointegrating vector by OLS is greater than that by adopting a multivariate approach $\left(\sigma_{\beta_{OLS}} > \sigma_{\beta_{VECM}}\right)$.

3.4.4 Johansen Full Information Maximum Likelihood Procedure

Consider the following Vector Autoregressive (VAR) model applied to the consumer usage equation of (3.3.13) with 4 endogenous variables in the vector z_t and the number of lags in the VAR, p, equal to 4:

$$z_t = \Pi_1 z_{t-1} + \Pi_2 z_{t-2} + \Pi_3 z_{t-3} + \Pi_4 z_{t-4} + \varepsilon_t, \tag{3.4.9}$$

where: $z_t = (m_t, v_t, \hat{p}_{c,t}, \tilde{p}_{c,t})'$

$\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}, \varepsilon_{4t})'$

$$\Pi_1 = \begin{pmatrix} \pi_{11.j} & \pi_{12.j} & \pi_{13.j} & \pi_{14.j} \\ \pi_{21.j} & \pi_{22.j} & \pi_{23.j} & \pi_{24.j} \\ \pi_{31.j} & \pi_{32.j} & \pi_{33.j} & \pi_{34.j} \\ \pi_{41.j} & \pi_{42.j} & \pi_{43.j} & \pi_{44.j} \end{pmatrix} \quad for\ j = 1, 2, 3, 4$$

Note ε_t is a vector of serially uncorrelated errors, such that $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_s') = 0\ \forall\ t \neq s$ and $E(\varepsilon_t \varepsilon_s') = \Sigma$.

The above VAR given by equation (3.4.9) can be repararamterised to give the general expression for the VECM with p lags:

$$\Delta z_t = -\Pi z_{t-1} + \sum_{s=1}^{p-1} \Phi_s \Delta z_{t-s} + \varepsilon_t \tag{3.4.10}$$

where $\Pi = (I - \sum_{s=1}^{p} \Pi_s)$ and $\Phi_s = -\sum_{j=s+1}^{p} \Pi_s$

The VECM expressed in equation (3.4.10) expresses all variables in terms of their differences, with the exception of the lag of the endogenous vector, z_{t-1}. Therefore, assuming all variables contained in equation (3.4.10) are integrated of order one, then in order for the system to be stationary, Πz_{t-1} must also be stationary.

As previously shown in Appendix B, the matrix Π can be decomposed into two matrices:

$$\Pi = \alpha\beta, \tag{3.4.11}$$

where: α is the speed of adjustment parameter to disequilibrium, and β is a matrix of long run coefficients such that the term βz_{t-1} in equation (3.4.10) represents up to $(n-1)$ cointegrating relationships.

There are three alternatives for which the VECM in equation (3.4.10) is stationary. Firstly, all variables are stationary. Secondly, the matrix Π contains $(n \times n)$ elements of zeros which can eventuate if no cointegrating relationship is observed between the endogenous variables in z_t. Lastly, there is at least one cointegrating relationship amongst the variables that cause Πz_{t-1} to become a stationary process. Assuming all variables are integrated of order one, which is a valid assumption with economic time series data, the r cointegrating vectors in β form r linearly independent combinations of the variables in z_t which are stationary. Only the cointegrating vectors in β enter equation (3.4.10) , which implies the last $(n-r)$ columns of α are effectively zero as each of the r cointegrating vectors in β are associated with at least one non-zero element. Consequently, testing for cointegration is equivalent to testing for the rank of Π. For example, if it is deduced that Π is of rank 2, then there exists two cointegrating relationships between the variables in z_t.

Johansen recommends estimating the VECM of equation (3.4.10) and performing a reduced rank regression to correct for short-run dynamics. This procedure involves estimating the following regressions:

$$\Delta y_t = A_1 \Delta y_{t-1} + A_2 \Delta y_{t-2} + ... + A_{k-1} \Delta y_{t-(k-1)} + R_{0t}, \tag{3.4.12}$$

$$y_t = B_1 \Delta y_{t-1} + B_2 \Delta y_{t-2} + ... + B_{k-1} \Delta y_{t-(k-1)} + R_{kt}, \tag{3.4.13}$$

The residuals obtained from estimating equation (3.4.12) and (3.4.13) are then used to calculate the product matrices given by:

$$S_{ij} = \frac{\sum_{t=1}^{T} R_{it} R'_{jt}}{N} \; for \; i, j = 0, k \tag{3.4.14}$$

The maximum likelihood estimates of β are then subsequently obtained as the eigenvectors corresponding to the r largest eigenvalues from solving the equation:

$$\left| \lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k} \right| = 0, \tag{3.4.15}$$

The n eigenvalues from equation (3.4.15) are then ordered in descending order. The test for cointegration is a test for the number of non-zero eigenvalues and the estimated cointegration space is the space spanned by the eignenvectors associated with these non-zero eigenvalues. The following test statistics are used to test the null hypothesis of r cointegration vectors against the alternative that $r+1$ exist:

- Trace Statistic:

$$\lambda_{trace} = -T \sum_{i=r+1}^{n} \log\left(1 - \hat{\lambda}_i\right) \ \ for \ r = 0, 1, 2..., n-1, \qquad (3.4.16)$$

- Max Eigenvalue Statistic:

$$\lambda_{\max} = -T \log\left(1 - \hat{\lambda}_{r+1}\right) \ \ for \ r = 0, 1, 2, ..., n-1, \qquad (3.4.17)$$

3.4.5 Impulse Response Functions

The VAR model specified by equation (3.4.9) can be expressed as a moving average (MA). The MA representation provides an opportunity to model the impact of a shock from one of the explanatory variables over time as it filters through the VAR representation of the consumer and merchant demand models. From an academic perspective, it would be interesting to know how shocks to one of the variables are filtered through both demand equations and how long it would take for equilibrium to be restored within the system.

Consider a more general specification of the VAR model with p lags and n variables:

$$z_t = \alpha + \Pi_1 z_{t-1} + \Pi_2 z_{t-2} + \Pi_3 z_{t-3} + ... + \Pi_4 z_{t-p} + \varepsilon_t, \tag{3.4.18}$$

As opposed to the previous VAR specification of equation (3.4.9), z_t is a $(n \times 1)$ vector of variables, α is a a $(n \times 1)$ vector of constants and ε_t is a $(n \times 1)$ vector of errors. Equation (3.4.18) can then be transformed by using lag operator notation:

$$\begin{aligned} z_t &= \alpha + \Pi_1 L z_t + ... + \Pi_p L^p z_t + \varepsilon_t \qquad (3.4.19) \\ z_t \left(I_n - L - L^2 - L^3 - ... - L^p\right) &= \alpha + \varepsilon_t, \\ \Pi(L) z_t &= \alpha + \varepsilon_t, \end{aligned}$$

In order to estimate impulse response functions (IRF), the vector z_t must be covariance stationary. To be considered covariance stationary, $E(z_t)$ and $E(z_t z_{t-j})$ must be independent of time. Assuming z_t is covariance stationary, expectations can be taken in equation (3.4.19).

$$\begin{aligned} E\left[z_t \left(I_n - L - L^2 - L^3 - ... - L^p\right)\right] &= E(\alpha) + E(\varepsilon_t), \qquad (3.4.20) \\ \mu \left(I_n - L - L^2 - L^3 - ... - L^p\right) &= \alpha, \\ \mu &= \left(I_n - L - L^2 - L^3 - ... - L^p\right)^{-1} \alpha, \end{aligned}$$

Hamilton (1994) shows a MA(∞) representation of the VAR can then be expressed as follows:

$$z_t = \mu + \varepsilon_t + \Upsilon_1 \varepsilon_{t-1} + \Upsilon_2 \varepsilon_{t-2} + \Upsilon_3 \varepsilon_{t-3} + \Upsilon_4 \varepsilon_{t-4} + ... \tag{3.4.21}$$

with $\Upsilon_j \left\{ {}^{I_n \text{ for } j=0}_{\Sigma_{k=1}^{j} \Phi_{j-k}\Pi_k \text{ for } j=1,2,..} \right.$

The Υ_j matrices signify impulse response functions. More specifically, row i, column l element of Υ_j identifies the impact of a unit increase in the lth variable's innovation at time t, ε_{lt}, on the ith element of z_t after j periods, holding all other innovations at all periods constant. Therefore, the ith row and lth column element of Υ_j as a function of j gives rise to a visual representation of the impulse response function$\left[\frac{\partial z_{i,t+j}}{\partial \varepsilon_{l,t}} = \Upsilon_{jl,j}\right]$.

The covariance stationary assumption is not expected to hold for all the variables contained in the vector z_t. Therefore, impulse response functions can not be estimated for MA representation of VAR equation (3.4.21). Assuming there is cointegration, impulse response functions can be formulated by estimating a VECM and by noting that:

$$\begin{aligned} \Pi_1 &= \Pi + \Phi_1 + I_n \\ \Pi_j &= \Phi_j - \Phi_{j-1} \\ \Pi_p &= -\Phi_{p-1} \end{aligned} \tag{3.4.22}$$

Expressions for Π and Φ_j were derived previously in section 3.4.4. However, the existence of serial correlation among the innovations, implies that we cannot assume that all other innovations are held constant for a given change in $\varepsilon_{l,t}$. Hence, it is anticipated a shock in the lth variable will be associated with shocks to the other variables in the system that will consequently affect the lth variable. Additionally, IRF's may be sensitive to the ordering of the variables in the VAR.

Pesaran and Shin (1998) provide an alternate estimator of impulse response functions that rectifies these shortcomings. Hence, the methodology suggested by Pesaran and Shin (1998) will be used to assess the impacts of shocks on the demand for merchant acceptance and consumer usage.

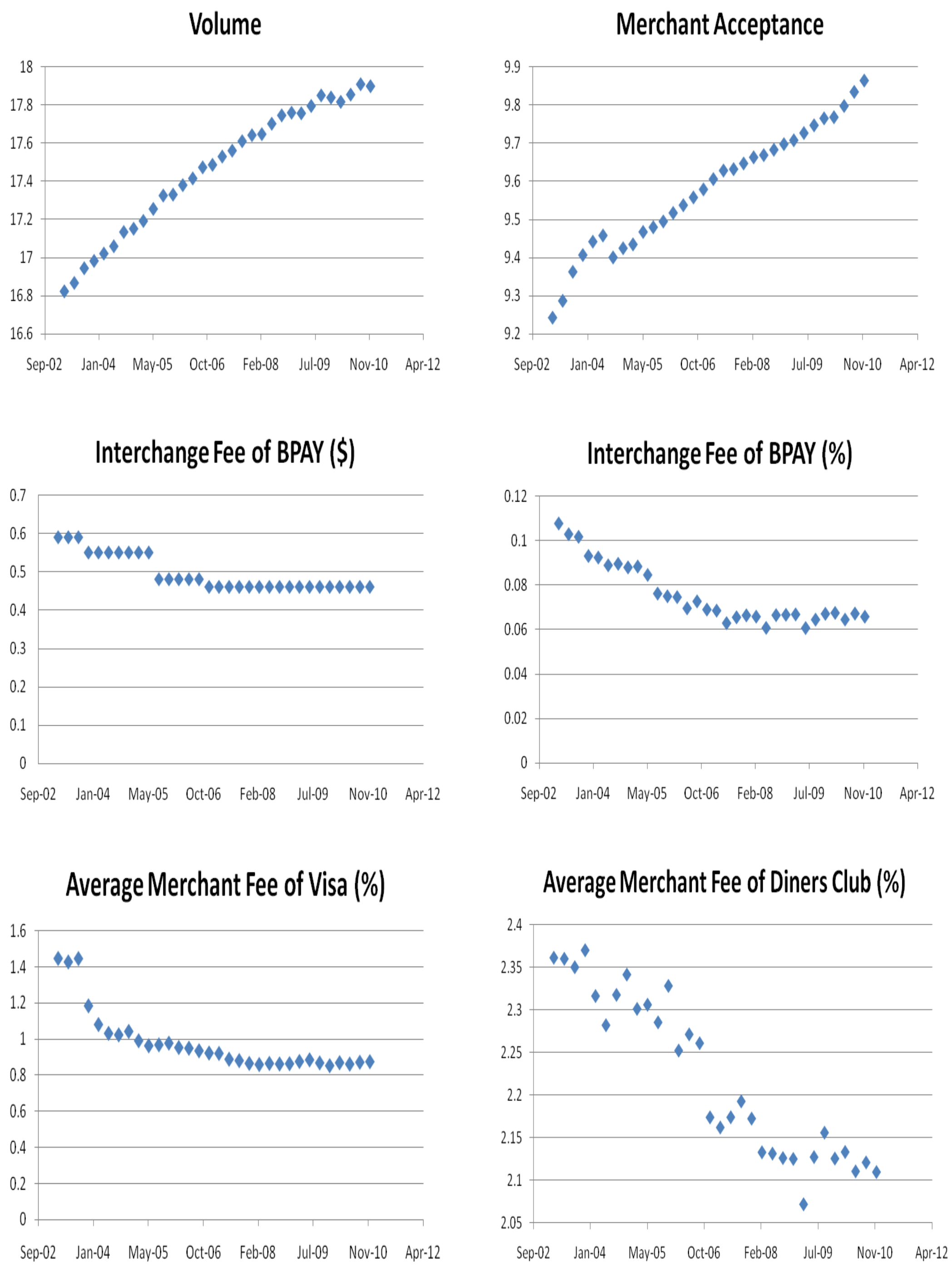
3.5 Results

3.5.1 Data Analysis

Prior to the implementation of unit root tests, the attributes of the time series data need to be determined. Figure 3.5.1 illustrates the behaviour of the five variables of interest from the consumer usage and merchant acceptance demand models of BPAY between March 2003 and December 2010. All variables in the sample are either upward trending or downward trending over time. The volume of transactions and merchant acceptance are increasing over time, in comparison with the vector of prices, which are decreasing over the same time period. Thus, all variables have either a long term increase or decrease over time, indicating non-stationary means.

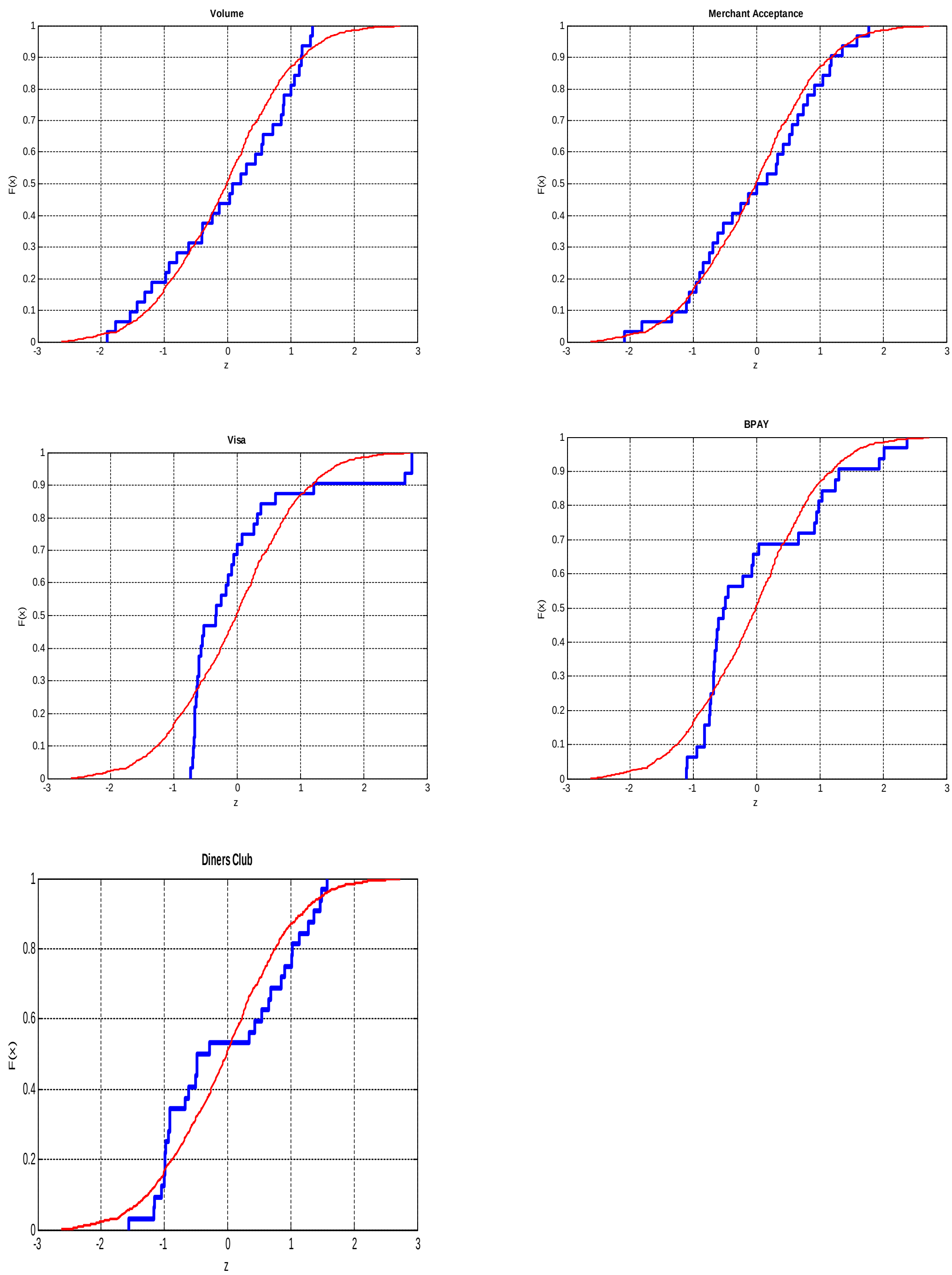
The decrease in the average merchant fee of Visa is a direct result of RBA intervention and pressure applied to four-party payment platforms to decrease the set interchange fee, and subsequently, reduce the cost imposed on merchants. However, as Diners Club is a three-party payment scheme that deals with merchants and consumers directly, the fall in its merchant fees can be directly attributed to competitive pressures to respond to the fall in merchant fees of Visa. The threat of merchants surcharging and rejecting the Diners Club payment method is reflected by the consistent decrease in merchant fees from 2.36% in March 2003 to 2.11% in December 2010. The interchange fee of BPAY varies four times over the sample period. In expressing the interchange fee as a percentage of the average value of a transaction in each respective quarter, the result is a steady decrease from March 2003 to June 2007. However, beyond June 2007 the interchange fee expressed as a percentage of the average value of a transaction has stabilised. Therefore, based on Figure 3.5.1, all univariate and panel unit root tests are applied with both a trend and drift term. Otherwise, all unit root tests are void. All variables have been transformed, as described in section 3.3.1.

Figure 3.5.1. Demand Equation Time-Series Variables.



3.5.1.1 Descriptive Statistics

Figure 3.5.1.1. Normalised cdf of All Time-Series Variables



To assess whether the variables in the respective demand functions were normally distributed, it is beneficial to graphically depict the standardised cumulative density function (cdf). Each variable was standardised and its cdf computed. Figure 3.5.1.1 plots the standardised cdf of each variable against the cdf of a standard normal random variable. The closer the blue curve (variable cdf) approximates the red curve (standard normal cdf), the greater the likelihood that the respective variable is normally distributed. Volume and merchant acceptance appear to be approximately normally distributed with their respective cdfs closely approximating that of the standard normal cdf. The probability of extreme outcomes for both variables is less than that given by a normal distribution, hence the distributions are expected to have negative kurtosis relative to the normal distribution. The skewness of both volume and merchant acceptance is also expected to be negative.

From the cdfs of volume and merchant acceptance, it can be seen that the distributions are centred above their respective means with the magnitude of the domain of the standardised values not symmetric around the origin ($z = 0$). Volume and merchant acceptance have a standardised domain of $-2 \leq z_v \leq 1.3$ and $-2.1 \leq z_m \leq 1.8$, respectively with $P(0 < z_v < 1.3) = P(-2 < z_v < 0) = 0.5$ and $P(0 < z_m < 1.8) = P(-2.1 < z_m < 0) = 0.5$. Therefore, relative to the normal distribution, there is negative skewness for volume and merchant acceptance as their respective distributions are centred above their means. The standardised cdfs of merchant fees for Visa, BPAY and Diners Club do not approximate the cdf of a standard normal random variable as well as that of volume of transactions and merchant acceptance. The distributions of merchant fees all appear to be asymmetric and have kurtosis that is different to that of the normal distribution.

Table 3.5.1.1. Descriptive Statistics.

Statistic	v_t	Δv_t	m_t	Δm_t	$\hat{p}_t$	p_t	$\tilde{p}_t$
Mean	17.46	3.35%	9.58	1.94%	0.98%	0.08%	2.22%
Std. Dev	33.35%	2.59%	16.10%	2.04%	16.84%	1.35%	9.51%
Jarque-Bera	2.44	1.47	1.04	55*	29.3*	4.9	3.25
Minimum	16.83	-0.02	9.24	-0.06	0.86	0.06	2.07
Maximum	17.91	0.08	9.86	0.08	1.45	0.11	2.37

Table 3.5.1.1 presents some descriptive statistics with volume and merchant acceptance expressed in logarithms and merchant fees as a percentage of the value of a transaction. Δv_t and Δm_t represent approximations of the respective growth rates of volume and merchant acceptance per quarter. Between March 2003 and December 2010, volume and merchant acceptance have increased at a average of 3.35% and 1.94% per quarter. Diners Club has had the highest merchant fee on average, followed by Visa and BPAY. This is consistent with a-prori expectations as Diners Club is not directly affected by the RBA intervention in regulating the interchange fee. Diners Club is a three-party platform and as such, sets merchant

fees directly. Therefore, Diners Club has been able to sustain high rewards for their consumers and this is reflected in high merchant fees. BPAY has the lowest interchange fee of the three platforms. This result for BPAY conforms with their strategy of offering no benefits to consumers using their platform. The lower the benefits made available by a platform to consumers per transactions, the lower the interchange fee charged and, consequently, the lower the merchant fee.

All variables with the exception of p_t, have high standard deviations. This attribute is desirable from a modelling perspective as the greater the variation in a variable, the more accurate its parameter estimate. The standard deviations of Δv_t and Δm_t are significantly lower in comparison with those of v_t and m_t. Therefore, modeling the demand equations via a cointegrating framework will induce greater information and accuracy in the estimated parameters.

BPAYs interchange fee is chosen to proxy for merchant fees. It is likely that scheme fees and charges (Figure 3.3.2) that the acquiring bank pays BPAY per transaction influences merchant fees and induces greater variation in merchant fees than interchange fees alone. However, such data on scheme fees and charges is proprietary in nature and is not available.

To assess normality, the Jarque-Bera (JB) test was employed. A requirement for inferential procedures in estimating a VAR is for its error term to be multivariate normal. A useful way in testing for multivariate normality is to assess whether each variable is normally distributed and if so, any linear combination of normally distributed random variables is also normally distributed and, as such, the assumption for normality is satisfied[2]. In Table 3.5.1.1 an asterisk denotes the rejection of the null hypothesis at the 5% level of significance (los). The test statistic is chi-squared distributed with two degrees of freedom with the critical value of the JB test at the 5% los being 5.99. All variables in the model appear to be normally distributed, with the exception of the merchant fees of VISA, where the null hypothesis of normality is rejected at the 5% los. This may indicate the error term from the VAR may not be multivariate normal. However, the JB test has poor finite sample properties, a characteristic of this data set. As a consequence, the JB test lacks size for this data which increases the probability of rejecting the null hypothesis when it is in fact true. Therefore, it is assumed the distribution of the error term is approximately multivariate normal, due to the small time dimension attached to the data set making make any normality testing procedure sensitive to type-1 error.

[2]Each variable being normally distribued is a necessary condition for multivariate normality but not a sufficient condition.

3.5.1.2 Correlations

Table 3.5.1.2. Correlation Matrix.

$\rho_{x,y}$	v_t	m_t	$\hat{p}_t$	p_t	$\tilde{p}_t$
v_t	1				
m_t	0.98	1			
$\hat{p}_t$	-0.85	-0.83	1		
p_t	-0.94	-0.89	0.92	1	
$\tilde{p}_t$	-0.95	-0.94	0.78	0.88	1

Table 3.5.1.2 presents the correlations between all the variables of interest. Network effects imply volume is highly correlated with merchant acceptance. Network effects is reflected by the results in Table 3.5.1.2 by observing the correlation between volume and merchant acceptance having a value of 0.98. Merchant fees have been decreasing over the sampling period due to intervention by the RBA and the competitive nature of the market. It is therefore not surprising to observe the high correlations between the merchant fees of the three respective platforms. The high correlations between the merchant fees of all platforms suggest, that the errors from panel unit root tests to evaluate whether the $(3T \times 1)$ vector $P = (p_t', \hat{p}_t', \tilde{p}_t')'$ contains a unit root, will be cross-sectionally dependent. Panel unit root tests that rely on the assumption of cross-sectionally independent errors are invalid if such an assumption is violated.

3.5.1.3 Univariate Unit Root Tests

Table 3.5.1.3. ADF and KPSS Unit Root Tests.

Variable	ADF	KPSS
v_t	3.13	0.2*
Δv_t	9.36*	0.16**
m_t	-2.42	0.06
Δm_t	-5.26*	*NA*
$p_{m,t}$	-1.77	0.2*
$\Delta p_{m,t}$	-5.79*	0.25*
$\hat{p}_{m,t}$	-2.44	0.17**
$\Delta \hat{p}_{m,t}$	-4.93*	0.11
$\tilde{p}_{m,t}$	-3.02	0.13
$\Delta \tilde{p}_{m,t}$	-5.85*	*NA*

Table 3.5.1.3 reports the test statistics from employing the ADF and KPSS test for each data series in the merchant acceptance and consumer usage demand equations. The Schwarz Information Criterion determined the number of lags to include in the respective tests. When appropriate, additional lags were used to handle serial correlation. The critical values for the ADF test at the 1% and 5% level of significance are -4.3 and -3.6 respectively. Whilst, the critical values for the KPSS test at the 1% and 5% level of significance are 0.22 and 0.15 respectively. A (*) and (**) represents a rejection of the null hypothesis at the 1% and 5% level of significance, respectively. The results indicate that all

variables are $I(1)$ based on the ADF test. However, results using the KPSS test imply only $\hat{p}_{m,t}$ contains a unit root. Hence, not all variables in the sample are found to be $I(1)$ across both univariate unit root tests.

The ADF test is more reliable in small samples in comparison with the KPSS test, hence more weight will be placed upon its conclusions to determine the time series properties of the data. As cointegration requires that the variables are integrated of the same order, based on the ADF tests, a cointegration vector may be found as all the variables in the consumer usage and merchant acceptance equations of (3.3.13) and (3.3.14) are of the same order. However, to gain more power in our testing procedure, we examine unit root tests in a panel data context by stacking the pricing variables, where $P_t = (p_{m,t}, \hat{p}_{m,t}, \tilde{p}_{m,t})$. The dimensions of this pricing panel is $n = 3$, $t = 32$.

3.5.1.4 Panel Unit Root Tests

Table 3.5.1.4 Panel Unit Root Tests.

Variable	*LLC*	*Baltagi*	*IPS*	*Fisher-ADF*	*Hadri*	*Pesaran*
P_t	-2.42*	0.9	-0.5	6.7	4.2*	-2.1
ΔP_t	*NA*	-6.6*	-7.67*	51.2*	3.2*	-4.6*

Results from the panel unit root tests on both the level of the price vector and first differences are presented in Table 3.5.1.4. An (*) and (**) represents a rejection of the null hypothesis at the 1% and 5% level of significance. All tests with the exception of Hadri (2000) have a null hypothesis of a panel unit root. All panel unit root tests with the exception of Hadri (2000) and LLC, do not meet the a-prori of all the levels of the variables being $I(1)$. This is in contrast to the Baltagi (2000), Fisher-ADF, IPS and Pesaran (2007) tests, which conclude the price vector, P_t, in the panel is $I(1)$ with the null hypothesis not being rejected, while the null hypothesis is rejected when P_t is expressed in first differences.

In order to determine whether the panel unit root tests that rely on the errors being cross-sectionally independent are valid, the errors need to be examined using a Breusch and Pagan (1980) (BP) test. From the results of Table 3.5.1.4 with two out of the six tests not concluding that the variable of interest is $I(1)$, it is expected that the errors will be cross-sectionally dependent.

To examine whether the errors are cross-sectionally independent, it is sufficient to evaluate the residuals of individual ADF tests of each pricing variable by employing a BP test. The test statistics are chi-squared distributed with 3 degrees of freedom. The critical value at the 5% level of significance is 7.82. The observed LM statistic of the BP test is 5.5, hence the null hypothesis of the errors being cross-sectionally independent is not rejected at the 5% los. The implications of the BP test on all panel unit root tests employed, with the exception of Pesaran (2007), is that the test statistics will converge to their asymptotic distribution, rendering the tests valid.

Four out of a possible six panel unit root tests indicate there is strong support for all the pricing variables being $I(1)$, with the null hypothesis not being rejected for all variables in their levels. On the other hand, the null hypothesis is rejected at the 5% level of significance when the pricing variable is expressed in first differences. Thus, there is support for all variables being $I(1)$, and the VECM can be estimated to evaluate whether a cointegrating vector exists in the respective demand equations.

3.5.2 Johansen Maximum Likelihood Procedure

The Johansen cointegration tests satisfied all the assumptions relating to the residuals being serially uncorrelated, homoscedastic and normally distributed. The lag length was chosen to maximise the degrees of freedom in estimation. Accordingly, the residuals were subject to a LM test for serial correlation. In the event that the residuals were serially correlated, additional lags were added to the model to solve the problem.

Table 3.5.2. Johansen Cointegration Tests

Demand Equation	λ_{Max} $H_0 : rank = r$			λ_{Trace} $H_0 : rank = r$		
	$r = 0$	$r = 1$	$r = 2$	$r = 0$	$r = 1$	$r = 2$
Consumer Usage	36.1*	18.9	9.6	66.4*	30.2*	11.3
Merchant Acceptance	85.5*	45.7	29.6	39.8*	16.0	14.7

Table 3.5.2 contains the Johansen cointegration test results for the respective demand equations. The cointegrating rank is determined by a sequential testing

procedure, which initially starts with examining $H_o : r = 0$. A large test statistic is evidence against the H_o and assuming it is rejected, $H_o : r \leq 1$ is evaluated. The process continues until H_o is not rejected and the cointegrating rank attained is the number of cointegrating vectors present in the VAR.

The critical values differ between the consumer usage and merchant acceptance demand equations due to the presence of dummy variables in the demand equation for consumer usage. The critical values for the λ_{Max} and λ_{Trace} test statistics at the 5% level of significance in testing for cointegration in the consumer usage demand equation are 27.6 and 47.9 for $H_o : r = 0$, 21.1 and 29.8 for $H_o : r \leq 1$ and 14.3 and 15.5 for $H_o : r \leq 2$ respectively. For the merchant acceptance demand equation, critical values for the λ_{Max} and λ_{Trace} test statistics at the 5% level of significance in testing for cointegration are 33.9 and 69.8 for $H_o : r = 0$, 27.6 and 47.9 for $H_o : r \leq 1$ and 21.1 and 29.8 for $H_o : r \leq 2$, respectively. An $(*)$ represents a rejection of the null hypothesis at the 5% level of significance.

The results indicate there is one cointegrating vector in the consumer usage demand equation based on the maximum eigenvalue (λ_{Max}) test statistic and two cointegrating vectors based on the Trace (λ_{Trace}) test statistic at the 5% los. At the 1% los, both test statistics suggest that there is only one cointegrating vector with a p-value of 0.045 associated with $\lambda_{Trace} = 30.2$. Hence the null hypothesis of more than one cointegrating vector can not be rejected. As the rejection of the null hypothesis that $r = 1$ is marginal based on λ_{Trace}, the confidence in rejecting the null hypothesis and the corresponding theory that there is only one cointegrating vector is minimal. The test based on the λ_{Max} does not reject the null hypothesis that $r = 1$. It does so with high confidence, with the p-value associated with the test statistic of 18.9 being 0.1.

The consequences of establishing cointegration are vital and considering the poor finite sample properties of the Johansen Maximum Likelihood Procedure, a more cautious approach in selecting the los is required. Therefore, a los of 1% seems more appropriate, thereby concluding there is one cointegrating vector in the consumer usage demand equation.

Based on the two test statistics, there appears to be a single cointegrating vector in the merchant acceptance demand equation. At both the 1% and 5% los, the null hypothesis that $r = 0$ is rejected based on the λ_{Max} and λ_{Trace} test statistics. In comparison, the non-rejection of the null hypothesis that $r = 1$ at the 1% and 5% los. Therefore, it can be concluded that there is evidence for a single long-run relationship between all variables in the consumer usage and merchant acceptance demand equations of (3.3.13) and (3.3.14).

3.5.3 Vector Error Correction Model

3.5.3.1 Consumer Usage

$$\begin{aligned}
\Delta v_t &= 0.02 + 0.034X_1 - 0.12e_{t-1} - 0.14\Delta v_{t-1} + \qquad (3.5.1) \\
& \quad (3.19) \quad (3.83) \quad (-2.15\ ) \quad (-0.89) \\
& \quad 0.34\Delta m_{t-1} + 0.012\Delta \hat{p}_{t-1} + 0.006\Delta \tilde{p}_{t-1} + \varepsilon_t \\
& \quad (1.46) \quad (0.14) \quad (0.05) \\
R^2 &= 0.54,\ \textit{LM-Stat} = 20.8, F_{H0:\beta_i=0\ \forall\ i} = 4.58
\end{aligned}$$

The ECM given by equation (3.5.1) partially presents the VECM estimated. For the sake of brevity, only the ECM of consumer usage is presented in this subsection. The estimated model is statistically valid, with errors being serially uncorrelated, the errors being normally distributed and the parameters being jointly statistically significant. Autocorrelation was tested up to three lags, by applying a LM test (Breusch–Godfrey test). The p-value associated with the observed test statistic of 20.8 is 0.19, thereby not rejecting the null hypothesis of the errors being independent at the 5% los. The critical F-statistic to test whether the model is empirically valid at the 5% los is approximately 2.55. The observed F-statistic is 4.58, thereby rejecting the null hypothesis and concluding that the model is empirically valid. In addition, there is no evidence of the errors being heterosckedastic. As the data set consists of multiple time series variables, observing that the errors are homoscedastic is expected. The model fits the data well with 54% of the variation in the growth of volume being explained by the model. Equation (3.5.2) below provides an autoregressive distributed lag model (ADRL) representation of equation (3.5.1).

$$\begin{aligned} v_t &= 1.76 + 0.34X_1 + 0.34m_t + 0.012\hat{p}_t + 0.006\tilde{p}_t - 0.27m_{t-1} \\ &\quad +0.116\hat{p}_{t-1} + 0.084\tilde{p}_{t-1} + 0.91v_{t-1} + 0.14v_{t-2} + \varepsilon_t \end{aligned} \tag{3.5.2}$$

In establishing that the model is empirically valid by observing that the errors follow the classical assumptions of linear regression and testing for the joint significance of the parameters, the parameters can be used to assess the expected impact of various variables on consumer usage. To account for the seasonality in the data, the model initially had three deterministic dummy variables that represent the March, June and September quarters. The base quarter was the December quarter. However, the t-statistics observed for the March and June quarters were 1.6 and 0.27 respectively, which implies at the 5% los that both the March and June quarters are no different to that of the December quarter. Thus, only the September dummy variable was included in the model as it attained a observed t-statistic of 3.83, which implies it is statistically significant at the 5% los with a p-value of approximately zero. Therefore on average, in the September quarter the volume of transactions is 3.4% greater than that of other quarters, holding all other factors constant.

The parameters of the variables that are expressed as rates of change provide the short-run impacts on consumer usage. A one percent increase in merchant acceptance is expected to increase volume by 0.34%, other things being held constant. Whereas a 1% increase in the benefits of transacting on the Visa or Diners Club platform is expected to increase volume by 0.012% and 0.006% respectively, other factors being held constant. The impact on volume from an increase in the benefits of an alternative platform is expected to be negative. However, as the magnitude of the increase in volume is small in magnitude, it can be argued in the short-run there is no impact on volume from a change in the benefits of alternative platforms as consumers take time to adjust to billing payment habits and to become members of the respective credit card platforms.

$$
\begin{aligned}
v_t &= 14.9 + 0.55 m_t - 1.07 \hat{p}_t - 0.75 \tilde{p}_t \qquad (3.5.3) \\
\alpha &= 0.12
\end{aligned}
$$

The long-run relationship between consumer usage, merchant acceptance and the benefits of transacting on the Visa and Diners Club platform is given by equation (3.5.3). All prior conditions regarding the signs of the parameters are satisfied. The first condition of network effects in the Bill Payments market is met with the parameter of merchant acceptance being positive. More specifically, a one percent increase in the merchant acceptance of BPAY is expected to increase consumer usage by 0.55% in the long run. As expected, the impact of alternative platforms to BPAY increasing the benefits to consumers has a long-run negative effect on volume. A one percent increase in the benefits to the usage of the Visa and Diners Club platform is expected to decrease consumer usage on the BPAY platform by 1.07% and 0.75% respectively, other things being held constant.

The results indicate that the growth of the BPAY platform has been driven by the actions of competing platforms. The magnitude of the parameter associated with Visa is almost twice that of merchant acceptance. The downward trend in interchange fees and, thereby, the fall in benefits to consumers, has had a positive effect on the volume of transactions placed with BPAY. The Bill Payments market appears competitive, with consumers being price sensitive to other platforms in the Bill Payment market, as reflected by the magnitude of the cross-price elasticities of VISA and Diners Club.

To increase future volume BPAY has two options; increase merchant acceptance, or provide benefits (additional value) to consumers relative to alternative platforms. Increasing merchant acceptance increases the demand of consumers to use the BPAY platform, as predicted by the network effects that are expected to exist in the Bill Payments market. It is important to observe that the cross-price elasticities of Visa and Diners Club for the consumers of BPAY can be considered symmetric, in that these are benefits relative to BPAY. For example, the RBA notes in the 2010 Payments System Annual Report that the average benefit provided to consumers that hold a standard credit card that offers rewards is approximately 0.67% of the value of the transaction. Assuming BPAY offers consumers modest benefits, such as 0.1% of the value of the transaction, the attractiveness of the Visa platform falls and is equivalent to Visa directly decreasing the benefits to consumers by 0.1%.

Figure 3.5.3.1. The Disequilibrium Error of Consumer Usage.

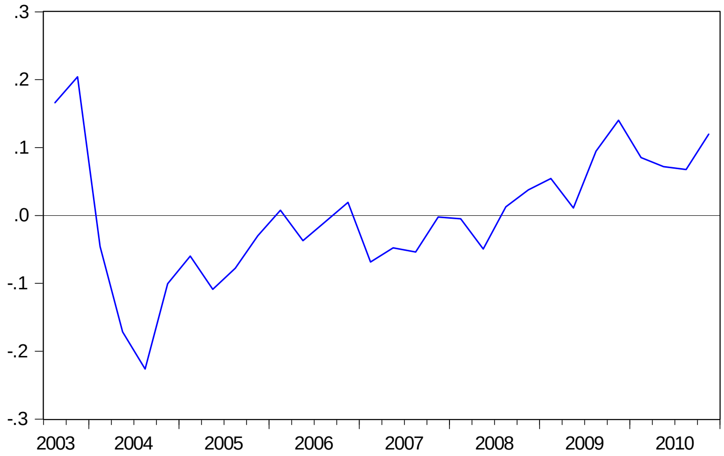


Figure 3.5.3.1 illustrates the disequilibrium error given by the long-run relationship of consumer usage ($e_t = v_t - (14.9 + 0.55m_t - 1.07\hat{p}_t - 0.75\tilde{p}_t)$). The adjustment coefficient is quite low at 0.12, meaning 12% of the disequilibrium error is made up in the following period. Hence, consumer usage does not adjust quickly back to equilibrium following deviations from the equilibrium relationship. Figure 3.5.3.1 indicates, that for the past several quarters, consumer usage has consistently been greater than what is predicted by the cointegrating relationship. It is expected that in future periods, the disequilibrium error will mean revert and consequently, consumer usage will be less than that predicted by the cointegrating relationship of equation (3.5.3).

3.5.3.2 Merchant Acceptance

$$\begin{aligned}
\Delta m_t &= 0.02 - 0.41e_{t-1} - 0.16\Delta m_{t-1} - 0.03\Delta v_{t-1} - 0.53\Delta p_{t-1} - \\
&\quad (3.2) \quad (-5.7) \quad (-1.03) \quad (-0.25) \quad (-0.64) \\
&\quad 0.18\Delta \hat{p}_{t-1} - 0.12\Delta \tilde{p}_{t-1} + \varepsilon_t \\
&\quad (-2.8) \quad (-1.34) \\
R^2 &= 0.65,\ \textit{LM-Stat} = 17, F_{H0:\beta_i=0\ \forall\ i} = 6.4
\end{aligned} \tag{3.5.4}$$

Equation (3.5.4) is the first of a system of equations estimated via a VECM. There is no evidence of autocorrelation and the model as a whole is valid, with the null hypothesis associated with $\beta_i = 0\ \forall\ i$ being rejected at the 5% los. Autocorrelation was tested up to two lags, by applying a LM test. The p-value associated with the observed test statistic of 17 equaled 0.84, implying the errors are not serially correlated. The critical F-statistic to test whether the model is empirically valid at the 5% los is 2.55, while the observed F-statistic given by equation (3.5.4) is 6.4. Hence the model is empirically valid. The data fits the model well with 65% of the variation in the growth of merchant acceptance being explained by the model. Equation (3.5.5) provides the ADRL representation of the merchant acceptance ECM above.

$$\begin{aligned}
m_t &= -1.91 - 0.03v_t - 0.53p_t - 0.18\hat{p}_t - 0.12\tilde{p}_t + 0.76v_{t-1} \\
&\quad -0.05p_{t-1} + 0.6\hat{p}_{t-1} + 0.32\tilde{p}_{t-1} + 0.43m_{t-1} + 0.16m_{t-2} + \varepsilon_t
\end{aligned} \tag{3.5.5}$$

The short-run changes in merchant acceptance are all negative to changes in volume and the merchant fees of BPAY, Visa and Diners Club. It is expected that a one per cent increase in consumer usage is expected to decrease merchant acceptance initially by 0.03%, other things being equal. An increase of one percent in the merchant fee of BPAY is expected to decrease merchant acceptance by 0.53%, other things being constant. Additionally, a one percent increase in the merchant fees of Visa and Diners Club are expected to decrease merchant acceptance by 0.18% and 0.12% respectively, holding all other factors constant.

$$\begin{aligned} m_t &= -4.69 + 0.75v_t - 0.27p_t + 0.53\hat{p}_t + 0.27\tilde{p}_t + \varepsilon_t \qquad (3.5.6) \\ \alpha &= 0.41 \end{aligned}$$

The long-run relationship between merchant acceptance, volume and the merchant fees of all platforms is given by equation (3.5.6). In the long-run, a one percent increase in volume is expected to increase merchant acceptance by 0.75%. Hence, as the coefficient of volume is positive, the second condition for the existence of network effects is satisfied in the Bill Payments market. Thus, it can be concluded that volume increases with merchant acceptance from equation (3.5.3) and merchant acceptance increases with volume, as shown by equation (3.5.6).

A second observation can be made in reference to the magnitude of the network effect parameters, given by equations (3.5.3) and (3.5.6). Currently, the allocation of costs between end-users in the Bill Payments market is weighted towards the merchant. The results given by the two demand functions of consumer usage and merchant acceptance support the decision of payment platforms to have merchants subsidise the cost of consumers using the platform. This is supported by the parameter of volume in equation (3.5.6) being greater than the parameter of merchant acceptance in equation (3.5.3). Therefore, consumers have a greater influence on merchant acceptance than does merchant acceptance on consumer usage. Thus, the bill payment platform is directed to entice consumers to use the platform in order to increase merchant acceptance. The platform does so by having merchants subsidise the cost of consumers using the platform. For example, BPAY has no usage and joining fees for consumers to pay their bills, however merchants are required to pay merchant fees and a cost for subscribing to the platform.

Its interesting to observe that the short-run effect of an increase in the merchant fee of BPAY is much greater than the long term effect of such a change. The immediate impact of increasing the merchant fee of BPAY by 1% is an expected decrease of 0.53%, other things being held constant. However the long-term change in merchant acceptance is a decrease of 0.27%. The increase in merchant acceptance after the initial fall from an increase in merchant fees suggest merchants react quickly to changes in the cost of payment instruments. This result may be explained by the expected network effects within the Bill Payments market and competition between merchants to attract marginal consumers.

Marginal consumers are a segment of a market that views merchants in a selected industry equally and differentiates according to the payment options available. In a competitive market, merchants price goods and services at similar prices that don't vary considerably across merchants. Merchants therefore try to attract consumers by offering their preferred payment method; thereby providing incentives for merchants that didn't offer BPAY to join the platform and capture those marginal consumers. Consequently, the merchant that withdraws from the BPAY platform may re-join the platform to attract those marginal consumers back to the merchant.

The cross-price elasticity of demand for merchant acceptance shares similar attributes to that of the price elasticity of BPAY. There is a clear distinction between the short-run and long-run impact on merchant acceptance from a change in merchant fees. The initial effect of a increase in the merchant fees of Visa and Diners Club is a fall of 0.18% and 0.12% in the merchant acceptance of BPAY. Whereas the long-run effect of a increase in the merchant fees of Visa and Diners Club is an increase in the merchant acceptance for the BPAY platform of 0.53% and 0.27%, respectively. This slight fall in demand may be in response by merchants to decrease the overall costs of payment instruments made available to consumers in the short term. However, merchants that may not have offered BPAY previously will have incentives to do so due to the cost of offering the more expensive alternative, VISA and Diners Club, has risen. Therefore, over the long-run, merchant acceptance will increase. From a competitive perspective, if a merchant in a competitive market does not offer a particular platform that others in the market are offering, the perceived value of that merchant declines and marginal consumers will purchase elsewhere. Thus, those platforms that initially left the BPAY platform due to the rise in the merchant fees of Visa and Diners Club may re-join the platform.

The price and cross-price elasticities provide a gauge to how competitive the Bill Payments market is. The magnitude of the cross-price elasticity of Visa is almost double that of the price elasticity of BPAY. This implies Visa is a dominant platform in the market for bill payments. Merchants in aggregate are therefore more sensitive to changes in pricing from Visa compared to that of BPAY in deciding whether to join the BPAY platform.

Figure 3.5.3.2. The Disequilibrium Error of Merchant Acceptance

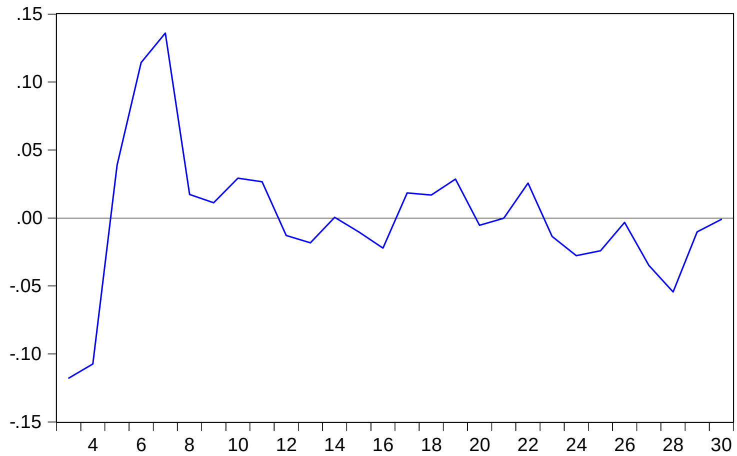


Figure 3.5.3.2 illustrates the disequilibrium error of the long-run relationship of merchant acceptance for the BPAY platform. The disequilibrium error is

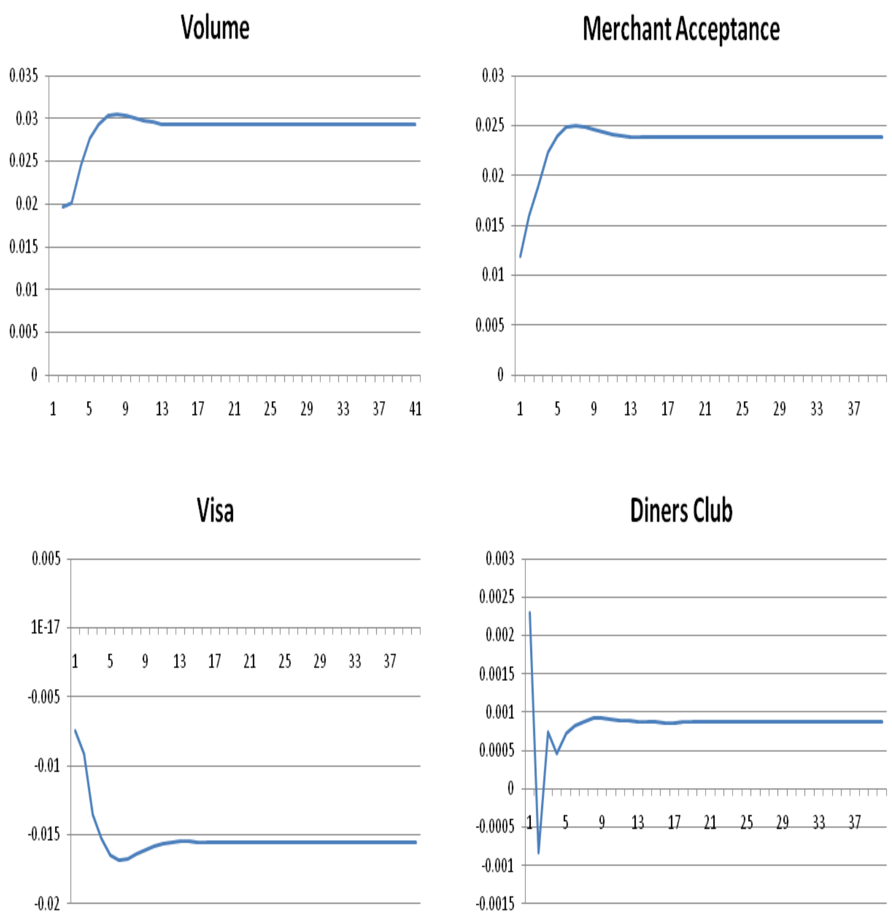
the difference between the actual merchant acceptance and that predicted by merchant demand equation (3.5.6). The adjustment coefficient is reasonably high at 41%, hence 41% of the disequilibrium error is made up in the following period. Thus, merchant acceptance adjusts quickly back to equilibrium following deviations from the equilibrium relationship. Between June 2006 and December 2010 the disequilibrium error has been quite low, fluctuating around zero. Thus implying, no shocks have occurred in recent times and the model has explained variations in merchant acceptance well.

3.5.4 Impulse Response Functions

Following the establishment of cointegration in the demand functions of consumer usage and merchant acceptance, impulse response functions (IRF) were estimated from their respective VECM. The IRF of consumer usage and merchant acceptance were estimated from the VECM of equations (3.5.1) and (3.5.4), respectively. The graphs to follow provide a visual representation of the partial lagged response of consumer usage and merchant acceptance to a unit standard deviation innovation from all variables that are contained in their respective demand functions, and themselves.

3.5.4.1 Consumer Usage

Figure 3.5.4.1. IRF of Consumer Usage.

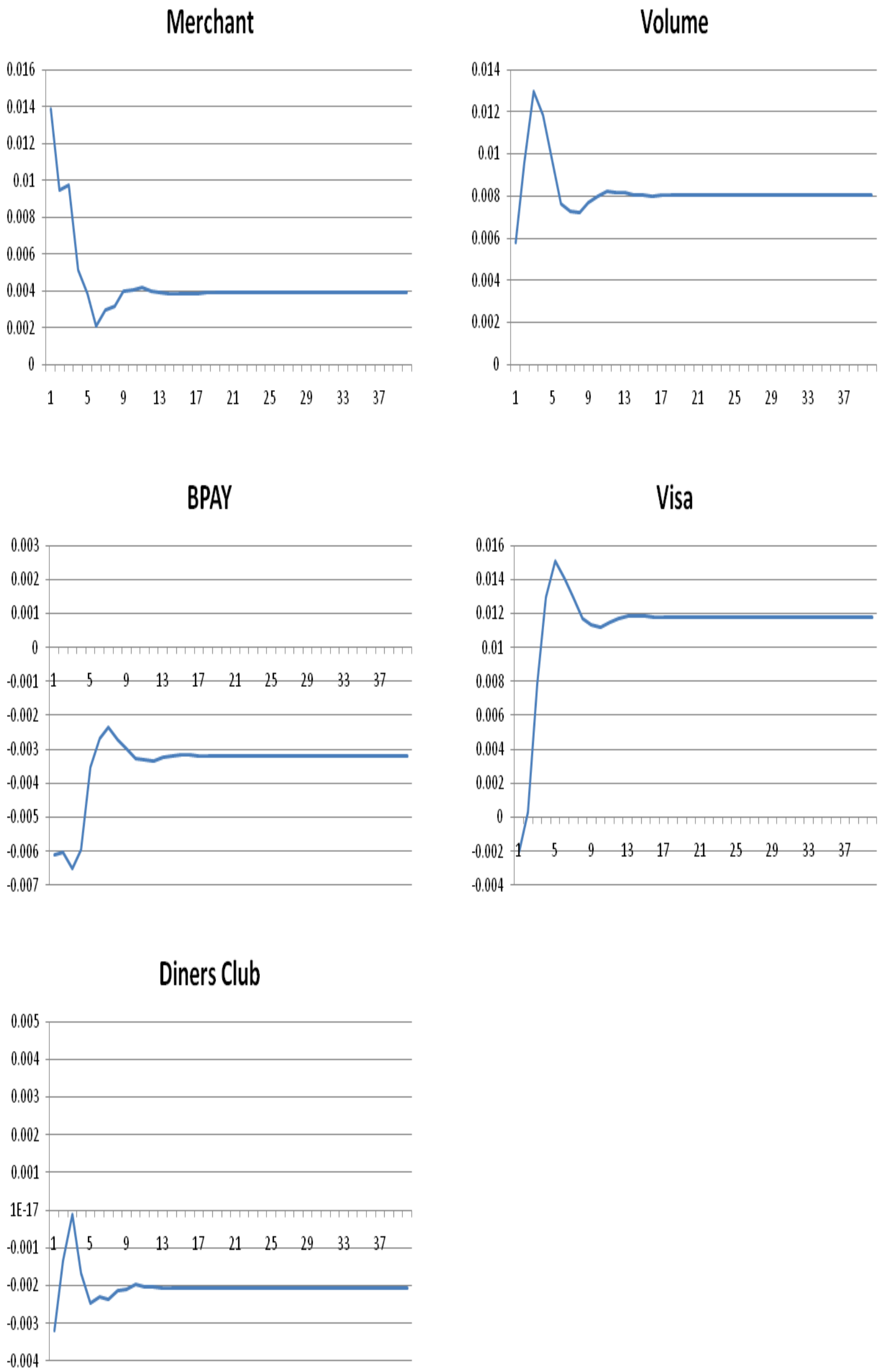


The impulse response functions of consumer usage to shocks to the variables that define its demand equation are shown in Figure 3.5.4.1. The effects of the majority of shocks on consumer usage die out after 12 quarters, as displayed by the flattening of the curves of all variables in Figure 3.5.4.1. Consumer usage displays a positive and permanent increase of approximately 3% from a shock to its own series. An innovation to merchant acceptance results in a long-run increase of 2.4%. More interesting is the response of consumer usage from a shock to the benefits of Visa and Diners Club. Consumer usage falls 1.7% over 8 quarters, before improving 0.13% over the next successive 6 quarters. The response of consumer usage to a shock to Visa then dies out after the 14^{th} quarter for an overall decrease of 1.57% in volume. This suggests consumers react strongly to the increase in benefits from Visa and act irrationally by placing a sub-optimal volume of transactions on the Visa platform. As the ability of consumers to realise the benefits relies on the capability of a consumer to be a transactor (not pay interest on required payments), consumers increase their usage of the BPAY platform after changing payment habits. In doing so, consumers decrease their minimum repayments on their credit card and benefit from the increase in rewards being offered by the Visa platform.

In contrast, a shock to Diners Clubs benefits cause consumer usage to rise, then fall sharply in the first two quarters. Consumer usage then improves for the following 12 quarters for an overall increase of 0.09%. The positive response to consumer usage from a shock to the benefits of Diners Club is against prior expectations, however, the magnitude of the increase is not substantially different from zero. This is consistent with the views of BPAY, who view Visa as their main competitor.

3.5.4.2 Merchant Acceptance

Figure 3.5.4.2. IRF of Merchant Acceptance.



The merchant acceptance impulse response functions to shocks to the variables that define its demand equation are shown by Figure 3.5.4.2. The effects of the majority of shocks on merchant acceptance become negligible after 16 quarters, as shown by the flattening of the curves of all variables in Figure 3.5.4.2. There is a immediate increase of 1.4% in merchant acceptance following an innovation to its time-series. Following the first quarter, merchant acceptance consistently falls from the initial increase for the next 8 quarters and has a long-run effect of increasing merchant acceptance by 0.4%. Over the first 3 quarters, merchant acceptance displays a sharp increase of 1.3% from a shock to consumer usage. The response of merchant acceptance to the shock of consumer usage then falls for the next 5 quarters before stabilising for an overall increase of 0.8%. Of more note is the response of merchant acceptance to shocks to the merchant fee of BPAY and its competitors in the Bill Payments market.

An innovation to the merchant fee of BPAY is consistent with the findings from the VECM; merchant acceptance decreases sharply following an increase to merchant fees, however, following this fall, merchant acceptance improves over the long-term for a smaller overall decrease in merchant acceptance. It decreases by 0.65% in the first 3 quarters following a shock then improves 0.4% between quarters 4 and 7, before stabilising for an overall decrease of 0.33%. Thus, there is support for the hypothesis that merchants will react too quickly to movements in the merchant fees of BPAY and leave the platform, and subsequently re-join to offer the platform to consumers. The response of merchant acceptance to an innovation to the merchant fees of Visa meets prior expectations with a positive overall response. Following the shock to the merchant fees of Visa, merchant acceptance rises sharply by 1.5% over 5 quarters, then stabilisers in successive quarters for an overall increase of 1.18% in merchant acceptance over the long term. Once again, the effect of a shock to Diners Clubs merchant fees is against prior expectations, with merchant acceptance falling. However, the long term effect of such a fall is small in magnitude.

3.6 Conclusion - Competition in the Bill Payments Market

This paper has developed a model for consumer usage and merchant acceptance to test for the presence of network effects in the Bill Payments market. Using data from the RBA and proprietary data provided by BPAY between March 2003 and December 2010, Johansen's maximum likelihood procedure was employed to estimate and test for a cointegrating vector in the demand models of consumer usage and merchant acceptance. In comparison to earlier studies, like that of Rysman (2007), the endogeneity that network effects imply in two-sided markets is accounted for by the vector autoregressive framework. Additionally, this is the first study to estimate price elasticities and cross-price elasticities of a platform in the Bill Payments market.

The results indicate the existence of a network effect between consumer usage and merchant acceptance in the Bill Payments market. Consumer usage is more valuable to the BPAY platform as consumers' effect on merchants is greater than that of merchant acceptance on consumer usage. There is also a clear distinction between short-run and long-run effects from changes to the variables that define the consumer usage and merchant acceptance demand models. The pricing elasticities indicate the Bill Payments market is competitive, with consumers and merchants reacting strongly to changes in the benefits and merchant fees. The next logical step is to investigate the drivers of transactions for consumers and merchant acceptance from a microeconomic perspective.

Chapter 4

Survival Analysis of BPAY Consumers

4.1 Introduction

The market for payments in Australia has undergone substantial change over the last decade. Central bank intervention, internet connectivity and technological advancement has accelerated the use and the growing importance of electronic payment methods. The market for bill payments is a derivative of the market for payments. The benefits of the adoption of electronic payment methods for the economy is documented in Humphrey et al. (2003). Cost savings and time costs associated with substituting paper based payment instruments with electronic methods, such as that of BPAY and direct debit, are significant and the speed of adoption is not uniform across different countries.

Milne (2006) develops a model that seeks to describe the differences in adopting technological advancement in the market for payments in Eastern European countries, United States of America and the United Kingdom. The take-up of a new technology requires banks to coordinate and share costs in establishing the platform. As such, what is predicted by the model is a highly concentrated banking market with coordination amongst the banks and shared profit from the efficiencies that arise. However, there has been insufficient research into the factors at the individual level that determine whether adoption of a new technology will be successful.

The motivation of this paper is to examine the risk factors of individuals transacting on an innovative technological bill payments platform. BPAY is a dominant participant in the Bill Payments market and is a service that is owned by the major banking institutions in the Australian market. A unique data set allows the transactions of an individual on the BPAY platform to be followed over time. Survival analysis is employed to examine to key demographics that contribute to individuals leaving the BPAY platform by applying semiparametric models of Cox (1972). In doing so, the hypothesis that technology adoption influences the probability of adopting an alternative technology is also evaluated, as in Hayashi and Klee (2003), by estimating the effect that having a credit card has on an individual leaving the BPAY platform.

4.2 Methodology and Modelling Framework

4.2.1 Data

Daily transaction level data was obtained from BPAY and the Commonwealth Bank of Australia (CBA) between May 2004 and October 2006. Every transaction on the BPAY platform is matched to the unique payer reference number such that a consumers payment patterns can be tracked over time. In addition, the demographics attached to the individual were supplied by the CBA. In order for an individual to be included in the data set, a CBA bank account must be held, otherwise the individual is not observed.

The data consists of over 100 million recorded transactions being made by approximately 1.8 million individuals. The transaction related data include the payment method, payment value and merchant. Demographic information of the individuals in the data include gender, date account opened, credit card ownership, location and the bank segment of which the individual is part of within the bank.

The characteristics of the data set are unique in the sense that an individual can be tracked for 30 months. In that time individuals firstly initiate the use of the BPAY platform by making their first payment, either by phone or internet banking. Of main interest is when an individual ceases to transact on the BPAY platform. To establish the context of the data to be examined, the characteristics of the Bill Payments market needs to be reflected in the data to ensure the survival analysis results are not misleading.

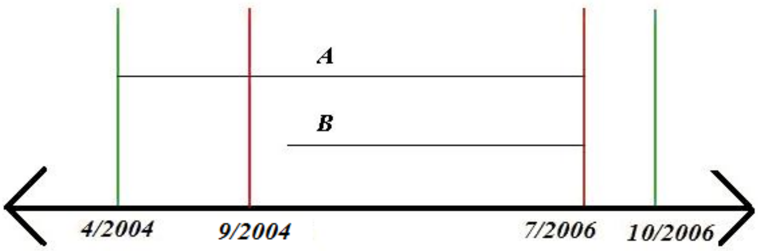
The Bill Payments market in Australia is highly concentrated. The platforms available to consumers include paying with cash at a post office, direct debit, credit cards and BPAY. The frequency of receiving a bill is dependent on the nature of the goods and services obtained. Infrequent bills include payments for council rates, parking fines, car insurance premiums and tax payments. Whereas, bills that consumers receive more frequently include telephone line rental, credit card usage and mobile phone carrier charges. Thus, to classify an individual as leaving the BPAY platform, a time period of at least three months needs to considered in the decision rule in order to classify such an individual. Accordingly, this is the classification period used in this paper. The frequency of the data to be considered is monthly. Consequently, the timing of the bill payment is highly accurate and the issue of interval censoring is not applicable to estimation.

Event-time analysis depends heavily on the order of the data and the time period attached to an individual is highly important. Given that an observation window of 30 months is available, the issue of left truncation needs to be addressed such that results are not overly biased. Left truncation arises from delayed entry and, as a result, length of time any individual has been using the BPAY platform cannot be accurately deduced. Hence, the ordering of the individuals based on their survival time renders event-time analysis subject to bias.

To overcome the issue of left truncation, it has been suggested in the literature to remove all those individuals that have the potential for delayed entry. The resulting hazard ratios of the Cox (1972) semiparametric model can then be considered unbiased, but there will be a loss of efficiency as some individuals will be removed from the data set. Given the lowest frequency a bill can be received is six months and to have results that are efficient, only those individuals that place their first transaction on the BPAY platform after the fifth month will be included in the data set and be analysed using survival analysis models and methods. An individual that ceases to use the BPAY platform over the last three months of the data is classified as having left the BPAY platform. Even though billing frequencies may be as long as six months, it is highly unlikely no other bills would be received with greater frequency.

Figure 4.2.1 illustrates the procedure of refining the data set to include only those individuals that fit the decision rule criteria. Both consumer "A" and "B" place their last transaction on the BPAY platform during July 2006 and be classified as having left the platform. However, only consumer "B" will be included in the data set as this consumer places their first transaction on the platform in the observational window. Consumer "A" places his first transaction at the beginning of the observational window. Even though it will be more efficient to include individual "A", it can't be determined that this is indeed their first transaction and, therefore, they are excluded from the analysis. All individuals that place their first transactions on the BPAY platform between April 2004 and August 2004 will be removed from the data set, resulting in a loss of efficiency in estimating the hazard parameters. Unfortunately, this loss of efficiency is unavoidable.

Figure 4.2.1. Data Selection



The introduction of systematically identifying new customers on the BPAY platform and customers that have left the platform is subject to two limitations. Firstly, an individual may hold multiple bank accounts. At any given time, an individual can choose to pay bills via the BPAY platform with a financial institution other than the CBA. Such an individual will be classified as having left the BPAY platform if the individual does not transact over the BPAY platform in the last three months. Secondly, the bank account of a given individual may be closed over the observational window. And lastly, an account may be idle if an individual becomes deceased during the time period. The limitations attached to the data set in identifying individuals that leave the BPAY platform causes the number of individuals estimated to have left the platform to be over-estimated.

However, given the large size of the data set, such resultant biases in the results will be minimal.

The motivation of the paper is to model the time individuals use the BPAY platform. The demographics attached to the individuals in the data allow the use of semiparametric models to be estimated to determine the influence of covariates on survival time. The Cox (1972) proportional hazard (PH) model will be used to model time to failure as provided by the following hazard function:

$$h(t, x, \beta) = h_0(t) \exp(x'\beta) \tag{4.2.1}$$

where $h_0(t)$ = Baseline hazard function

x' = Vector of covariates

β = Vector of regression coefficients.

Two alternatives of model (4.2.1) are used to model failure time. The first is a fixed Cox (1972) PH model incorporating only fixed covariates. The second model is a mixed Cox (1972) PH model incorporating both time-varying and fixed covariates. The fixed covariates and time-varying variables to be included in equation (4.2.1) are listed below:

- Age
- Gender
- State

- Banking segment. This includes retail, premium banking and wealth management
- Dummy variable indicating whether an individual holds a credit card with the bank
- Number of transactions processed in the current period.

Table 4.2.1. BPAY Consumer Classifications

Total Transactions	**Definition**
1	One-Off
1 to 8	Very Light
9 to 48	Light
49 to 120	Average
121 to 240	Medium
> 240	Heavy

Table 4.2.1 classifies the categorical variable that serves as the proxy for the number of transactions placed on the BPAY platform of individuals. The proxy for the number of transactions placed by individuals is utilised in the fixed model. Whereas, the number of transactions placed per month on the BPAY platform is incorporated in the alternative model to be estimated.

Individuals that placed a single payment on the BPAY platform are removed prior to modelling. One-off payers cannot be considered to be BPAY consumers as such behaviour suggests that the payer has no alternative payment platforms from which to pay their bill.

4.2.2 Survival Analysis

Survival analysis involves analysing the time to an event of interest. Modelling time to an event presents challenges to conventional regression methods in determining the impact covariates have on survival time. For example, consider the following ordinary least squares (OLS) approach to analyse survival time:

$$t_j = \beta_0 + \beta_1 x_j + \varepsilon_j,\ j = 1, 2, ..., n \tag{4.2.2}$$

$$t_j \sim N\left(\beta_0 + \beta_1 x_j, \sigma^2\right)$$

where t_j is time to an event and x_j is the sole explanatory variable. The assumption of normality in the errors is incorrect. Firstly, if the risk of the event occurring is constant over-time, an exponential distribution is more appropriate. Secondly, time to an event is strictly positive, whereas the normality assumption places a positive probability on negative values of time. Thus, an entirely different modelling framework is required in event time analysis.

The unique characteristics of event time analysis have developed models and methods that can be classified as either nonparametric, parametric or semi-parametric. Equation (4.2.2) is completely parameterised, the effect of time is assumed normal and the covariate variable, x_j, is a linear function of time.

Specifying different distributions for time results in different covariate parameter estimates in a parametric model, a cause for concern as parameter estimates associated with the explanatory variables is the primary interest in modelling event time data in this thesis. Whereas, semiparametric modelling imposes no restriction on the distribution of time and time serves no purpose other than the ordering of the data. The covariates in equation (4.2.2) however, are parameterised by assigning a functional form.

Nonparametric models make no assumption regarding the distribution of time and explanatory variables. For example, the Kaplan and Meier (1958) method to estimate the probability of survival past a certain time or to compare the survival probability of each qualitative covariate is popular in the literature. Nonparametric and semiparametric methods account for censoring and other characteristics unique to survival data.

In a parametric model, assuming the distribution for the time to event or failure time is correctly specified, the parameter estimates associated with the covariates are more efficient than the semiparametric or nonparametric approach. The time between failures is informative in a parametric model, whereas time simply orders the survival data in a semiparametric model. However, due to the risk of incorrectly specifying the distribution of failure time, a semiparametric approach is employed.

4.2.3 Describing the distribution of failure times

4.2.3.1 The Survival and Hazard Functions

Let T be a nonnegative random variable denoting the time to the event. Rather than referring to $T's$ probability density function, $f(t)$, or the cumulative density function, $F(t) = \Pr(T \leq t)$, it is convenient to refer to $T's$ survivor function.

The survivor function is simply the complement of the cumulative distribution function (cdf) of T:

$$\begin{aligned} S(t) &= 1 - F(t) \qquad (4.2.3) \\ &= \Pr(T > t) \end{aligned}$$

The survivor function reports the probability of surviving beyond time t, or the probability that there is no failure event prior to t. The function is equal to 1 at $t = 0$ and decreases to 0 as $t \to \infty$. The survivor function is a monotone, nonincreasing function of t.

The density function, $f(t)$, can be obtained from either $S(t)$ or $F(t)$:

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - S(t)) = -S'(t) \qquad (4.2.4)$$

The hazard function, $h(t)$, also known as the conditional failure rate, is the instantaneous rate of failure. It is the limiting probability that the failure event occurs in a given interval, conditional upon the subject having survived to the beginning of that interval, divided by the length of the interval with $1/t$ units.

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{4.2.5}$$

The hazard rate can vary from 0 with no risk at all, to ∞, with certainty of failure at that instant. Over time, the hazard rate can increase, decrease, remain constant, or take any unique shape through time. It is the underlying process of the event of interest that determines the shape of the hazard function.

Given one of the four functions that describe the probability distribution of failure times, the other three are completely determined. The cdf of the hazard function is given by:

$$\begin{aligned} H(t) &= \int_0^t h(u)\,du \\ &= \int_0^t \frac{f(u)}{S(u)} du \\ &= -\int_0^t \frac{1}{S(u)} \left\{ \frac{d}{du} S(u) \right\} du \\ &= -\ln\{S(t)\} \end{aligned} \tag{4.2.6}$$

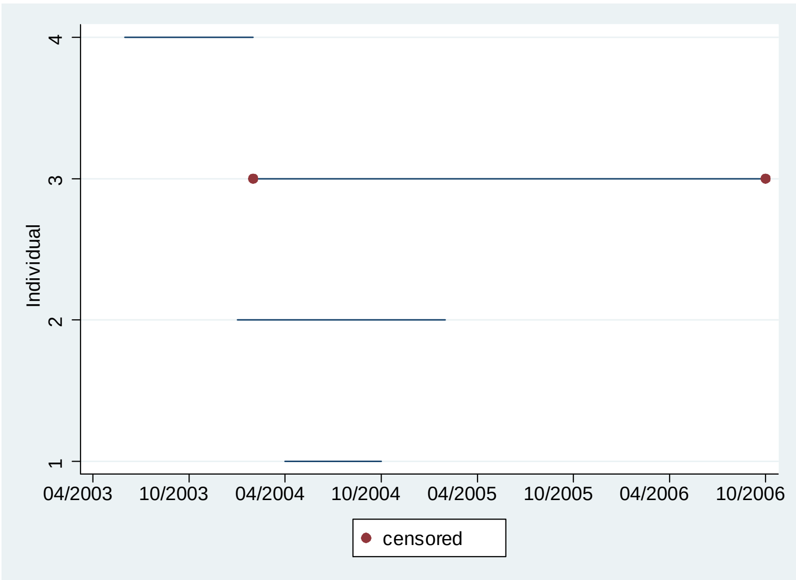
The cumulative hazard function measures the total amount of risk that has been accumulated up to time t. There is an inverse relationship between the accumulated risk and survival. We can now conveniently write:

$$\begin{aligned} S(t) &= \exp\{-H(t)\} \\ F(t) &= 1 - \exp\{-H(t)\} \\ f(t) &= h(t)\exp\{-H(t)\} \end{aligned} \tag{4.2.7}$$

4.2.4 Characteristics of Survival Time Data

Censoring and truncation are common characteristics of survival time data. To distinguish among the different types of censoring and truncation that apply to the BPAY consumer transaction data, it is important to define the onset of risk and the observational window.

Figure 4.2.4 (a). Snapshot of Consumers in Calender Time



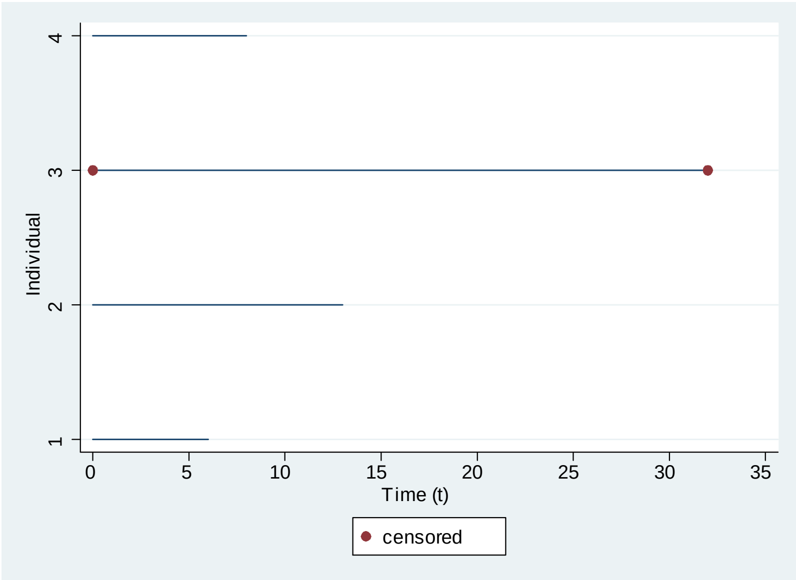
All transactions by CBA customers placed on the BPAY network are observed between April 2004 and October 2006. The onset of risk, $t = 0$, is defined as the moment an individual places his first transaction on the BPAY platform. After which, the individual is at risk of leaving the BPAY platform.

The four individuals captured by figure 4.2.4 (a) characterise the different types of censoring and truncation that is observed in the data set. The first individual places his first transaction on the BPAY platform in April 2004, and subsequently stops using the platform by October 2004. Then the first individual is fully observed in the data set. It is assumed for individual one that, if no further transactions are captured in the data set, they have left the platform. There may be instances of individuals changing financial institutions and being classified as leaving the BPAY platform. In these cases, being classified as right censored is more appropriate. However, information surrounding whether the departure of an individual from the study is attributed to competing events such as death, changing financial institutions, or merging bank accounts is unknown. Consequently, the study is limited in the sense that the rate of failure, number of individuals leaving the BPAY platform, will be over-estimated.

On the other hand, individual two's first transaction is not observed in the data set. That person's first transaction is placed on the BPAY platform in January 2004, before he/she leaves the platform during February 2005. Individual two is classified as being left-truncated as only those individuals whose event times lie within the observation window are completely observed. Individual three is the most common case represented in the data set. Individual three places his first transaction prior to April 2004, and continues to place transactions until the end of the study. The individual is both left-truncated and right censored as the event of interest, leaving the BPAY platform, is not observed prior to October 2006. Right censoring is most common in the data as an individual may enter at any time prior to October 2006, but are not observed beyond that point in time. The predetermined end time of the observation window is the primary cause of the large number of right censored individuals in the study. Individual four places their first and last transaction on the BPAY platform prior to the observational window and is not observed in the data set. Such cases are known as right-truncated data.

Figure 4.2.4 (b) denotes the analysis times, t, of all the individuals in figure (4.2.4). The onset of risk, $t = 0$, begins once the first transaction is placed on the BPAY platform. Classifying individuals in event time as opposed to calendar time is crucial in conducting semiparametric modelling. Arranging data in calendar time assigns explanatory power to time in the hazard functions of all individuals. It is plausible to assume that given a set of covariates of several individuals, their instantaneous risk of leaving the platform at any given point in time is constant within the 30 month observation window. In addition, the quantity and quality of the covariates available in the data set further diminishes the need of analysing the data in calendar time. Consequently, assigning a role to time.

Figure 4.2.4 (b). Snapshot of Consumers in Event Time.



4.2.5 Modelling Event Time Data

Modelling event time data is typically done in the following form:

$$h_j(t, x, \beta) = f\left(t, \beta_0 + x_j'\beta_x\right) \tag{4.2.8}$$

The hazard for individual j is some function $f(\cdot)$ of $\beta_0 + x_j'\beta_x$, with the hazard being systematically affected by several covariates via the row vector x_j', and the β_x column vector of regression coefficients. Alternatively, there is the option to model event times directly:

$$t_j = \beta_0 + x_j'\beta_x + \varepsilon_j$$

or

$$\ln(t_j) = \beta_0 + x_j'\beta_x + \varepsilon_j$$

There is a one-to-one mapping relationship from distributions to hazard functions. Consequently, the distributional assumption chosen for ε_j, implies the nature of the hazard function. However, event times are commonly modelled by equation (4.2.9) in order to incorporate semiparametric models:

$$h_j\left(t, x, \beta\right) = somefunction\left(h_0\left(t\right), \beta_0 + x_j^{'}\beta_x\right) \tag{4.2.9}$$

where $h_0\left(t\right)$ is called the baseline hazard function. That is, the hazard for individual j is $somefunction\left(\cdot\right)$ of the common baseline hazard of all individuals, modified by the set of covariates for individual j, x_j'.

4.2.5.1 Semiparametric Regression Models

A semiparametric model can be expressed as follows:

$$h\left(t, x, \beta\right) = h_0\left(t\right) r\left(x, \beta\right) \tag{4.2.10}$$

The baseline hazard function, $h_o\left(t\right)$, characterises how the hazard function changes as a function of survival time. Whereas, the function, $r\left(x, \beta\right)$, incorporates information about a set of covariates x and the resulting impact on the hazard function via the regression coefficients β and functional form $r\left(\cdot\right)$. The functional form, $r\left(\cdot\right)$, is chosen such that $h\left(t, x, \beta\right) > 0$. The choice of $r\left(\cdot\right)$ may depend on the characteristics of the data. Three specific forms detailed by Feigl and Zelen (1965) include:

1. $r\left(x, \beta\right) = 1 + x'\beta$,

2. $r\left(x, \beta\right) = \frac{1}{1+x'\beta}$

3. $r(x,\beta) = \exp(x'\beta)$.

The first two functional forms place restrictions on the value of regression coefficients, β, to ensure $h(t,x,\beta) > 0$. The third functional form ensures $r(x,\beta) > 0$ and places no undue restrictions on β. It is the functional form suggested by Cox (1972), as displayed below:

$$h(t,x,\beta) = h_0(t)\exp(x'\beta) \tag{4.2.11}$$

From equation (4.2.11), the ratio of the hazard functions for a pair of individuals with covariate values denoted by x_1 and x_0 is:

$$\begin{aligned} HR(t,x_1,x_0) &= \frac{h_0(t)\exp(x_1\beta)}{h_0(t)\exp(x_0\beta)} \\ &= \frac{h_0(t)\,r(x_1,\beta)}{h_0(t)\,r(x_0,\beta)} \\ &= \exp(\beta(x_0 - x_1)) \end{aligned} \tag{4.2.12}$$

The hazards of the pair of individuals are multiplicatively related and constant over time. The baseline hazard is cancelled in the numerator and denominator of expression (4.2.12). Consequently, expressing the hazard function as equation (4.2.11) is known as the proportional hazards or relative risk model. The Cox (1972) model is flexible in the sense it allows for the presence of fixed covariates, time-varying covariates and time-varying regression coefficients.

The Cox (1972) hazard function results in the following survival function:

$$
\begin{aligned}
S(t|x) &= \exp\left[-\int_0^t h(t|x)\,dt\right] \\
&= \exp\left[-\int_0^t h_0(t)\exp(x'\beta)\,dt\right] \\
&= \left\{\exp\left[-\int_0^t h_0(t)\,dt\right]\right\}^{\exp(x'\beta)} \\
&= S_0{}^{\exp(x'\beta)}
\end{aligned}
\tag{4.2.13}
$$

where S_0 is an arbitrary baseline survivor function corresponding to $x' = 0$. The above implies that:

$$
\ln[-\ln S(t|x)] = x'\beta + \ln[-\ln S_0] \tag{4.2.14}
$$

The logarithms of the negative logarithm of the survival functions of T, for a set of covariates x, are parallel. This relationship can be used to verify if the proportional hazard assumption is correct.

4.2.5.2 **Cox** (1972) **Proportional Hazard Model**

For simplicity, assume the covariate values of x_i are fixed. Let the event time, covariate and censoring variables be denoted as (t_i, x_i, c_i). The likelihood function is derived by isolating the impact of the vectors $(t_i, x_i, 1)$ and $(t_i, x_i, 0)$ on the likelihood function in a Bernoulli fashion. The former vector refers to individuals that have left the BPAY platform with exact survival time known to be t. Therefore, the likelihood function for these individuals will be given by the density function $f(t, \beta, x)$ for a given a set of covariate values, x_i. For the vector $(t_i, x_i, 0)$, the individual is censored and the survival time is at least t_i units. Their probability is given by the survival function $S(t, \beta, x)$. Assuming the survival times are independently and identically distributed, the full likelihood function is given by multiplying the respective contributions of the observed vectors. This is given by:

$$f(t, \beta, x)^{c_i} S(t, \beta, x)^{1-c_i} \tag{4.2.15}$$

where $c_i = 1$ for an individual that leaves the BPAY platform and $c_i = 0$ for a censored observation. As the observations are assumed to be independent, the likelihood function is the product of the expression (4.2.15) over the entire sample and is denoted by,

$$l(\beta) = \prod_{i=1}^{n} \left\{ [f(t_i, \beta, x_i)]^{c_i} [S(t_i, \beta, x_i)]^{1-c_i} \right\}, \tag{4.2.16}$$

or by the log likelihood function,

$$L(\beta) = \sum_{i=1}^{n} \{ c_i \ln [f(t_i, \beta, x_i)] + (1 - c_i) \ln [S(t_i, \beta, x_i)] \}$$

The monotone shape of the log function will yield the same vale for β as that for the likelihood function in equation (4.2.16). However, it is computationally efficient to maximise the log likelihood. The procedure to obtain values of the maximised log likelihood function (MLE) involves taking the derivative of (4.2.16) with respect to β, setting the partial derivative to zero and solving for β. Rearranging equation (4.2.5), yields:

$$f(t, x, \beta) = h(t_i, \beta, x_i) \times S(t_i, \beta, x_i)$$

It follows from equation (4.2.5) and (4.2.16) that:

$$L(\beta) = \sum_{i=1}^{n} \{c_i \ln[h_o(t_i)] + c_i x_i \beta + \exp(x_i \beta) \ln[S_0(t_i)]\} \tag{4.2.17}$$

Full information likelihood requires that we maximise (4.2.17) with respect to the unknown parameter of interest, β, and the unspecified baseline hazard and survival functions. Kalbfleisch and Prentice (2002) discuss in detail why it is not possible to use the log-likelihood function in (4.2.17). Cox (1972) suggests finding estimates of the value of β by forming partial likelihood functions. The definition and some properties of partial likelihood functions, along with the estimation of the survival function are explained in detail in Appendix C

4.2.5.3 Regression Diagnostics

Any model must be assessed to determine whether the assumptions underlying the model are satisfied and the model best fits the data. The PH model is no different, the primary assumption of constant proportional hazards over time must be satisfied, otherwise the results can be misleading. Other aspects of the model to be checked include examining the functional form of the covariates, the leverage on covariate coefficients of individuals in the study and the prediction accuracy of individuals leaving the BPAY platform.

The combination of data, model and likelihood framework utilised to estimate parameter values, β, causes the definition attached to a residual difficult in comparison with OLS or binary regression. For example in time to event analysis, for those individuals that are censored, there is no alternative for the observed minus predicted residual used as in OLS. There is no obvious method to define a residual in a PH model due to censoring and the use of partial likelihoods to obtain parameter estimates. As a consequence, several residuals are proposed by the literature to evaluate different measures of model adequacy in a PH Model.

In summary, Cox-Snell residuals are used to examine how the model fits the data overall. Score and scaled form residuals can be used to assess leverage of specific individuals on parameter estimation, while martingale residuals are useful in determining the optimal functional form of the covariates in the model. Scaled Schoenfeld residuals are commonly used to evaluate whether the proportional hazards assumption is correct.

Residuals Schoenfeld (1982) was one of the first to propose a residual to accompany the PH model. Schoenfeld residuals are based on the individual contributions to the derivative of the log partial likelihood. Assume there are n independent observations of time, p covariates and a censoring indicator provided by the vector (t_i, x_i, c_i), with $i = 1, ..., n$, and $c_i = 1$ for uncensored observations and zero otherwise. According to equation $(C.6)$, the derivative for the k^{th} covariate is expressed as:

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{j=1}^{k} c_i \left[x_{jk} - \frac{\sum_{\ell \in R(t_j)} x'_{\ell k} \exp\left[x'_\ell \beta\right]}{\sum_{\ell \in R(t_j)} \exp\left[x'_\ell \beta\right]} \right] \tag{4.2.18}$$
$$= \sum_{j=1}^{k} c_i \left[x_{jk} - \bar{x}_{w_i k}\right]$$

The estimator of the Schoenfeld residual for the i^{th} subject on the k^{th} covariate is derived from equation (4.2.18) by substituting the partial likelihood estimator of the parameter, $\hat{\beta}$:

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{j=1}^{k} c_i \left[x_{jk} - \frac{\sum_{\ell \in R(t_j)} x'_{\ell k} \exp\left[x'_\ell \hat{\beta}\right]}{\sum_{\ell \in R(t_j)} \exp\left[x'_\ell \hat{\beta}\right]} \right] \tag{4.2.19}$$
$$= \sum_{j=1}^{k} c_i \left[x_{jk} - \hat{x}_{w_i k}\right]$$

As the solution to β involves setting the value of equation (4.2.18) to zero, the sum of the Schoenfeld residuals is zero. Additionally, the Schoenfeld residuals are equal to zero for all individuals that are censored.

Grambsch and Therneau (1994) propose scaling the Schoenfeld residuals by an estimator of its variance and this results in a residual with greater diagnostic power than their unscaled counterparts. Let the vector of p Schoenfeld residuals for the i^{th} subject be expressed as $\hat{r}'_i = \left(\hat{r}'_{i1}, ..., \hat{r}'_{ip}\right)$, where $\hat{r}'_{ik}$ is the estimator provided by equation (4.2.19). Let the estimator of the $(p \times p)$ covariance matrix of the residuals for the i^{th} subject be denoted by $\hat{V}\left(\hat{r}_i\right)$, with the estimator missing if the individual is censored. The vector of scaled residuals is given by:

$$\hat{r}_i^* = \left[\hat{V}\left(\hat{r}_i\right)\right]^{-1} \hat{r}_i \tag{4.2.20}$$

Grambsch and Therneau (1994) suggest using an approximation to equation (4.2.20) in estimating scaled Schoenfeld residuals. The intuition underlying the approximation involves the observation that the distribution of $\hat{V}\left(\hat{r}_i\right)$ is relatively constant over time. This observation is true whenever the distribution of the covariates are similar in different risk sets. This assumption simplifies the estimator for the scaled Schoenfeld residuals and is provided below:

$$\hat{r}_i^* = mV\left(\hat{\beta}\right)\hat{r}_i \tag{4.2.21}$$

where m is the number of events.

The counting process formulation of time to event analysis derives the next set of residuals used in evaluating the PH model. Assume a sole individual with covariates x is observed from time zero. The counting process representation of the PH model counts whether the event occurs at time t and is denoted by:

$$N\left(t\right) = \Lambda\left(t, x, \beta\right) + M\left(t\right) \tag{4.2.22}$$

where $N\left(t\right)$ is the count that represents the observed part of the model that is equal to zero until the event occurs and equals one thereafter, $\Lambda\left(t, x, \beta\right)$ is the systematic component of the model and $M\left(t\right)$ is the error component. Thus, function $N\left(t\right)$ attains its maximum value at the end of follow-up and is equal to its censoring indicator variable. Hosmer et al. (2008) show the systematic component of the model is equal to the cumulative hazard at time t under a PH model. Hence, equation (4.2.22) can be expressed as the following:

$$N(t) = H(t, x, \beta) + M(t) \tag{4.2.23}$$

The function $M(t)$ in equation (4.2.23), under suitable mathematical assumptions, is a martingale and is the error component of the model with zero mean. The most appropriate time to compute the martingale residual is at the end of follow up and substituting the partial likelihood estimator of β yields:

$$\hat{M}_i(t) = c_i - \hat{H}\left(t, x, \hat{\beta}\right) \tag{4.2.24}$$

In an alternative derivation of the martingale residuals using the counting process approach, the next set of residuals is obtained. Assume for simplicity that there are tied event times and the value of the baseline hazard and cumulative baseline hazard at time t_i are provided by:

$$h_0(t_i) = \frac{c_i}{\sum\limits_{j \in R(i)} \exp\left(x_j'\beta\right)} \tag{4.2.25}$$

$$H_0(t_i) = \sum_{t_j \leq t_i} h_0(t_j) \tag{4.2.26}$$

It follows that the derivative in (4.2.18) may be expressed as:

$$\sum_{i=1}^{n} x_{ik}\left[c_i - H(t_i, x, \beta)\right] \tag{4.2.27}$$

The score residuals are obtained by transforming the martingale residuals of equation (4.2.27). The derivation for the score and scaled score residuals are provided in Hosmer et al. (2008). The computational formula for the vector of score residuals, $\hat{L}_i = \left(\hat{L}_{i1}, ..., \hat{L}_{ip}\right)$,is given by:

$$\hat{L}_{ik} = c_i \left(x_{ik} - \hat{x}_{w_i k}\right) - x_{ik} \hat{H}\left(t_i, x, \hat{\beta}\right) + \exp\left(x_i' \beta\right) \sum_{t_j \le t_i} \hat{x}_{w_j k} \frac{c_j}{\sum_{l \in R_j} \exp\left(x_l' \beta\right)} \quad (4.2.28)$$

The scaled score residual are defined as follows:

$$\hat{L}_i^{*\prime} = V\left(\hat{\beta}\right) \hat{L}_i \quad (4.2.29)$$

Evaluating the Proportional Hazard Assumption The PH assumption allows the covariates to completely describe the risk factors of the survival time of individuals without parameterising the effect of time. Assuming there is a binary covariate, x_j, the PH assumption describes the ratio of the hazards for $x_j = 1$ and $x_j = 0$ to be constant through time. It is the deviation from this constant relationship that is used in examining the PH assumption.

Grambsch and Therneau (1994) consider the following model to test the assumption of PH:

$$\begin{aligned} \ln\left[h\left(t, x, \beta\right)\right] &= \ln\left[h_0\left(t\right)\right] + x'\beta_j\left(t\right) \quad (4.2.30) \\ \beta_j\left(t\right) &= \beta_j + \gamma_j g_j\left(t\right) \end{aligned}$$

where $g_j\left(t\right)$ is some function of time. Equation (4.2.30) allows the effect of any covariate to vary over time. In the event that the PH is incorrect, the parameter γ_j in equation (4.2.30) is statistically significant and different from zero. As an alternative test, Grambsch and Therneau (1994) deduce that the scaled Schoenfeld residuals in equation (4.2.20) or their approximation have the following mean for the j^{th} covariate:

$$E\left[r_j^*\left(t\right)\right] \tilde{=} \gamma_j g_j\left(t\right) \tag{4.2.31}$$

Equation (4.2.31) suggests graphing the scaled Schoenfeld residuals against $g_j\left(t\right)$ can provide an informal test of the PH assumption. Assuming that PH is correct, the graph should have zero slope. Grambsch and Therneau (1994) propose a formal test by formulating a generalised least squares estimator of the coefficients and a score test of the hypothesis that $\gamma_j = 0$. A variety of functions for $g_j\left(t\right)$ have been suggested and these include $g_j\left(t\right) = \ln\left(t\right)$, $g_j\left(t\right) = t, g_j\left(t\right) = H_0\left(t\right)$ and $g_j\left(t\right) = rank\left(t\right)$. Simulated results in Quantin et al. (1996) and Ng'andu (1997) show the functional form given by $g_j\left(t\right) = \ln\left(t\right)$ yields a test with associated power that is as high as alternative functions of time.

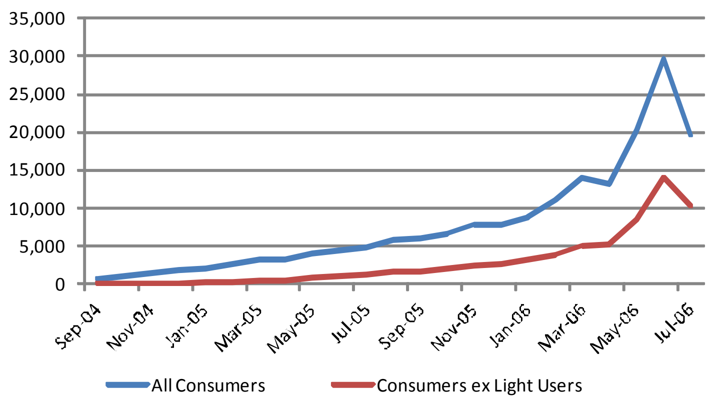
4.3 Results

A total of $581,063$ individuals remain in the data set once individuals are filtered, as explained in section (4.2.1) and one-off payers are removed. Overall, $10,334,634$ transactions were recorded between September, 2004 and July, 2006, while $179,922$ payers left the BPAY platform during the study.

4.3.1 Descriptive statistics

Figure[1] 4.3.1 (a) illustrates the rise in the number of consumers leaving the BPAY platform between September 2004 and July 2006. There is a consistent upward trend in the number of payers leaving the BPAY platform between September 2004 and March 2006. However, there is an substantial increase in the attrition of BPAY payers towards the end of the observational period. The number of payers that left the platform rose from $13,110$ in October 2004 to $29,723$ in June 2006.. A similar pattern of consumers leaving the BPAY platform when light users are excluded from the sample is also observed.

Figure 4.3.1 (a). Attrition of BPAY payers

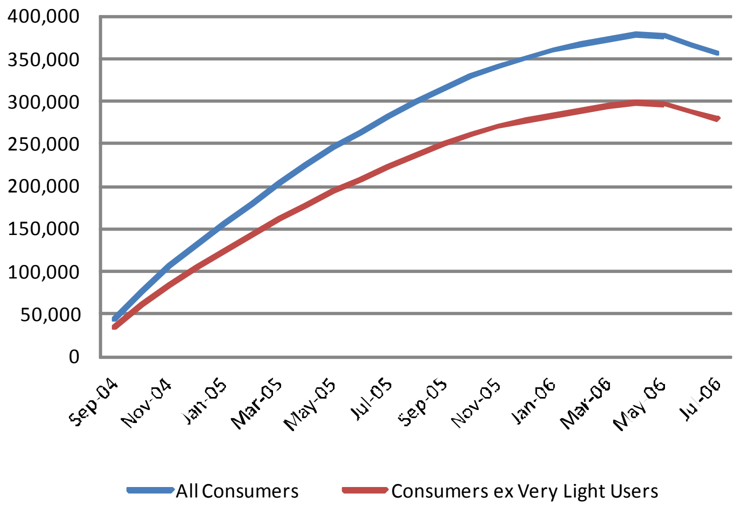


The upward trend in the number of payers leaving the BPAY platform can be explained by the increase in the number of consumers in the BPAY market. This is reflected in figure 4.3.1 (b), with the number of active payers in the sample increasing seven fold between April 2006 and June 2006. A similar pattern is

[1] Further descriptive statistics can be found in Section 5.3.1 in Chapter 5.

observed with all consumers excluding those belonging in the very light category with the number of active payers increasing from 34,000 to 280,000.

Figure 4.3.1 (b). Number of Active BPAY payers



The rate of attrition amongst active payers of the BPAY platform is reflected in figure 4.3.1 (c). The proportion of active BPAY consumers leaving the platform was relatively stable between September 2004 and January 2006, consistently between 1% − 2%. However, the rate of attrition amongst active BPAY payers started increasing rapidly beyond January 2006, reaching a peak of 8% in June 2006. This sudden increase may be just a one-off as it is also apparent in consumers excluding those belonging in the very light user category. Hence, the construction of consumers leaving the BPAY platform does not suffer from severe measurement error.

Figure 4.3.1 (c). Proportion of Active payers leaving the BPAY platform

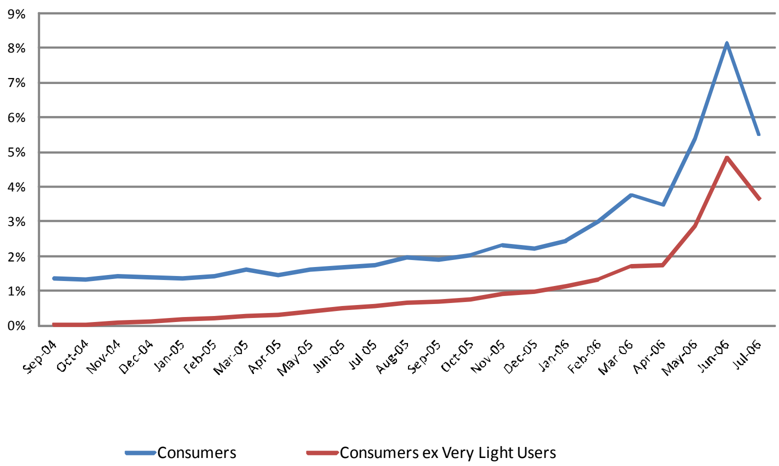
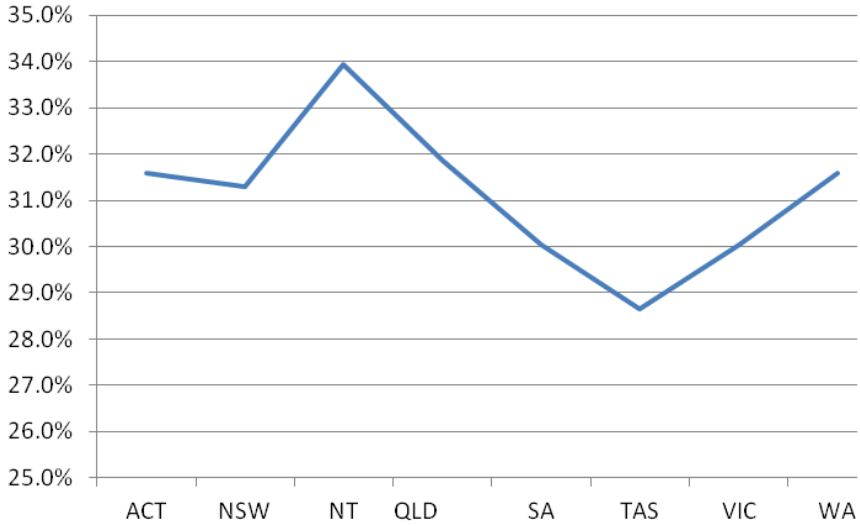


Figure 4.3.1 (d) denotes the rate that payers leave the BPAY platform between states. Northern Territory has the highest proportion of BPAY consumers that leave the BPAY platform with 34% of payers ceasing to further transact on the platform during the study. Tasmania, on the other hand, has the lowest rate of attrition with 28.6% of individuals leaving the BPAY platform. The remaining states have similar percentages of payers leaving the platform, with attrition rates of between 30% and 31.6%, respectively.

Figure 4.3.1 (d). Rate of Attrition of BPAY payers between States



The percentage of attrition amongst BPAY payers is shown by figure 4.3.1 (e). Unsurprisingly, there is an inverse relationship between usage of the BPAY platform and the percentage of attrition within consumer usage categories. Very light and light users have the highest rates of attrition with 49% and 22% of payers leaving the platform, respectively. While, average, medium and heavy payers of the BPAY platform have relatively low rates of attrition with 6%, 3% and 2%, respectively.

Figure 4.3.1 (e). Rate of Attrition of BPAY payers between Usage Categories

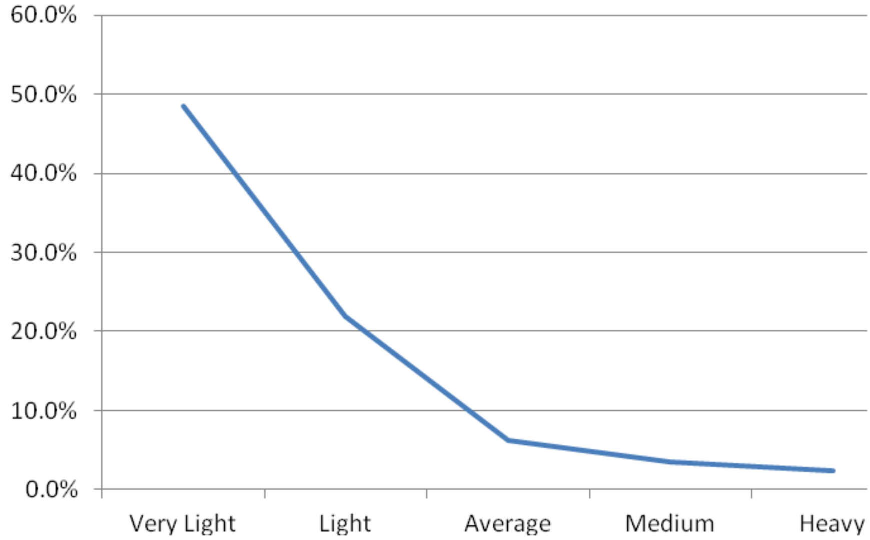
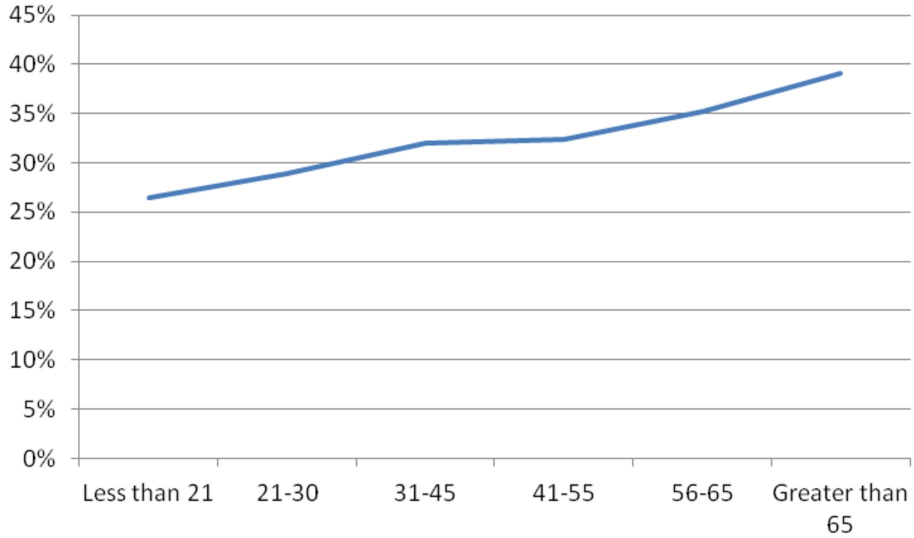


Figure 4.3.1 (f) depicts the relationship between age and the rate of attrition of BPAY payers. There is an upward trend between age and the percentage of payers within that given age category to leave the BPAY platform. Young BPAY users that are aged less than 21 have the lowest rate of attrition with 27%. Whereas, individuals aged $56 - 65$ and greater than 65 have the highest rates of attrition with 35% and 39%, respectively.

Figure 4.3.1 (f). Rate of Attrition of BPAY payers amongst Age Categories



Further descriptive statistics are provided in chapter 6.

4.3.2 Cox (1972) Proportional Hazard Model

Two Cox (1972) proportional hazard models are estimated and differ through the proxy used to identify consumer usage. The fixed and mixed Cox (1972) proportional hazard models can be defined by the following set of equations:

$$h(t,x,\beta) = h_0(t)\exp\begin{pmatrix} age_i\beta_1 + \Sigma_{j=1}^{4}\delta_j Usage_j + \Sigma_{k=1}^{7}\alpha_k State_j + \beta_2 M_i + \\ \Sigma_{m=1}^{2}\lambda_m Segment_m + \beta_3 Card_i \end{pmatrix} \tag{4.3.1}$$

$$h(t,x,\beta) = h_0(t)\exp\begin{pmatrix} \beta_1 Trans_{it} + age_i\beta_2 + \Sigma_{k=1}^{7}\alpha_k State_j + \beta_3 M_i \\ +\Sigma_{m=1}^{2}\lambda_m Segment_m + \beta_4 Card_i \end{pmatrix} \tag{4.3.2}$$

The base hazard, $h_o(t)$, differ by the transaction proxy used in equation (4.3.1) and (4.3.2). The base hazard for equation (4.3.1) comprises of the following:

$$\begin{aligned} age &= 15 \\ Usage &= \text{Very light payer on the BPAY platform.} \\ State &= \text{ACT.} \\ Gender &= \text{Female.} \\ Segment &= \text{Premium Banking.} \\ Card &= \text{No credit card.} \end{aligned}$$

The base hazard for the mixed model does not contain the variable usage. The base value for the transactions variable, $Trans_{it}$, is one. Otherwise, the base hazards of equation (4.3.1) and (4.3.2) are identical.

All variables were introduced and described at the start of section (4.2.1). The fixed model given in equation (4.3.1) uses consumer type, introduced in section (5.3.1.2), to incorporate a covariate for the number of transactions placed on the BPAY platform. Table 5.3.1.2 (a) in section (5.3.1.2) defines consumer types by total transactions placed on the BPAY platform. The greater the frequency of usage of the BPAY platform, the less likely a payer will leave the BPAY platform. Hence, it is expected that the parameter value associated with the very light dummy variable will be greater than that of the light payer, other things being held constant. This characteristic is expected to hold true as individuals are compared across usage categories, from very light to heavy.

The mixed model defined by equation (4.3.2) has a time-varying variable to incorporate consumer usage. It is hypothesised that the greater the number of transactions an individual places on the BPAY platform in any given month, the less likely the individual will leave the BPAY platform. Hence, the parameter value associated with the transaction variable is expected to be negative.

A-priori expectations concerning the value for the parameter value for age is positive. It is anticipated that younger individuals will appreciate the ease of use and convenience of paying bills through internet banking. The security concerns of younger individuals that have grown up adopting new technologies and embracing the internet into every aspect of their lives will be lower in comparison to that of older Australians. In addition, older individuals may not be aware of all the benefits of using BPAY (such as BPAY view) and having the option of scheduling bill payments as they may not be as 'tech savy' as the young.

The 2006 Census of Population and Housing report, commissioned by the Australian Bureau of Statistics (ABS), provides the expected parameter values of the location dummies in equations (4.3.1) and (4.3.2). It can be hypothesised that the greater the internet connectivity of the state in which an individual resides, the lower the parameter value associated with the location dummy variables. Findings in the ABS report suggests that the ACT had the highest internet connectivity with 75%. The ACT is followed by NSW, VIC, QLD and

WA with approximately 64% of the states having access to the internet. TAS had the worst rate of internet access with 55%. The ranking of states in relation to broadband connectivity is identical to that of internet access, with ACT and TAS being the most and least connected, respectively. Thus, it is expected that the parameter values in the Cox (1972) proportional hazard models will be all positive as ACT has the highest internet connectivity. Hence, the likelihood of an individual leaving BPAY will be lower for an individual that resides in ACT, in comparison with alternative states. In addition, it is expected TAS will have the highest parameter value as it is the least connected state to the internet.

The expected parameter values for the gender and banking segment is not as clear. Banking segment categories may proxy for wealth, with individuals in the wealth management category being the most wealthy, followed by individuals that are categorised in premium banking and retail payers. However, the contribution of wealth to risk of leaving the BPAY platform is not clear. It can be hypothesised that women generally like routine and are, therefore, more resistant than men in changing habits. Thus, it can be expected that the parameter value for the gender dummy variable will be positive as men are more likely to change billing payment habits than woman.

Individuals that own credit cards are expected to be less likely to leave the BPAY platform than individuals that don't have a credit card. Individuals with a credit card are more likely to pay their bills electronically and be more technologically savy than individuals with no credit card facility attached to their bank account. In addition, individuals with a credit card can be considered less likely to change banks, thereby decreasing the bias created from over-estimating the number of individuals leaving the BPAY platform.

4.3.2.1 Fixed Covariate Model

Table 4.3.2.1 (a). Fixed Model Proportional Hazard Regression Results

Variable	β_i	σ_{β_i}	z	$p > \|z\|$	95% C.I	
Age_i	0.02	0.0001	10.9	0.00	0.001	0.002
$Light_i$	−2.64	0.005	−508	0.00	−2.65	−2.63
$Average_i$	−4.64	0.017	−270	0.00	−4.67	−4.6
$Medium_i$	−5.4	0.071	−77	0.00	−5.57	−5.29
$Heavy_i$	−5.87	0.29	−20	0.00	−6.43	−5.3
NSW_i	0.07	0.16	4.5	0.00	0.4	0.11
NT_i	0.16	0.31	5.2	0.00	0.99	0.22
QLD_i	0.09	0.17	5.1	0.00	0.05	0.12
SA_i	0.11	0.19	5.8	0.00	0.07	0.15
TAS_i	−0.004	0.02	−0.2	0.86	−0.05	0.04
VIC_i	0.05	0.016	3.1	0.00	0.02	0.08
WA_i	0.13	0.018	7.3	0.00	0.09	0.16
$Male_i$	0.01	0.005	2.3	0.02	0.001	0.02
$Retail_i$	0.06	0.009	6.8	0.00	0.04	0.08
$Wealth_i$	0.85	0.04	20.5	0.00	0.77	0.93
$Card_i$	−0.1	0.005	−21.3	0.00	−0.11	−0.1

Table 4.3.2.1 (a) summarises the regression output of the Cox (1972) proportional hazards model defined by equation (4.3.1). All parameters are statistically

significant and different from zero at the 5% level of significance, with the exception of the dummy variable for TAS. The majority of the parameters satisfy the a-priori expectations established in section (4.3.2).

The estimated parameter for age is strictly positive, implying the older an individual, the greater the likelihood of leaving the BPAY platform, holding other factors constant. The usage frequency dummy variables estimated coefficients also adhere to a-priori expectations. There is an inverse relationship between total usage of the BPAY platform and the hazard rate of an individual, as reflected by the decrease of value in the estimated coefficients of the total usage dummy variable. For example, the estimated parameter for the total usage dummy variable Heavy is less than that for Medium, and so on.

Results from the 2006 ABS census suggested that the estimated parameters for all state dummy variables should be positive, as the ACT had the highest internet connectivity in 2006. The greater the internet connectivity, the greater the convenience and cost savings of consumers in obtaining the internet due to competition between telecommunication providers. All estimated parameter values for the state dummy variables are positive, with the exception of TAS. Estimated coefficient values of the state dummy variables imply residents in the NT and WA are most at risk of leaving the BPAY platform. The estimated parameters for NT and WA are the highest amongst all states with values of 0.16 and 0.13, respectively. Residents in the ACT and VIC are the least at risk of leaving the BPAY platform.

The estimated parameter estimate for gender satisfies the a-priori expectation that males are more likely to leave the BPAY platform. The positive value attained for the coefficient of the male dummy variable increases the instantaneous probability of an individual leaving the BPAY platform, in the case that an individual is male. However, the estimated parameters for segment do not satisfy a-priori expectations. Segment was thought as a proxy for wealth, however, the positive values attached to the coefficient of individuals categorised as of both retail and wealth management is conflicting.

The estimated coefficient of the card dummy variable satisfies the hypothesis outlined earlier in section 4.3.2. Individuals owning a credit should be more comfortable with using a payment platform like BPAY. They are more likely to use internet banking, of which BPAY is an extension of the existing services offered as part of online banking. The negative value attached to the coefficient of the credit card dummy verifies this hypothesis.

Table 4.3.2.1 (b) . Fixed Model Estimated Hazard Ratios.

Variable	**Hazard Ratio**		95% *CI*	
	All	Ex VL	Upper	Lower
Age_i	1.0018	$1,002$	1.0015	1.002
$Light_i$	0.071	NA	0.071	0.072
$Average_i$	0.01	0.13	0.009	0.01
$Medium_i$	0.004	0.06	0.004	0.005
$Heavy_i$	0.003	0.04	0.002	0.005
NSW_i	1.08	1.06	1.04	1.11
NT_i	1.17	1.21	1.1	1.24
QLD_i	1.09	1.08	1.05	1.13
SA_i	1.12	1.09	1.08	1.16
TAS_i	0.999	0.95	0.95	1.04
VIC_i	1.05	1.02	1.02	1.09
WA_i	1.14	1.14	1.1	1.18
$Male_i$	1.011	1.002	1.002	1.02
$Retail_i$	1.06	1.06	1.05	1.08
$Wealth_i$	2.33	4.8	2.15	2.53
$Card_i$	0.9	0.9	0.9	0.91

The hazard ratio and their corresponding 95% confidence interval of all consumers for the fixed Cox (1972) PH model are provided in Table 4.3.2.1 (b) of all consumers. For robustness, the hazard ratio's of the all consumers excluding very light (VL) users is also estimated. The hazard ratio provides an intuitive measure to quantify the risks of an individual leaving the BPAY platform. The hazard ratio for a variable is equal to the exponential of the estimated parameter listed in Table 4.3.2.1 (*a*).

The interpretation of the hazard ratio plays a similar role as the odds ratio in logistic regression. However, the hazard ratio is a measure of rates as opposed to odds in logistic regression. There is no material change in the interpretation of the hazard ratio between all consumers and the sample that excludes very light users of the BPAY platform. The 95% confidence interval of the majority of variables in the total sample includes the point hazard ratio parameter estimates of all variables in the sub-sample, with the exception of the wealth dummy variable. In addition, all parameters are statistically significant once very light users are excluded.

The hazard ratio for a consumer being categorised as a light user is 0.071. Such a light user can be interpreted as being 0.071 times more likely to leave the BPAY platform at any point in time in comparison with an individual who is a very light user of the BPAY platform. Alternatively, the hazard ratio of the individual that is classified as light can be interpreted as being 92.9% less likely to leave the BPAY platform at any point in time in comparison with an individual that is a very light user of the BPAY platform. The smaller the standard errors of the estimated parameters in Table 4.3.2.1 (a), the smaller the confidence intervals (CI) and the greater the accuracy of the hazard ratios listed in Table 4.3.2.1 (b). There is 95% confidence that a light user is between 92.8% and 92.9% less likely to leave the BPAY platform at any point in time in comparison to a very light user of the BPAY platform.

Residents located in the ACT have the lowest risk of leaving the BPAY platform at any point in time, as confirmed by the positive hazard ratios of all states and territories in Table 4.3.2.1 (b). In order of region, VIC, NSW and QLD, share similar hazard ratio values of between $1.05 - 1.09$. Hence, an individual residing in VIC is 5% more likely to leave the BPAY platform in comparison to an individual in ACT. Residents in the NT are the most likely to leave the BPAY platform with a hazard ratio of 1.17. Thus, individuals in the NT are 17% more likely to leave the platform in comparison to an individual in the ACT. The segment dummy variables indicate individuals being part of the wealth management arm of CBA are most at risk of leaving the BPAY platform. In relation to premium banking customers, individuals in the wealth management segment are 2.33 times more likely to leave the BPAY platform. Meanwhile, individuals that are part of the retail segment are 6% more likely than premium banking customers to stop using the BPAY platform.

The hazard ratio of variables that have a continuous scale can be interpreted in much the same way as categorical variables. The hazard ratio of 1.0018 for age implies that for a unit increase in age, the individual is 1.0018 times more likely to leave the BPAY platform at any point in time. Thus, the hazard ratio of 1.0018 for age suggests that for two individuals that have an x-year age difference, the older individual is 1.0018^x more likely to leave the BPAY platform. Additionally, the confidence intervals increases to the power of the age difference, x, in comparing two different individuals. For example, an individual that is aged 30 is 1.0933 times more likely to leave the BPAY platform at any point in time in comparison with an individual aged 25. Alternatively, an individual aged 30 is 9.3% more likely to leave the BPAY platform in comparison with an individual aged 25.

The standard error for the estimated coefficient for males is larger than that for individuals owning a credit card, as illustrated by the wider confidence intervals for the hazard ratio for males in relation to individuals owning a credit card. Other things being equal, males are 1.011% more likely to leave the BPAY platform at any point in time in comparison to females. In addition, there is 95% confidence that males are between 0.2% and 2% more likely to leave the BPAY platform at any point in time in comparison to female payers. Whereas, individuals holding a credit card are 10% less likely to leave the BPAY platform at any point in time in comparison with payers that don't hold a credit card. There is 95% confidence that individuals with a credit card are between 9% and 10% less likely to leave the BPAY platform at any point in time.

Testing the Proportional Hazard Assumption The assumption of proportional hazards is vital to the interpretation of covariates in the Cox (1972) hazards model. Two methods have been proposed to evaluate the assumption of proportional hazards. Firstly, scatterplots of the scaled Schoenfeld residuals are graphed versus functions of time for each covariate. Any linear trend that is apparent in the residuals indicates that the assumption of proportional hazards is incorrect. The second method to evaluate the assumption of proportional hazards is the procedure outlined by Grambsch and Therneau (1994). A series of score tests are suggested by Grambsch and Therneau (1994) to evaluate whether the proportional hazards assumption is incorrect by incorporating functions of time in the Cox (1972) hazards model, as outlined in section 4.2.5.3.

Figure 4.3.2.1. Scatterplot of Scaled Schoenfeld residuals for Age.

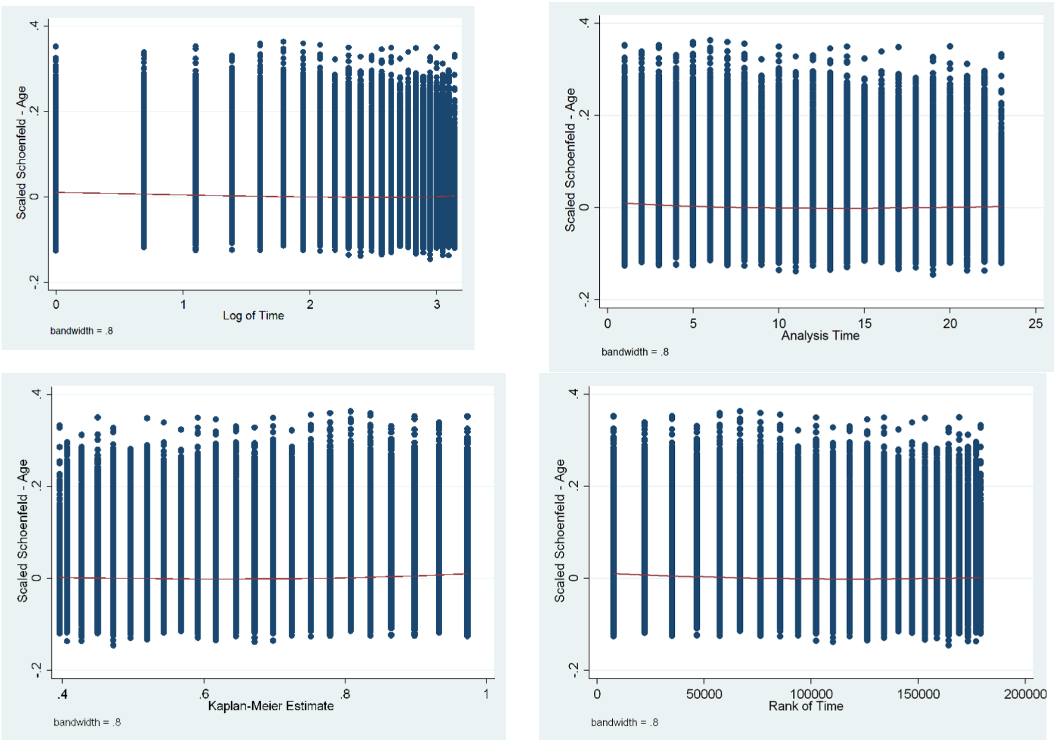


Figure 4.3.2.1 displays the scaled Schoenfeld residuals for age as a function of the logarithm of time, time, Kaplan-Meier estimate and rank of time. Any trend that is observed by the red curve in each plot of Figure 4.3.2.1 indicates there may be a violation of the proportional hazard assumption. All four plots of the scaled Schoenfeld residuals as a function of time do not appear to contain any substantial deviations from the proportional hazard assumption as the smoothed red curve in all four plots appears relatively horizontal, indicating no trend. However, Hosmer et al. (2008) notes that graphs are difficult to interpret to identify violations of the assumption of proportional hazards and any violations of the assumption can be subtle and hard to detect.

Table 4.3.2.1 (c). Score Test for Proportional Hazards

		$g(t)=t$		$g(t)=\ln(t)$		$g(t)=\hat{S}_{KM}(t)$		$g(t)=r(t)$	
Variable	**df**	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Age_i	1	29	0.00	431	0.00	252	0.00	305.3	0.00
$Light_i$	1	8404	0.00	6663	0.00	8370	0.00	8151	0.00
$Average_i$	1	1923	0.00	1053	0.00	1838	0.00	1624	0.00
$Medium_i$	1	123	0.00	65	0.00	118	0.00	103	0.00
$Heavy_i$	1	7	0.01	3.7	0.06	6.3	0.01	5.7	0.02
NSW_i	1	1.5	0.22	0.98	0.32	1.5	0.23	1.3	0.26
NT_i	1	3	0.08	2.4	0.13	3.1	0.08	2.9	0.09
QLD_i	1	5	0.02	4	0.045	5.2	0.03	5	0.03
SA_i	1	4	0.049	2.6	0.11	3.8	0.05	3.4	0.06
TAS_i	1	1.3	0.26	0.1	0.73	1.1	0.3	0.8	0.38
VIC_i	1	3.9	0.05	3.9	0.05	4.1	0.04	4.1	0.04
WA_i	1	4.5	0.03	3.1	0.08	4.4	0.04	3.9	0.05
$Male_i$	1	7.5	0.01	4.2	0.04	7.2	0.01	6.6	0.01
$Retail_i$	1	0.1	0.78	2	0.15	0.01	0.9	0.1	0.08
$Wealth_i$	1	3.8	0.05	7.3	0.01	3.8	0.05	5	0.03
$Card_i$	1	3.3	0.07	16	0.00	3.5	0.06	7.4	0.01
Global	16	10,044	0.00	8016	0.00	9983	0.00	9695	0.00

Table 4.3.2.1 (c) evaluates the score test based on the Schoenfeld residuals using four functions of time $\left[t, \ln(t), \hat{S}_{KM}(t), rank(t)\right]$. There appears to be evidence, for all four functions of time, of the hazard being not proportional in age and the indicators for consumer usage (Light, Average, Medium, Heavy) at the 5% level of significance (los). The p-value associated with the score test for both age and the consumer usage dummies are all less than 0.02, with the exemption of *Heavy* with a function of time given by $g(t) = \ln(t)$. Thus, there is evidence the null hypothesis of proportional hazards is rejected in age and consumer usage.

The score tests for the state dummy variable largely support the assumption of proportional hazards. With the exception of QLD, all state indicator variables have evidence for proportional hazards with at least two of the four score tests of each state dummy variable not rejecting the null hypothesis of the hazard being proportional. For example, NSW and NT have p-values of greater than 0.05 for all four specifications of time in the score test, hence the null hypothesis of proportional hazards can not be rejected at the 5% los.

Segment and the credit card indicator variables also support the assumption of proportional hazards with at least two of the four score tests not rejecting the null hypothesis of proportional hazards at the 5% los. The Retail dummy variable supports the assumption of proportional hazards with p-values of $0.78, 0.15, 0.9$ and 0.08 for the four score tests with time being specified as

$t, \ln(t), \hat{S}_{KM}(t), rank(t)$, respectively. The evidence for proportionality in the hazards for Wealth and Card dummy variables is not clear. The p-values associated with the proportionality test for Wealth are $0.05, 0.01, 0.05$ and 0.03, respectively. Hence, evidence for rejecting the null hypothesis of proportionality is marginal with the null hypothesis being rejected for two of the four tests. The results of hypothesis tests evaluating the assumption of proportional hazards for the Card dummy variable is similar to that of the wealth indicator variable. The p-values associated with the tests of proportional hazard for Card are $0.07, 0.00, 0.06$ and 0.01, respectively. Therefore, the assumption of proportional hazards in Card cannot be rejected with certainty at the 5% los, with two of the four tests do not reject the null hypothesis.

The global test for proportional hazards is rejected comfortably at the 5% los for all functions of time listed in Table 4.3.2.1 (c), with p-values of approximately zero in all four cases. The strong non-proportionality of age and the consumer usage indicator variables strongly influences this outcome. However, it is more important when considering tests of proportionality to consider covariate specific tests in order to identify the cause of the nonproportionality. The majority of the interpretations, therefore, hold true for all variables in the fixed model. However, interpreting the hazard ratio for age and the consumer usage dummy variables should be done with due care as the test suggested by Grambsch and Therneau (1994) reveals evidence of nonproportional hazards.

4.3.2.2 Mixed Covariate Model

Table 4.3.2.2 (a). Mixed Model Proportional Hazard Regression Results

Variable	β_i	σ_{β_i}	z	$p > \|z\|$	95% C.I	
$Transactions_{it}$	−0.28	0.002	−147	0.00	−0.28	−0.27
Age_i	0.006	0.0002	36	0.00	0.0057	0.006
NSW_i	−0.002	0.016	−0.15	0.88	−0.034	0.03
NT_i	0.09	0.031	3.1	0.002	0.034	0.15
QLD_i	0.02	0.17	1.3	0.19	−0.011	0.05
SA_i	−0.025	0.02	−1.3	0.19	−0.06	0.013
TAS_i	−0.1	0.022	−4.7	0.00	−0.15	−0.061
VIC_i	−0.047	0.016	−2.9	0.004	−0.08	−0.015
WA_i	0.025	0.018	1.1	0.16	−0.01	0.06
$Male_i$	0.03	0.005	6.8	0.00	0.023	0.04
$Retail_i$	0.04	0.01	4.6	0.00	0.024	0.06
$Wealth_i$	1.14	0.04	27.7	0.00	1.07	123
$Card_i$	−0.11	0.005	−24	0.00	−0.13	−0.11

The estimated parameters of the mixed model, along with the standard errors, p-values and 95% confidence interval (CI) are presented in Table 4.3.2.2 (a). In the mixed model, four of the seven state dummy variables are not statistically different from zero at the 5% los, with only TAS being statistically insignificant. Additionally, only the signs of NT, QLD and WA meet a-priori expectations of positive coefficients. From section 4.3.2, the ACT has the highest internet connectivity among states and territories and is the reference state for all location dummy variables. All location coefficients are expected to be positive as it is hypothesised the greater the internet connectivity, the greater the access of payers in using BPAY and the lower the cost of being connected to the internet. The indicator variables for gender, segment and credit card are all statistically significant at the 5%, los with parameters for gender and credit card meeting a-priori expectations in relation to the sign of the coefficient.

Table 4.3.2.2 (b). Mixed Model Estimated Hazard Ratios.

Variable	**Hazard Ratio**		95% CI	
	All	Ex VL	Upper	Lower
$Transactions_{it}$	0.756	0.9	0.75	0.76
Age_i	1.006	1.02	1.005	1.01
NSW_i	0.998	1.03	0.97	1.03
NT_i	1.1	1.17	1.04	1.17
QLD_i	1.02	1.05	0.99	1.06
SA_i	0.98	1.03	0.94	1.01
TAS_i	0.9	0.89	0.86	0.94
VIC_i	0.95	0.97	0.92	0.99
WA_i	1.025	1.1	0.99	1.06
$Male_i$	1.033	1.01	1.02	1.04
$Retail_i$	1.04	1.14	1.02	1.06
$Wealth_i$	3.15	5.3	2.9	3.41
$Card_i$	0.89	0.85	0.88	0.9

The hazard ratios of the parameters associated with equation (4.3.2) are provided in Table 4.3.2.2 (b) and correspond to all consumers being in the sample. The hazard ratio's corresponding to the sub-sample of individuals not in the very light (VL) category is also included. All parameters are statistically significant in the sub-sample. The parameters estimated of the sub-sample lie outside the 95%

confidence interval of the overall sample for a few variables, but the conclusions are identical to that of the overall sample that includes all consumers in the discussion that follows.

The time-varying variable, that is, the number of transactions in the current period, has a hazard ratio of 0.756. Thereby, for every additional transaction placed on the BPAY platform, the individual is 24.4% less likely to leave the BPAY platform. The confidence interval of transactions is also small, with a range of 0.01. This places more confidence on the parameter associated with transactions in the mixed model of equation (4.3.2). The hazard ratio of 1.006 for age implies that for a unit increase in age, the individual is 1.006 times more likely to leave the BPAY platform at any point in time. Thus, an individual that is aged 35 is 1.03 times more likely to leave the BPAY platform at any point in time in comparison with an individual aged 30. Alternatively, an individual aged 35 is 3% more likely to leave the BPAY platform in comparison with an individual aged 30. This compares with a 9% chance of leaving the BPAY platform in the fixed model.

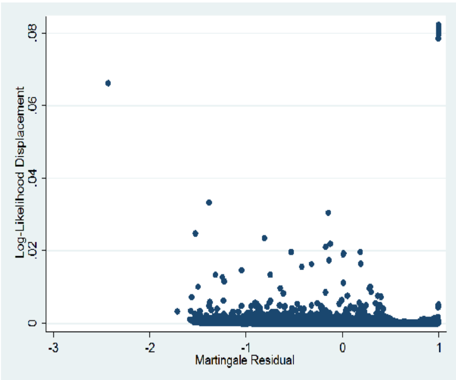
As was the case in the fixed model, the standard error for the estimated coefficient for males is larger than that for individuals owning a credit card. This is illustrated by the wider confidence intervals for the hazard ratio for males. Males are 3.3% more likely to leave the BPAY platform at any point in time in comparison to females. In addition, there is 95% confidence that males are between 2% and 4% more likely to leave the BPAY platform at any point in time in comparison to female payers. Whereas from 4.3.2.1 (*b*), the fixed model estimated a 1.1% increase in probability of males leaving the BPAY platform, an increase of three fold. However, there is minimal change in the likelihood of individuals holding a credit card leaving the BPAY platform in the mixed model in comparison to the fixed model. Individuals that have a credit card are 11% less likely to leave the BPAY platform at any point in time in comparison with payers that don't hold a credit card. Further, there is 95% confidence that individuals with a credit card are between 10% and 12% less likely to leave the BPAY platform at any point in time.

The hazard ratio for segment of 1.04 for retail suggests an individual that is a member of the retail segment is 4% more likely to stop using the BPAY platform to pay their bills in relation to a payer that is in the premium banking segment. Meanwhile, an individual that is in the wealth management segment is 3.15 times more likely to leave the BPAY platform in comparison with an individual within the premium banking segment.

4.3.2.3 Leverage and Goodness of fit

Figure 4.3.2.3.Log-Likelihood Displacement

(a) Fixed Model



(b) Mixed Model

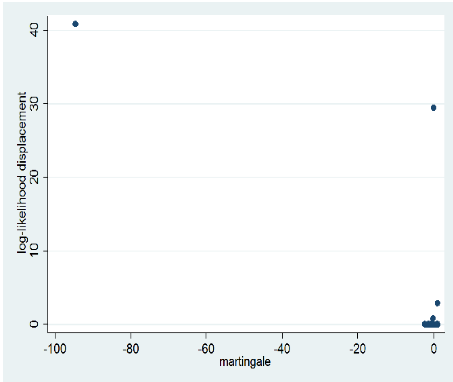


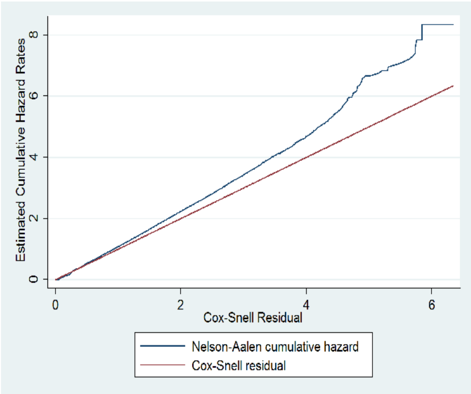
Figure 4.3.2.3 (a) illustrates the sample of residuals attached to all individuals in the data set and their accompanying impact on the log-likelihood of the fixed model given by equation (4.3.1). There appears to be a few outliers in the sample of $179,922$ individuals. No individuals were omitted from the sample following their identification as outliers. To determine whether individuals with high leverage had a material effect on the estimated parameters in the fixed model proportional hazard regression results (Table 4.3.2.1 (a)), the individuals that were identified as outliers in the sample were omitted and the fixed model was estimated once again. The outliers did not have a substantial effect on the parameter estimates.

The mixed model has only two outliers that are of note. However, the outliers in comparison to the fixed model are far greater in magnitude as shown by Figure 4.3.2.3 (b). They are expected to impact on the parameters estimated in Table

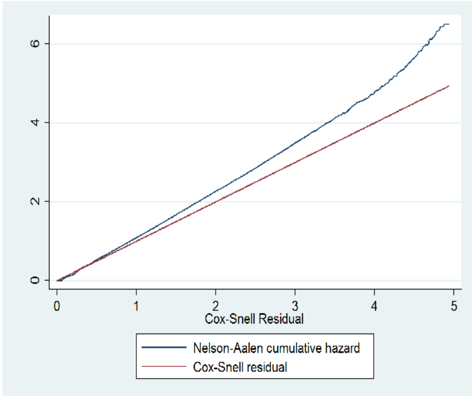
4.3.2.2 (a). The mixed model was estimated once more without the outliers and the parameters changed slightly. As a consequence, the outliers were omitted.

Figure 4.3.2.3. Cox-Snell Residual Plot

(c). Fixed Model



(d). Mixed Model



To examine goodness of fit, the Nelson-Aalen estimator of the cumulative hazard of residuals is estimated for both the fixed and mixed models. The better the fit of the model, the closer is the Nelson-Aalen cumulative hazard function to the 45 degree line. Both models appear to fit the data similarly. However, the fixed model has large variability in the tail of the estimate of the cumulative hazard function. Based on the Cox-Snell residual plots, the mixed model fits the data well with the Nelson-Aalen cumulative hazard function trending closer to the line of best fit.

4.4 Conclusion - Survival Analysis of BPAY Consumers

Technology is having an ever increasing role in altering the demand of several payment platforms throughout the globe. The demise of traditionally based payment platforms for their electronic counterpart has translated to substantial cost saving for the economy and improved the efficiency of payment systems. This thesis has sought to measure the factors that influence the rate of adoption of a recent technology-based product in the market for payments at the consumer level. This provides further evidence for the Hayashi and Klee (2003) finding that consumers are more likely to adopt new technology to pay bills if they have done so in the past.

By exploiting a unique data set which details the demographics and transactions of an individual over a 30 month observational window, survival analysis techniques are employed to quantify the risks of individuals leaving the platform. Cox (1972) models are estimated in the fixed and mixed framework with the transaction covariate varying in both instances to determine the most appropriate model. Results suggest support for the Hayashi and Klee (2003) finding in the market for bill payments with individuals having a credit card 10% and 12% less likely to leave the BPAY platform at any point in time. Also of note is males are approximately 3% more likely to leave the BPAY platform at any point in time, while there is a geographical influence on whether an individual adopts a innovative bill payments platform.

This thesis is the first of its kind to study the adoption of a new payments platform. Findings have implications for central banks and business as they try to provide incentives to consumers to change their payment platform preferences to more efficient platforms that seek to increase the welfare of all participants in the economy. There are limitations that may drive some bias in the results, mainly due to the difficulty in observing the correct entry and exit points of new payers. However, the benefits of exploring the transactions of individuals over time in a survival analysis framework outweigh any weakness in constructing the sample.

Chapter 5

Online Billing Payment Adoption

5.1 Introduction

The wide spread availability and integration of the internet in everyday life has increased the acceptance of electronic payment instruments in recent years and the adoption and use of electronic payments has provided many benefits to consumers and the economy as a whole. Cost savings associated with automating transactions and the tangible benefits of rewards, availability of credit and time saved associated with not being required to withdraw cash, has made electronic payments a popular alternative to traditional paper-based payment methods. This chapter seeks to determine the influence demographics and credit card holding on the adoption of the BPAY platform.

Regulators have the incentive to increase the share of transactions made via electronic means to improve the efficiency of the payments system. Bolt et al.

(2010) estimate approximately 50 Million Euros in cost savings to the overall payments system in the Netherlands can be achieved by abolishing surcharging on debit cards for transactions below the 11.6 Euro threshold. The authors argue the cost of processing a debit transaction has decreased over time as larger value transactions are placed on its network and such a surcharge does reflect the marginal cost of processing a debit transaction relative to cash. BPAY is an innovate platform that allows consumers to electronically pay their bills through internet or phone banking. The platform has enjoyed enormous success in the Australian bill payments market, enjoying a market share of more than 30%.

Literature exploring the adoption of new technology in the payments market to date has been limited. Hayashi and Klee (2003) suggest technology leads to other technologies. This study makes use of a proprietary data set which details the demographics of individuals and the transactions made on the BPAY platform over time. However, credit cards are a direct competitor to BPAY in the market for bill payments. Therefore, the exact influence of credit card holding on the adoption of a new technology-based platform is unknown. This study seeks to clarify this research question by empirically determining the effect of credit card holding on BPAY usage.

5.2 Methodology

To measure the influence that credit card membership has on usage, payers are classified into five categories that rank all individuals according to their usage of the BPAY platform. This specification is chosen as this is how BPAY classifies their consumers as they seek to measure their market for usage. Moving consumers up the category usage ranking will improve the volume on the BPAY platform. Hence an ordinal specification is most appropriate, although some information may be lost.

The dependent variable is thereby a categorical random variable and multinomial models and methods are most appropriate to utilise in this setting. Prior to considering multinomial models, it is beneficial to consider its simple alternative, that is, discrete response models for which the dependent variable is a binary random variable.

A linear probability model (LPM) can be used to estimate a discrete response model by assuming the random error term equals $(1 - x'\beta)$ with probability $x'\beta$, and $-x'\beta$ with probability $(1 - x'\beta)$. The LPM is specified as:

$$\Pr\left(y = 1|x\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + +\beta_k x_k \qquad (5.2.1)$$

The parameters of equation (5.2.1) can be interpreted as the change in probability of a chance event given a one-unit increase in the explanatory variable. In the event that the explanatory variable is a binary variable, the corresponding parameter estimate is the change in the probability when $x = 1$ in comparison to $x = 0$.

The linear probability model suffers from a few short comings, one of which is the issue of heteroscedasticity and this is demonstrated below:

$$\begin{aligned} V(y|x) &= E\left(y^2|x\right) - \left[E(y|x)\right]^2 \\ &= x'\beta - (x'\beta)^2 \\ &= x'\beta(1 - x'\beta) \end{aligned} \tag{5.2.2}$$

Equation (5.2.2) implies that the variance of the dependent variable is a function of the level of the covariates, violating the assumption of homoscedasticity under the assumption that the values of all covariates are non-zero. Thus, the linear probability model produces standard errors that are biased and inconsistent and can't be used for inference testing. The other main limitation of the linear probability model is that it assumes constant marginal changes in the dependent variable regardless of the level of the covariates in the model. However, such an attribute of a model with a dependent variable is not desirable as constant marginal changes will drive the predicted probability to be greater than one or less than zero.

5.2.1 The Probit and Logit Models

Consider a dependent variable, y, whose outcome is 1 if an event occurs and 0 otherwise. Examples of such binary variables include participation in the labour force, going on an overseas trip, or purchasing a new car. The value of the dependent variable, y, can be viewed as an outcome from an underlying latent variable model. Equation (5.2.3) below describes such a model.

$$y^* = x'\beta + e, y = 1\left[y^* > 0\right] \tag{5.2.3}$$

where x is matrix of covariates, β a vector of coefficients and e the random error term that is continuously independent and identically distributed around zero.

The value of y can be regarded as the observed outcome of a random variable, which takes a value of one when the index variable is greater than zero, and zero otherwise. Thus, the following can be deduced based on the specifications of the index model above:

$$\begin{aligned} E\left(y|x\right) &= 1 \times \Pr\left(x'\beta + e > 0\right) + 0 \times \left(1 - \Pr\left(x'\beta + e > 0\right)\right) \qquad (5.2.4) \\ &= \Pr\left(x'\beta + e > 0\right) \\ &= \Pr\left(e > -x'\beta\right) \\ &= \Pr\left(e \leq x'\beta\right) \\ &= G\left(x'\beta\right) \end{aligned}$$

The distribution of the error term in equation (5.2.4) determines whether a probit or logit model is formulated. A probit or logit model is derived by assuming the error term, e, follows either a standard normal distribution, or a standard logistic distribution, respectively. Thus, for a probit and logit model, this implies the following:

$$G(x'\beta) \equiv \Phi(x'\beta) \equiv \int_{-\infty}^{x'\beta} \phi(v)\,dv \tag{5.2.5}$$
$$\phi(x'\beta) = \frac{1}{\sqrt[2]{2\pi}} \exp\left(\frac{-(x'\beta)^2}{2}\right)$$

$$G(x'\beta) = \Lambda(x'\beta) \equiv \frac{\exp(x'\beta)}{1+\exp(x'\beta)} \tag{5.2.6}$$

where equations (5.2.5) and (5.2.6) refer to probit and logit models, respectively. The interpretation of marginal effects in the probit and logit models, in comparison with the LPM, differ greatly. The partial derivatives of the LPM and index models are given by:

$$\frac{\partial \Pr(y=1)}{\partial x_j} = \beta_j \tag{5.2.7}$$

$$\frac{\partial \Pr(y=1)}{\partial x_j} = g(x'\beta)\,\beta_j \tag{5.2.8}$$
$$\text{where } g(z) = \frac{\partial G}{\partial z}(z)$$

The marginal change in the probability of the LPM is given by equation (5.2.7). The estimated coefficient of the covariate gives the marginal change in the probability of the event occurring and is constant for all values for x_j. Whereas, equation (5.2.8) shows the partial derivative of x_j on $\Pr(y=1)$ varies in the probit and logit models. The presence of $g(x'\beta)$ in equation (5.2.8) causes the marginal change in the $\Pr(y=1)$ to be dependent upon the values of all other covariates. The probability density function for both the logit and probit models are positive for increasing values of x as the cumulative distribution function of both error term distributions are strictly increasing. Thus, whether the partial derivative will be positive or negative will depend upon the sign of the estimated parameter β, much like the LPM model.

5.2.1.1 Maximum Likelihood Estimation

Conditional maximum likelihood estimation is used to estimate the probit and logit model. The probability of an event occurring or not occurring can be conveniently written as:

$$f\left(y|x_i:\beta\right) = \left[G\left(x_i\beta\right)\right]^{y_i}\left[1-G\left(x_i\beta\right)\right]^{1-y_i} \tag{5.2.9}$$

for $y_i = 1$ or $y_i = 0$. The likelihood and log-likelihood functions for the logit model based on the density of y_i in equation (5.2.9) are given by:

$$\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} \left[G(x_i\beta)\right]^{y_i} \left[1 - G(x_i\beta)\right]^{1-y_i} && (5.2.10)\\
\log\left[L(\beta)\right] &= \sum_{i=1}^{n} y_i \ln\left[G(x_i\beta)\right] + (1-y_i)\ln\left[1 - G(x_i\beta)\right]\\
&= \sum_{i=1}^{n} y_i \ln\left[\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right] + (1-y_i)\ln\left[\frac{1}{1+\exp(x_i'\beta)}\right]\\
&= \sum_{i=1}^{n} y_i \ln\left[\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \div \frac{1}{1+\exp(x_i'\beta)}\right] + \ln\left[\frac{1}{1+\exp(x_i'\beta)}\right]\\
&= \sum_{i=1}^{n} y_i x_i'\beta + \ln\left[\frac{1}{1+\exp(x_i'\beta)}\right]\\
&= \sum_{i=1}^{n} y_i x_i'\beta - \ln\left[1+\exp(x_i'\beta)\right]
\end{aligned}$$

Equation (5.2.10) details the derivation of the log-likelihood function of the logit model. The score function and Hessian conditional on x_i is derived from equation (5.2.10) and are specified as:

$$s_i(\beta) = \frac{\partial \log\left[L_i(\beta)\right]}{\partial\beta} = \sum_{i=1}^{n} x_i y_i - \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \qquad (5.2.11)$$

$$\begin{aligned}
I(\beta) &= E\left(-\frac{\partial^2 \log\left[L_i(\beta)\right]}{\partial\beta\partial\beta'}\right) && (5.2.12)\\
&= \sum_{i=1}^{n} \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} x_i' x_i
\end{aligned}$$

The score function given by equation (5.2.11) is non-linear in β. The the Newton-Raphson, or the scoring method, are commonly used to find estimates of the vector, β.

The log-likelihood, score function and Hessian conditional on x_i of the probit model can be derived in a similar way by substituting the standard normal cumulative distribution function in place of the logit cumulative distribution function. These are summarised below:

$$\log\left[L\left(\beta\right)\right] = \sum_{i=1}^{n} y_i \ln\left[\Phi\left(x_i'\beta\right)\right] + \left(1-y_i\right)\ln\left[1-\Phi\left(x_i'\beta\right)\right] \tag{5.2.13}$$

$$s\left(\beta\right) \equiv \frac{\phi\left(x'\beta\right)x_i'\left[y_i - \Phi\left(x_i'\beta\right)\right]}{\Phi\left(x_i'\beta\right)\left[1-\Phi\left(x_i'\beta\right)\right]} \tag{5.2.14}$$

$$I\left(\beta\right) = \frac{\left[\phi\left(x'\beta\right)\right]^2 x_i'x_i}{\left\{\Phi\left(x_i'\beta\right)\left[\left[1-\Phi\left(x_i'\beta\right)\right]\right]\right\}} \tag{5.2.15}$$

5.2.2 Ordered Probit and Logit Model

The variable categorising the frequency of usage of BPAY payers can be classified as an ordered discrete random variable. A positive relationship exists between the value of the categorical variable and the frequency of BPAY usage. The model most appropriate in this study corresponds to an ordered response model.

Consider a latent variable, y^*, that represents BPAY usage:

$$y^* = x'\beta + e \tag{5.2.16}$$

where x, is a matrix of covariates, β the corresponding coefficients of the explanatory variables and assume $e \sim NID\left(0,1\right)$. Note that the matrix of covariates, x, does not contain a constant term. There is a identification issue relating to estimating the threshold parameters along with the constant term, hence the constant is omitted. In general terms, the thresholds of y^* and the accompanying values of y are written as:

$$\begin{aligned} y &= 0 \text{ if } y^* \leq \alpha_1 \qquad (5.2.17) \\ y &= 1 \text{ if } \alpha_1 < y^* \leq \alpha_2 \\ y &= 2 \text{ if } \alpha_2 < y^* \leq \alpha_3 \\ &\vdots \\ y &= J \text{ if } y^* > \alpha_J \end{aligned}$$

The conditional probability of y given x, based on expression (5.2.17), is specified as:

$$\begin{aligned} P(y=0|x) &= P(y^* \leq \alpha_1|x) = P(x'\beta + e \leq \alpha_1) = \Phi(\alpha_1 - x'\beta) \qquad (5.2.18) \\ P(y=1|x) &= P(\alpha_1 < y^* \leq \alpha_2|x) = \Phi(\alpha_2 - x'\beta) - \Phi(\alpha_1 - x'\beta) \\ &\vdots \\ P(y=J|x) &= P(y^* > \alpha_J|x) = 1 - \Phi(\alpha_J - x'\beta) \\ \therefore\ & P(Z_{ij}) = \Phi(\alpha_j - x_i'\beta) - \Phi(\alpha_j - x_i'\beta) \\ \text{where } Z_{ij} &= 1 \text{ if } y_i \text{ falls in the } j^{th} \text{ category and } Z_{ij} = 0 \text{ otherwise} \end{aligned}$$

Based on expression (5.2.18), the likelihood and log-likelihood functions of the ordered probit are given by:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \prod_{j=1}^{m} \left[\Phi(\alpha_j - x_i'\beta) - \Phi(\alpha_j - x_i'\beta)\right]^{Z_{ij}} \qquad (5.2.19) \\ \log[L(\beta)] &= \sum_{i=1}^{n} \sum_{j=1}^{k} Z_{ij} \log\left[\Phi(\alpha_j - x_i'\beta) - \Phi(\alpha_j - x_i'\beta)\right] \end{aligned}$$

The log-likelihood of the ordered logit can be formulated by replacing Φ with Λ, and with means and variances equal to $(0,1)$ and $\left(0, \frac{\pi^2}{3}\right)$, respectively. Prior to stating the first and second partial derivatives of the log-likelihood function of equation (5.2.19), note the following:

$$\begin{aligned}
Y_{i,j} &= \alpha_j - x_i'\beta \\
\phi_{i,j} &= \phi\left(\alpha_j - x_i'\beta\right) \\
\frac{\partial \Phi(x)}{\partial x} &= \phi(x) \\
\frac{\partial \phi(x)}{\partial x} &= -x\phi(x)
\end{aligned}$$

Kronecker delta: $\delta_{j,k} = 1$ if $j = k$ and 0 otherwise

The first and second partial derivatives of the log-likelihood function are stated below:

$$\begin{aligned}
\frac{\partial \log\left[L(\beta)\right]}{\partial \beta} &= \sum_{i=1}^{n}\sum_{j=1}^{m} Z_{ij}\frac{\phi_{i,j-1} - \phi_{i,j}}{\phi_{i,j} - \phi_{i,j-1}}x_i^{'} \qquad (5.2.20) \\
\frac{\partial \log\left[L(\beta)\right]}{\partial \alpha_k} &= \sum_{i=1}^{n}\sum_{j=1}^{m} Z_{ij}\frac{\delta_{j,k}\phi_{i,j} - \delta_{j-1,k}\phi_{i,j-1}}{\Phi_{i,j} - \Phi_{i,j-1}}x_i^{'}
\end{aligned}$$

$$\frac{\partial^2 \log\left[L\left(\beta\right)\right]}{\partial\beta\partial\beta'} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{Z_{ij}}{\left(\Phi_{i,j}-\Phi_{i,j-1}\right)^2} \times \begin{bmatrix} \left(\Phi_{i,j}-\Phi_{i,j-1}\right)\left(Y_{i,j-1}\phi_{i,j-1}-Y_{i,j}\phi_{i,j}\right) \\ -\left(\phi_{i,j}-\phi_{i,j-1}\right)^2 \end{bmatrix} x_i'x_i \tag{5.2.21}$$

$$\frac{\partial^2 \log\left[L\left(\beta\right)\right]}{\partial\beta\partial\alpha_k} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{Z_{ij}}{\left(\Phi_{i,j}-\Phi_{i,j-1}\right)^2} \times \begin{bmatrix} \left(\Phi_{i,j}-\Phi_{i,j-1}\right) \\ \times\left(Y_{i,j}\phi_{i,j}\delta_{j,k}-Y_{i,j-1}\phi_{i,j-1}\delta_{j-1,k}\right) \\ -\left(\phi_{i,j}-\phi_{i,j-1}\right)\left(\phi_{i,j}\delta_{j,k}-\phi_{i,j-1}\delta_{j-1,k}\right) \end{bmatrix} x_i'$$

The implied assumption in the ordered logit and probit model is that the parameters of the covariates are homogenous across all values of y_i. The parallel regression assumption can be demonstrated as follows by stating the probabilities that $y \leq z$:

$$\begin{aligned} \Pr\left(y \leq 1|x\right) &= F\left(\mu_1 - x'\beta\right) \\ \Pr\left(y \leq 2|x\right) &= F\left(\mu_2 - x'\beta\right) \\ &\vdots \\ \Pr\left(y \leq z|x\right) &= F\left(\mu_J - x'\beta\right) \end{aligned} \tag{5.2.22}$$

Equation (5.2.22) illustrates that an ordered probit or logit model is equivalent to $(J-1)$ binary regressions with all the coefficients of the explanatory variables homogenous across the all $(J-1)$ regressions. Hence for the parallel regression assumption to hold and an ordered probit or logit model to be suitable, all k coefficients across $(J-1)$ binary regressions should be relatively equal.

Brant (1990) develops a procedure based on the Wald test that evaluates the parallel regression assumption on each covariate separately. In the event that the parallel regression assumption is not true, alternative models need to be considered that allow for parameters to differ across alternative values of y.

5.2.3 The Generalised Ordered Logit Model

The generalised ordered logit model, with no assumption made regarding the parameters being homogenous across the $(J-1)$ binary regressions is shown below:

$$\begin{aligned} \Pr(Y_i > j) &= g\left(X\beta_j\right) \\ &= \frac{\exp\left(\alpha_j + X_i\beta_j\right)}{1 + \left\{\exp\left(\alpha_j + X_i\beta_j\right)\right\}} \end{aligned} \tag{5.2.23}$$

where $j = 1, 2, ..., J-1$ and J is the number of categories of the ordinal dependent variable, y. From equation (5.2.23) it can be deduced that the probabilities that y will take for different values of j are equal to:

$$\begin{aligned} \Pr(Y_i = 1) &= \frac{\exp(\alpha_1 - x'\beta_1)}{1 + \exp(\alpha_1 - x'\beta_1)} \\ \Pr(Y_i = j) &= \frac{\exp\left(\alpha_j - x'\beta_j\right)}{1 + \exp\left(\alpha_1 - x'\beta_j\right)} - \frac{\exp\left(\alpha_{j-1} - x'\beta_{j-1}\right)}{1 + \exp\left(\alpha_{j-1} - x'\beta_{j-1}\right)}, j = 1, ..., J-1 \\ \Pr(Y_i = J) &= 1 - \frac{\exp(\alpha_j - x'\beta_J)}{1 + \exp(\alpha_1 - x'\beta_J)} \end{aligned}$$

The generalised logit model collapses to the standard logistic binary regression model when $M = 2$ and is equivalent to a series of binary logit models with the values of the dependent variable combined. The following condition must be true to ensure that $0 \leq \Pr(y = j|x) \leq 1$:

$$\left(\alpha_j - x'\beta_j\right) - \left(\alpha_{j-1} - x'\beta_{j-1}\right) \geq 0 \tag{5.2.24}$$

5.2.4 Data

The data to be applied corresponds to the original data set in section 4.2.1. The sample consists of daily transaction data of all individuals with bank accounts linked to CBA over a two and a half year period. A cross-section of the data set has been formed by aggregating the total number of transactions made by individuals and constructing a categorical dummy variable based on table 5.3.1.2. It is assumed all consumers do not operate commercially and adoption is measured by frequency of use. Ideally it would be based on the proportion of all bills paid but this information is not available. The variables to be applied in the model are also found in section 4.2.1, and are reproduced below for convenience:

- $Usage_i, i = 1, ..., 6$: Usage categorical variable ranging in values from one-off payers to heavy users of the BPAY platform
- Age
- Gender

- State
- Banking segment. This includes retail, premium banking and wealth management
- Dummy variable indicating whether an individual holds a credit card with the bank

The motivation for including the variables above is summarised by section 4.3.2 as they also influence the usage of the BPAY platform by the same reasoning. This is verified by the literature as demographics are typically included to explain usage of payment platforms across individuals. Age is expected to have a positive parameter, as well as credit card ownership. Younger individuals have grown up with technology and are more inclined to use the BPAY platform. Older individuals also recieve more bills due to taking on the payments of their children and household, hence it is expected that age will have a positive parameter estimate. Credit card ownership implies individuals are confident in using an electronic means of transaction and hence, its parameter is expected to be positive. The estimated parameters for state dummy variables will be expected to be related to internet connectivity. The greater the internet connectivity, the more inclined an individual is to use BPAY. Whereas customer segment is a proxy for wealth, the more wealthy the individual the more bills are expected an individual to have as consumption increases. Females are expected to have a higher usage of the BPAY platform, as they typically run the household finances hence a positive parameter estimate, as detailed in section 4.3.2.

5.3 Results

5.3.1 Descriptive Statistics

5.3.1.1 BPAY Volume Market Share of Payer Financial Institutions

Prior to the implementation of the model, it is necessary to ensure the sample of individual's tracked over time are representative of all BPAY consumers, to ensure that inferences are valid for all BPAY consumers.

Table 5.3.1.1. Volume Statistics of BPAY Member Banking Institutions (%)

Statistic	*ANZ*	*CBA*	*CRU*	*NAB*	*STG*	*WBC*
$\bar{x}$	15	19.7	10.2	16	10	14.7
$\sigma_{\bar{x}}$	1.1	1.5	0.7	0.3	0.6	0.9
$Max\,(x)$	16.17	23.8	11.2	16.9	11.2	16.4
$Min\,(x)$	13	18.1	7.9	15.2	8.9	13.5

Table 5.3.1.1 lists the top contributors to monthly BPAY volume from the financial institutions of consumers between July 2002 and February 2009. Following the recent takeover of St George Bank (STG) by Westpac (WBC), the total volume accounted by the big four is 76%. The CBA accounts for the highest proportion of total volume with 19.7% of volume being processed for payers. It is of note that Cuscal[1] (CRU), accounting for 10.2% of volume, is the next

[1] Credit Union Services Corporation Australia Limited

biggest contributor following the big four banking institutions. There is also little change over time in the variability of the proportion of total volume that is transacted by each bank with standard deviations of between 0.3% and 1.53%, respectively.

Consumers with CBA as their banking institution account for approximately a fifth of all BPAY transactions. With low volatility between member banks of BPAY in relation to the total number of transactions received by consumers, data sourced by the CBA is representative of all BPAY consumers due to its significant market share of BPAY transactions.

5.3.1.2 Data Summary Statistics

The number of transactions processed by BPAY via CBA totalled approximately 74.1 million. The data set has been reduced by matching each consumer in their data set with their corresponding demographics. The first filter applied to the data was to remove any individuals in the data for which their accompanying demographic related data was missing. In addition, the focus of this paper is on Australian consumers, as such, all individuals in the data set that are based overseas have been removed. The remaining number of consumers in the data set totals 1.53 million.

Figure 5.3.1.2 (a). Total Monthly Transactions and Consumers

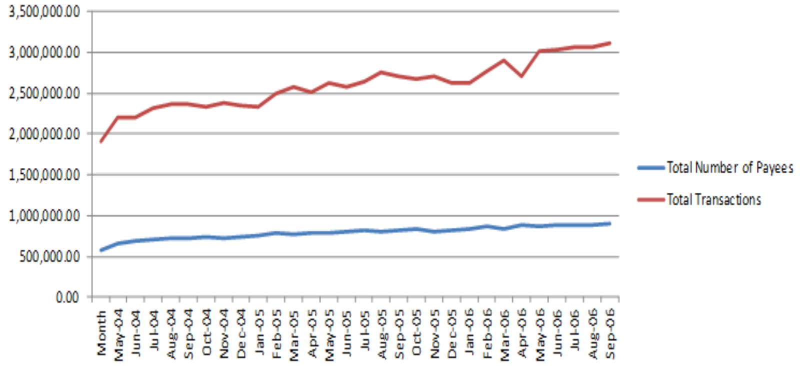


Figure 5.3.1.2 (a) illustrates the upward trend in the number of transactions processed by BPAY consumers and the total number of consumers of the platform. During the observation window, transactions have grown from 1.9 million in May 2004 to 3.1 million in October 2006. Due to the seasonal characteristics of the bill payment market, month on month growth figures may provide a more accurate representation in the growth of the BPAY platform amongst CBA consumers. The growth in the transactions processed in October 2006 has grown 15% in comparison with the same month of the previous year and 30% in comparison with the total number of transactions processed in October 2004.

Figure 5.3.1.2 (b). Average Transaction Value

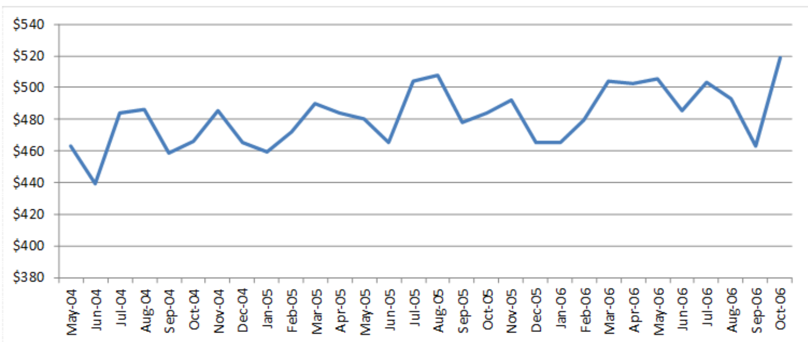


Figure 5.3.1.2 (*b*) depicts the average transaction value (ATV) between May 2004 and October 2006. The ATV has been steadily increasing over time, in addition, the ATV appears to be seasonal. Typically, the ATV is at its highest for the year in August and lowest in the month of June. The growth in the ATV in October 2006 in relation to the same month of the previous year is 7.2% and 11.2% in comparison with the ATV in October 2004.

Table 5.3.1.2 (*a*). BPAY Consumer Classifications

Total Transactions	**Definition**
1	One-Off
1 to 8	Very Light
9 to 48	Light
49 to 120	Average
121 to 240	Medium
> 240	Heavy

By segmenting individual BPAY consumers by their frequency of usage, information can be revealed in relation to their demographics and transaction characteristics. Table 5.3.1.2 (*a*) categorises BPAY consumers according to the number of transactions placed over the observational window of 30 months. Figure 5.3.1.2 (*c*) illustrates the proportion of total transactions that is accounted for by each type of BPAY consumer that is listed in table 5.3.1.2 (*a*). Average BPAY consumers account for the largest proportion of all BPAY transactions with 42% of total volume. Medium and light usage BPAY consumers account for 28% and 20% respectively, with the remaining classifications of consumers accounting for 10%.

Figure 5.3.1.2 (*c*). Total Transactions and Consumer Activity

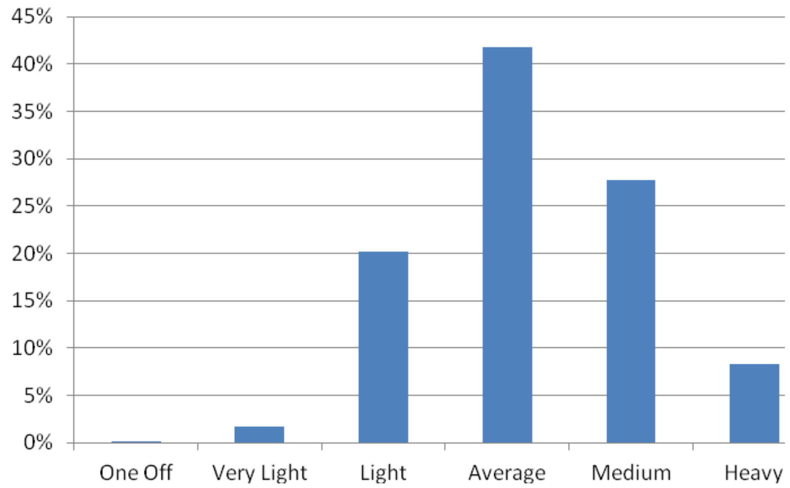
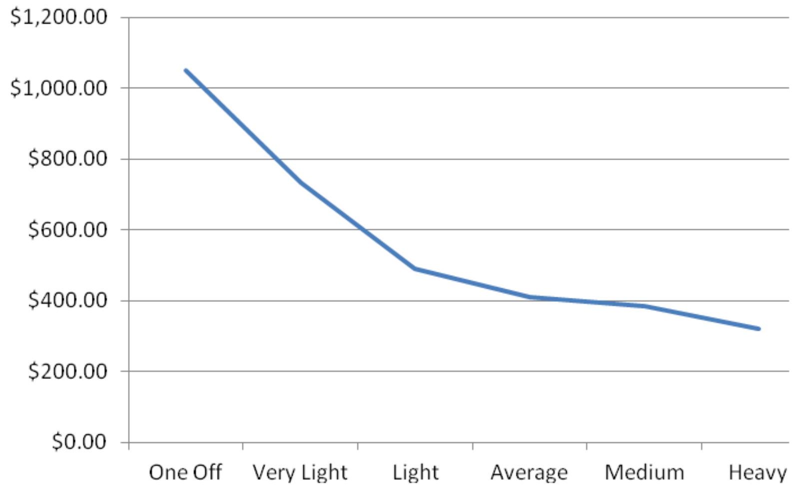


Figure 5.3.1.2 (*d*) details the ATV of different types of BPAY consumers. There appears to be a decreasing relationship between the frequency of BPAY usage and the ATV. The ATV of one-off and very light consumers is $1050 and $733, with medium and heavy users of the BPAY platform having an ATV of $384 and $322, respectively. This suggests there is heavy consumer usage for the smaller ATV's with usage diminishing as the ATV increases.

Figure 5.3.1.2 (*d*). ATV and Consumer Activity



The percentages-per-type of consumer activity is shown by figure 5.3.1.2 (*e*). Light users account for the majority of consumers in the sample representing 39% of all individuals, followed by average and very light transactors with 26% and 19%, respectively. Heavy transactors and one-off users account for the least number of consumers with 1% and 7%, respectively.

Figure 5.3.1.2 (*e*). Consumer Activity

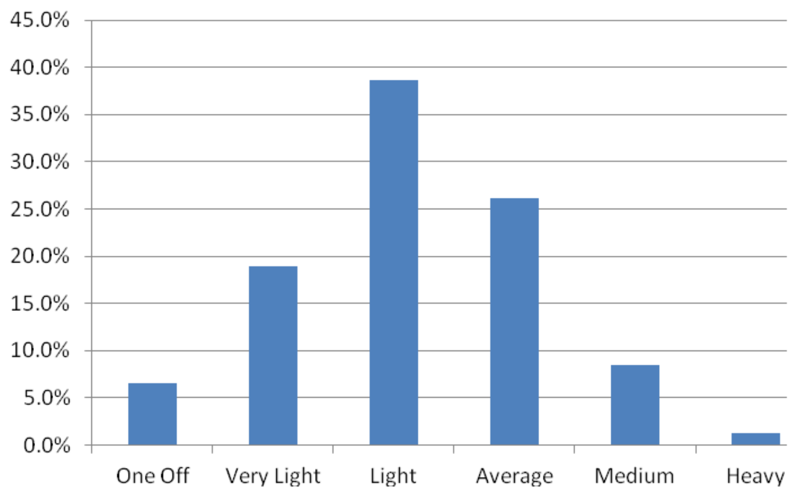


Table 5.3.1.2 (b) lists the average number of bills paid over the observational window per consumer activity classification. The average heavy user paid 324 bills using the BPAY platform, an average of approximately 11 per month. The average medium user of the BPAY platform averaged $5 - 6$ transactions per month. Average users of the BPAY platform typically pay $2 - 3$ bills per month and light users average just under 1 bill per month. The average very light user only pays bills once every $7 - 8$ months using the BPAY platform.

Table 5.3.1.2 (b). Average Number of transactions per Consumer Classification.

Consumer Type	**Average Number of Transactions**
One Off	*Not Applicable*
Very Light	4
Light	25
Average	77
Medium	159
Heavy	324

Covariates of Consumer Activity The demographics attached to the individuals in the data can be utilised to reveal differences in consumer behaviour. The covariates to be used in the multinomial model will be used to infer information related to consumer activity. Figure 5.3.1.2 (f) illustrates the relationship between the state in which an individual is located and consumer type.

Overall, New South Wales (NSW) accounts for 33% of all consumers in the sample, followed by Victoria (VIC) with 30%, Queensland (QLD) with 20%, South Australia (SA) with 5%, Australia Capital Territory (ACT) with 2%, Tasmania (TAS) with 2% and the Northern Territory (NT) with 1%. Relative to consumer activity, NSW has the highest share of one-off consumers with 34%. Whereas, Victoria has the greatest percentage of heavy users, accounting for 35%. It is worth noting that Victoria's share of consumer activity increases from one off users to heavy users, while the opposite is true for NSW.

Figure 5.3.1.2 (f). Location and Consumer Activity

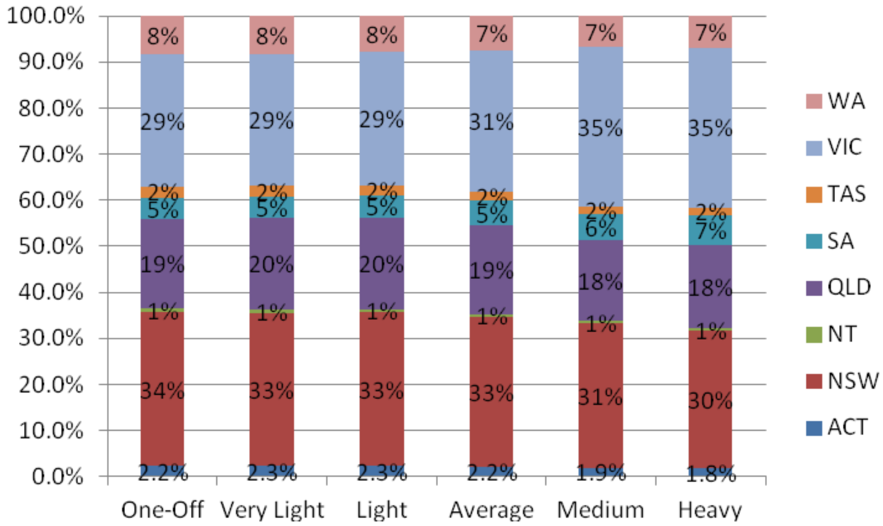


Figure 5.3.1.2 (g) indicates that 51% of the sample is female and 49% male. The split of male and female in the various consumer categories is split fairly evenly for the first three categories. However, there is a slightly greater proportion of females as consumer usage increases. Heavy users have a noticeably higher percentage of female users with 54%. Female users account for a slight

majority in the average and medium categories.

Figure 5.3.1.2 (g). Gender and Consumer Activity

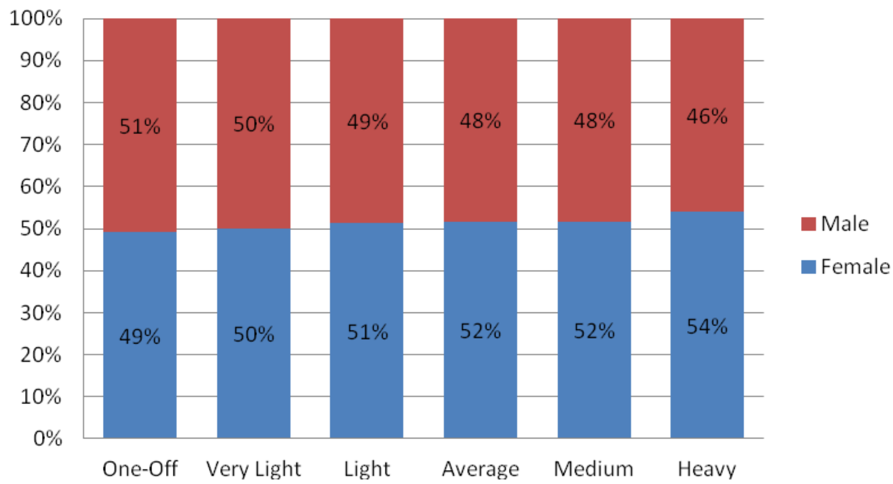
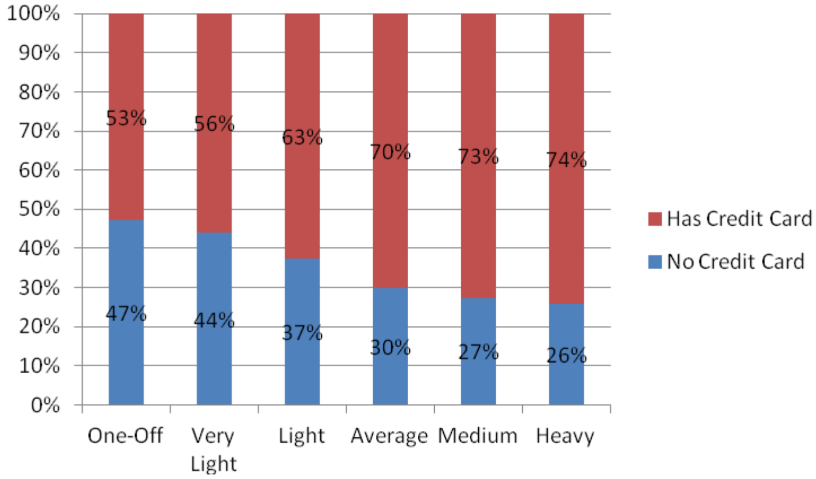


Figure 5.3.1.2 (h) provides a visual representation of the proportion of consumers that have a credit card facility attached to their bank accounts for each respective transactor category. Overall, 36% of all individuals do not have a credit card facility, in comparison with 64% of individuals who have access to a credit card. There is an increasing relationship between consumer usage and having a credit card facility attached to a bank account. The proportion of individuals holding a credit card rises along the categories of consumer usage. Individuals that own credit cards may be more accepting of using BPAY to handle bill payments as opposed to traditional over-the-counter methods. As reflected by Figure 5.3.1.2 (h), categories of higher usage have greater proportions of credit card ownership in comparison with lower categories of consumer usage.

Figure 5.3.1.2 (h). Credit Card Facility and Consumer Activity



The proportion of consumers in the retail banking, premium banking and wealth management segments per consumer activity classification is shown in Figure 5.3.1.2 (i). Individuals in the retail banking segment account for the majority of the sample with 91%, followed by premium banking and wealth management with 8.8% and 0.2%, respectively. It is interesting to note that the share of premium banking individuals grows as consumer activity increases from below 8% for one-off, very light and light users to 16% for heavy users.

Figure 5.3.1.2 (*i*). Bank Segment and Consumer Activity.

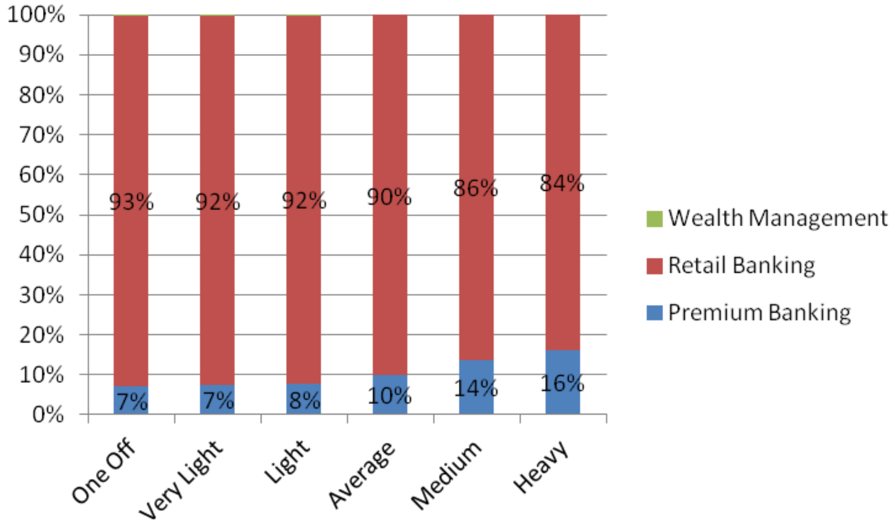


Table 5.3.1.2 (*c*) provides a breakdown of the data into six age brackets. Middle age consumers and adults characterise the majority of the data with the age brackets 21−30, 31−40 and 41−55 representing 83.1% of all individuals. The 21 − 30 age group is most represented in the data with 31.3% of all individuals. Whereas, the elderly and young account for the smallest proportion of all payers in the sample with 3.9% and 4.2%, respectively.

Additionally, the ATV and share of total transactions by age group are presented in Table 5.3.1.2 (*c*). There appears to be a positive relationship between ATV and age. The ATV increases consistently from $227 in the 0 − 20 age bracket to $560 in the 56 − 65 age bracket before slightly decreasing to $553 for individuals older than 65. In relation to the share of total transactions processed by age group, individuals in the 31 − 40 and 41 − 55 age groups are over represented with 61% of all transactions processed while only accounting for 51% of all individuals in the sample. Whereas, individuals in the $0-20$ and $21-30$ age groups while combing for 35.2% of all individuals in the sample, only account for 24% of total volume.

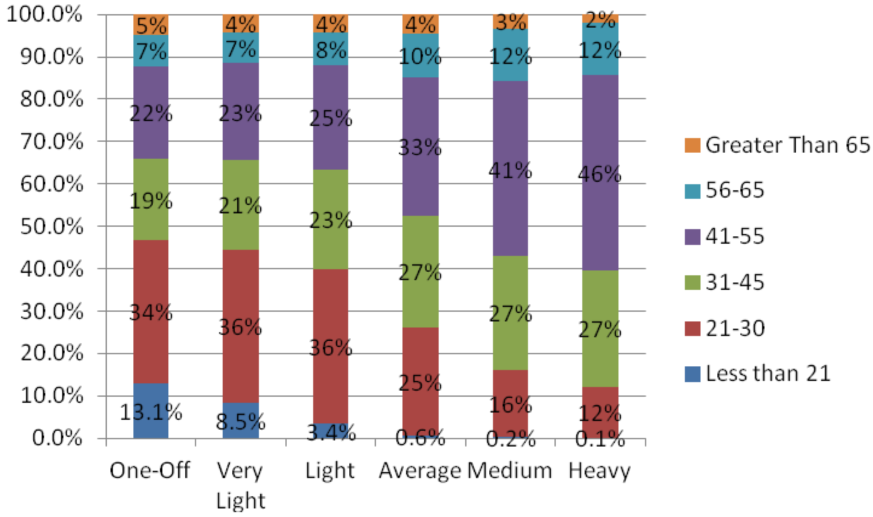
Table 5.3.1.2 (c). Description of Age Distribution.

Age Bracket	**Individuals** (%)	**ATV** ($)	**Share of Transactions** (%)
$0-20$	3.9	227	1
$21-30$	31.3	260	23
$31-40$	23.9	385	26
$41-55$	27.9	494	35
$56-65$	8.8	560	11
≥ 65	4.2	553	4

Figure 5.3.1.2 (j) provides a graphical illustration of the breakdown of consumer activity into different age groups. Consumers that are less than 21 over-represent as a proportion of total individuals that use BPAY platform the one time. As a group, individuals less than 21 account for only 3.9% of the sample, however, 13.1% of all individuals that use the BPAY platform once are less than 21. Another interesting feature of the data is the age groups that represent the higher categories of consumer activity. In relation to medium and heavy BPAY consumers, the age group $41-55$ is over-represented with 41% and 46%, dramatically higher then their share of 27.9% all individuals. Whereas, individuals in the $21-30$ age bracket are marginally under-represented with 27% in

the medium and heavy categories, in comparison with a share of 31.3% of all individuals.

Figure 5.3.1.2 (j). Consumer Activity and Age.



In the regression results to follow, the base characteristics of an individual are to be a female aged 15, resident of the ACT, classified by the CBA as being within the premium banking segment and does not hold a credit card.

A-priori expectations concerning the value for the parameter value for age is positive. It is anticipated that older individuals will have more bills outstanding than younger individuals. Thus, older individuals will be expected to transact more often on the BPAY platform. As discussed in section 4.3.2, the ACT is followed by NSW, VIC, QLD and WA in regards to internet activity with the state of TAS having the worst internet connectivity of all. Other things being equal, it is expected that the better the internet connectivity, the higher the usage of the BPAY platform. Hence, it is expected that all parameters of the state indicator variables will be negative as the ACT is the base case category in all models estimated.

The expected influence of gender on BPAY usage is unclear. In the event a household lead is a female in regards to managing the finances of the household, it is expected that the coefficient of gender will be negative. Assuming females administer all of the payments of bills, the number of bills paid in total. Therefore, the number of transactions placed is expected to be more than males. Banking segment categories may proxy for wealth, with individuals in the wealth management category being the most wealthy, followed by individuals that are categorised in premium banking and retail payers. It is expected that the greater the disposable income, the greater the consumption of an individual and the higher the amount of bills outstanding, other things being equal. Therefore, it is expected the parameter of retail will be negative and that for wealth management will be positive, as premium banking is the base category. Lastly, the coefficient of credit card holding is expected to be positive as suggested by Hayashi and Klee (2003) who postulate that prior adoption of a technological product influences the take-up of an innovative new technology-based product.

5.3.2 Ordered Logit Model

Table 5.3.2. Ordered logit

$Usage_i$	β_i		95% *Confidence Interval*	
Age_i	0.18	(0.00)	0.018	0.018
NSW_i	0.05	(0.00)	0.025	0.065
NT_i	−0.1	(0.00)	−0.14	−0.062
QLD_i	0.05	(0.00)	0.03	0.07
SA_i	0.17	(0.00)	0.15	0.019
TAS_i	−0.13	(0.00)	−0.16	−0.1
VIC_i	0.16	(0.00)	0.14	0.18
WA_i	0.01	(0.54)	−0.01	0.03
$Male_i$	−0.08	(0.00)	−0.09	−0.076
$Retail_i$	−0.22	(0.00)	−0.23	−0.21
$Wealth_i$	−0.63	(0.00)	−0.7	−0.56
$Card_i$	0.39	(0.00)	0.38	0.4
α_1	−1.9	(0.00)	−1.94	−1.89
α_2	−0.32	(0.00)	−0.34	−0.3
α_3	1.39	(0.00)	1.36	1.41
α_4	3.07	(0.00)	3.05	3.1
α_5	5.2	(0.00)	5.2	5.26

The results of the ordered logit model are presented in Table 5.3.2. The terms in the brackets refer to the p-values of the respective estimated parameter. The estimated cut-off values which define the ordered logit models y_i varies from 1 to 6 are given by the coefficients of α_i. Overall, the model has a relatively low R^2 of 1.4% and the p-value corresponding to the null hypothesis that all parameters in the model are equal to zero is 0.00, hence the model as a whole is statistically significant at the 5% level of significance (los). All parameters are statistically significant at the 5% los, with the exception of the binary variable that represents an individual from Western Australia.

The a-priori expectation of the impact of credit card holding on usage is satisfied with the coefficient of holding a card being positive. This indicating that holding a credit card is positively correlated with higher categories of the consumer usage dependent variable. Age is also positively correlated with higher categories of the consumer usage variable, as are individuals residing in NSW, QLD, SA VIC and WA. Whereas, individuals from NT, TAS and those categorised as retail and wealth management individuals are negatively related to consumer usage. Gender also has a strong influence on the extent of usage on the BPAY platform with males not being associated in higher categories of consumer usage, in comparison to females.

The ordered logit model assumes that the proportional odds assumption holds for every variable in the model, that is, the impact of all covariates on the dependent variable is constant over all values of y_i. In practise, this assumption is typically violated. Consequently, further models need to be considered to determine the influence of the covariates on consumer usage. The Brant test of proportional hazards is employed to test the assumption of proportional odds through global and covariate specific tests.

5.3.2.1 Brant Test

Table 5.3.2.1. Brant Test

$Variable$	χ^2		df
Age_i	1712	(0.00)	4
NSW_i	16	(0.00)	4
NT_i	16	(0.01)	4
QLD_i	6	(0.22)	4
SA_i	87	(0.00)	4
TAS_i	17	(0.00)	4
VIC_i	127	(0.01)	4
WA_i	14	(0.00)	4
$Male_i$	77	(0.00)	4
$Retail_i$	1120	(0.00)	4
$Wealth_i$	111	(0.00)	4
$Card_i$	108	(0.00)	4
$Overall$	4297	(0.00)	48

Table 5.3.2.1 provides the results of the Brant test of whether the parameters of equation (5.2.18) are relatively equal over sequential logit models that vary according to the value of the categorical dependent variable. Overall, the global test for the assumption of proportional odds that is implied by the ordered logit model is rejected at the 5% los, with the p-value associated with such a test

being approximately equal to zero. All covariates with the exemption of the dummy variable indicating an individual resides in QLD, reject the null hypothesis of proportional odds. Hence, the results of the ordered logit model are void following th rejection of the assumption of proportional odds. Consequently,a generalised ordered logistic model is estimated that allows for the assumption to be relaxed for those covariates that do not satisfy the assumption of proportional odds.

5.3.3 Generalised Ordered Logit Model

Table 5.3.3 (a). Generalised Logit Model

	$Usage_i = 1$	$Usage_i = 2$	$Usage_i = 3$	$Usage_i = 4$	$Usage_i = 5$
	β_i	β_i	β_i	β_i	β_i
Age_i	0.01	0.01	0.02	0.02	0.02
NSW_i	0.17*	0.04	0.05	0.08	0.01*
NT_i	−0.11	−0.09	−0.12	−0.06*	0.15
QLD_i	0.05	0.05	0.05	0.05	0.05
SA_i	0.08	0.12	0.2	0.27	0.34
TAS_i	−0.11	−0.08	−0.16	−0.17	−0.17
VIC_i	0.07	0.11	0.19	0.3	0.26
WA_i	−0.03*	−0.01*	0.02*	0.04	0.03*
$Male_i$	−0.1	−0.09	−0.07	−0.09	−0.18
$Retail_i$	−0.09	−0.1	−0.22	−0.41	−0.53
$Wealth_i$	−0.25	−0.43	−0.93	−1.3	−1.12
$Card_i$	0.41	0.41	0.39	0.33	0.36

The results of the generalised logit model are displayed in Table 5.3.3 (a). The parameters associated with the QLD binary variable is constant across all values of the consumer usage categorical dependent variable and is consistent with the assumption of proportional odds not being rejected by the Brant test.

The results of Table 5.3.3 (a) can be interpreted as estimated parameters from individual binary logit models with the dependent variable collapsed into two categories. For example, results related to the dependent variable taking the value of 1 in Table 5.3.3 (a) refer to a logit model with the dependent variable being coded as 0 if $Usage_i = 1$ and 1 if $Usage_i = 2, ..., 6$. The second component of the results corresponding to $Usage_i = 2$ refer to another binary logit model estimated with the dependent variable being coded as 0 if $Usage_i = 1, 2$ and 1 if $Usage_i = 3, ..., 6$. The positive parameter coefficients in table 5.3.3 (a) imply that higher values of the covariates are associated with higher values of the categorical dependent variable. Hence increasing usage of the BPAY platform.

The generalised logit model achieved an R^2 of 1.5% and a p-value associated with a null hypothesis that all parameters are not significantly different from zero of approximately 0.00, implying the model is empirically valid. Whites standard errors that are robust to heteroscedasticity are used and an asterisk $(*)$ denotes a coefficient that was not significant at the 5% los. The parameters of the majority of covariates are statistically significant at the 5% level, with the exception of a few parameters per panel being insignificant. The parameters associated with the state indicator variables for WA, NSW and NT are statistically insignificant at the 5% los for four, two and one panels out of the six estimated, respectively.

The influence of age on the usage of BPAY platform is positive with the coefficient of age being positive for all five panels. The impact of age also gets larger across cut-off points, as the magnitude of the coefficient increases with the value of the dependent variable, $Usage$. Hence, older individuals are more likely to be members of the higher categories of usage.

The reference location in the consumer usage model is the ACT and all binary variable coefficients signify the impact of location on consumer usage in comparison to residents from the ACT. Overall, in comparison to individuals from the ACT, residents in NSW, QLD, SA and VIC are more likely to be members of the higher consumer usage categories. Whereas, individuals from NT and TAS are more likely to be associated with the lower categories of consumer usage, given their negative coefficient values. The influence of SA and VIC gets larger across the cut-off points of consumer usage, hence there is a higher probability of individuals from these states being members of the higher categories of usage. The opposite is true for NSW, while the influence of consumer usage across the categories for QLD is consistent. However, the effect of consumer usage by residents of NT and TAS is negative with the magnitude of its impact on progressing to higher levels of usage decreasing across cut-off points. Hence, individuals from the NT and TAS are less likely to be medium or heavy users of the BPAY platform than individuals from other states.

Gender and Customer segment are relevant covariates in explaining the likelihood of an individual being placed in the 6 categories of BPAY usage. Other things being equal, males use the BPAY platform less than females as signified by the negative estimated coefficients across the 5 panels. The influence of gender, while fairly consistent over the first four categories, gets stronger at the last cut-off points of BPAY usage. That is, males are far less likely being classified as an average to heavy user than a light user of the BPAY platform. Similarly, with reference to premium banking customers, retail and wealth management individuals are less likely to be high users of the BPAY platform. The impact of being members of the retail and wealth managements is negative and it gets larger across cut-off points. This suggests individuals are far less likely to be members of the higher BPAY usage categories.

The effect on credit card holding on consumer usage is positive and relatively uniform across the 5 panels of the generalised ordered logit models estimated. This satisfies a-priori expectations concerning the influence of prior technology adoption and the consumption of an innovative technological product. The parameter value does not vary substantially over cut-off points, with the coefficients value ranging from 0.33 to 0.41. To assist in the interpretation of the results in Table 5.3.3 (a), odd ratios are computed below.

Table 5.3.3 (b) Generalised Logit Odds Ratios

	$Usage_i = 1$	$Usage_i = 2$	$Usage_i = 3$	$Usage_i = 4$	$Usage_i = 5$
	β_i	β_i	β_i	β_i	β_i
Age_i	1.01	1.01	1.02	1.02	1.02
NSW_i	1.02	1.04	1.05	1.08	1.01
NT_i	0.9	0.92	0.88	0.94	1.16
QLD_i	1.06	1.06	1.06	1.06	1.06
SA_i	1.09	1.12	1.22	1.31	1.4
TAS_i	0.9	0.92	0.85	0.84	0.85
VIC_i	1.08	1.11	1.2	1.35	1.29
WA_i	0.97	0.99	1.02	1.04	1.03
$Male_i$	0.91	0.91	0.93	0.91	0.83
$Retail_i$	0.92	0.9	0.8	0.66	0.59
$Wealth_i$	0.78	0.65	0.39	0.27	0.33
$Card_i$	1.51	1.5	1.48	1.39	1.43

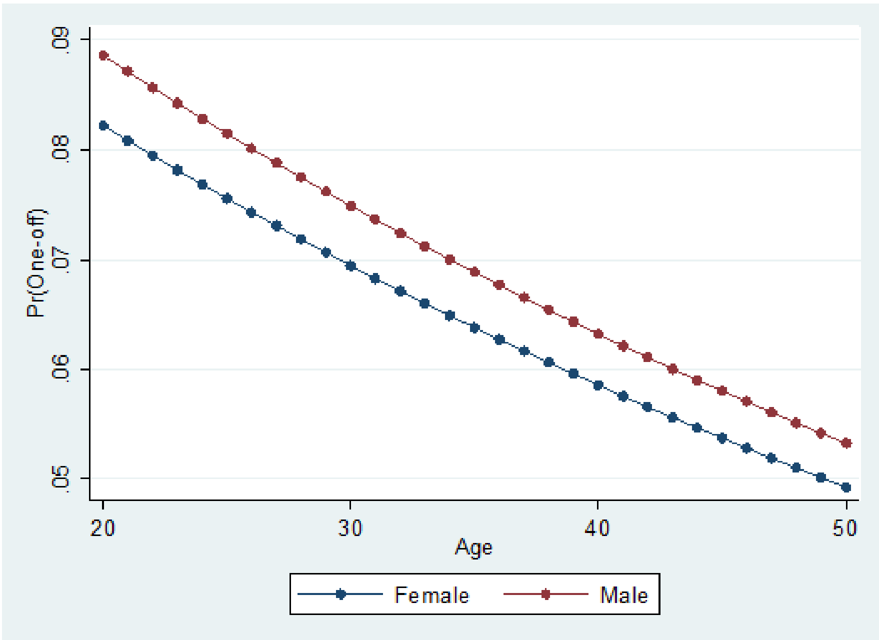
Table 5.3.3 (b) portrays the odds ratio of every covariate across the five panels estimated for the generalised ordered logit model. It is essentially the exponential of the coefficients estimated in Table 5.3.3 (b). *Age* is the only continuous explanatory variable in the model and, as such, has a slightly different interpretation to that of the other covariates present. In relation to the results of

the first panel, $Usage_i = 1$, for every unit increase in the age of an individual, the odds of an individual being placed in a higher usage category than $Usage_{i1} = 1$ improve by 1%. Alternatively, comparing a 20 year old to that of a 40 year old, the odds of an individual that is 40 years old being placed in a higher usage category than $Usage_{i1} = 1$ improve by 22%. A similar interpretation can be applied to the remaining coefficients for age in the other panels. For example, in comparing the odds of a 20 year old to that of a 40 year old to be placed in the heavy category of consumer usage, $Usage_i = 6$, the odds of the 40 year old improve by 50%.

The discrete variables in the model have a simple interpretation in comparison to the continuous explanatory variables. In relation to the state indicator variables and focusing on the results of the first panel, the odds that an individual born in NSW will be placed in a usage category other than that of a one-off user improve by 2%, in comparison to an individual born in the ACT. However, the odds decrease by 10% if the individual were born in the NT. The greatest odds ratio value amongst the location dummy variables is recorded by SA in the fifth panel. Hence, there is an improvement in the odds of 40% if an individual is placed in the heavy user category as opposed to a lower usage category when comparing an individual from SA to an individual from the ACT.

The impact of gender on the odds ratio do not differ greatly across the five panels estimated, except for the last panel estimated. The odds of a male to be in a higher usage category than one-off decreases by 9% in relation to a female, other variables being held constant. The impact of gender is most pronounced in the last panel. Assuming all other variables are held constant, the odds decrease by 17% for males being in the highest category of BPAY usage. Figure 5.3.3 (a) shows the impact that gender has on the probability of being classified as a One-off payer. For all values of age, a female is less likely to be a one-off platform payer of BPAY.

Figure 5.3.3 (a). Probability of being an One-Off payer: Gender



A-priori expectations are largely satisfied concerning the influence of bank segment on platform usage, assuming it proxies for wealth. The least wealthy classification is retail and, other things being equal, such an individual should consume less and have lower amounts of bills outstanding. This observation is confirmed by the model. For example, in regards to the first panel, the odds of an individual in the retail segment is 8% less likely to be in usage categories greater than a one-off user in comparison to an individual that is a premium banking customer. The results of the most wealthy individuals of the three segments, wealth management, are against a-priori expectations concerning the impact of wealth on platform usage. The results suggest that there is a negative relationship between consumer usage and being part of the wealth management segment. For example, in regards to the last panel, the odds of an individual being a heavy user of the BPAY platform, $usage_i = 6$, decrease by 67% if they are regarded as individuals in the wealth management sector in comparison with someone from premium banking.

The improvement of the odds of individuals holding a credit card ranges from 43% to 51% of being in a higher consumer usage category across the five panels. The influence is observed most in the first panel. For example, the odds of an individual being in a usage category that is greater than $Usage_i = 1$, light user, improve by 51% if the consumer has a credit card facility attached to their bank account. Thus, the influence of early technology adoption on the adoption of alternative new technology is confirmed by the model.

Figure 5.3.3 (*b*). Probability of being an One-Off payer: Card Holding

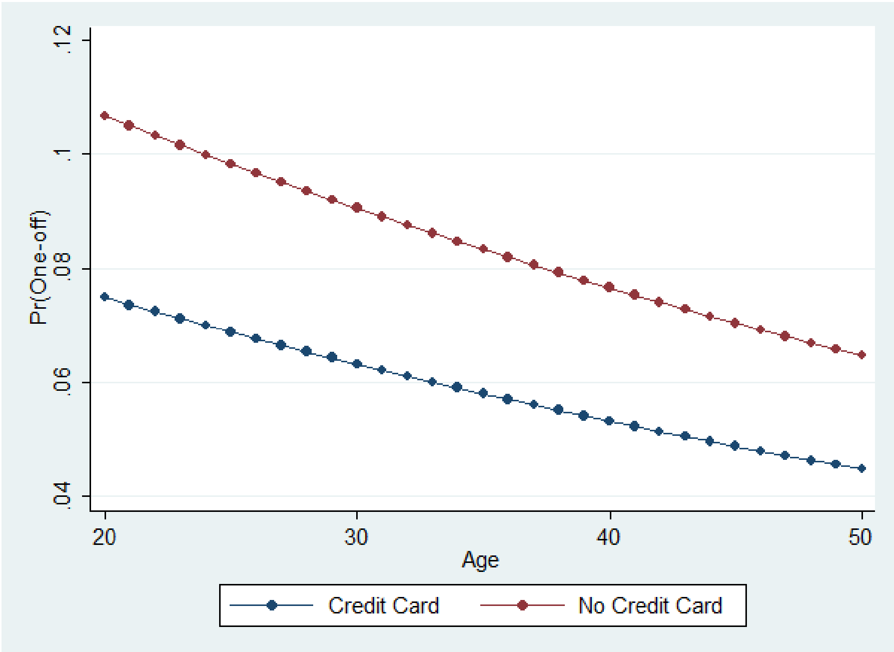


Figure 5.3.3 (*b*) graphically illustrates the impact that holding a credit card has on the probability of being classified as a One-off payer. For all values of age, an individual that holds a credit card is less likely to be a one-off platform user of BPAY.

5.4 Conclusion - Online Billing Payment Adoption

The motivation of Chapter 5 was to establish whether a link existed between the usage of the BPAY platform by consumers with the adoption of prior payment method technologies, given by credit card holding. Unlike Hayashi and Klee (2003), there is an added layer of complexity as credit cards are another payment instrument individuals can use in the Bill Payment market. Hence, the impact of credit card holding on the adoption of a new technology in the Bill Payment market is unknown and answered empirically in this paper.

Using a proprietary data set that consisted of transactions of CBA customers on the BPAY platform between May 2004 and October 2006, along with their corresponding demographics, generalised linear ordered logit models are applied to determine the impact of credit card holding on consumer usage. The results of Chapter 5 is consistent with Hayashi and Klee (2003) hypothesis that the adoption of a technology based payment instrument is influenced by the usage of prior technologies. Other findings include a positive relationship between the age of a consumer and being in a high category of usage and gender, with females more likely to be in the higher categories of usage than males.

Chapter 6

Conclusion

The motivation underlying my research was to model the Bills Payment market in which the BPAY platform operates. In doing so, an understanding of the dynamics of this market can be gained to assist central banks and businesses in their provision of incentives to consumers to change their payment preferences to more efficient platforms that have an ongoing effect of increasing the welfare of all participants in the economy.

The three research questions addressed in this dissertation are given by the following:

- An examination of the existence of network effects in the Bill Payments market
- Determining the role that demographics, transaction related characteristics and credit card holding has on a new consumer abandoning the BPAY

platform

- Investigating the influence of prior technology adoption on the consumer usage of BPAY.

Chapter 3 of my thesis addresses the first research question. A macroeconomic model was developed to model the demand of consumers and merchants to test for the presence of network effects in a two-sided, four-party bill payments market. The endogeneity that network effects imply in two-sided markets is modelled within a vector autoregressive framework. Price elasticities affecting BPAY and cross-price elasticities with competing platforms, along with short-run dynamics associated with changes in consumer demand and merchant acceptance, was also modelled within a vector error-correction framework and by impulse response functions. The results are summarised below:

- Consumers are more valuable to the BPAY platform as their effect on merchants is greater than that of merchant acceptance on consumer usage
- There is a clear distinction between short-run and long-run effects from changes to the variables that define consumer usage and merchant acceptance demand models
- Pricing elasticities indicate the Bill Payment market is competitive, with consumers and merchants reacting strongly to changes in benefits offered to consumers and fees for merchants

- Approximately 8 quarters are needed following an innovation for equilibrium to be re-established in the consumer usage and merchant acceptance demand model.

A few assumptions were made regarding the model and the data. The assumption of a perfectly competitive issuing and acquiring market simplified the consumer usage and merchant acceptance demand models. Secondly, data relating to the merchant fees of BPAY and the benefits of competing credit card schemes was not directly available. However, given these limitations, Chapter 3 explores areas of the payments market which previously has been untouched due to the lack of proprietary data made available by the private sector.

Technology has driven change in the majority of sectors in the economy. In relation to the payments market, the cost savings eventuating from shifting to an automated technology-based payments system from a paper-based payments system are substantial. Chapter 4 identifies factors that contribute to new consumers of BPAY to leave the platform. A Cox (1972) proportional hazards model was applied to quantify the risks of an individual leaving the platform. The main findings include:

- Individuals who have a credit card are up to 12% less likely to leave the BPAY platform at any point in time

- Males are approximately 3% more likely than females to leave the BPAY platform at any point in time

- There is a geographical influence on whether an individual adopts a innovative bill payments platform such as that offered by BPAY.

The limitations of Chapter 4 stem from the assumptions made regarding the formation of the data set. Identifying new consumers of BPAY given the sample was based on a number of assumptions. Assigning a fixed arbitrary observational window to remove the biases induced by left-censoring decreased the dataset under observation greatly. Furthermore, assuming an individual has left the BPAY platform from not observing any other transactions on the account induces error if an individual has another bank account and uses BPAY, as does death or another member of the household begins making bill payments for an extended period of time. However, due to the large sample size of the initial data set, such biases are not expected to weaken the conclusions reached in Chapter 4.

Chapter 5 of my thesis aggregates the data set used in Chapter 4 over individuals to determine the impact of prior technology adoption, given by credit card holding, on the consumer usage of the BPAY platform. A generalised ordered logit model was applied on the data once consumers were grouped into usage bands to quantify the effect of credit card holding on each level of consumer group. Results indicate the following:

- Consumer usage is positively influenced by age with older individuals more likely to be in the higher usage categories

- As in Chapter 4, there is a geographical influence on consumer usage
- Support for the hypothesis that consumer usage is positively influenced by credit card holding.

In conjunction with BPAY, data availability has enabled research questions raised to be addressed. This dissertation has contributed to extending the literature in the market for payments at both the macroeconomic and microeconomic level. Although given the limitations discussed, the findings provide a base for further research to study the Bill Payments market with applications to other two-sided markets which share similar attributes. Another future research avenue may be to determine the impact of RBA intervention on merchant fees and consumer fees more directly. The RBA has introduced several reforms in recent years and it will be of interest to isolate whether decreases in interchange fees were due to competitive pressure or regulatory changes.

Appendix

A Panel Unit Root Tests

A.1 Tests with a Common Unit Root Process

A.1.1 Levin, Lin and Chu (2002)

The model LLC consider to test for a panel unit root is given by:

$$\Delta y_{it} = \rho y_{it-1} + \sum_{j=1}^{p_i} \beta_{ij} \Delta y_{it-j} + \alpha_i + \delta_i t + \varepsilon_{it,} \tag{A.1}$$

where the null and alternative hypothesis are given by H_o and H_{1a} respectively. To test for a panel unit root, LLC recommend standardising and removing the deterministic components of Δy_{it} and y_{it-1} to attain an estimate of ρ in the following pooled regression:

$$\Delta \tilde{y}_{it} = \rho \tilde{y}_{it-1} + \nu_{it}, \tag{A.2}$$

The method for implementing the LLC test is summarised in the following steps:

1. Run an ADF test for each cross-section to determine the maximum number of lags p_i required for each cross-section using an information criterion.

2. To attain the orthogonalised residuals of Δy_{it} and y_{it-1}, perform the following auxiliary regressions:

$$\Delta y_{it} = \sum_{j=1}^{p_i} \beta_{ij} \Delta y_{it-j} + \alpha_i + \delta t_i + u_{it,}, \tag{A.3}$$

$$y_{it-1} = \sum_{j=1}^{p_i} \beta_{ij} \Delta y_{it-j} + \alpha_i + \delta t_i + \eta_{it-1,} \tag{A.4}$$

3. The estimated residuals from equation $(A.3)$ and $(A.4)$ are then standardised to account for differences in variances across cross-sections, as shown below:

$$\tilde{u}_{it} = \frac{\hat{u}_{it}}{\hat{\sigma}_{ui}}, \tag{A.5}$$

$$\tilde{\eta}_{it-1,} = \frac{\hat{\eta}_{it-1}}{\hat{\sigma}_{ui}},$$

4. An estimate of the ratio of long-run to short-run standard deviations is computed. Under the null hypothesis, the long-run variance of equation $(A.1)$ is given by

$$\hat{\sigma}_{yi}^2 = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y_{it}^2 + 2 \sum_{L=1}^{\bar{K}} \omega_{\bar{K}L} \left[\frac{1}{T-1} \sum_{t=2+L}^{T} \Delta y_{it} \Delta y_{it-L} \right], \tag{A.6}$$

where $\bar{K}$ is a truncation lag that can be data dependent. For a Bartlett kernel, $\omega_{\bar{K}L} = 1 - \left(\frac{L}{\bar{K}+1}\right)$. For each cross-section i, the ratio of the long-run standard deviation is estimated by $\hat{s}_i = \frac{\hat{\sigma}_{yi}}{\hat{\sigma}_{\varepsilon i}}$. The average standard deviation is then estimated by $S_N = \frac{1}{N} \sum_{i=1}^{N} \hat{s}_i$.

5. The regression denoted by equation $(A.2)$ is then run to obtain an estimate of ρ. The regression is based on $N\tilde{T}$ data observations, where $\tilde{T} = T - \bar{p} - 1$ and $p = \frac{1}{N}\sum_{i=1}^{N} p_i$.

Compute the adjusted *t-statistic*:

$$t_{\rho}^{*} = \frac{t_{\rho} - N\tilde{T}\hat{S}_N\hat{\sigma}_v^{-2}\hat{\sigma}\left(\hat{\rho}\right)\mu_{m\tilde{T}}^{*}}{\sigma_{m\tilde{T}}^{*}}, \tag{A.7}$$

where $t_{\rho} = \frac{\hat{\rho}}{\hat{\sigma}(\hat{\rho})}$, $\hat{\sigma}\left(\hat{\rho}\right)$ is the standard error of $\hat{\rho}$, $\mu_{m\tilde{T}}^{*}$ and $\sigma_{m\tilde{T}}^{*}$ are the mean and standard deviation adjustments that depend on the specification of deterministic terms in the model and are tabulated by LLC. The LLC test statistic t_{ρ}^{*} is asymptotically distributed as a standard normal random variable and the requirements for t_{ρ}^{*} to achieve its limiting distribution is $\frac{\sqrt{N_T}}{T} \rightarrow 0$, where N_T emphasises that the cross-section dimension N is an arbitrary monotonically increasing function of T. Alternatively, $N \rightarrow \infty$ with $\frac{N}{T} \rightarrow 0$, such that N is small enough relative to T. In addition to the specification of the deterministic elements of the data generating process and the selection of the number of lags p_i, a kernel choice needs to be specified in the computation of $\hat{S}_N$. LLC recommend using their test for panels where $10 < N < 250$ and $25 < T < 250$. Monte Carlo Simulations performed by LLC show that the normal distribution provides a good approximation to the empirical distribution of t_{ρ}^{*}.

A.1.2 Baltagi (2000)

Baltagi (2000) investigates the local power of the LLC and IPS test statistics against a sequence of local alternatives. Baltagi (2000) results indicate that both tests suffer a persistent loss of power if individual specific trends are included. This is a result of correcting for the bias of the test statistic as it is not centered at zero, as reflected by the negative term in the test statistic of LLC in equation $(A.7)$. Baltagi (2000) proposes a test statistic that does not require a bias adjustment and has more power in comparison with the IPS and LLC tests if a deterministic trend is included in the test.

To implement Baltagi (2000), the first three steps are identical to that of the LLC test, however no trend term or individual specific intercept is included in the regression of equation $(A.3)$ and $(A.4)$. The following steps are then required to compute the test statistic:

- Transform the standardised residuals $\tilde{u}_{it}$ and $\tilde{\eta}_{it-1}$ using the forward orthogonalisation transformation advocated by Arellano and Bover (1995):

$$\begin{aligned} u^*_{it} &= \sqrt{\frac{T-t}{T-t+1}}\left(\tilde{u}_{it} - \frac{\tilde{u}_{it+1} + ... + \tilde{u}_{iT}}{T-t}\right), \qquad \text{(A.8)} \\ \eta^*_{it-1} &= \tilde{\eta}_{it-1} - \tilde{\eta}_{i1} - \frac{t-1}{T}\tilde{\eta};\ \text{with trend and intercept} \\ \eta^*_{it-1} &= \tilde{\eta}_{it-1} - \tilde{\eta};\ \text{with intercept, no trend} \\ \eta^*_{it-1} &= \tilde{\eta}_{it-1};\ \text{no trend or intercept} \end{aligned}$$

Estimate the pooled regression:

$$u^*_{it} = \rho\eta^*_{it-1} + v_{it}, \qquad \text{(A.9)}$$

The t-statistic for the null hypothesis of a panel unit root illustrated previously by H_o in equation (3.4.3) is shown to have a limiting standard normal distribution.

A.1.3 Hadri (2000)

In contrast to the LLC and Baltagi (2000) tests, Hadri (2000) is a test of stationarity. Hadri (2000) proposes a residual-based Lagrange Multiplier (LM) test and is essentially an extension of the univariate KPSS test to the panel data context.

Hadri (2000) considers the following data generating process:

$$\begin{aligned} y_{it} &= r_{it} + \beta_i t + \varepsilon_{it}, \qquad (A.10) \\ r_{it} &= r_{it-1} + u_{it}, \end{aligned}$$

Equation $(A.10)$ decomposes the data series into its deterministic and random components. It is assumed that $\varepsilon_{it} \sim IIN\,(0, \sigma_\varepsilon^2)$, $u_{it} \sim IIN\,(0, \sigma_u^2)$ and $E(\varepsilon_{it} u_{it}) = 0$ $for\ all\ i, t$. The null hypothesis of stationarity corresponds to the random component of the data series, r_{it}, having a variance of zero. Equation $(A.10)$ can be expressed using backward substitution of the stochastic component of the data series into the following:

$$y_{it} = r_{i0} + \beta_\iota t + e_{it}, \text{ where } e_{it} = \sum_{t=1}^{t} u_{it} + \varepsilon_{it}, \qquad (A.11)$$

The Null and Alternative Hypotheses Hadri (2000) proposes are shown below:

$$H_o \quad : \quad \lambda = 0, \tag{A.12}$$

$$H_1 \quad : \quad \lambda > 0,$$

where $\lambda = \frac{\sigma_u^2}{\sigma_\varepsilon^2}$. The test statistic is a one-sided LM-Statistic and utilises residuals, $\hat{e}_{it}$, from equation $(A.11)$.

$$LM_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}S_{it}^2}{\hat{\sigma}_\varepsilon^2}, \tag{A.13}$$

where $S_{it} = \sum_{j=1}^{t}\hat{e}_{ij}$ and $\hat{\sigma}_\varepsilon^2$ is a consistent estimator of σ_ε^2 under the Null Hypothesis, such as that given by equation $(A.14)$. It should be noted that in finite samples equation $(A.14)$ should be corrected for the degrees of freedom:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2, \tag{A.14}$$

Hadri (2000) also derives an alternative LM test that can account for hederoscedasticity across individuals:

$$LM_2 = \frac{1}{NT^2}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{S_{it}^2}{\hat{\sigma}_{\varepsilon i}^2}\right), \tag{A.15}$$

The test statistic proposed is provided below:

$$Z = \frac{\sqrt{N}\,(LM - \xi_i)}{\zeta_i} \text{ for } i = 1, 2 \tag{A.16}$$

where ξ_i and ζ_i are constants and whose value depends on whether a trend has been included in the test. Monte Carlo simulations were conducted to derive first and second moments of the test statistic and, by applying the Lindeberg-Levy Central Limit Theorem, the test statistic is distributed as a standard normal random variable. A disadvantage associated with the test proposed by Hadri (2000), is that it suffers from severe size distortions when the null is close to the alternative of a unit root, which is a feature of unit root tests that involve a null of stationarity. Therefore, caution needs to be exercised in using the test proposed by Hadri (2000). In addition, Hlouskova and Wagner (2005) find the Hadri (2000) test performs very poorly in small samples due to the asymptotic values for the mean and variance of the KPSS statistic being used in standardising the test statistic. It should be noted the hederoscedasticity consistent test statistic will be used in this thesis.

A.2 Tests with Individual Unit Root Processes

A.2.1 Im, Pesaran and Shin (2003)

IPS propose a computationally simple test for a panel unit root through a likelihood-based framework, with the alternative hypothesis allowing for stationary and a non-stationary series to exist in a panel. In addition, the IPS test can account for residual serial correlation. The estimated equation is as follows:

$$\Delta y_{it} = \rho_i y_{it-1} + \sum_{k=1}^{K_i} \beta_{ik} \Delta y_{it-k} + \varepsilon_{it}, \tag{A.17}$$

Equation $(A.17)$ is essentially a panel augmented Dickey Fuller test with heterogeneity imposed on the coefficients. The IPS test is conducted after the demeaning and detrending of the data series. IPS test the null hypothesis of a unit root by computing the average of N individual t-statistics associated with the DF or ADF test on each cross-section with the assumption that each individual t-statistic is independently and identically distributed. The test statistic is provided below:

$$t_{IPS} = \frac{\sqrt{N}\left(\bar{t} - \frac{1}{N}\sum_{i=1}^{N} E\left[t_{iT}|\rho_i = 0\right]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} V\left[t_{iT}|\rho_i = 0\right]}}, \tag{A.18}$$

Monte Carlo simulations are run to estimate the first and second moments of the average t-statistic. By the Lindeberg-Levy Central Limit Theorem, t_{IPS} follows a standard normal distribution as $T \rightarrow \infty$, followed by $N \rightarrow \infty$, sequentially under the null hypothesis. The rejection of the null hypothesis implies some cross-sections in the panel do not contain a unit root. However, due to the heterogeneous nature of the alternative hypothesis, the test does not specify which cross-sections are stationary. The disadvantages associated with the IPS test include the condition of a balanced panel if tabulated critical values are to be used. In addition, the lag length of all individual ADF tests needs to be the same.

A.2.2 Fisher-ADF

The Fisher tests assume that the p-values of univariate unit root tests associated with each cross-section are independent. The Fisher result is that the p-values from N independent tests are distributed uniformly between zero and one, and thereby the test statistic (γ) is distributed as a Chi-Square random variable with $2N$ degrees of freedom under the null hypothesis as $T_i \to \infty$ for all $i \in N$.

$$\gamma = -2 \sum_{i=1}^{N} \log (p_i) \sim \chi_{2N}, \tag{A.19}$$

Equation $(A.19)$ is the test statistic used in the Fisher type tests advocated by Maddala and Wu (1999) and Choi (2001) to be applied to panel unit root tests. It has the advantage over IPS that only T needs to approach infinity for a fixed N, for the test statistic to approach its limiting distribution rather than both N and T. Other advantages of the Fisher type test compared to the IPS test is that it can be based on any unit root test to compute p-values. It does not require a balanced panel and different lag lengths can be used in the individual ADF tests. However, simulation is required to yield the p-values of the individual unit root tests.

B Problems With Estimating An ECM With More Than 2 Variables

Let $z_t = [y_{1t}, y_{2t}, x_t]'$ and allow all three variables in z_t to be potentially endogenous:

$$\begin{aligned} z_t &= A_1 z_{t-1} + ... + A_k z_{t-k} + u_t \\ u_t &\sim IN(0, \Sigma) \end{aligned} \tag{B.1}$$

The above Vector Autoregression (VAR) can be represented as a Vector Error Correction Model (VECM) given by $(B.2)$:

$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + ... + \Gamma_{k-1} \Delta z_{t-k+1} + \Pi z_{t-1} + u_t \tag{B.2}$$

where $\Gamma_i = -\left(I - \Sigma_{i=1}^{i} A_i\right)$ with $i = 1, 2, .., k-1$ and $\Pi = -\left(I - \Sigma_{i=1}^{k} A_i\right)$. The (3×3) Π matrix can be decomposed further into the following:

$$\Pi = \alpha\beta \tag{B.3}$$

where α represents the speed of adjustment to disequilibrium and β denotes the cointegrating vectors. Setting the lag length to two, the VECM can be expressed as:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \\ \Delta x_t \end{bmatrix} = \Gamma_1 \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \\ \Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{23} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ x_{t-1} \end{bmatrix} + u_t \tag{B.4}$$

Taking the VECM in equation $(B.4)$ and considering just the ECM of the first endogenous variable, it can be seen in equation $(B.5)$ below that it is not possible to obtain an estimate of neither cointegrating relationships.

$$\begin{aligned}\Delta y_{1,t} &= \Gamma_1 \Delta y_{1,t-1} + \alpha_{11}\left(\beta_{11} y_{1,t-1} + \beta_{21} y_{2,t-1} + \beta_{31} x_{t-1}\right) + \\ & \quad \alpha_{12}\left(\beta_{12} y_{1,t-1} + \beta_{22} y_{2,t-1} + \beta_{32} x_{t-1}\right) + u_t\end{aligned} \tag{B.5}$$

The OLS estimator computes a weighted average of the two cointegrating relationships as it can't uniquely distinguish one long-run relationship from the other. In the event that there is only one cointegrating relationship between the endogenous variables, more efficient estimates can be obtained using the multivariate approach as all the information in the system is used. A single cointegrating vector in equation $(B.4)$ restricts all the elements in the second column of α to equal zero.

C Definition And Some Properties of Partial Likelihood

Let an observation on a random vector Y corresponding to the data have density function $f(y;\theta,\beta)$. The vector of parameters of interest is given by β and the nuisance parameter is denoted by θ. In relation to the Cox (1972) model, $h_0(\cdot)$ is the nuisance function of equation (4.2.11). Suppose that Y is mapped into a set of variables $A_1, B_1, ..., A_m, B_m$ in a one-to-one transformation, and let $A^{(j)} =$

$(A_1, .., A_j)$ and $B^{(j)} = (B_1, .., B_j)$. In addition, let the joint density of $A^{(m)}, B^{(m)}$ be denoted by:

$$\prod_{j=1}^{m} f\left(b_j | b^{(j-1)}, a^{(j-1)}; \theta, \beta\right) \prod_{j=1}^{m} f\left(a_j | b^{(j)}, a^{(j-1)}; \beta\right) \tag{C.1}$$

The second term in equation $(C.1)$ is termed the partial likelihood of β based on $\{A_j\}$ in the sequence $\left\{A_j, \beta_j\right\}$. The number of terms m could be fixed or allowed to vary.

Thus, based on equation $(C.1)$, the partial likelihood is:

$$L(\beta) = \prod_{j=1}^{m} f\left(a_j | b^{(j)}, a^{(j-1)}; \beta\right) \tag{C.2}$$

The partial likelihood shown above in equation $(C.2)$ cannot generally be given any probability interpretation as either the conditional or the marginal probability of any event. However, in many instances it can be used like an ordinary likelihood for purposes of large-scale estimation in that the usual asymptotic property formulas and properties associated with the likelihood function and likelihood estimation still apply.

C.1 Partial Likelihood for β

The partial likelihood framework can be applied to the Cox (1972) model with right censoring present in the data. Suppose that the sample comprises of k uncensored failure times $t_1 < ... < t_k$ and assume there are no tied event times. The remaining $n - k$ individuals are right censored. Let j denote the individual leaving the BPAY platform at t_j, and let and x_ℓ denote the covariate vectors for the ℓth individual. From section (C), let B_j specify the censoring information in $[t_{j-1,}t_j)$ plus the information that one individual leaves the BPAY platform in the interval $[t_j, t_j + dt_j)$. Let A_j specify that individual j fails in $[t_j, t_j + dt_j)$. The j^{th} term in the partial likelihood $(C.2)$ is:

$$L_j(\beta) = f\left(a_j | b^{(j)} a^{(j-1)}; \beta\right)$$

Conditioning event $b^{(j)}$, $a^{(j-1)}$ specifies all the censoring and failure information in the trial up to time t_j^- and also provides the information that a failure occurs in $[t_{(j)}, t_{(j)} + dt_{(j)})$. Assuming independent censoring, it follows that:

$$L_j(\beta) = \frac{h(t_j; x_j)\, dt_j}{\sum_{\ell \in R(t_j)} h(t_j; x_\ell)\, dt_j} \tag{C.3}$$

where $R(t)$ is the set of individuals at risk of leaving the BPAY platform at time t^-, just prior to t. Thus $R(t)$ consists of all individuals who are still placing billing transactions through BPAY and are still under observation at time t^-. Define as risk sets, $Y_i(t) = 1[i \in R(t)]$, where $1[\cdot]$ is the indicator function, the j^{th} term in the partial likelihood is given by:

$$L_j(\beta) = \frac{h(t_j; x_j)\, dt_j}{\sum_{\ell=1}^{n} Y_l(t_j)\, h(t_j; x_\ell)\, dt_j}$$

The Cox (1972) model, as shown by equation (4.2.11), simplifies equation $(C.3)$ since the baseline hazard $h_0(t_j)\, dt$ cancels in the numerator and denominator. The product over j then gives the partial likelihood for β,

$$\prod_{j=1}^{k} \frac{\exp\left[x_j' \beta\right]}{\sum_{\ell \in R(t_j)} \exp\left[x_\ell' \beta\right]} \tag{C.4}$$

It is important to note that in constructing the partial likelihood of β in $(C.4)$, no information in relation to β can be obtained for where no consumers leave the BPAY network in the intervals (t_{j-1}, t_j), $j = 1, ..., k+1$, $t_0 = 0$ and $t_{k+1} = \infty$. As $h_0(t)$ is not specified, it suggests a failure-free interval (t_{j-1}, t_j) can yield minimal information about β. Assuming information relating to the distribution of $h_0(t)$ is known, there would be contributions to the inference about β from the intervals with no failures. Equation $(C.4)$ is transformed into its logarithmic form to more easily solve for β:

$$L(\beta) = \sum_{j=1}^{k} \left\{ x_j'\beta - \ln \left[\sum_{\ell \in R(t_j)} \exp\left[x_j'\beta\right] \right] \right\} \tag{C.5}$$

The maximum likelihood estimate, $\hat{\beta}$, from equation $(C.5)$, can be obtained as a solution to the vector equation

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= \sum_{j=1}^{k} \left\{ x_j' - \frac{\sum_{\ell \in R(t_j)} x_\ell' \exp\left[x_\ell'\beta\right]}{\sum_{\ell \in R(t_j)} \exp\left[x_\ell'\beta\right]} \right\} \\ &= \sum_{j=1}^{k} \left\{ x_j' - \sum_{\ell \in R(t_j)}^{k} w_{j\ell}(\beta)\, x_\ell' \right\} \end{aligned}$$

$$U(\beta) = \partial \log L / \partial \beta = \sum_{j=1}^{k} \left[x_j' - \bar{x}_{w_i} \right] = 0 \tag{C.6}$$

where

$$\bar{x}_{w_i} = \sum_{\ell \in R(t_j)} w_{j\ell}(\beta)\, x_\ell'$$

and

$$w_{j\ell}(\beta) = \frac{\exp\left[x_\ell'\beta\right]}{\sum_{m \in R(t_j)} \exp\left[x_m'\beta\right]} \tag{C.7}$$

Similarly, the observed information matrix is

$$l(\beta) = -\frac{\partial \log L}{\partial\beta\partial\beta'} = \sum_{j=1}^{k} V(\beta, t_j) \tag{C.8}$$

where

$$V(\beta, t_j) = \sum_{\ell\in R(t_j)} [x'_\ell - \Xi(\beta, t_j)]^{\otimes 2}\, w_{j\ell}(\beta)$$

is the covariance matrix of x_ℓ under the distribution $(C.7)$. As a consequence, $I(\beta)$ is typically positive definite for all β, the log likelihood is strictly concave and the estimate $\hat{\beta}$ is typically unique. The value $\hat{\beta}$ that maximises $(C.4)$ can be obtained by a Newton-Raphson iteration utilising $(C.6)$ and $(C.8)$.

Asymptotic results completely analogous to those for parametric likelihoods apply under quite general conditions. In the absence of ties, the asymptotic distribution of $\hat{\beta}$ is normal with mean β and estimated covariance matrix $I\left(\hat{\beta}\right)^{-1}$. Thereby, inference based on the l^{th} component β_l of β can be based on the asymptotic result:

$$\hat{\beta}_l - \beta_l \approx N\left(0, \hat{I}^{ll}\right)$$

where $\hat{I}^{ll}$ is the (l, l) element of $I\left(\hat{\beta}\right)^{-1}$. Likelihood ratio tests can be based on the partial likelihood, and the score statistic $U(\beta_0)$ can be used to test $\beta = \beta_0$ with χ^2 and normal asymptotic results.

C.2 Estimating the Survival Function

C.2.1 Kaplan-Meier estimator

The Kaplan and Meier (1958) estimator is a nonparametric method to estimate a survival function. The estimator at any point in time is obtained by multiplying a sequence of conditional survival probability estimators. Each conditional probability estimator is obtained from the number of individuals in the risk set and the number of individuals leaving the BPAY platform during the i^{th} period and is written as:

$$h_i = 1 - \frac{d_i}{n_i}$$

where n is the number in the risk set and d is the number of failures. Subjects that leave the BPAY platform contribute to the number at risk until they leave the platform, from which point they contribute to the value of the numerator. Individuals that are censored contribute to the number at risk until they are lost to follow up. The estimated survivor function can be expressed as follows:

$$\begin{aligned} S_t &\equiv \Pr(T > t) \\ &= \Pr(T > t | T > t-1) \Pr(T > t-1) \\ &= (1 - h_t) S_{t-1} \\ &= \Pi_{s=1}^{t} (1 - h_s) \end{aligned}$$

A convenient feature about the Kaplan-Meier estimator is that no adjustments are required for tied failure times in estimating the survivor function.

C.2.2 Cox (1972) Baseline Survivor Function

The expression for the survival probability that accompanies the Cox (1972) model is given by equation (4.2.13). This implies that once estimates of the regression coefficients are estimated, an estimate of the baseline survivor function is needed to compute the survival probability, given a set of covariates. A likelihood-based approach, which assumes that the hazard is constant between observed survival times, is applied to estimate the baseline survivor function. Lawless (2003) provides in-depth details which are summarised below.

The framework of the likelihood approach is to replicate the procedure of the Kaplan and Meier (1958) estimator of the survival function. The key point in that development is the use of the quantity $\hat{\alpha}_i = 1 - \frac{d_i}{n_i}$ as an estimator of the conditional survival probability at observed survival time t_i. The Kaplan-Meier estimator of the survival function is the product of estimators of the individual conditional survival probabilities. The expression for the conditional survival probability that leads to this estimator is $\alpha_i = S(t_i)/S(t_{i-1})$. Let $\alpha_{0i} = S(t_{0i})/S_0(t_{i-1})$, and it follows that the survival probability is:

$$\frac{S\left(t_{(i)}, x, \beta\right)}{S\left(t_{(i-1)}, x, \beta\right)} = \frac{\left[S_0\left(t_{(i)}\right)\right]^{\exp(x'\beta)}}{\left[S_0\left(t_{(i)}\right)\right]^{\exp(x'\beta)}} = \left\{\frac{S_0\left(t_{(i)}\right)}{S_0\left(t_{(i)}\right)}\right\}^{\exp(x'\beta)} = \alpha_{0i}^{\exp(x'\beta)} \quad \text{(C.9)}$$

Maximum likelihood methods are employed, conditional on the partial likelihood estimator of the regression coefficients, $\hat{\beta}$, in the model. To simplify the notation, we let $\hat{\theta}_l = \exp(x'\beta)$, and the estimator of the conditional baseline survival probability is obtained by solving the equation:

$$\sum_{i \in D_i} \frac{\hat{\theta}_l}{1 - \alpha_{0i}^{\hat{\theta}_l}} = \sum_{i \in R} \hat{\theta}_l \tag{C.10}$$

where R_i refers to the individuals in the risk set at ordered observed survival time $t_{(i)}$ and D_i denotes the individuals in the risk set with survival times equal to $t_{(i)}$. If there are no tied survivals, D_i contains one individual and the solution to $(C.10)$ is:

$$\hat{\alpha}_{0i} = \left[1 - \frac{\hat{\theta}_l}{\sum_{i \in R} \hat{\theta}_l}\right]^{\hat{\theta}_l^{-1}} \tag{C.11}$$

In the event of tied survival times, expression $(C.10)$ is solved by iterative methods. The estimator of the baseline survival function is the product of the individual estimators of the conditional baseline survival probabilities:

$$\hat{S}_0(t) = \prod_{t_{(i)} \leq t} \hat{\alpha}_{0i} \tag{C.12}$$

where $\hat{\alpha}_{0i}$ is the solution to expression $(C.10)$. Alternatively, the baseline hazard function can be estimated as:

$$\lambda_0\left(t_{(i)}\right) = 1 - \hat{\alpha}_{0i}$$

Bibliography

Arellano, M. and Bover, O. (1995). "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68(1), 29–51.

Baltagi (2000). "The local power of some unit root tests for panel data." *In B. Baltagi (ed.), Nonstationary Panels, Panel Co-integration, and Dynamic Panels: Advances in Econometrics*. Elsevier Science.

Berry, S. T. (1994). "Estimating discrete-choice models of product differentiation." *RAND Journal of Economics*, 25(2), 242–262.

Bolt, W., Jonker, N., and van Renselaar, C. (2010). "Incentives at the counter: An empirical analysis of surcharging card payment and payment behaviour in the netherlands." *Journal of Banking & Finance*, 34, 874–885.

Borzekowski, R., Elizabeth, K. K., and Shaista, A. (2008). "Consumers' use of debit cards: Patterns, preferences, and price response." *Journal of Money, Credit and Banking*, 40(1), 149–172.

Brant, R. (1990). “Assessing proportionality in the proportional odds model for ordinal logistic regression.” *Biometrics*, 46(4), 1171–1178.

Breusch, T. S. and Pagan, A. R. (1980). “The lagrange multiplier test and its applications to model specification in econometrics.” *Review of Economic Studies*, 47(1), 239–53.

Brits, J. and Winder, C. (2005). “Payments are no free lunch.” *Occasional Studies*, De Nederlandsche Bank 3(2), 874–885.

Ching, A. and Hayashi, F. (2010). “Payment card rewards programs and consumer payment choice.” *Journal of Banking & Finance*, 34(2), 1773–1787.

Choi, I. (2001). “Unit root tests for panel data.” *Journal of International Money and Finance*, 20(2), 249–272.

Cox, D. R. (1972). “Regression models and life-tables.” *Journal of the Royal Statistical Society Series B Methodological*, 34(2), 187–220.

Dawson, A. and Hugener, C. (2006). “A new business model for card payments.” *Research discussion papers*, Diamond Management & Technology Consultants, Inc.

Dickey, D. A. and Fuller, W. A. (1979). “Distribution of the Estimators for Autoregressive Time Series With a Unit Root.” *Journal of the American Statistical Association*, 74(366), 427–431.

EC (2006). "Interim report 1: Payment cards." *Report No. 2006*, European Commission.

EIM (2007). "Point-of-sale payments in the netherlands: Costs and revenue of merchants." *Report No. 2007*, De Nederlandsche Bank.

Engle, R. F. and Granger, C. W. J. (1987). "Co-integration and error correction: Representation, estimation, and testing." *Econometrica*, 55(2), 251–76.

Feigl, P. and Zelen, M. (1965). "Estimation of exponential survival probabilities with concomitant information." *Biometrics*, 826–838.

Gardner, G. and Stone, A. (2009). "Competition between payment systems." *Rba research discussion papers*, Reserve Bank of Australia.

Grambsch, P. and Therneau, T. (1994). "Proportional hazards tests in diagnostics based on weighted residuals." *Biometrika*, 81(3), 515–526.

Guthrie, G. and Wright, J. (2007). "Competing payment schemes." *Journal of Industrial Economics*, 55(1), 37–67.

Hadri, K. (2000). "Testing for stationarity in heterogeneous panel data." *Econometrics Journal*, 3(2), 148–161.

Harris, R. D. F. and Tzavalis, E. (1999). "Inference for unit roots in dynamic panels where the time dimension is fixed." *Journal of Econometrics*, 91(2), 201–226.

Hayashi, F. and Keeton, W. R. (2012). "Measuring the costs of retail payment methods." *Economic Review*, (Q II), 1–41.

Hayashi, F. and Klee, E. (2003). "Technology adoption and consumer payments: Evidence from survey data." *Review of Network Economics*, 2(2), 8.

Heckman, J. J. (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." *Annals of Economic and Social Measurement, Volume 5, number 4*, NBER Chapters, National Bureau of Economic Research, Inc, 475–492.

Hlouskova, J. and Wagner, M. (2005). "The performance of panel unit root and stationarity tests: Results from a large scale simulation study." *Economics Working Papers ECO2005/05*, European University Institute.

Hosmer, D., Lemeshow, S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. J. Wiley, 2nd edition.

Huffman, W. E. and Mercier, S. (1991). "Joint adoption of microcomputer technologies: An analysis of farmers' decisions?." *The Review of Economics and Statistics*, 73(3), 541–546.

Humphrey, D., Willesson, M., Lindblom, T., and Bergendahl, G. (2003). "What does it cost to make a payment?." *Review of Network Economics*, 2(2), 7.

Humphrey, D. B. (2010). "Retail payments: New contributions, empirical results, and unanswered questions." *Journal of Banking & Finance*, 34(8), 1729–1737.

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). "Testing for unit roots in heterogeneous panels." *Journal of Econometrics*, 115(1), 53–74.

Kalbfleisch, J. and Prentice, R. (2002). *The statistical analysis of failure time data*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. J. Wiley, 2nd edition.

Kaplan, E. and Meier, P. (1958). "Nonparametric estimation from incomplete observations." *Journal of the American statistical association*, 457–481.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root?." *Journal of Econometrics*, 54(1-3), 159–178.

Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley-Interscience, New York.

Levin, A., Lin, C.-F., and James Chu, C.-S. (2002). "Unit root tests in panel data: asymptotic and finite-sample properties." *Journal of Econometrics*, 108(1), 1–24.

Maddala, G. S. and Wu, S. (1999). "A comparative study of unit root tests with panel data and a new simple test." *Oxford Bulletin of Economics and Statistics*, 61(0), 631–52.

Milne, A. (2006). "What is in it for us? network effects and bank payment innovation." *Journal of Banking & Finance*, 30(6), 1613–1630.

Ng'andu (1997). "An empirical comparison of statistical tests for assessing the proportional hazards assumption of cox's model." *Statistics in Medicine*, 16(6), 611–626.

Pesaran, H. H. and Shin, Y. (1998). "Generalized impulse response analysis in linear multivariate models." *Economics Letters*, 58(1), 17–29.

Pesaran, M. H. (2007). "A simple panel unit root test in the presence of cross-section dependence." *Journal of Applied Econometrics*, 22(2), 265–312.

Quantin, C., Moreau, T., Asselain, B., Maccario, J., and Lellouch, J. (1996). "A regression survival model for testing the proportional hazards hypothesis." *Biometrics*, 52(3), 874–885.

Rochet, J.-C. and Tirole, J. (2003). "Platform competition in two-sided markets." *Journal of the European Economic Association*, 1(4), 990–1029.

Rochet, J.-C. and Wright, J. (2010). "Credit card interchange fees." *Journal of Banking & Finance*, 34(8), 1788–1797.

Rysman, M. (2004). "Competition between networks: A study of the market for yellow pages." *Review of Economic Studies*, 71(2), 483–512.

Rysman, M. (2007). "An empirical analysis of payment card usage." *Journal of Industrial Economics*, 55(1), 1–36.

Schoenfeld, D. (1982). "Partial residuals for the proportional hazards regression model." *Biometrika*, 69(1), 239–241.

Schuh, S. and Stavins, J. (2010). "Why are (some) consumers (finally) writing fewer checks? the role of payment characteristics." *Journal of Banking & Finance*, 34(2), 1745–1758.

Simon, J., Smith, K., and West, T. (2009). "Price incentives and consumer payment behaviour." *RBA Research Discussion Papers rdp2009-04*, Reserve Bank of Australia.

Simon, J., Smith, K., and West, T. (2010). "Price incentives and consumer payment behaviour." *Journal of Banking & Finance*, 34(8), 1759–1772.

Sims, C. A. (1980). "Macroeconomics and reality." *Econometrica*, 48(1), 1–48.

Wang, Z. (2010). "Market structure and payment card pricing: What drives the interchange?." *International Journal of Industrial Organization*, 28(1), 86–98.

Zinman, J. (2009). "Debit or credit?." *Journal of Banking & Finance*, 33(2), 358–366.