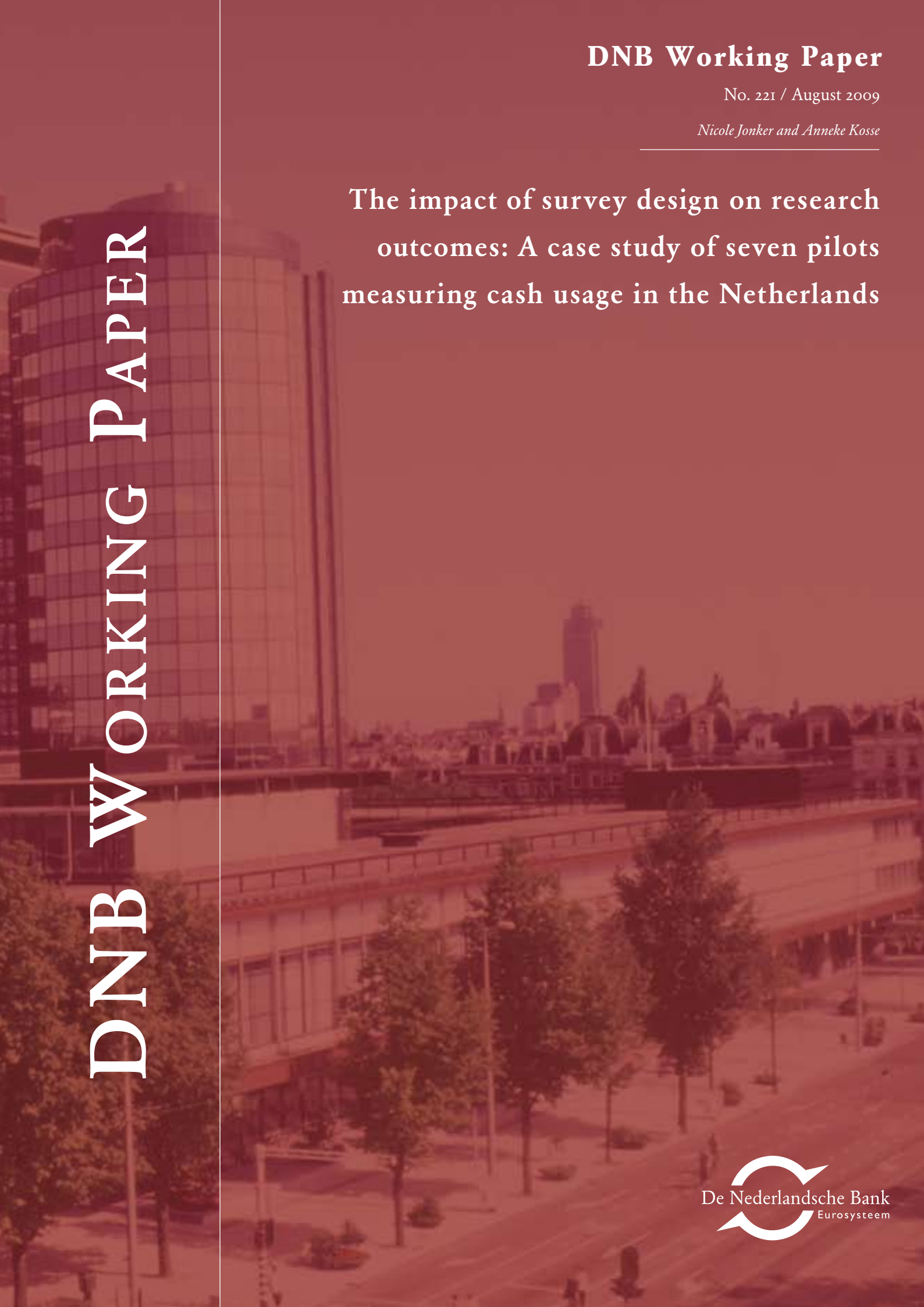
DNB Working Paper

No. 221 / August 2009

Nicole Jonker and Anneke Kosse

The impact of survey design on research outcomes: A case study of seven pilots measuring cash usage in the Netherlands

DNB WORKING PAPER



De Nederlandsche Bank
Eurosysteem

The impact of survey design on research outcomes: A case study of seven pilots measuring cash usage in the Netherlands

Nicole Jonker and Anneke Kosse *

* Views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank.

Working Paper No. 221/2009

August 2009

De Nederlandsche Bank NV
P.O. Box 98
1000 AB AMSTERDAM
The Netherlands

The impact of survey design on research outcomes: A case study of seven pilots measuring cash usage in the Netherlands [*]

August 2009

Nicole Jonker and Anneke Kosse

Cash and Payment Systems Division

De Nederlandsche Bank

Abstract:

We employ a unique dataset of transaction records to analyse whether results of consumer surveys are influenced by the survey set-up. We have asked more than 5,000 consumers to report their payments using seven different data collection methods. The results of the seven pilot studies were validated against actual payments data from retailers and the owner of the Dutch debit card scheme. The results of both the validation exercise and econometric analyses reveal that both the data collection mode and the length of the registration period significantly influence consumers' registration of payments. Measurement errors are minimised when consumers are asked to use a self-reported transaction diary for one single day.

Keywords: cash, efficiency, payment behaviour, survey design, measurement error
JEL-codes: C42, D12, E41

[*] Contact information: Nicole Jonker, phone: +31-20-5242759, e-mail: n.jonker@dnb.nl or Anneke Kosse, phone: +31-20-5242827, e-mail: j.c.m.kosse@dnb.nl, address: De Nederlandsche Bank, Cash and Payment Systems Division, P.O. Box 98, 1000 AB Amsterdam, The Netherlands. The authors like to thank Wilko Bolt, Hans Brits, Maarten van Rooij, Peter Vlaar and seminar participants at DNB, the Payment habits 2010+ workshop organised by the Bank of Finland and the Central Bank Forum for their comments. We also like to thank Richard Heuver and Mustapha Setta for excellent research assistance. The views expressed in this paper are ours and do not necessarily reflect those of the Nederlandsche Bank or the European System of Central Banks. All remaining errors are our own.

1 INTRODUCTION

Sample-based surveys are an accepted research technique among researchers in various fields, such as economists, physicians, psychologists, sociologists and anthropologists. Using questionnaires and self-reported diaries, researchers analyse specific features of populations and social phenomena. Despite their widespread use, there is an ongoing, lively debate on the quality of survey data. All sample-based surveys are subject to various types of error causing results to deviate from the actual population figures.

Whereas there is a wealth of literature on the magnitude and nature of survey errors in health, expenditure and income research, there is hardly any literature on the errors arising in payment research. Existing literature suggests that surveys on consumers' payment behaviour are sensitive to measurement error and other types of survey error. In the Netherlands, estimates on the annual number of cash payments range between 3 and 7 billion. Studies vary widely in data collection methods and show evidence of survey errors. Which methodology is most accurate remains unclear, however, since actual cash usage data are hard to obtain.

Having accurate information on the number of cash payments is relevant for several reasons. It is essential for monitoring the substitution process of cash by cards and for assessing the efficiency of the payment system. Especially for cost efficiency studies, an accurate estimate of the number of cash payments is of vital importance (see e.g. Brits and Winder, 2005; Bergman *et al.*, 2007; Gresvik and Haare, 2009). Moreover, the number of cash payments reflects the demand for cash to be used in point-of-sale (POS) transactions. This is useful information for central banks, which are responsible for the production and distribution of banknotes.

The aim of this paper is to assess the size and nature of the survey error arising in consumer surveys and to analyse the impact of several survey characteristics on estimates of cash and debit card usage. To this end, we set up an extensive survey to create a unique set of transaction data collected from 5,400 consumers in seven simultaneously held pilot studies, each one using a different research methodology. Transaction data supplied by retailers and the owner of the Dutch debit card scheme allow us to validate the survey results with actual payments data. We employ the pilot results to examine what research design reduces overall survey error best. We focus especially on the impact of survey design on biases resulting from incomplete recall, telescoping, social desirability and selection bias.

Both the validation exercise and econometric analyses show that consumer surveys may be suitable for collecting accurate data on cash usage. Survey set-up clearly influences the survey estimates. Evidence is found that respondent registration of cash payments benefits from the usage of self-reported diaries as opposed to retrospective recall surveys. However, the survey period should be short in order to keep respondents motivated to complete the diary accurately. The source from which the sample is drawn is of no influence: Internet panels provide the same results as telephone panels. This study adds to the existing literature in that it provides an explanation for the variation between existing cash estimates. Due to its

extensive and solid approach, this study may also be useful for researchers of other research fields interested in measuring consumer expenditure or the incidence of everyday routine consumer events.

The structure of this paper is as follows. Section 2 provides an overview of the literature on survey errors and measurement errors. Section 3 formulates and discusses six research questions regarding the factors that may influence the quality of survey data. These questions form the basis of the set-up of the seven pilot studies in which consumers are asked to report all their payments, each pilot using a different survey method. Section 4 validates the results from the pilot studies against actual payment data provided by Currence (the owner of the Dutch debit card scheme) and by retailers. Section 5 introduces the Zero Inflated Negative Binomial (ZINB) count data model that we use to explain the number of payments registered per person per day and also presents the estimation results. Section 6 summarises and concludes.

2 LITERATURE REVIEW

2.1 Overall survey error

Sample based surveys, such as questionnaires, self-reported diaries and experiments, are accepted research techniques among researchers in various fields. Yet they are all subject to various types of errors that may cause results to deviate from reality, such as sampling errors, coverage errors, non-response errors and measurement errors. Sampling errors arise from the process by which a sample is selected from among the population, while coverage errors arise where certain population groups are overlooked by the sampling process. Non-response errors relate to the failure of sampled persons to respond, and measurement errors occur when surveys do not measure what they purport to measure. Sampling errors can be quantified from the data on the basis of confidence levels. However, because survey data do not provide evidence of the other types of error, the overall survey error is hard to estimate. Some studies have tried to estimate its size and nature by comparing different studies and assuming one to reflect reality (see e.g. Ahmed *et al.*, 2006 and Gibson and Kim, 2007). Others use validation data gathered from administrative sources which may themselves suffer from inaccuracies (see e.g. Estelami and Lehmann, 2001 and Lynn *et al.*, 2004).

There is little literature on survey errors in payment research. Existing evidence suggests that surveys investigating consumer payment behaviour are sensitive to errors. In the Netherlands, several consumer surveys have been conducted to estimate the number and relative share of cash payments. Results have varied widely, however. In Boeschoten (1992) respondents were asked about their current transaction behaviour by means of a written questionnaire. According to this study, 64% of all transactions were paid in cash. GfK/Currence (2005-2007) asked respondents to keep a paper transaction diary for seven days. This study generated estimates of between 2.7 and 3.2 billion cash transactions per year and a cash share between 65% and 67%. According to TNS Nipo (2003), cash usage is considerably higher: 7.6 billion transactions per year. In this study respondents were asked to fill in a detailed online questionnaire about their current transaction behaviour. The respondents were drawn from the TNS Nipo Capi@home panel often used for

electronic surveys. These examples of differing cash usage estimates indicate the presence of survey error. Which estimate is most accurate remains unclear, however, since actual cash usage data are hard to obtain.

Payment surveys may be especially sensitive to coverage, sampling and measurement errors. Lately, online household panels have become common research instruments since they allow for relatively fast and low-cost data collection. These panels may cause coverage error, however, since people without Internet access may be excluded from participation. This may lead to underrepresentation of certain groups of people with specific payment behaviour patterns, such as the elderly and low-income groups. In addition, online panels may introduce selection bias. People participating in such 'modern' types of panel may also have a relatively positive attitude towards electronic means of payment. However, measurement errors may be the most common and problematic type of error to arise in payment surveys.

2.2 Measurement error

There is a great deal of literature on various types of measurement error, each one distorting the survey results in a different way. In general, measurement errors tend to be related to incomplete recall, social desirability or telescoping.

Measurement error – incomplete recall

Respondents participating in payment research may have difficulty in recalling their past payment behaviour accurately. The extent to which respondents are able to recall their payment behaviour may be affected by the number of days over which they have to report their payments. The further respondents have to go back in history, the harder they may find it to correctly recall all payments. The literature is unanimous in that there is a negative relationship between the length of time between the interview and the event and the share of events reported (see e.g. Sudman and Bradburn, 1973; Linton, 1982).

Remarkable (salient) events are events that are of more importance in a person's life and likely to be easier to remember. It is therefore to be expected that low-value payments will be easily forgotten, because of their high frequency and low budget impact. Since low-value purchases are most often paid in cash (see e.g. Bounie and François; 2006, Borzekowski *et al*., 2007), payment surveys may be especially vulnerable to underreporting of cash payments. This corresponds to assertions by Gibson and Kim (2007) and Sudman and Bradburn (1973) on the likelihood of frequent food purchases and small expenditures being omissed.

Krosnick and Alwin (1987) state that respondents try to satisfy the interviewer with a minimum level of effort and suggest that a higher response burden increases the risk of underreporting. This burden can be reduced by asking information which can be easily retrieved from existent sources (such as bank statements, bills etc), by shortening the participation period, or by providing clear instructions, detailed questions, a recall list and incentives (see e.g. Sudman and Bradburn, 1973 and Gibson and Kim, 2007). Additionally, the mode of data collection may also affect consumers' survey responses. Compared to face-to-face interviews and recall surveys, self-reported diaries have been shown to generate more accurate and detailed information and to suffer less from heaping and rounding (Sudman and Lannom, 1980; Scott and Okrasa,

1998; Gibson, 2002; Battistin *et al*., 2003; Ahmed *et al*., 2006). There are no clear guidelines as to how long a diary period should be. Diary surveys lasting several days are attractive for collecting many observations in a relatively easy way. They may, however, place a considerable burden on the respondents, which could seriously affect their compliance and hence the validity of the results. Tincello *et al*. (2007) show that multiple-day diaries are sensitive to "diary fatigue" (loss of commitment and accuracy after a couple of days) and "diary despair" (immediate loss of commitment and accuracy due to the prospect of completing a diary for several days) and therefore lead to underreporting. Ahmed *et al*. (2006) report similar phenomena, especially for small items and purchases from convenience stores.

Measurement error – social desirability

Sometimes respondents may recall past events in great detail but still not report them. This could be related to social desirability. People may be reluctant to reveal their true behaviour when they perceive this behaviour to be disapproved of by others. Bound *et al*. (2001) find that survey measures of benefit receipts are subject to underreporting due to conscious suppression. Conversely, respondents may invent 'good' behaviour, resulting in overreporting. Social desirability may play an important role in payment surveys. Respondents can have various reasons for reluctance to report all their transactions. Such 'secret' transactions are often made in cash because of its guaranteed anonymity. Therefore, payment surveys may suffer especially from the underreporting of actual cash usage. On the other hand, consumers in the Netherlands are increasingly encouraged to use debit cards. This may create a tendency to satisfy the researcher by reporting relatively many debit card payments resulting in an upward bias of debit card usage.

Measurement error – telescoping

Another phenomenon causing measurement error is telescoping. Telescoping refers to respondents being mistaken about the exact moment in time an event occurred and may result in underreporting (see e.g. Bound *et al*., 2001; Lynn *et al*., 2004). Telescoping can also lead to overreporting. Brown *et al*. (1985) find evidence of saliency being related to forward telescoping: events which are clearly remembered are found to be perceived as occurring more recently, leading to overreporting. Sudman and Bradburn (1973) share this finding and demonstrate that the use of aided recall, where respondents are confronted with possible answer possibilities, may encourage telescoping.

Telescoping can also come in in payment surveys and contaminate the results. This may be particularly true for salient payments, such as high value purchases, leading to overreporting of debit card usage. Multiple-day diary surveys may be especially vulnerable to telescoping errors. "Diary fatigue" and "diary despair" may bring respondents to postpone updating the diary until the end of the survey period, thus needing to remember back several days. This possible connection between diary surveys and telescoping has not yet been tackled in the different survey research fields.

2.3 Impact of survey participation on consumer behaviour

Even in the absence of survey errors, consumer surveys may still yield questionable results, as when respondents depart from normal behaviour due to participation in the survey. Participants in payment diary surveys may (unintentionally) make more payments or pay differently than usual. One reason to do so could be the social desirability effect, while keeping a payment diary may also reduce respondents' expenses, since it provides a clear insight into respondents' own payment behaviour, triggering them to adjust it.

3. METHODOLOGY

In this paper, we use seven different consumer survey methods to assess which method yields the narrowest gap between true and stated payment behaviour, and to assess the impact of research methods on consumers' registration of payments. This section presents the design of the seven pilot studies we conducted and introduces our research questions. Also, it presents the key statistics on the pilot studies, which provide a first idea as to whether and how pilot design may influence respondents' registration of payments. We use this information in section 5 to estimate the impact of survey design on the number of payments recorded by the respondents and to test our research questions.

3.1 Pilot study set-up and introduction to the research questions

Section 2 presented several potential sources of survey error arising in consumer surveys. They refer to the way the sample is drawn, the usage of a diary for event registration, the length of the registration and reference period, the type of questionnaire used to collect information retrospectively, and the saliency and social desirability of the type of reported events. In order to investigate how consumers' stated payment behaviour is influenced by these methodological factors, and to assess what method provides the most reliable estimates on the number of cash payments, we design and conduct seven pilot studies.

Table 1: Summary set-up of pilot studies

	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6	Pilot 7
One-day registration							
One-week registration							
Transaction diary							
End-of-day Internet survey							
End-of-day telephone survey (sample from Internet panel)							
End-of-day telephone survey (sample from telephone database)							
Interim reminder after 3-4 days							

Pilot 1 is considered as the base pilot. In this pilot, consumers are asked beforehand to record their payments during one specific day. They receive a diary to record, for each transaction they make, the amount involved, the payment instrument used and the type of spending place of each transaction made. At the end of the day, they are asked to fill in an Internet survey on all the payments made during the day, using their diary as a reminder. The other six pilots are variations on the base pilot (see Table 1). The pilots are designed in such a way that answers can be found to six research questions:

Q1: Does payment registration using a transaction diary in addition to a retrospective online survey (pilot 1) lead to a different recording of payments compared to merely filling in an online survey (pilot 3)?
Q2: Does filling in an online questionnaire at the end of the day, in addition to keeping a transaction diary (pilot 1) lead to different payment recording compared to merely filling in a transaction diary (pilot 4)?
Q3: Does a retrospective telephone survey (pilot 2) lead to different payment recording compared to a retrospective online survey (pilot 1)?

By these three questions we intend to establish what mode of data collection encourages consumers most to register all their payments. Payment registration by means of a retrospective questionnaire only may lead to underreporting of payments because of incomplete consumer recall. On the other hand, due to the personal contact or the ability to add highly specific questions, telephone interviews or online questionnaires may help consumers to recall payments they initially forgot to register. However, telephone interviews may be sensitive to social desirability error because of the personal contact with the interviewer, leading to either under- or overreporting certain payments.

Q4: Does sampling respondents from an Internet panel (pilot 2) lead to different results compared to sampling respondents from a database of people who are used to being interviewed by phone (pilot 5)?

This research question addresses the issue of whether online panels tend to be biased in favour of electronic payments. People participating in such ‘modern’ types of panel may have a stronger than average preference for electronic means of payment and therefore use cash less.

Q5: Does the length of the registration period affect the recording of payments by consumers (pilot 4 versus pilot 7)?
Q6: Does a telephone reminder after three or four days influence payment registration of consumers participating in a one-week diary survey (pilot 6 versus pilot 7)?

These research questions focus on the impact the length of the registration period may have on consumers' payment registration. The longer the registration period, the higher the burden on the respondent will be. This may lead to underreporting due to diary fatigue and diary despair, or to overreporting of 'salient' payments due to telescoping. Interim reminders may possibly reduce time-related measurement errors.

3.2 Data collection

All pilots were conducted from Thursday August 30 until Wednesday September 29, 2007. In addition, the pilots are equal with respect to the types of payment included (see table A.2 in the Appendix for a list) and the research population (Dutch consumers aged 12 to 75). We kept these three factors constant across the pilots in order to rule out any possible variation in the estimation results due to differences in these factors. Pilots 1 to 5 are based on a one-day payment registration survey and carried out by the research bureau TNS Nipo. Each day, a different group of respondents reported their payments. Pilots 6 and 7 are one-week diary surveys in which respondents were asked to report their transactions for one entire week, starting on Thursdays. These one-week pilots were conducted by research bureau GfK[1]. Because the respondents in pilots 1 to 5 participate for one day only, the samples in these pilots are larger than those in pilot 6 and pilot 7, in order to ensure sufficient numbers of observations and to enable us to make reliable statistical comparisons.[2]

3.3 Descriptive statistics on registered payments

Key statistics of pilot studies

Table 2 shows the key statistics on the seven pilot studies regarding the number of respondents and the number of recorded payments. The number of respondents varies between 494 in pilot 6 to 1077 in pilot 5. Of the respondents participating in the one-day surveys, more than 75% recorded at least one payment. In pilots 6 and 7, 87% of the respondents recorded at least one payment during their registration week. The total number of reported payments over the entire research period varies between 1,279 in pilot 3 and 4,369 in pilot 7. Due to the longer registration period, the number of payments in pilots 6 and 7 is much higher than that in pilots 1 to 5.

The statistics in Table 2 are based on raw, unweighed data. In order to make fair comparisons across the pilots, the data have been reweighed in order to represent the entire Dutch population aged between 12 and 75 years (see Appendix 2). Table 3 presents the weighted payment estimates, decomposed by payment instrument. The number of reported cash payments per person per day is 1.5 to 3.3 times the

[1] It might be possible that not all the differences between the TNS Nipo and GfK pilots are due to the differences in registration period. A part might also relate to the type of research agency used. In order to minimise the impact of employing two research agencies, meetings were organised to discuss the set-up of the pilots. This enabled the research agencies to align their research methods (sample selection, data-collection, diaries, etc.).

[2] Previous research conducted by TNS Nipo revealed that sample sizes of 600 respondents are needed in order to make a difference of 0.1 in the average number of payments per person a day statistically significant at the 5% level.

Table 2: Key statistics on pilots studies (unweighed data)

	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6	Pilot 7
Number of respondents	845	1017	638	1077	831	499	494
Share of respondents who paid at least once	80%	75%	76%	81%	79%	87%	87%
Number of payments recorded	1926	2129	1279	2427	1827	4368	4321

number of reported debit card transactions. Average transaction amounts vary between EUR 17.79 in pilot 1 to EUR 26.93 in pilot 3. The average value of a debit card transaction is 2.0 to 3.7 times that of a cash transaction. These figures show that there are substantial differences between the different pilots in the number and value of registered payments and in the relative reported usage of cash compared to the debit card.

Distribution of transaction numbers

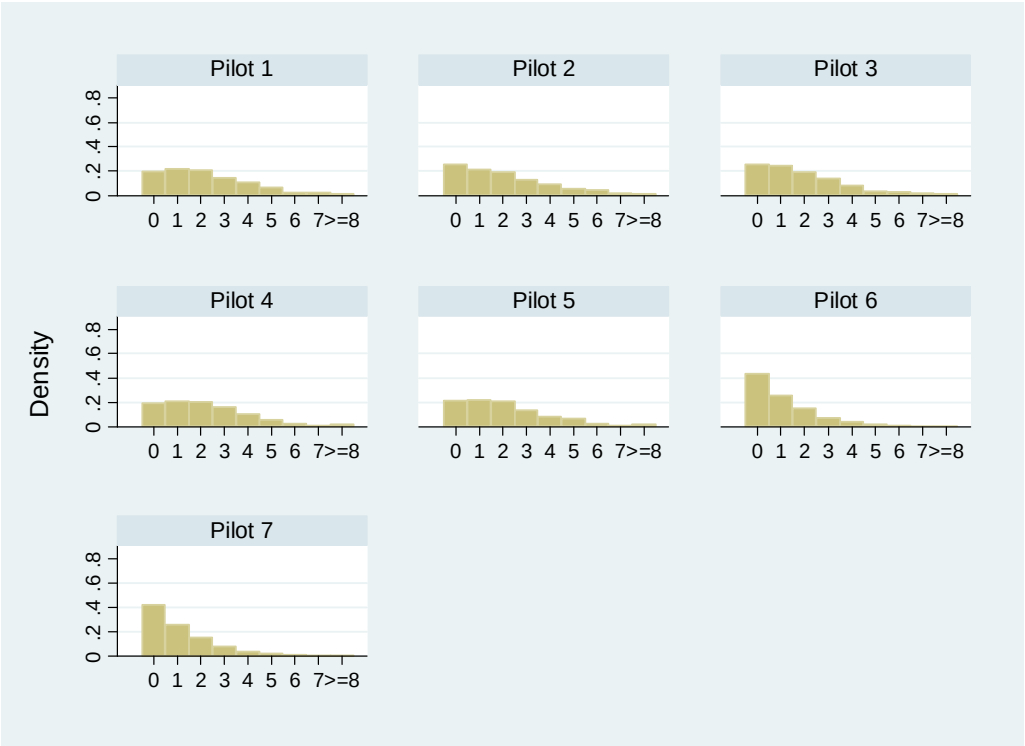
The primary focus of this paper is on *numbers* of transactions. The daily number of transactions may either be zero or take on a positive integer value. Graph 1a depicts, for each pilot separately, the distribution function of the total number of registered payments per respondent per day. It provides a first insight into why the different pilots differ in the number of recorded transactions. The densities of zero, one and two payments are almost equal in pilots 1 through 5. The densities decline gradually for higher numbers of

Table 3: Key statistics on pilots studies (reweighed data)

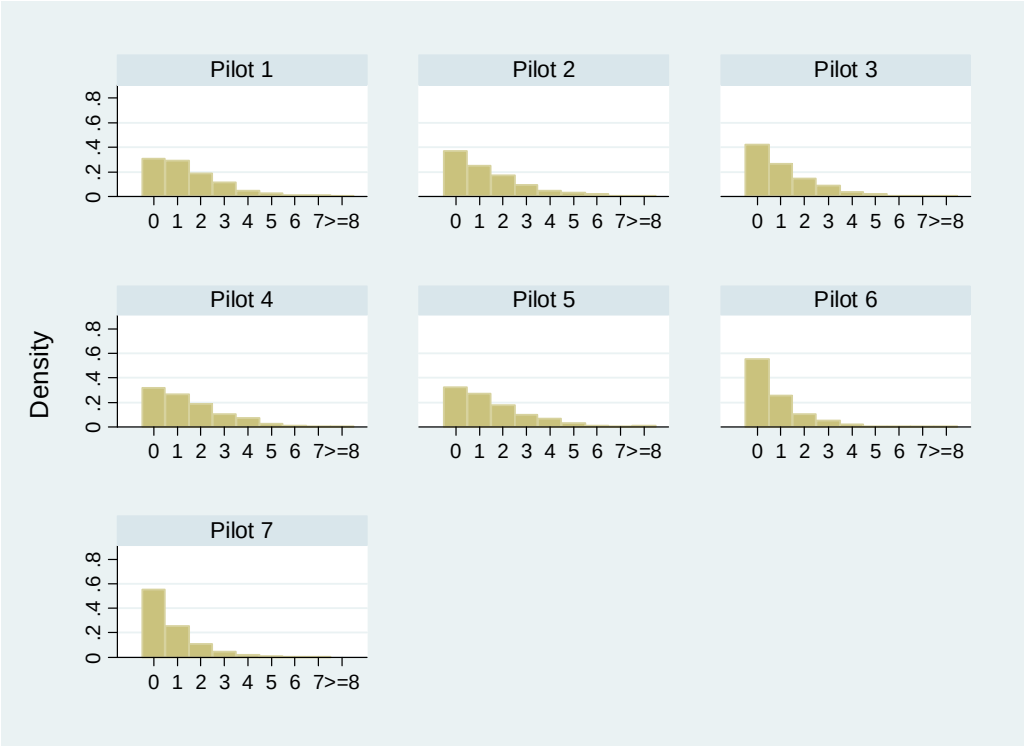
	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6	Pilot 7
Avg. number of payments (pppd) *	2.33	2.20	1.99	2.28	2.23	1.10	1.10
σ^2 number of payments (pppd)	3.90	3.89	3.55	3.81	4.03	2.16	2.11
Avg. number of cash payments (pppd)	1.54	1.42	1.14	1.54	1.60	0.75	0.73
σ^2 number of cash payments (pppd)	2.53	2.49	2.02	2.70	3.00	1.37	1.26
Avg. number of debit card payments (pppd)	0.65	0.64	0.70	0.58	0.55	0.31	0.30
σ^2 number of debit card payments (pppd)	0.98	1.07	1.02	0.87	0.89	0.50	0.50
Avg. transaction amount (in EUR)	17.79	19.09	26.93	18.70	19.58	22.60	23.95
Avg. transaction amount cash (in EUR)	10.92	10.68	18.88	11.85	13.42	13.95	17.61
Avg. transaction amount debit card (in EUR)	37.19	39.66	38.44	36.72	40.05	39.33	37.02

* pppd = per person per day

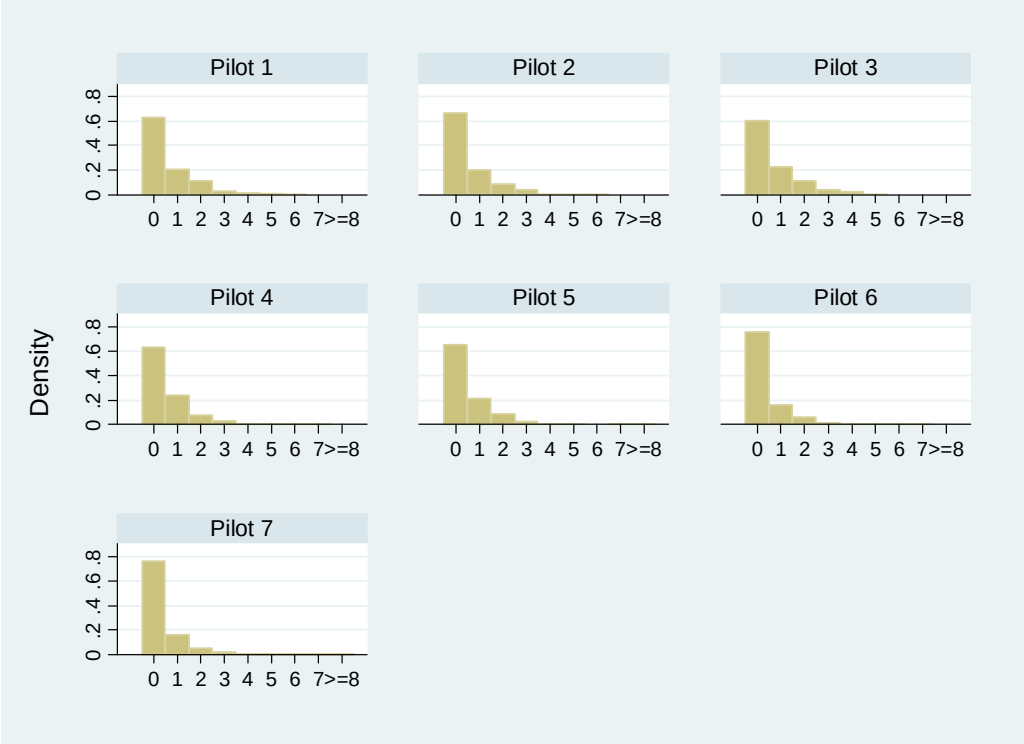
Graph 1a: Number of recorded payments per respondent per day



Graph 1b: Number of recorded cash payments per respondent per day



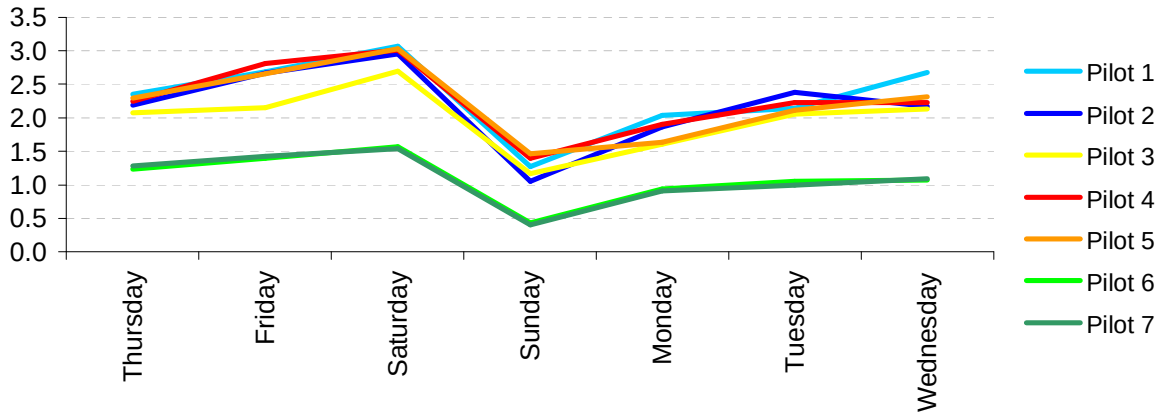
Graph 1c: Number of recorded debit card payments per respondent per day



payments. The distribution functions of pilots 6 and 7 look different, with a relatively high share of zero reported payments. When focusing on cash payments only (Graph 1b), the differences between the densities of zero, one and two payments become larger. Whereas all pilots show a declining density function, they differ in steepness. Pilots 1, 2, 4 and 5 decline gradually, whereas the density functions of pilots 6 and 7 are relatively steep, with a high incidence of zero recorded payments. With respect to the number of recorded debit card payments (Graph 1c), all pilots show steeply declining density functions, with densities of 0.6 or higher for zero recorded payments.

The descriptive statistics and distributional graphs reveal that data on the number of payments recorded per person per day follow a discrete distribution taking on non-negative integer values only. Since zero and low positive values occur most frequently and values higher than eight are rare, count data regression models are to be preferred over linear regression models to explain consumers' daily payments registration. The standard Poisson model, which assumes equality of the mean and the variance, may be too restrictive for our analysis as the average variances in the number of recorded payments exceed the average means. In addition, a high share of respondents reported zero payments which may indicate that the data suffer from the 'zero inflation' problem. This holds for both cash and debit card payments. These outcomes suggest that a suitable alternative for the standard Poisson model may be the Zero Inflated Negative Binomial model which takes into account the fact that many people reported no payments at all and the fact that the variance in the number of registered payments exceeds the average number of payments (see also Section 5).

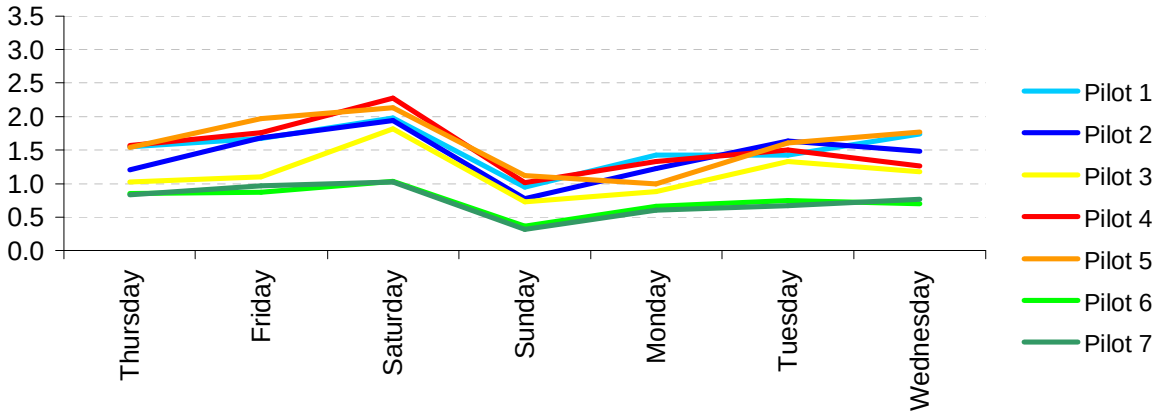
Graph 2a: Average number of payments recorded per person per day



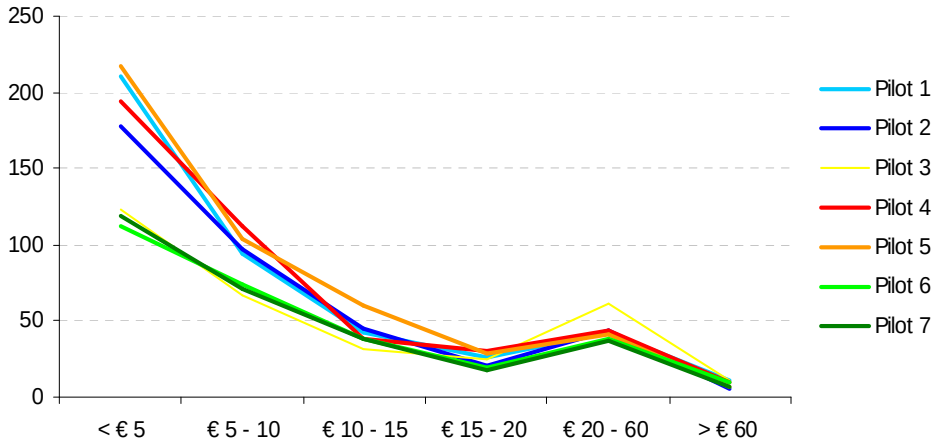
Week pattern of registered payments

Graph 2a shows that the average daily number of payments registered varies over the week. Most payments are made on Fridays and Saturdays, whereas on Sundays, when most shops are closed in the Netherlands, the fewest payments are made. From Monday onwards, the number of payments increases steadily again. For all days of the week, the estimates of the one-week pilots (pilots 6 and 7) lie below the estimates of pilots 1 to 5. The interim reminder used in pilot 6 seems to lead to a slightly higher registration of payments during the second half of the week. Of all the one-day pilots, it is pilot 3 (the pilot without diary registration) that generates the lowest number of recorded payments for almost all days. When focussing on cash payments, a very similar picture emerges (see Graph 2b). The main difference is the large gap between pilot 3 and the other one-day pilots.

Graph 2b: Average number of recorded cash payments per person per day



Graph 3: Total number of recorded payments by transaction size
(in millions)



Distribution of payments by transaction size

Graph 3 presents, for each of the seven pilots, the number of payments recorded by transaction size. In all pilots, the number of payments declines sharply with the transaction size. About 1/3 of all transactions have transaction values of EUR 5 or less. The number of registered payments below EUR 5 is highest for pilots 1, 2, 4 and 5 and lowest for pilots 3, 6 and 7. The differences between the pilots converge with increasing transaction values. For transaction sizes over EUR 15, the pilots yield very similar numbers of payments. This finding suggests that the influence of pilot design on consumers' payments recording varies with the transaction size. This may point to the presence of incomplete recall errors related to the saliency of specific payments.

4. VALIDATION OF PILOT RESULTS

Section 3 showed that the registration of payments differs per pilot. In this section, we validate the pilot results, in order to assess to what extent they correspond to actual payment behaviour in the retail trade. The results will reveal which pilots correspond well with information from the retail trade and which suffer most from measurement error. Two validation sources were used. First, actual transaction data on debit card payments for the month of September 2007 were employed, furnished by Currence. Furthermore, we used information from a sector study by EIM (2007). This study is unique in that it contains reliable estimates on cash usage, based on information from 31 large retail chains and 350 small and medium-sized retailers. Our assessment of the performance of the seven pilots is focussed on cash payments. Yet we also report test results with respect to debit card payments.

There are two caveats to keep in mind when comparing the pilot results with real transaction data. First, the pilot results and the information supplied by the retailers may not match, even assuming that all respondents registered their payments accurately. One reason for this is that the retailers' data include

payments made by consumers who are excluded from the pilot studies, such as people older than 75 years, children, tourists and non-Dutch residents. These people may use cash relatively more often than our respondents. Therefore, the pilots may underestimate actual cash usage and overestimate actual debit card usage. Second, some inequalities may be due to a difference in reference periods. The pilot data were collected in September 2007, whereas the retailer data refer to 2006. Given the ongoing substitution of cash by card payments, it seems only natural that the pilot results with respect to the number of cash payments, the cash share and the average cash transaction amounts, are somewhat lower than the retail figures.

Table 4 presents the results of the validation tests. We performed one-sample t-tests in order to establish how well the pilot results reflect the actual payments data[3]. The average transaction sizes found in pilots 4 and 5 do not differ significantly from the actual figure of EUR 18.60. Nor does the average cash transaction size found in these pilots and in pilot 2 differ significantly from the information provided by EIM. Especially the EUR 9.71 in pilot 5 is very close to the average amount of EUR 9.60 reported by the retailers. The other pilots overestimate the average transaction size and especially the average transaction size of cash payments, pointing at possible measurement errors related to incomplete recall of small cash transactions. The average debit card sizes in pilots 1, 2, 3 and 6 do not differ statistically from the actual size. The other pilots, however, show significantly lower sizes.

The average number of payments recorded per person per day is overestimated in the one-day registration pilots, but underestimated in the one-week pilots. Focussing on cash payments only, the results of pilots 1, 2 and 5 do not differ significantly from the information provided by EIM. The respondents in pilot 4 reported significantly more cash payments, however, whereas the respondents in pilots 6 and 7 underreported the number of cash payments. The overreporting in pilot 4 may signal survey errors related to

Table 4: Results validation tests

Variable	Retail data	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6	Pilot 7
Avg. transaction size (EUR)	18.60	21.34*	23.19*	27.34*	19.32	18.02	23.07*	23.98*
- cash (EUR)	9.60	12.26*	11.18	14.40*	11.50	9.71	12.82*	15.51*
- debit card (EUR)	41.55	38.30	41.98	38.11	36.66*	35.52*	38.88	36.41*
Avg. number of payments pppd	1.00	1.26*	1.30*	1.12*	1.32*	1.23*	0.70*	0.72*
- cash	0.72	0.75	0.77	0.57*	0.84*	0.80	0.44*	0.46*
- debit card	0.26	0.47*	0.50*	0.51*	0.43*	0.41*	0.26	0.24*
Share of cash payments	0.72	0.64*	0.59*	0.49*	0.66*	0.67*	0.61*	0.63*
Share of debit card payments	0.27	0.34*	0.39*	0.46*	0.30*	0.32*	0.36*	0.33*

* denotes a significant difference at the 5% level between the pilot and the validation source. Pilot results are based on weighted data.

[3] The retailer data on cash payments are partly based on scanner data from large retail chains and partly on survey estimates among small and medium-sized enterprises. Therefore two-sample t-tests may have been preferable above one-sample t-tests. Unfortunately, the EIM study does not provide information on standard errors. One-sample t-tests are more conservative than two-sample t-tests. Some tests which turn out to be significant could have appeared insignificant in a two-sample test. However, we do not expect the main results to change much when using two-sample t-tests. The differences between the information from retailers and from consumers are often large, and the p-values of the equality tests are well below the critical value of 0.05.

social desirability or faulty sampling, whereas the underreporting in the two one-week pilots again indicates the presence of incomplete recall error. The picture is completely reversed for debit card payments. Pilot 6, and to some extent pilot 7, correspond very well with the actual average number of debit card transactions. The one-day registration pilots, however, all overestimate the actual number. The overestimation is lowest in pilot 5 and highest in pilot 2. This failure of the one-day pilots to correctly estimate the number of debit card payments may signal measurement errors related to social desirability. Since the participants of these pilots are asked to record their transactions for one day only, the probability of registering a debit card transaction is relatively small. Therefore, due to the nationwide campaigns to encourage debit card usage, respondents may have felt 'obliged' to 'invent' a debit card payment in order to satisfy the researcher. The much higher probability of reporting a debit card payment in the one-week pilots may have reduced the need for 'invention'. In addition, the participants in the one-week pilots had more time to check their debit card payments on their bank statements.

Since all pilots either overestimate the number of debit card payments or underestimate cash usage, they all underestimate the share of cash and overestimate the share of debit cards in total payments. Pilot 5 approximates the share of cash best, whereas pilot 4 has the debit card share closest to reality. The shares found in pilot 3 diverge most from the information provided by EIM.

Summarising, the results indicate that the design of pilots 2 and 5 are most suitable for assessing cash usage, the primary focus of this study. The registered payments in pilot 5 do not differ significantly from our validation sources with respect to the average transaction amount, the average cash transaction amount and the average number of cash payments recorded per person per day. Pilot 2 also reflects reality well with respect to average cash transaction size and to the average number of cash transactions. The results show that underreporting errors due to incomplete recall or social undesirability of (certain) cash transactions are fairly modest in pilots using a one-day transaction diary. The deviation falls within the 95% confidence interval and part of the downward difference may be due to the exclusion of payments made by children, the elderly or other frequent cash users excluded from our samples. Furthermore, the good performance of pilot 2 suggests that online panels do not, or no longer, differ significantly from 'traditional' panels with respect to cash usage.

The estimation results of pilot 5 reveal that, once reweighed to reflect the Dutch population aged between 12 and 75, Dutch consumers made 456 million cash payments in the Netherlands in September 2007, of which 220 million in the retail trade and 72 million in the hotel and catering industry. 46 million cash payments did not take place in a business context but referred to cash payments to family, friends or charities. Assuming that Dutch consumers' payment behaviour in September is fairly representative of their payment behaviour during the entire year suggests that the number of cash payments by the Dutch in 2007 amounted to 5.5 billion, of which 4.9 billion concerned payments by consumers to businesses. [4]

[4] Information about monthly cash withdrawals, which are mainly used to make POS payments, and debit card transactions support the assumption that the month September represents the entire year fairly well with respect to the number of payments made by consumers in the Netherlands.

5 ECONOMETRIC MODEL

5.1 The Zero Inflated Negative Binomial model

In order to analyse the impact of survey design on consumers' payments registration, the data gathered in the seven pilot studies are used to estimate an econometric model. The event of interest is the number of (cash) payments recorded per person per day, denoted by y_i for observation i (i=1,…, n). As shown in section 3.3 this number may be either zero or take on a positive integer value. The majority of respondents in the seven pilot studies reported zero, one or two transactions per day. Therefore, common least squares regression models are not suitable to use, since they ignore this restricted character of the dependent variable, leading to significant deficiencies (Cameron *et al.*, 1998). Because of this, we use a count data model, in which the dependent variable is restricted to be a nonnegative random variable taking on discrete values only. The benchmark model for count data is the Poisson distribution. An important property of the Poisson distribution is the equidispersion assumption of the mean and variance being equal. However, in our pilot results (see section 3.3), the variance of the count event exceeds its average, i.e. the count data are overdispersed.[5] Negative binomial count models are less restrictive and therefore offer a good alternative. They allow the variance to exceed the conditional mean and can be used to analyse tail behaviour more thoroughly than the Poisson model. A second reason to depart from the standard Poisson regression is the 'zero inflation' problem. The survey statistics presented in section 3.3 show a frequency of zeros above what would be consistent with the standard Poisson model (Mullahy, 1986; Lambert, 1992). 'Zero' observation counts may arise from different sources. A 'zero' can be recorded by those who did not make any purchase that day and by those who forgot or refused to report payments. Respondents may also report zero payments if they (almost) never use the specific payment means in question. The important thing is that the reasons for reporting zero payments may differ from the determinants of the total number of payments recorded where respondents registered at least one transaction. So the payments data, including both zero and positive counts, may in reality be a mixture of two data generating processes, of which the mass is concentrated at zero. One process always generates zeros, whereas the other process stems from the negative binomial distribution and may generate both zero and positive counts. The results of a Bernouilly trial determine which process is appropriate:

$$y_i \sim \begin{cases} 0 & \text{with probability } \varphi_i \\ f(y_i \mid x_i) & \text{with probability } 1-\varphi_i \end{cases} \tag{1}$$

with probabilities

$$P(Y_i = y_i \mid x_i, z_i) = \begin{cases} \varphi(\gamma' z_i) + (1-\varphi(\gamma' z_i))\, g(0|x_i) & \text{if } y_i = 0 \\ (1-\varphi(\gamma' z_i))\, g(y_i|x_i) & \text{if } y_i > 0_i \end{cases} \tag{2}$$

[5] An indication of the magnitude of overdispersion can be obtained comparing the sample mean and variance of the dependent count variable. If the sample variance is more than twice the sample mean, data are likely to be overdispersed, even after inclusion of regressors. A sound practice is to estimate both Poisson and negative binomial models and to perform a LR test or a Wald test (Cameron *et al.*, 1998) as we have done in the remainder of this section.

Function *g* denotes the negative binomial model and x_i a vector of variables explaining the number of reported payments. We assume that the probability φ_i comes from the logistic distribution and depends on a vector of explanatory variables z_i that explain the likelihood of failure to report any payments at all.

5.2 Measurement error

Measurement error in count data will introduce biases when these data are used for parameter estimations. Following Cameron and Trivedi (1998) we denote y^t as the unobserved true number of payments made. Instead of y^t, we only observe the reported number of payments y^0. The difference between them, the error term, is denoted by ν. The observation-specific error terms are assumed to be independent from each other and to come from the same Normal distribution.

$$y^0 = y^t + \nu \tag{3}$$

Following Cameron et al (1998), one simple model is

$$y^t|\mu \sim \mathrm{P}[\mu] \tag{4}$$

$$\nu|\gamma \sim \mathrm{P}[\gamma], \tag{5}$$

where the true counts are distributed with mean μ, ν is distributed with mean γ and y^t and ν are independently distributed. This implies that

$$y^o|\mu,\gamma \sim \mathrm{P}[\mu+\gamma]. \tag{6}$$

Assume that each payment has a constant probability π $(0 \le \pi \le 1)$ of being reported, then the expected number of recorded events y^o is $E[y^o|\mu,\pi] = \mu\pi$, where $E[y^t|\mu] = \mu$ is the true mean and π is the average recording probability. Then y^o is a downward-biased estimate of the true mean of the payments μ. Therefore, accounting for this type of error may result in improved inferences about estimation results. This could be done by using simple nonlinear regression to parameterise μ and π. In previous research, zero inflated count models have been used in order to deal with the overreporting of zeros.[6] We therefore follow this approach and use the ZINB model to analyse the number of recorded (cash) payments.

5.3 Explanatory variables used in the ZINB model

In our case, the ZINB model contains two sets of explanatory variables: the explanatory variables stored in the vector *x* determining whether an observation falls in the 'always zero payments' process, and the explanatory variables stored in the vector *z* explaining the number of payments recorded per day, if any.

Because of the limited information available on the individuals in the samples and since the same factors may well govern both processes, we decided to include the same socio-demographic variables in *x* as we did

[6] The results of the validation tests reveal that underreporting of payments especially holds for cash payments. Debit card payments are not under- but overreported. A zero deflated model to analyse the number of debit card payments may therefore be more suitable. In this paper we still employ a zero inflated model, because it does tackle another important issue, namely that the population may consist of heterogeneous groups of people with different preferences for specific payment instruments. It might be interesting for future research to further analyse this issue and to explore the possibilities of zero deflated models and finite mixed models.

in *z*.[7] The socio-demographic variables used are dummies indicating the respondents' age, educational level, gender and residential region. We also include a dummy variable indicating whether the respondent has Internet access at home. This variable reflects the electronic-mindedness of respondents' households. In addition, dummies are included indicating the day the transaction is reported. Last but not least, we include dummy variables indicating the pilot study in which the respondents participated, in order to assess the impact of the design of the pilot on payment recording after having corrected for all other variables. The reference pilot used is pilot 5: the one which performs best in the validation test with cash transaction data provided by retailers. If pilot design significantly influences the 'zero payments process', this would suggest that somehow by its design the pilot encourages or discourages respondents to register their payment behaviour. If pilot design influences payment registration on account of the social desirability factor or their ability to recall their payments, the pilot dummies will appear to be significant in the 'negative binomial' process determining the number of payments recorded.

5.4 Estimation results

The influence of pilot design

Table A.3 presents the results of the ZINB models estimated to assess the impact of the pilots' design and consumers' personal characteristics on consumers' payments registration. The model is estimated for all payments recorded (the general model), as well as for cash and debit card payments separately and for specific transaction values. The estimated parameters α, which relax the equidispersion property of the standard Poisson model, are all significantly different from zero. This indicates the presence of excess zero registrations and supports our choice for the ZINB count data model.

Once corrected for the standard socio-demographic variables and the registration day (see Appendix 3 for a discussion of estimation results for these variables), the design of the pilots significantly affects consumers' registration of payments. This holds for both the registration of zero payments and the number of payments recorded. The regression results of the general model reveal that the respondents in pilots 6 and 7 (the one-week pilots) reported significantly fewer transactions than those participating in pilot 5. No significant differences are found between pilot 5 and the other one-day pilots at the 5% level of significance. The results for the transaction size-specific models show that the gap between the two one-week pilots on the one hand and pilot 5 on the other is largest for small transaction sizes. While for transaction sizes above EUR 15, pilots 6 and 7 still yield significantly fewer payments, the relative differences do become smaller. Pilots 6 and 7 also yield significantly more zero payments at the 10% level of significance. These results support the conclusions in the literature that one-week diaries may lead to underreporting of non-salient events, such as small purchases. This may stem from diary fatigue or even diary exhaustion given the relatively high number of zero payments.

[7] We also run regressions with different variables stored in *x* and *z*. However, this hardly altered the estimation results.

Respondents in pilot 3 also reported significantly fewer low-value payments below EUR 15 than those in pilot 5. However, they reported significantly more payments from EUR 15 upwards. Apparently, using a retrospective recall survey only may lead to recall errors relating to small transactions. The overestimation of more expensive purchases may point to errors caused by heaping and rounding where respondents did recall some of the small purchases but, forgetting the exact transaction amounts, rounded them upwards. These findings support the conclusions in the literature that recall surveys suffer more from heaping and rounding than surveys using a diary.

The results of the model estimated for cash payments only are in line with the overall findings: participants in pilots 3, 6 and 7 reported significantly fewer cash payments than those in pilot 5. One-week diary surveys and retrospective recall surveys are more vulnerable to incomplete recall error than one-day diary surveys. The impact of the pilot used on the recorded number of debit card transactions is less evident. The results do not indicate any pilot effect on the number of debit card transactions registered by those who recorded at least one payment. Debit card payments are thus less sensitive to incomplete recall errors. However, significantly fewer zero debit card transaction counts were reported in Pilot 5 than in the one-week surveys. This may suggest that one-day surveys, irrespective of the survey mode used, are more sensitive to certain forms of social desirability bias than one-week surveys.

Research questions

In order to answer the research questions formulated in section 3, the ZINB estimation results can be analysed in further detail using statistical hypothesis testing. The test results are summarised in Table A.4.

In order to answer the first research question '*Does a transaction diary in addition to a retrospective online survey lead to different recording of payments compared to solely filling in an online survey?*' we compare the results of pilot 1 (transaction diary and Internet survey) with the results of pilot 3 (Internet survey only). The hypothesis that pilot 1 respondents are as likely to have registered zero payments as pilot 3 respondents cannot be rejected. The data do, however, reject the hypothesis that pilot 3 respondents reported as many payments as those in pilot 1; they reported significantly fewer (cash) payments below EUR 15. No significant differences could be found for payments above EUR 15. These results show that incomplete recall errors relating to small value payments can significantly be reduced by adding a transaction diary.

The second research question is '*Does filling in an additional online questionnaire at the end of the day lead to different payment recording compared to solely filling in a transaction diary?*' The idea behind this question is that people who keep a diary and fill in an online survey may, when filling in the questionnaire, recall transactions which they forgot to record in their diary. In order to test this hypothesis, we compare the results of pilot 1 (transaction diary and additional end of day survey) with the results of pilot 4 (just transaction diary). The test results reveal that neither hypothesis, i.e. that of no difference in having recorded zero payments and that of no difference in the number of recorded payments, can be rejected. Perhaps the nature of the 'end-of-day' questionnaire affects consumers' payment recording. The third research question

therefore reads as follows: *'Does a retrospective telephone survey lead to different payment recording compared to a retrospective online survey?'* Using a telephone survey might make respondents recall more transactions than using an Internet survey. That is, the interviewer can stimulate respondents to do their best to remember all their payments and can specifically ask for transactions which are easily forgotten. Therefore we compare the results of pilot 1 (transaction diary and Internet survey) with the results of pilot 2 (transaction diary and telephone survey). Again, neither hypothesis can be rejected. Nor was the joint test significant. Both questions 2 and 3 indicate that additional questionnaires, regardless of their nature, do not significantly reduce possible recall errors arising in transaction diaries.

In order to answer the fourth question, '*Does sampling from an Internet panel lead to different results compared to sampling from a database with people who are usually interviewed by phone*?', the results of pilot 2 (Internet panel) are compared with the results of pilot 5 (telephone database). The test results do not show any sign of electronic bias in pilot 2 for either small or large value payments[8]: neither the equal zero payment counts hypothesis nor the equal payment counts hypothesis can be rejected. These findings indicate that consumers participating in online surveys do not, or no longer, differ significantly from the overall population regarding their registration of cash usage.

The fifth research question is *'Does the duration of the registration period affect the recording of payments?* Based on the conclusions from the literature, we expected that respondents participating in a multiple-day diary survey are more likely to postpone registration of their transactions (due to diary fatigue), increasing the probability of incomplete recall error, or even to stop recording transactions altogether (due to diary exhaustion), leading to too many zero transaction counts. To test for this hypothesis, we compare the results of pilot 5 (one-day diary) with the results of pilot 7 (one-week diary). The test results reveal that the duration of the registration period significantly influences consumers' registration of zero and low-value cash payments. Both the hypothesis of no difference in having recorded zero cash payments and the hypothesis of no difference in the number of recorded cash payments, is rejected by the data. These results confirm that one-week surveys are indeed vulnerable to both diary fatigue and diary exhaustion.

The last question runs as follows: '*Does a telephone reminder after 3 or 4 days affect consumers' payments registration when participating in a one-week survey?*' An interim reminder such as the one given in pilot 6 might stimulate respondents to continue recording their payments accurately during the entire registration period. However, the test results show no significant differences between pilot 6 and pilot 7. Neither the hypothesis of no difference in having recorded any payment, nor the hypothesis of no difference in the number of recorded payments, is rejected by the data.[9]

Summarising, the results of the statistical hypothesis tests reveal that it is to be recommended to use a one-day diary when collecting information on cash payments.[10] This research method reduces measurement

[8] No significant differences were found for the individual transaction categories, whereas the pilot results did differ on an overall level.

[9] We also performed tests using data from the second half of the registration period only after the reminder was received. The results remained the same, however.

[10] In this paper we compare one-day registration with one-week registration. It may be interesting to examine registration of 2-3 days as well. Longer registration periods have the advantage of collecting more information against lower cost than one-day registration

error considerably and is likely to be preferred as well for collecting information on the incidence of other ordinary daily events. The added value of additional surveys at the end of the day, by phone or online, is low, although they do not harm the results either. We did not find evidence of any electronic bias reflected by a more frequent use of debit cards among respondents participating in online panels.

6. CONCLUSION

In this study we employ a unique dataset of transaction records to investigate the extent to which transaction data provided by consumers are influenced by the survey set-up, and to assess what method is best suited to estimate the total number and share of cash payments in an economy. For this, we conducted seven different pilot studies among more than 5,000 consumers, each one using a different survey method. On top of this, we collected actual payment data from retailers and from the owner of the Dutch debit card scheme. Due to our extensive dataset, we are able to mutually compare the pilot results using the Zero Inflated Negative Binomial count data model and to validate the survey results against real transaction data. The results of the validation exercise and the econometric analyses show that the methods used to gather payment information from consumers and the duration of the survey period significantly affect consumers' ability to recall and record their purchases. In addition, the results suggest that the recording of some payments, such as debit card payments, may be sensitive to social desirability bias.

The results show that using a transaction diary for one single day is preferred for assessing the value and number of cash payments. An additional retrospective questionnaire does not measurably stimulate people to recall payments which were initially missed. We also demonstrate that a one-week registration method leads to a significant increase of incomplete recall error and of zero observations, indicating the occurrence of diary fatigue and diary exhaustion. Especially small transactions up to EUR 15 turn out to be vulnerable to incomplete recall error. The 'postponing' and 'abandoning' effect of the one-week method does not lessen when an interim reminder is provided. Finally, no proof is found that using an Internet survey may introduce pro-electronic bias reflected in an overestimation of debit card usage.

This study adds to the existing literature in that it provides a thorough explanation for the variation between existing cash estimates. It underlines the importance of designing consumer surveys in such a way that recall errors and other types of survey errors are minimised. Our conclusions may not just hold for payments research. Due to its extensive and solid approach, it may also be useful for research focusing on consumer expenditures or on the incidence of common, non-salient events in general.

periods and may reduce part of the social desirability errors the one-day survey suffers with respect to the registration of debit card payments.

REFERENCES

Ahmed, N., M. Brzozowski and T.F. Crossley (2006), Measurement errors in recall food consumption data, Working Paper 06/21, The institute for fiscal studies, Cambridge.

Battistin, E., R. Miniaci and G. Weber (2003), What do we learn from recall consumption data? The Journal of Human Resources, vol. XXXVIII, 2.

Bergman, M., G. Guibourg and B. Segendorf (2007), The Costs of Paying-Private and Social Costs of Cash and Card Payments, Sveriges Riksbank Working Paper Series no. 212.

Biemer, P.P. and L.E Lyberg (2003), Introduction to survey quality, New York, Wiley.

Boeschoten, W.C. (1992), Currency Use and Payment Patterns, PhD thesis, Universiteit van Amsterdam**.**

Borzekowski, R., E. K. Kiser and S. Ahmed (2007), Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response, Journal of Money, Credit and Banking, (forthcoming).

Bound, J., C. Brown and N. Mathiowetz (2001), Measurement error in survey data in Heckman J.J. and Leamer E. (ed.s), Handbook of Econometrics vol. 5, Amsterdam: Elsevier Science.

Bounie, D. and A. François (2006), Cash, Check or Bank Card? The Effects of Transaction Characteristics on the Use of Payment Instruments, mimeo.

Brits, J.H. and C.C.A. Winder (2005), Payments are no free lunch, DNB Occasional Studies, vol. 3 no. 2, De Nederlandsche Bank NV, Amsterdam.

Brown, N.R., L.J. Rips and S.K. Shevell (1985), The subjective dates of natural events in very-long-term memory, Cognitive Psychology, Vol. 17, pp. 139-177.

Cameron, A.C., and P.K. Trivedi (1998), Regression analysis of count data, Econometric Society Monographs No. 30, Cambridge University Press.

EIM (2007), Het Toonbankbetalingsverkeer in Nederland. kosten en opbrengsten van toonbankinstellingen in kaart gebracht (POS retail payments in the Netherlands: Costs and Revenues for Retailers). Zoetermeer.

Estelami, H. and D.R. Lehmann (2001), The impact of research design on consumer price recall accuracy: an integrative review, Journal of the Academy of Marketing Science, pp. 29-36.

GFK/Currence (2005, 2006, 2007), Monitor Consumptieve Toonbankbetalingen in Nederland, Currence, Amsterdam.

Gibson, J. (2002), Why does the Engel Method Work? Food demand, economies of size and household survey methods, Oxford Bulletin of Economics and Statistics, Vol. 64(4), pp. 341-360.

Gibson, J. and B. Kim (2007), Measurement error in recall surveys and the relationship between household size and food demand, American Journal of Agricultural Economics, Vol. 89(2), pp. 473-489.

Groves, R. M. (1989), Survey errors and survey costs, Wiley, New York.

Gresvik, O. and H. Haare (2009), Costs in the Norwegian payment system, Norges Bank Staff memo 2009, no.4.

Krosnick, J.A., and D.F. Alwin (1987), An evaluation of a cognitive theory of response-order effects in survey measurement, Public Opinion Quarterly, Vol.51(2), pp. 201-219.

Linton, M. (1982), Transformations of memory in everyday life, in U. Neisser (Ed.), Memory Observed, p. 77-91 , W.H. Freemand and Co., New York.

Lynn, P., A, Jäckle, S.P. Jenkins and E. Sala (2004), The impact of interviewing method on measurement error in panel survey measures of benefit receipt: evidence from a validation study, ISER Working papers, no. 2004-28, Institute for Social & Economic Research, University of Essex, Colchester.

Mullahy, J. (1986), Specification and Testing of Some Modified Count data Models, Journal of Econometrics, Vol. 33, pp. 341-365.

Scott, K. and W. Okrasa (1998), Analysis of Latvia diary experiment, World Bank Development Research Group, Washington DC.

Sudman, S., and N. Bradburn (1973), Effects of time and memory factors on response in surveys, Journal of the American Statistical Association, Vol. 68(344), pp. 805-815

Sudman, S., and L. Lannom (1980), Health care surveys using diaries, DHHS Publication No. PHS 80-3279, Washington, DC: National Center for Health Survey Research.

Tincello, D.G., K.S. Williams, M. Joshi, R.P. Assassa, and K.R.Abrams (2007), Urinary diaries, Obstetrics and Gyneacology, Vol. 109(2), Part 1, pp. 227-280.

TNS Nipo (2003), Contante Transacties in Nederland (Cash transactions in the Netherlands), mimeo, Amsterdam.

APPENDIX 1 WEIGHTING AND EXTRAPOLATION METHOD

In order to guarantee the representativeness of the samples, the pilot results have been weighted to correct for possible biases related to the characteristics of the sample:

a) The one-day diary and questionnaire-only samples have been weighted by total observations per day, in order to prevent that the days with many participants would weight more heavily compared to the days with less participants. The same correction has been made to the two one-week diary samples on a weekly-basis.
b) The one-day diary and questionnaire-only samples have been weighted by gender, age, education, region and Internet usage, using the Gouden Standard and TNS NIPO data on Internet usage. Weighting by income, position in the family and activity appeared not to be necessary.
c) The one-week diary samples have been weighted by gender, age, education, family situation, family size, position in the family, activity, region and credit card possession, using the Gouden Standaard and TOF Particulier.
d) Finally, a correction has been made to all samples with respect to non-buyers. Experience has shown that part of the non-response is caused by persons believing that participation is not useful because the did not make any payments during the registration period. Therefore, in order to prevent any possible underestimation, the samples are adjusted, based on the assumption that on any particular day, 35% of the total population do not make any transaction.

In order to approximate the total number of transactions made in the Netherlands, the weighted survey results have been extrapolated to the total Dutch population aged between 12 and 75 years who were able to make any purchase during the specified research period (11.4 million). To this end, the total Dutch population aged between 12 and 75 years (12.8 million) is cut down by the number of people that are expected not to be able to make any transaction in the Netherlands in September 2007 (1.4 million).[11]

[11] The sum of disabled persons (0.5 mln), people living in an institution/special home (0.2 mln), the seriously physically and mentally handicapped living at home (0.2 mln) and people being abroad (0.5 mln) (Source: TNS Nipo).

APPENDIX 2 TABLES

Respondents were asked to record all the payments they made. The payments included payments in the retail trade, the catering industry, at fuelling stations and at vending or ticket machines, admission for cultural or recreational events (museum, cinema, exhibition, swimming pool, etc) and non-business payments to other persons. Table A1 shows a complete listing of the spending purposes distinguished in this study.

Table A.1. Spending purposes

1 Supermarket
2 Retail trade: food, beverages and tobacco
3 Retail tradel fashion and shoes
4 Retail trade non food (excl fashion and shoes)
5 Home furnishing, home improvement, white goods
6 Department stores
7 Vending and ticket machines
8 Fuelling stations
9 Hotel and catering industry
10 Street vending
11 Amusement, culture, sports
12 Transport
13 To family members, friends, colleagues or acquaintances
14 Charity
15 Service providers (incl. travel agencies)
16 Other

Table A.2 Payments made in Dutch retail trade, referring to Sept. 2007

Number of payments	342 million
Avg. transaction amount	EUR 18.40
Avg. number of payments pppd	1.0
Number of cash payments	246 million
Avg. transaction amount cash	EUR 9.60
Avg. number of cash payments pppd	0.72
Share of cash payments on all payments	0.719
Number of debit card payments	91 million
Avg. transaction amount debit card	EUR 41.55
Avg. number of debit card payments pppd	0.265
Share of debit card payments	0.26

Source: EIM (2007) and Currence website (www.currence.nl).

Table A.3: Estimation results ZINB model by payment instrument and transaction size
Non-negative counts

	All payment instruments				Cash payments				debit card payments		
Range transaction size	No limit	€ 0-5	€ 5-15	>=€15	No limit	€ 0-5	€ 5-15	>=€15	No limit	<€15	>=€15
p1	0.052	0.014	-0.028	0.099	-0.034	-0.030	-0.101	-0.176	0.149	0.344	0.107
p2	0.031	-0.021	-0.040	-0.094	-0.061	-0.092	-0.067	-0.280	-0.008	0.186	-0.073
p3	-0.036	-0.401*	-0.196*	0.201*	-0.261*	-0.420*	-0.405*	0.116	0.164	0.227	0.194
p4	0.078	0.066	-0.013	0.059	0.029	0.037	-0.030	-0.261	0.018	0.267	0.021
p6	-0.440*	-0.646*	-0.542*	-0.368*	-0.547*	-0.645*	-0.585*	-0.463	-0.164	0.177	-0.234
p7	-0.475*	-0.665*	-0.571*	-0.465*	-0.626*	-0.704*	-0.624*	-0.547	-0.248	0.065	-0.327
Tuesday	0.190*	0.240*	0.102	0.127*	0.223*	0.218*	0.183*	0.152	0.160	0.260	0.247
Wednesday	0.308*	0.282*	0.210*	0.340*	0.256*	0.265*	0.229*	0.181	0.336*	0.183	0.536
Thursday	0.306*	0.259*	0.254*	0.375*	0.275*	0.208*	0.250*	0.472	0.300*	0.261	0.394
Friday	0.466*	0.282*	0.425*	0.583*	0.434*	0.271*	0.483*	0.410	0.496*	0.107	0.699
Saturday	0.550*	0.313*	0.574*	0.619*	0.552*	0.355*	0.636*	0.693	0.506*	0.118	0.636
Sunday	-0.280*	-0.262*	0.179	-0.544*	-0.250	-0.169	0.205	-0.232	-0.535*	-0.388	-0.441
Female	0.198*	0.068	0.320*	0.096	0.210*	0.119*	0.336*	0.058	0.158*	-0.001	0.100
Region North	-0.030	0.003	-0.057	-0.118	0.044	0.060	-0.077	0.018	-0.114	-0.146	-0.244
Region East	-0.090*	-0.091	-0.105*	-0.159*	-0.025	-0.002	-0.077	-0.163	-0.146	-0.097	-0.218
Region South	-0.060	-0.107*	-0.114*	0.020	-0.051	-0.063	-0.079	0.247	-0.082	-0.162	-0.116
Age 12-17	-0.435*	0.216	-1.223*	-0.883	-0.400*	0.200	-1.380*	-1.774	0.178	1.336	0.202
Age 18-29	-0.229*	-0.105	-0.348*	-0.156	-0.473*	-0.318*	-0.608*	-0.527*	0.278*	1.057	0.124
Age 30-39	-0.169*	-0.194	-0.222*	-0.018	-0.354*	-0.285*	-0.451*	-0.030	0.183	0.825	0.091
Age 40-49	-0.114*	-0.086	-0.237*	0.187*	-0.241*	-0.176*	-0.400*	-0.045	0.297*	0.767	0.361
Age 50-59	-0.138*	-0.101	-0.226*	0.010	-0.228*	-0.162*	-0.329*	-0.239	0.171	0.777	0.153
Education:											
Primary ed.	-0.055	-0.413*	0.227	-0.021	0.091	-0.354	0.336*	0.210	0.059	-0.225	-0.210
Lower sec. ed.	0.079	-0.120	0.263*	0.052	0.127	-0.037	0.306*	0.185	0.023	0.160	0.009
Inrterm. sec. ed.	0.092	-0.010	0.273*	0.064	0.158*	0.025	0.301*	0.211	0.046	0.081	0.033
Higher voc. ed.	0.045	-0.116	0.146*	0.065	-0.028	-0.143	0.111	0.007	0.057	0.130	0.062
Internet at home	0.173*	0.100	0.328*	0.114	0.240*	0.209*	0.310*	0.212	0.184	0.011	0.150
Constant	0.397*	-0.269	-0.948*	-0.693*	0.058	-0.494*	-1.122*	-1.445	-1.052*	-2.392*	

Table A.3 continued: Estimation results on zero inflation

	All payments				Cash payments				Debit card payments		
Range transaction size	No limit	€ 0-5	€ 5-15	>=€15	No limit	€ 0-5	€ 5-15	>=€15	No limit	<€15	>=€15
p1	12.005	1.821	-0.320	-0.349	13.267*	2.255	-0.471	-1.912	0.268	1.164	-0.147
p2	12.788	1.174	0.307	-1.800	13.913*	1.215	-0.323	-1.142	-0.310	1.111	-1.217
p3	12.882	2.738	0.444	-3.965	14.617*	3.643	-0.192	-0.791	-0.628	0.707	-0.739
p4	12.509	2.828	-0.599	-0.561	14.750*	3.520	-0.762*	-2.704	-0.276	1.323	-0.472
p6	14.538	5.091	0.862	0.901*	16.459*	4.907	0.556	-0.147	1.753*	2.566*	1.172
p7	14.373	4.473	0.961	0.630	15.983*	4.287	-0.723	-0.047	1.591*	2.281*	1.068
Tuesday	0.707	-0.018	-2.107	0.623	0.868	-0.301	14.254*	0.269	0.663	0.627	1.152*
Wednesday	1.130*	0.980	11.028*	0.852	1.254	0.762	13.863*	0.133	0.427	0.111	1.146*
Thursday	0.524	0.133	-3.064	0.610	0.701	-0.021	2.262	0.665	0.013	-0.120	0.586
Friday	0.572	0.497	-2.505	-0.280	1.106	0.201	14.239*	-0.876	-0.310	-0.589	0.195
Saturday	0.767*	1.029	11.503*	-0.922	1.112	0.692	12.885	-0.082	-0.316	-0.661	-0.225
Sunday	2.103*	2.049*	15.742*	-0.043	2.158*	1.879*	18.610*	0.529	1.293*	1.854*	1.284
Female	-0.134	-1.112*	0.840*	-1.168*	-0.449	-0.956*	1.394	-0.873	-0.607*	-0.955*	-0.793
Region north	0.253	0.627	0.332	-0.786	0.906	0.731	-0.069	-0.002	-0.377	-0.466	-0.978*
Region east	0.265	0.450	0.121	-0.286	0.849	0.841	0.291	-0.342	-0.129	-0.001	-0.424
Region south	-0.032	-0.801	-0.528	0.348	-0.834	-0.654	-0.694	0.836	0.132	0.106	-0.177
Age 12-17	1.974*	3.020*	-1.014	5.001	2.299	2.334	-1.517	-1.300	2.844*	3.042*	4.172*
Age 18-29	0.230	0.163	-0.986*	3.347	-0.496	-0.597	-1.062	0.555	0.765	1.039	1.543
Age 30-39	-0.189	-0.444	-1.223*	2.523	-0.400	-0.503	-1.524	1.677	-0.353	0.532	0.192
Age 40-49	-0.031	0.390	-1.103*	3.536	0.089	0.152	-1.618	0.979	0.542	0.997	1.435
Age 50-59	-0.499	-0.224	-0.755	1.547	-0.586	-0.293	-1.185	-0.361	0.280	1.633	0.526
Education											
Primary ed.	-0.014	-13.808*	-2.721	1.467	0.111	-13.392*	-3.745	0.036	0.972	-0.443	0.682
Lower sec. ed.	0.070	0.733	-0.886	-0.450	0.219	0.469	-0.680	-0.322	-0.355	-0.011	-0.386
Interm. Sec. ed.	-0.084	1.031	-0.627	-0.473	0.452	0.541	-0.584	-0.091	-0.436	-0.413	-0.368
Higher voc. ed.	-0.517	-2.012	-0.554	-1.047	-2.278	-2.245	-0.542	-0.767	-0.711	-0.370	-0.677
Internet at home	-0.328	0.343	-0.162	-1.289	0.004	0.450	-0.054	0.008	-0.987	-1.357	-0.819
Constant	-16.507	-7.658*	-15.091*	-2.888	-19.331*	-6.988	-17.394	-1.111	-1.013	-1.114	-1.360
α*	0.250	0.480*	0.438*	0.297*	0.430*	0.508*	0.684*	0.470*	0.507*	0.677*	0.464
No. obs.	11461	11461	11461	11461	11461	11461	11461	11461	11461	11461	11461
Zero obs.	4053	7420	7588	7409	5457	7835	8451	9557	8185	10180	8929
LogL	-18332	-10861	-10288	-10486	-15354	-10016	-8664.2	-5988.0	-9507.1	-4536.7	-7538.8

Standard errors are adjusted for clusters in persons, * denotes a significant difference at the 5%-level

Table A.4: Impact of pilot on registration of payments, by payment instrument and transaction size

Research question	Pilots involved	Test on. equality of	All payments				Cash payments				Debit card payments		
			All	< EUR 5	EUR 5-15	>EUR 15	All	<EUR 5	EUR 5-15	>EUR 15	All	<=EUR 15	>EUR 15
3	p1 vs p2	Positive counts	P=0.65	P=0.58	P=0.85	P=0.01	P=0.68	P=0.40	P=0.67	P=0.84	P=0.13	P=0.53	P=0.08
		Zero count	P=0.20	P=0.61	P=0.34	P=0.13	P=0.68	P=0.49	P=0.87	P=0.72	P=0.51	P=0.95	P=0.26
		joint	P=0.27	P=0.81	P=0.58	P=0.01	P=0.67	P=0.65	P=0.92	P=0.55	P=0.28	P=0.50	P=0.19
1	p1 vs p3	Positive counts	P=0.09	P=0.00	P=0.03	P=0.25	P=0.01	P=0.01	P=0.00	P=0.53	P=0.87	P=0.74	P=0.42
		Zero count	P=0.19	P=0.43	P=0.31	P=0.20	P=0.30	P=0.29	P=0.78	P=0.45	P=0.32	P=0.79	P=0.39
		joint	P=0.03	P=0.00	P=0.03	P=0.32	P=0.00	P=0.00	P=0.00	P=0.66	P=0.48	P=0.95	P=0.27
2	p1 vs p4	Positive counts	P=0.52	P=0.44	P=0.82	P=0.56	P=0.23	P=0.40	P=0.37	P=0.57	P=0.14	P=0.75	P=0.40
		Zero count	P=0.41	P=0.26	P=0.66	P=0.68	P=0.29	P=0.17	P=0.76	P=0.85	P=0.24	P=0.83	P=0.50
		joint	P=0.66	P=0.48	P=0.86	P=0.83	P=0.31	P=0.35	P=0.54	P=0.84	P=0.29	P=0.62	P=0.69
4	p2 vs p5	Positive counts	P=0.57	P=0.74	P=0.57	P=0.21	P=0.43	P=0.19	P=0.45	P=0.75	P=0.95	P=0.38	P=0.57
		Zero count	P=0.32	P=0.63	P=0.53	P=0.15	P=0.00	P=0.79	P=0.74	P=0.30	P=0.71	P=0.13	P=0.24
		joint	P=0.35	P=0.80	P=0.66	P=0.00	P=0.00	P=0.35	P=0.75	P=0.51	P=0.90	P=0.32	P=0.51
5	p4 vs p7	Positive counts	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.75	P=0.05	P=0.37	P=0.12
		Zero count	P=0.00	P=0.22	P=0.00	P=0.03	P=0.14	P=0.46	P=0.16	P=0.61	P=0.00	P=0.09	P=0.02
		joint	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.88	P=0.00	P=0.01	P=0.00
5	p5 vs p7	Positive counts	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.29	P=0.08	P=0.82	P=0.12
		Zero count	P=0.28	P=0.10	P=0.09	P=0.18	P=0.00	P=0.40	P=0.60	P=0.98	P=0.01	P=0.00	P=0.12
		joint	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
6	p6 vs p7	Positive counts	P=0.42	P=0.81	P=0.60	P=0.12	P=0.23	P=0.48	P=0.63	P=0.76	P=0.38	P=0.62	P=0.31
		Zero count	P=0.51	P=0.11	P=0.79	P=0.41	P=0.26	P=0.13	P=0.87	P=0.91	P=0.60	P=0.56	P=0.72
		joint	P=0.68	P=0.22	P=0.80	P=0.30	P=0.34	P=0.32	P=0.75	P=0.40	P=0.66	P=0.85	P=0.53

APPENDIX 3 The influence of socio demographic variables on payment registration

The estimation results presented in A.3 do not only provide insight into the influence of pilot design on payment registration, but also into the extent to which day and personal characteristics influence consumers' registration of payments. The number of reported payments varies a lot during the week. The reference weekday used in the model is Monday. The results show that fewest payments are reported on Sundays. In addition, many respondents report zero payments on Wednesdays and on Saturdays. This is quite surprising, since on Saturdays most people are free from work and have time to go shopping. Moreover, the total number of registered payments is relatively large on Saturdays (followed by Fridays). This apparent contradiction might possibly indicate that consumers either go shopping on Saturdays in an extensively manner, or shop their little needs before or after the weekend to avoid the bustle.

The results show that women report significantly more payments than men. This holds for both cash payments and debit card payments. Women also report significantly more zero debit card transactions. Age also influences the number of recorded payments: recorded total payments and cash payments decline with age and the youngest respondents record significantly more zero debit card transaction counts. Moreover, respondents with a lower or intermediate educational level report significantly more (cash) payments with a transaction size between EUR 5 and EUR 15 than those having a university degree. There are some regional differences as well: people living in the eastern and southern part of the Netherlands register significantly fewer payments of particular transaction amounts than people living in the western part. Finally, the estimation results show that respondents who have Internet access at home report significantly more cash transactions of up to EUR 15 than those who have not. However, corresponding to our findings from the hypothesis tests, no evidence is found that they record more debit card transactions.

It is difficult to conclude whether these results point at particular forms of measurement error being related to personal characteristics or whether they reflect differences in purchase and payment behaviour. Further research is recommended.

Previous DNB Working Papers in 2009

No. 198 **Peter ter Berg**, Unification of the Fréchet and Weibull Distribution
No. 199 **Ronald Heijmans**, Simulations in the Dutch interbank payment system: A sensitivity analysis
No 200 **Itai Agur**, What Institutional Structure for the Lender of Last Resort?
No 201 **Iman van Lelyveld, Franka Liedorp and Manuel Kampman**, An empirical assessment of reinsurance risk
No. 202 **Kerstin Bernoth and Andreas Pick**, Forecasting the fragility of the banking and insurance sector
No. 203 **Maria Demertzis**, The 'Wisdom of the Crowds' and Public Policy
No. 204 **Wouter den Haan and Vincent Sterk**, The comovement between household loans and real activity
No. 205 **Gus Garita and Chen Zhou**, Can Open Capital Markets Help Avoid Currency Crises?
No. 206 **Frederick van der Ploeg and Steven Poelhekke**, The Volatility Curse: Revisiting the Paradox of Plenty
No. 207 **M. Hashem Pesaran and Adreas Pick**, Forecasting Random Walks under Drift Instability
No. 208 **Zsolt Darvas**, Monetary Transmission in three Central European Economies: Evidence from Time-Varying Coefficient Vector Autoregressions
No. 209 **Steven Poelhekke**, Human Capital and Employment Growth in German Metropolitan Areas: New Evidence
No. 210 **Vincent Sterk**, Credit Frictions and the Comovement between Durable and Non-durable Consumption
No. 211 **Jan de Dreu and Jacob Bikker**, Pension fund sophistication and investment policy
No. 212 **Jakob de Haan and David-Jan Jansen**, The communication policy of the European Central Bank: An overview of the first decade
No. 213 **Itai Agur**, Regulatory Competition and Bank Risk Taking
No. 214 **John Lewis**, Fiscal policy in Central and Eastern Europe with real time data: Cyclicality, inertia and the role of EU accession
No. 215 **Jacob Bikker**, An extended gravity model with substitution applied to international trade
No. 216 **Arie Kapteyn and Federica Teppa**, Subjective Measures of Risk Aversion, Fixed Costs, and Portfolio Choice
No. 217 **Mark Mink and Jochen Mierau**, Measuring Stock Market Contagion with an Application to the Sub-prime Crisis
No. 218 **Michael Biggs, Thomas Mayer and Andreas Pick**, Credit and economic recovery
No. 219 **Chen Zhou**, Dependence structure of risk factors and diversification effects
No. 220 **W.L. Heeringa and A.L. Bovenberg**, Stabilizing pay-as-you-go pension schemes in the face of rising longevity and falling fertility: an application to the Netherlands