

# Consumers' Recording Behaviour in Payment Diaries – Empirical Evidence from Germany

Tobias Schmidt, Deutsche Bundesbank

29.08.2014

**How to cite this article:** Schmidt, T. (2014). Consumers' Recording Behaviour in Payment Diaries – Empirical Evidence from Germany. *Survey Methods: Insights from the Field*. Retrieved from <http://surveyinsights.org/?p=4563>

---

## Abstract

In this paper we analyse whether the recording behaviour of consumers keeping a payment diary for several days changes over the diary period. Using data from a large study on the payment behaviour of German consumers we find that individuals tend to report a higher number of transactions on the first day of the diary period than on subsequent days. Contrary to existing literature we find that the number of small cash payments recorded does not decrease during the one-week diary period. Our findings indicate that short diaries may be enough to reflect adequately the payment behaviour of consumers. Longer diaries do not seem to suffer from serious biases and are especially helpful when it comes to analysing subgroups of payment types or rare events.

## Keywords

[diary studies](#), [payment behaviour](#), [survey design](#)

## Acknowledgement

The author thanks Heinz Herrmann, Ulf von Kalckreuth, Carlos Arango, Heike Woerlen and Kylie Smith and an anonymous referee for excellent feedback and comments. The views expressed in this paper are those of the author and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

## Copyright

© the authors 2013. This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License \(CC BY-NC-ND 3.0\)](#) 

## 1. Introduction

Knowing consumers' payment behaviour is important for central banks as well as participants in the payment system. Surveys on payment behaviour therefore have a long tradition in several central banks around the globe. In recent years, questionnaire-based surveys have been complemented with diaries in which respondents register information on individual payments for a certain period (see e.g. Bagnall et al. 2014; Jonker et al., 2012). An open issue among survey designers in this area is whether longer or shorter diaries, in terms of the number of days of recording, deliver superior data quality. With the data we have at hand, we can contribute to this discussion by investigating whether longer diaries suffer from biases and distortions. The data we use to address this question are from a large representative sample survey of adult consumers, commissioned by the Bundesbank in 2008 ("Zahlungsverhalten in Deutschland"). The dataset comprises data from over 2,000 individuals on roughly 25,500 transactions. The diary lasts seven days, enabling us to analyse recording patterns within this time frame.

Our results show that respondents do report less payments as the diary period progresses, but the decline is less pronounced than expected, i.e. the average number and value of recorded transactions per consumer per day does not decrease dramatically over the one-week diary period. Between days three and seven no significant reduction in recorded payments and their value is observed. Contrary to existing literature we also find that the number of small (cash) payments recorded does not decrease from one day to the next during the one-week diary period. The distribution of payments over transaction values is similar for each of the seven days. We interpret these findings as indicating that shorter diaries may be enough to adequately reflect the main features of payment behaviour of consumers. However, longer diaries do not seem to suffer from serious biases and are especially helpful when it comes to analysing subgroups of payment types or rare events. Also the precision of the estimates increases with longer diaries, at small additional costs.

The next section will review the literature most closely related to our analysis. We will cite studies dealing with measurement errors in diaries and develop hypotheses for the empirical part of the paper. In Section 3 we describe the data before we present results in Section 4. Based on our results we will draw conclusions and provide recommendations for future (payment) diary studies.

## 2. Literature and Research Questions

In this section we review some of the papers on measurement error in surveys in order to develop questions and hypotheses to guide our empirical analysis below. We will focus on under-reporting in particular and not cover other types of survey or measurement errors.

### 2.1 Literature Review and Hypotheses

Our paper is closely related to a recent paper of Jonker and Kosse (2013) analysing payment diary data from the Netherlands. They conduct a true experiment by varying the diary length and the mode of diary (online, paper and pencil, etc.). Their main finding is that a one-day diary is least prone to measurement error. Although the Bundesbank did not conduct any experiments on the length of diaries, we are able to provide some indication of measurement errors in long versus short diaries by comparing the reporting behaviour of consumers for each day of the seven-day diary period.

A common measurement error in data collection based on diary surveys is fatigue or exhaustion, i.e. respondents tend to record fewer items or information as the diary period progresses. This phenomenon has been well documented for expenditure surveys (Crossley and Winter, 2014; Ahmed et al., 2006; McWhinney and Champion, 1974; Turner, 1961), a type of survey closely related to payment surveys. The drop in recorded expenditures is especially strong between the first and the second day: Stephens (2003) writes about the consumer expenditure data from the US Bureau of Labour Statistics: "An examination of the data indicates that expenditures are higher during the first diary week as well as on the first day of each diary week relative to the remaining days. These declines are likely attributable to 'survey fatigue'" (Stephens, 2003: 411). Similar results have been obtained for Canadian data (Ahmed et al., 2006; Statistics Canada, 2001). Turner (1961) argues that his findings that households report less expenditure on tea in week one of a two week diary, could be the result of survey fatigue. He also mentions survey exuberance as a possible cause for day one being different. He writes "[...] (a) the first-week records overstate expenditure, possibly by including purchases actually made in the immediately preceding period; and/or (b) the second-week records understate expenditure, possibly by omitting some purchases actually made." (Turner, 1961: 145). It is worth noting that there is some evidence that even the prospects of having to complete a week-long diary may bias the results and lead to non-cooperation (Sudman and Ferber, 1971) and underreporting (Tincello et al., 2007). Our first hypothesis is based on these findings from expenditure surveys and has two parts:

*Hypothesis 1: (i) The number and volume of payments recorded declines over the one-week period. (ii) The*

*drop is especially pronounced between day one and the following days.*

Our second hypothesis is motivated by the findings from the research on measurement errors in expenditure surveys as well. Ahmed et al. (2006) and Gibson and Kim (2007) document that consumers tend to forget about small purchases (“recall bias”) and underreport them in (expenditure) diaries. This specific underreporting is also the topic of an article by Alessie et al. (1990). They begin their abstract by stating that “In consumer expenditure surveys one often faces the problem that full information on small consumption expenditures is not available.” Since most small purchases are typically paid in cash (see e.g. Bagnall et al. 2014), this should lead to an underreporting of small cash transactions in payment diaries. Jonker and Kosse (2013) recently showed that this underreporting is stronger in payment diary surveys that last one week compared to diaries that run for only a few days. They argue that this is due to at least two effects: first, respondents postpone registration of payments (even for day one of the diary period) until the later days of the weekly diary – leading to higher recall bias, or second stop recording small payments completely. The latter argument is the basis for our second hypothesis; it links the underreporting pattern observed for consumer diaries with a time dimension.

*Hypothesis 2: The number of small transactions recorded in the weekly diary decreases over the one-week period.*

The studies cited above argue based on empirical and experimental evidence that people forget about purchases or they do not accurately record all their payments. From a theoretical point of view the behaviour of the respondents to not report all payments can be related to the theory on commitment/involvement (see Albaum et al., 1998 for a review of this theory and other response theories) borrowed from sociology. Albaum et al. (1998) write, that “The less the extend of involvement, the more behaviour productive of disorder (e.g., nonresponse, deliberate reporting of false information, etc.) is perceived as legitimate.” (Albaum et al., 1998:119) and argues that therefore data quality is related to the level of involvement or commitment. Usually it is assumed that the level of commitment and involvement is determined when the potential respondent decides on whether to participate or not, but it is feasible to assume that the level of commitment can change as the survey period progresses. One mechanism that may be relevant in this respect is that with her participation decision the respondent implicitly commits a certain amount of resources for the survey, e.g. time. Then with each question asked or transaction recorded, a certain amount of this “resource budget” is withdrawn. As the end of the diary period approaches, there will be fewer resources available for recording payments and some payments may be aggregated or just not reported to save time and resources. We are not aware that this later argument has been formalised somewhere, but it is related to the well-known fact that participating in a survey has costs for respondents (see e.g. Tullar et al. 1979 cited from Albaum et al, 1998).

## **2.2 Current practices with respect to diary length**

There is very little research on the optimal length of a payment diary (except for Jonker and Kosse, 2013) and, as a result, no clear recommendation on what length to use when designing a new study. We therefore finish this section by reporting some examples for the length of payment diaries actually implemented to give an impression of what survey designers in different countries currently think is an appropriate length. Payment diaries in Germany, France and Austria use a one-week period (see von Kalckreuth et al., 2014; Bounie and Francois, 2006, Mooslechner et al., 2006). Australia's Consumer Payments Use Study now also belongs to this group, as the Reserve Bank of Australia reduced the diary length from two weeks in 2007 to one week in 2010. There are at least two new surveys using short diaries: the 2009 Method of Payments Survey of the Bank of Canada (Arango and Welte, 2012) and the 2010 Boston Fed's Survey of Consumer Payment Choices (Foster et al., 2011), which both use three-day periods. To our knowledge, the Dutch National Bank's Payment Survey of 2010 is the only large-scale survey with the diary period restricted to a single day. At the other end of the spectrum is the UK Payments Council Survey with a four-week diary.

## **3. Data and Variables**

### **3.1 The Payment Survey of the Bundesbank – Design and Content**

We utilise data from a representative survey of adults living in Germany (see Hoffmann et al., 2009 for details and results). This survey, “Zahlungsverhalten in Deutschland” (Payment Behaviour in Germany), was conducted between April and June of 2008 by Ipsos on behalf of the Deutsche Bundesbank.

The sample for the survey was drawn using a random-route procedure developed by the Association of German Market and Social Researchers, the so called “ADM” (see Hoffmeyer-Zlotnik, 2003 on random route samples). This design implies that at a first stage small geographical areas are randomly drawn from the population of all geographical areas in Germany using pps-sampling. In a second step a random starting point is drawn which is linked to a specific pre-defined (walking) route the interviewer has to follow in order to reach the households. Finally within each household one interviewee is selected using the last birthday rule. When designing the random sample of areas (first stage), care was taken to ensure that consumers from all 16 federal states (Länder) were included in the gross sample.

The survey consists of two parts, a face-to-face CAPI interview and a drop-off paper and pencil diary. The latter was supposed to be completed by the respondents in the seven days following the face-to-face interview. To spread the interviews and collected diary data across the whole survey period, the addresses/random starting points were not given to the field at one point in time, but rolled out continuously, with new addresses/starting points in the field each day. This procedure insures that interviews are conducted on all days of the week (including weekends and holidays), which implies that payment diaries were also started on each day of the week.

While the interviews contained questions on topics like the ownership and usage of certain payment instruments, cash withdrawal behaviour and respondents' demographics, the diary collected information on individual transactions and specifically refers to direct payment transactions, i.e. all transactions apart from recurrent transactions like rent payments, insurance fees, telephone bills and utility bills. The information collected in the paper and pencil diary includes the euro amount of each transaction, the location where the transaction took place (17 different locations including "retail shops", "gas stations", "restaurants", "internet", etc.) and the payment medium used to settle it (cash and a list of ten cashless payment methods, e.g. debit cards, credit cards, internet payment services, mobile phone payments, fingerprint payment). The diary contains two pages for each day with space for up to eight transactions and some spare pages if more than eight transactions occur on one day. At the top of each page of the diary the respondents were asked to fill in the date and then all transactions pertaining to this date. The printed diary also contains a page with an example of how to fill in the diary and the interviewers explained the procedure of recording transactions to the respondents when they handed over the diary.

### 3.2 Descriptive Statistics on Transactions Recorded in the Diaries

Of the 3,612 individuals sampled, 2,272 completed the CAPI questionnaire and 2,218 the CAPI questionnaire and the diary. In sum, the 2,218 respondents completing the diary provide information on 25,677 transactions with a total value of 766,761 euro.

Table 1 Summary Statistics on Transactions Recorded in the Diary (unweighted)

	Mean	SD
Number of payments (pppd)	1.65	0.87
Number of cash payments (pppd)	1.33	0.78
Total transaction amount in euro (pppd)	49.39	107.36
Total transaction amount cash in euro (pppd)	28.98	95.58
Value of individual transactions in euro	30.62	154.14
Value of individual cash transactions in euro	22.44	157.07

Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"

Note: pppd = per person per day; all figures are calculated from the raw data and are unweighted

Table 1 provides some summary statistics on the transactions. The average value of an individual transaction recorded in the diary is 30.62 euro, with the median significantly lower at 12.20 euro (see Table 4 for the distribution of payments across different transaction values). Table 1 also shows that on average each person recorded 1.65 transactions per day, which add up to a total value of about 50 euro. Given the high prevalence of cash payments in Germany (Hoffmann et al., 2009), it is not surprising that most of the transactions are cash payments.

### 3.3 Variables used for Analysis

From the data we calculate the following indicators, which are used to analyse the impact of the length of the diary on respondents' recording behaviour:

1. euro amount of all transactions and number of transactions per person per day,
2. euro amount of all cash transactions and number of cash transactions per person per day,
3. share of cash payments in total payments recorded and total value taken out in cash per person per day,
4. number of transactions – by transaction size category,
5. share of transaction size categories in total number of transactions.

Transactions were excluded from the analysis of the cash shares, for which the value of the transaction or the information of whether it was a cash or non-cash transaction was missing. This data cleaning led to a reduction by 487 transactions.

## 4. Results

In this section we present results for the data from the diary by consumer and by transaction.

### 4.1 Payment Patterns by Consumer

To test our first hypothesis as to whether the number of payments decreases over the one week reporting period and in particular between day one and the following days, we analyse the diary data for each day by consumer. If there were a clear decreasing pattern from day one to the following days our hypothesis would be confirmed.

Table 2 shows summary statistics for the average value of direct payments conducted by a respondent. The figures show that respondents recorded higher aggregate amounts on day one than on the other days of the diary, confirming the second part of our hypothesis and the literature cited above.

Table 2 Number and Value of Transaction per Person by Day of Recording (unweighted)

	All payments recorded per person		CASH payments recorded per person		Share of CASH in total transactions per person	
	Number (1)	Value (2)	Number (3)	Value (4)	Number (5)	Value (6)
Day 1 of diary p.	1.998 (0.032)	58.64 (2.945)	1.580 (0.029)	34.18 (2.298)	0.808 (0.007)	0.758 (0.009)
Day 2 of diary p.	1.706 (0.028)	46.82 (2.709)	1.368 (0.026)	26.78 (2.266)	0.813 (0.008)	0.769 (0.009)
Day 3 of diary p.	1.623 (0.029)	49.05 (4.118)	1.315 (0.026)	27.55 (3.632)	0.826 (0.008)	0.782 (0.009)
Day 4 of diary p.	1.571 (0.027)	50.70 (5.689)	1.274 (0.025)	31.57 (5.492)	0.817 (0.008)	0.774 (0.009)
Day 5 of diary p.	1.563 (0.029)	46.64 (4.598)	1.268 (0.026)	28.26 (4.388)	0.818 (0.008)	0.777 (0.009)

Day 6 of diary p.	1.533 (0.028)	42.95 (4.285)	1.233 (0.025)	27.64 (4.158)	0.815 (0.008)	0.787 (0.009)
Day 7 of diary p.	1.583 (0.027)	50.90 (4.947)	1.283 (0.025)	26.87 (3.641)	0.816 (0.008)	0.773 (0.009)

Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"  
Notes: Standard errors in parenthesis.

A pair-wise comparison (t-tests) between the mean euro amount spent on each day reveals that the means for days two to seven do not differ at any of the conventional levels of significance. However, day one sticks out. The mean euro amount spent on day one is significantly different from all other days except for day 4 at least at the 95% level. The pattern observed for day one is confirmed for the number of transaction. For the latter indicator other combinations also reveal significant differences: all comparisons of day 2 with days 3 to 7 show up significant at the 99% level, as does the combination of day 3 and day 6. These descriptive statistics confirm our first hypothesis, i.e. the number and value of transactions declines over the diary period. The descriptive statistics indicate that the effect is confined to the first day or first two days in terms of the number of transactions. There is no further decline of the number and value recorded per person per day from day 3 to 7.

Given that a key goal of the survey conducted by the Bundesbank in 2008 was to assess the importance of cash in direct payment transactions (see Hoffmann et al., 2009), we also look at cash payments in particular. For cash payments the pattern is similar to the one observed for all payments. The number of payments recorded by an individual is higher for day one and two than for the following days (column 3). However, in terms of the value of cash payments (column 4) day one only differs significantly from days two and seven. All other combinations do not reach the 95% significance level. The average cash shares in terms of the number of transactions (column 5) calculated from the diary data do not significantly differ across days. Only the difference between day one and days 3, 5 and 7 for the average cash share calculated on the basis of the value of transactions are significant. This indicates that it is not only one type of transaction, i.e. cash transactions, which is responsible for the decline between days. On the contrary, the slight all be it insignificant increase from day one to day two indicates that less (high value) non-cash payments are recorded on days two to seven.

A multivariate analysis of consumers' recording behaviour provides further evidence on how the recording behaviour of respondents changes over the diary period. We are again investigating whether and how the day of recording influences the value and volume of transactions recorded by the respondents. The multivariate setting allows us to include variables in the regressions to control for factors that may influence the recording behaviour and/or the probability of a transaction occurring. We include the day of the week (*Monday to Sunday*), age (*P\_AGE*), gender (*P\_FEMALE*), education level (*P\_EDU\_UNI*, *P\_EDU\_HIGH*) and employment status (*P\_EMPLOYED*) of the respondent, household size (*HH\_SIZE*) and income (*HH\_INC*), an indicator for whether the person reporting is responsible for (grocery) shopping for his or her household (*P\_SHOPPER*) and whether the respondent lives in East Germany (*R\_EAST*).

The results confirm by-and-large our findings from the descriptive analysis for the number of transactions presented above. We find that the day of recording does have an influence on the number of all transactions recorded, even after we control for respondent characteristics and the day of the week. For the number of transactions the coefficients of day one and two are significant. An F-test on the equality of the two coefficients can be rejected, i.e. the coefficient on day 1 is significantly higher than on day two. Taken together this indicates that the number of transactions recorded decreases from day 1 to day 2 and from day 2 to day 3, but no further significant decline is observed for days 3 to day 7; partially confirming our first hypothesis. For the number of cash payments we find a similar result. The coefficients for all days of recording are significant, but pairwise comparisons tests show that only those for day 1 and day 2 differ from the remaining ones and the coefficient of day 1 is significantly higher than the coefficient for day 2. After controlling for respondents' characteristics and day of the week, the day of recording has – contrary to our first hypothesis – no effect on the value of the recorded transactions. Here the day of recording is insignificant for all days in the observation period. This finding is also true for cash payments.

Table 3 Coefficients from OLS Regressions without a Constant (unweighted)

VARIABLES	All payments recorded per person	CASH payments recorded per person

	Number	Value	Number	Value
Day 1	0.781***	-30.255	1.230***	33.033
	[0.278]	[40.657]	[0.252]	[38.916]
Day 2	0.531*	-42.839	1.045***	24.374
	[0.276]	[40.545]	[0.249]	[38.946]
Day 3	0.416	-39.865	0.960***	25.455
	[0.276]	[40.826]	[0.249]	[39.305]
Day 4	0.376	-38.737	0.937***	29.312
	[0.275]	[40.471]	[0.249]	[38.834]
Day 5	0.384	-41.911	0.936***	26.997
	[0.275]	[44.263]	[0.248]	[42.907]
Day 6	0.362	-47.158	0.911***	24.244
	[0.274]	[39.966]	[0.248]	[38.378]
Day 7	0.340	-43.700	0.907***	22.349
	[0.275]	[40.456]	[0.249]	[38.768]
Tuesday	-0.035	-8.976	-0.025	-7.122
	[0.034]	[6.998]	[0.032]	[6.275]
Wednesday	0.023	-2.079	0.046	-2.245

	[0.033]	[7.091]	[0.031]	[6.285]
Thursday	-0.078**	-9.988	-0.02	-6.807
	[0.035]	[6.782]	[0.032]	[6.098]
Friday	0.225***	5.503	0.171***	2.725
	[0.034]	[7.624]	[0.032]	[6.988]
Saturday	0.141***	6.658	0.121***	2.806
	[0.036]	[8.224]	[0.034]	[7.681]
Sunday	-0.659***	-28.509***	-0.496***	-16.832***
	[0.031]	[6.902]	[0.030]	[6.349]
P_AGE	-0.001	0.155	0.001	0.112
	[0.001]	[0.134]	[0.001]	[0.119]
P_EMPLOYED	0.073*	4.970	-0.03	-4.352
	[0.043]	[3.758]	[0.040]	[3.209]
P_EDU_HIGH	0.104*	3.264	0.055	4.02
	[0.056]	[7.059]	[0.049]	[6.727]
P_EDU_UNI	0.334***	20.355**	0.145*	6.27
	[0.081]	[9.182]	[0.076]	[8.417]
P_SHOPPER	0.291***	10.040***	0.190***	5.107**



	[0.054]	[3.232]	[0.050]	[2.203]
P_FEMALE	-0.156***	-4.027	-0.119***	-2.354
	[0.048]	[3.696]	[0.044]	[3.004]
R_EAST	0.045	13.405**	-0.128***	0.544
	[0.051]	[5.309]	[0.042]	[4.455]
HH_SIZE	0.041*	3.966***	0.034	2.397**
	[0.024]	[1.431]	[0.023]	[1.023]
HH_INC	0.135***	8.628	0.024	-1.032
	[0.040]	[5.288]	[0.037]	[5.014]
Observations	13,972	13,972	13,972	13,972
logl	-23,265	-92,053	-22,188	-90,196
R-squared	0.640	0.077	0.570	0.033

Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"

Notes: Clustered standard errors in brackets; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Taking the evidence from the analysis at the personal level together, we can partially confirm our first hypothesis. Respondents seem to record more payments on day 1 and day 2, than on the following day. However, there is no additional drop of recorded transactions in the later days of the diary. Furthermore, we find no evidence for a drop in the average value of recorded transactions over the recording period. This finding is at first glance related to our second hypothesis, if the average number of transactions per person is decreasing, but the average value is not, the average size of a transaction recorded is supposedly smaller at the beginning of the diary period (days 1 and 2) than at the end (days 3 to 7). This conclusion is not valid, because the analysis presented above includes "transactions of value zero", i.e. if a person does not conduct any transactions on a given day the number and value of this transaction would be included in the analysis above, bringing down the averages. In order to assess whether the number of low value transactions recorded ( $>0$  euros) is declining, these transactions cannot be included. To shed more light on this and to thoroughly test our second hypothesis, we turn to an analysis of individual payments in the next section

#### 4.2 The Structure of Individual Transaction over the Diary Period

Concerning the analysis on the level of individual transactions, our results paint a picture very much in concordance with the results for consumers presented above. Table 4 and Figure 1 show that more transactions are recorded on the first day of the diary than on the following days. Regardless of which day one picks, however, the distribution across size categories would have been the same, which is evidence against our

second hypothesis.

Table 4 Number of transactions for each day – by transaction size category (unweighted)

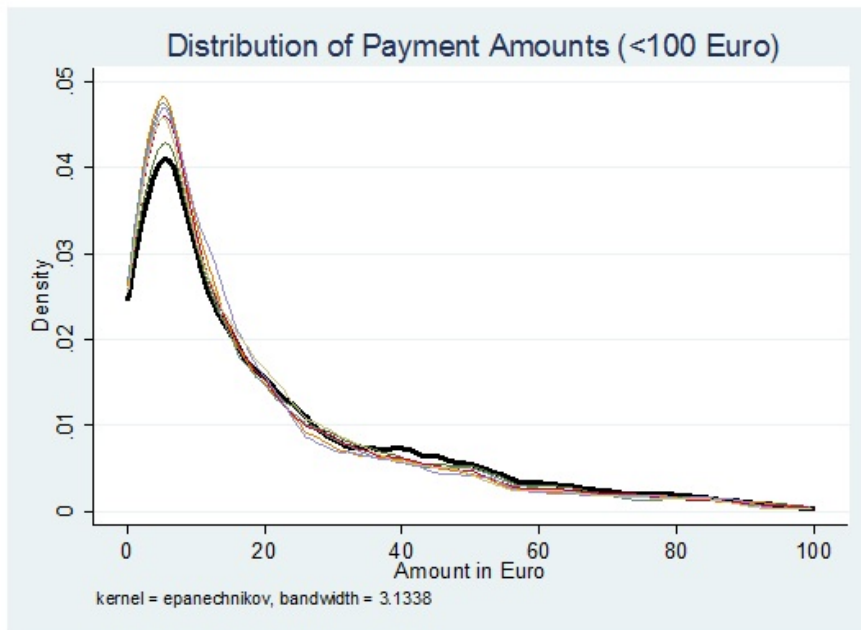
	below 5 euro	[5; 10[ euro	[10; 15[ euro	[15; 25[ euro	[25; 50[ euro	[50 ; 100[ euro	100 + euro	Total
Day 1	1,052 (24%)	722 (16%)	528 (12%)	679 (15%)	771 (17%)	519 (12%)	174 (4%)	4,445 (100%)
Day 2	914 (24%)	653 (17%)	462 (12%)	568 (15%)	636 (17%)	412 (11%)	112 (3%)	3,757 (100%)
Day 3	895 (25%)	682 (19%)	480 (13%)	508 (14%)	559 (16%)	328 (9%)	133 (4%)	3,585 (100%)
Day 4	891 (26%)	643 (18%)	428 (12%)	501 (14%)	559 (16%)	330 (9%)	142 (4%)	3,494 (100%)
Day 5	822 (24%)	652 (19%)	436 (13%)	503 (15%)	564 (16%)	341 (10%)	126 (4%)	3,444 (100%)
Day 6	814 (25%)	618 (19%)	469 (14%)	511 (15%)	481 (15%)	306 (9%)	111 (3%)	3,310 (100%)
Day 7	855 (24%)	634 (18%)	437 (12%)	561 (16%)	597 (17%)	302 (9%)	133 (4%)	3,519 (100%)
Total	6,243 (24%)	4,604 (18%)	3,240 (13%)	3,831 (15%)	4,167 (16%)	2,538 (10%)	931 (4%)	25,554 (100%)

Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"

Notes: Row percentages in parentheses. Includes all transactions for which transaction size is known.

Figure 1 below shows that the distribution of payments below 100 euros across transaction size (in euro) for each day is similar. However, day one (thick, black line) and day two (green line) are slightly different in that they have less distribution mass for payments of 5 to 10 euro and more for payments of 50 to 100 euro. A more detailed look at the distribution of transactions with values below 25 euro (Figure A1 in the appendix) reveals that, on these days, payments with a value of around 10 euro are less prevalent and those with values of 20 to 25 euro more prevalent. Similar results are obtained for cash payments only (see Figure A2 in the appendix), which is not surprising given that over 80% of transactions below 100 euro and over 90% of transactions below 25 euro are cash payments.

Figure 1 Distribution of payment values for transactions below 100 euro (unweighted)



Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"

Notes: For the sake of exposition we dropped the top 4% of the distribution. The density estimation for the full sample is available upon request.

A formal statistical test of the difference between the two distributions, the Wilcoxon-Mann-Whitney test leads to similar results. It shows (see Table A1 in the appendix) that only day one and, to a lesser degree, day two differ from the other days and here only with respect to larger payments. For payments of 25 euro and less, no significant differences (with two exceptions) show up. Taking these findings together we have to reject our second hypothesis, that the number of small transactions recorded decreases as the diary period progresses.

### 4.3 Quality Implications of Fatigue and Optimal Diary Length

We now turn to the quality implications of the observed structures. Table 2 above reports shares, amounts and standard deviations for separate days. The results show that regardless of which day one picks, the calculated average cash shares and associated standard deviations would have been very similar. In practice, however, the survey designer will be faced with the problem of deciding the "optimal" length of the survey based on cumulative results, unless the survey lasts only one day. This decision involves as Crossley and Winter (2014)

describe (for consumption diaries) a trade-off, because it has been shown that “Shorter recording periods will lead to less bias in the estimation of mean expenditures, but, because of infrequency, higher variance.”

For the cumulative shares based on the transaction value we find a small all be it significant decline when including additional days. The figures are (share/SD): 75.7%/38.9% (day 1 only), 72.4%/37.5% (days 1 and 2), 71.1%/35.8% (days 1 to 3), 69.6%/34.7% (days 1 to 4), 68.7%/33.8% (days 1 to 5), 68.3%/33.8% (days 1 to 6), 67.7%/32.7% (days 1 to 7). The estimated standard deviations decline as well, showing that the precision of the estimate increases, by 6 percentage points, between the hypothetical one-day and seven-day diary. Over the last three days of the diary period, the standard deviation decreases by only one percentage point.

For retail transactions and transactions at gas stations, the findings from the study can be validated against external data. Table 5 shows a comparison between actual retail data and evidence from our survey.

Table 5 Average Value of Debit Card Transactions – Retail Institution Data vs. Diary (unweighted)

	External Data: Mean	Full Diary: Mean	Full Diary: Median	Diary Day 1: Mean	Diary Day 2: Median
Retail 2008	61.50 euro	62.30 euro	44.23 euro	61.35 euro	47.35 euro
Gas stations 2008	43.17 euro	46.53 euro	46.12 euro	47.17 euro	45.91 euro

Sources: External data Rüter (2010) and diary data: author's calculations based on data from “Zahlungsverhalten in Deutschland 2008”

Overall, the data from the diary compare fairly well with data from the German Retail Institute (EHI). No different conclusion would have been drawn had the diary period stopped after a single day. The mean from the diary would have been even closer to the actual EHI data for retailers. For gas stations the difference widens, compared to the full diary period, but is only slightly higher (approx. 0.50 euro). These results provide further evidence that the payments recorded on day one of the diary period do not differ greatly from the payments recorded in the subsequent days and that short diaries may be enough to adequately capture this key payment indicator.

## 5. Conclusions

Using data from a large diary study on payment behaviour, we show that both the number and value of transactions recorded on average per person are higher on days one and two of the survey period compared to subsequent days. This is in support of our first hypothesis and confirms most of the existing literature. However, the distribution of payments on each day of the diary is very similar as is the cash share in terms of volume and value, with some exceptions for day one. This is contrary to previous findings in the literature and against our second hypothesis. Finally, calculating the cumulative cash transaction shares, by including extra days one at a time, a small but significant decline is observed when going from shorter time frames (e.g. days 1 and 2 combined) to longer ones (days 1 to 6 combined). Taking everything together, the reduction in the value and volume of transactions over the days seems to be just a “parallel” downward shift of the structure observed for day 1. At most this is true on average, however. The declining cumulative shares show that the payment pattern for each individual consumer can change over time.

The observed patterns are to some degree consistent with an interpretation along the lines of the “survey fatigue” literature, which states that respondents get fatigued and do not report all transactions as the diary continuous. The drop between days one and two and the following days would be consistent with this interpretation, the persistent structure for small (cash) payments is not. Alternative explanations should also be considered. Among them are survey exuberance and reactivity. It may be the case that day one recordings are subject to survey exuberance, ie. more payments are recorded on day one than actually should have been recorded for this day. Survey exuberance may be an issue if respondents have a tendency to postpone the start of recording payments until a day with at least one transaction. Actually we see a lower number of respondents with zero (cash) transactions on day one and two than on any of the following days, but our data does not allow us to test whether this is the result of fatigue on days three to seven or exuberance on day one (and two). The shift in the number and volume of transactions, but not the structure, could also be explained with reactivity. When filling out the diary during the first days of the diary period the respondents get feedback on their spending behaviour and spending volumes, which may induce them to spend less money on the following days.

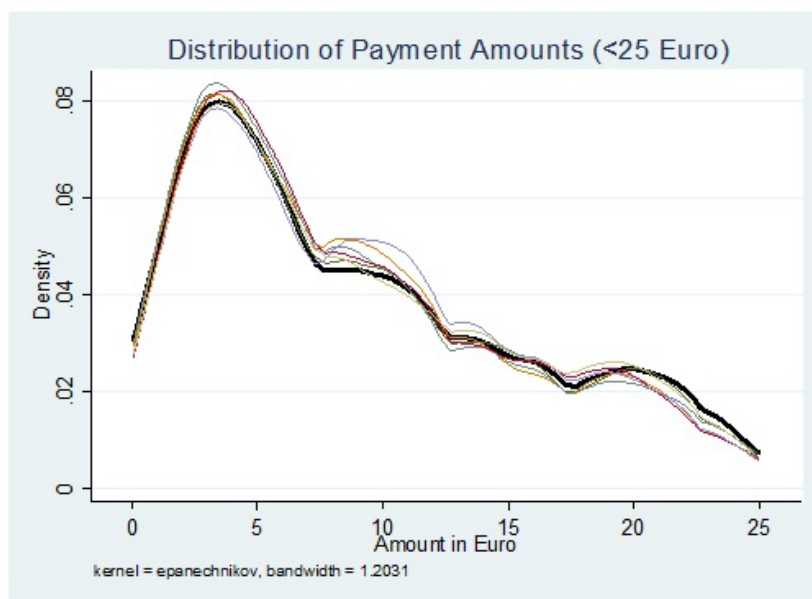
Questions on how tight a budget the persons are on or on whether they keep a regular expenditure diary could help to shed some light on this issue. Which of the three explanations is true is almost impossible to discern with the data we have at hand. Ideally one would be able to observe the true payment patterns for each individual and compare it with her or his recorded payments.

The main conclusion for diary design that can be drawn from our findings is that short diaries are enough to reflect consumers' payment behaviour adequately. This conclusion would be similar to that of Jonker and Kosse (2013). The analysis we present above is still at a fairly aggregate level, however. For example, we did not look at the payment behaviour of certain groups of consumers, such as older and younger groups. If the recording behaviour of these groups were to differ over the diary period, a short diary would not be enough. Sample size is also an issue if the behaviour of certain groups of customers or payments is under scrutiny. There may simply not be enough observations to conduct meaningful analysis on, e.g. gas station payments by old consumers, whereas a longer survey will yield enough data. To observe a longer "stream" of payments for each consumer may also be helpful for topics like the role of precautionary cash payments. The lack of sample size can be countered to some degree by increasing the number of respondents. Extending the diary period may be a more cost-effective way to increase the sample and the precision of the estimated indicators. Usually, recruiting additional respondents (especially in purely random samples) is more expensive than extending the diary for participating consumers.

Payment diaries are a relatively new method of data collection. They are closely linked to expenditure surveys, however, that have been around for several decades. An exchange on methodological issues regarding both types of surveys may result in better data for both – researcher of payment behaviour and research on consumption spending.

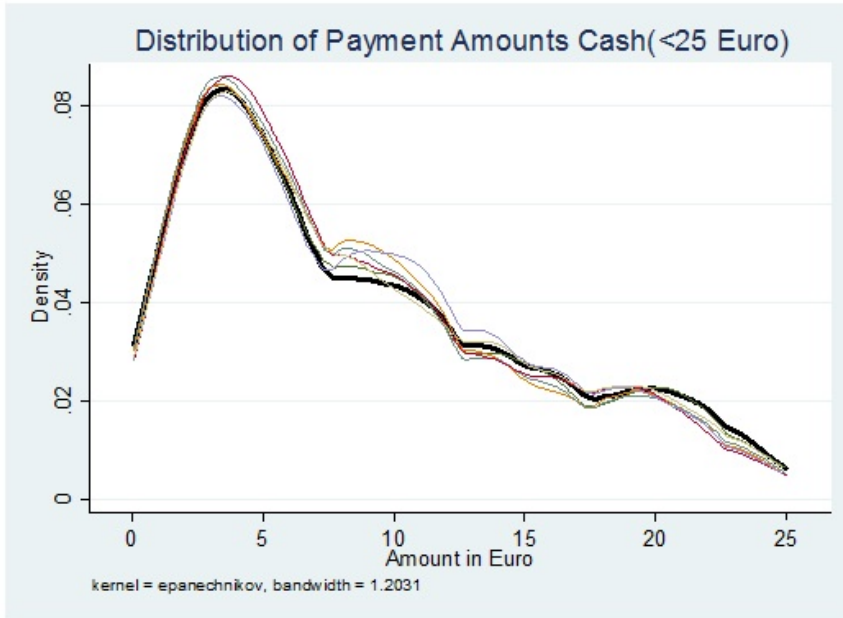
## 6. Appendix

Figure A1 Distribution of Payment Values for Transactions below 25 Euro (unweighted)



Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"  
 Note: Day one – thick black line

Figure A2 Distribution of Payment Values for Cash Transactions below 25 Euro (unweighted)



Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"  
 Note: Day one – thick black line

Table A1 Distribution tests – Median test

	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
--	-------	-------	-------	-------	-------	-------

Day 1	FS: ** <100: * <25:*	FS:*** <100: *** <25: -	FS:*** <100: *** <25: -	FS:*** <100: *** <25: -	FS:*** <100: *** <25: -	FS:*** <100: *** <25: -
Day 2		FS: * <100: *** <25:-	FS: - <100: ** <25:-	FS: * <100: - <25:-	FS: - <100: ** <25:-	FS: - <100: - <25:-
Day 3			FS: - <100: - <25:-	FS: - <100: - <25:-	FS: - <100: - <25:-	FS: - <100: - <25:-
Day 4				FS: - <100: - <25:-	FS: - <100: - <25:*	FS: - <100: - <25:*
Day 5					FS: - <100: - <25:*	FS: - <100: - <25:*
Day 6						FS: - <100: - <25:*

Source: Author's calculations based on data from "Zahlungsverhalten in Deutschland 2008"

Notes: FS: Full sample; <100: transactions of less than 100 euro; <25: transactions of less than 25 euro

Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

## References

1. Ahmed, N., M. Brozoowski, and T. F. Crossley. 2006. Measurement Error in Recall Food Consumption Data, IFS Working Paper 06/21, Cambridge, UK.
2. Albaum, G.S., F. Evangelista and N. Medina 1998. Role of Response Behavior Theory in Survey Research: A Cross-National Study. Journal of Business Research 42(2): 115-125.
3. Alessie, R., Gradus, R., and Melenberg, B. 1990. The problem of not observing small expenditures in a consumer expenditure survey. Journal of Applied Econometrics, 5(2): 151-166.
4. Arango, C. and A. Welte. 2012. The Bank of Canada's 2009 Methods-of-Payment Survey: Methodology and Key Results. Bank of Canada, Discussion Paper 12-6, Ottawa.
5. Bagnall, J., D. Bounie, K.P. Huynh, A. Kosse, T. Schmidt, S. Schuh and H. Stix. 2014. Consumer Cash Usage: A Cross-Country Comparison with Payment Diary Survey Data, DNB Discussion Paper No. 422, Amsterdam. available online: [http://www.bostonfed.org/payments2014/papers/Kim\\_P\\_Huynh.pdf](http://www.bostonfed.org/payments2014/papers/Kim_P_Huynh.pdf).
6. Bounie, D. and A. Francois (2006), Cash, Check or Bank Card? The Effects of Transaction Characteristics on the Use of Payment Instruments, Working Papers in Economics and Social Sciences ESS-06-05,

Telecom Paris Paris, France.

7. Crossley, T.F. and J. Winter. 2014, Asking Households About Expenditures: What have we learned? In: Carroll, C., Crossley T.F. and J. Sabelhaus (eds.) Improving the Measurement of Consumer Expenditures. NBER Ser. Stud. Income Wealth. NBER, Cambridge, MA.
8. Foster, K., E. Meijer, S. Schuh, and M. Zabek. 2011. The 2009 Survey of Consumer Payment Choice. Federal Reserve Bank of Boston, Public Policy discussion Paper No. 11-1, Boston.
9. 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* 89 (2), 473-489.
10. Hoffmann, A., H. Wörlen, A. Friedrich, N. Knaust, U. Von Kalckreuth, and T. Schmidt (2009), Payment behaviour in Germany – An Empirical Study of the Selection and Utilisation of Payment Instruments in the Federal Republic of Germany Deutsche Bundesbank, Frankfurt am Main, Germany.
11. Hoffmeyer-Zlotnik, J. H. (2003). New sampling designs and the quality of data. *Developments in applied statistics*. Ljubljana: FDV Metodoloski zvezki, 205-217.
12. Jonker, N. and A. Kosse. 2013. Estimating Cash Usage: The impact of survey design on research outcomes. *De Economist* 161(1): 19-44.
13. Jonker, N., A. Kosse and L. Hernández. 2012. Cash usage in the Netherlands: Who, where, when and whenever one wants? *DNB Occasional Study* 10(2), Amsterdam.
14. von Kalckreuth, Ulf, Tobias Schmidt and Helmut Stix. 2014. Choosing and Using Payment Instruments: Evidence from German Microdata. *Empirical Economics* 46(3): 1019-1055.
15. McWhinney, I. and H. E. Champion (1974), The Canadian Experience with Recall and Diary Methods in Consumer Expenditure Surveys, *Annals of Economic and Social Measurement* 3 (2), 411-435.
16. Mooslechner, P., H. Stix, and K. Wagner (2006), Wie wird in Österreich bezahlt? – Ergebnisse einer Untersuchung zur Struktur der Zahlungsmittelverwendung privater Haushalte im Kontext der geldpolitischen Analyse, *Geldpolitik & Wirtschaft* Q2/2006, 120-145.
17. Rüter, H. 2010. EHI-Research: Zahlungen und Kundenbindung per Karte. Presentation at the "EHI-Kartenkongress 2010" 4.5.2010, Messe Köln.
18. Statistics Canada. 2001. 1996 Food Expenditure Survey – Public-use Microdata Files, Income Statistics Division, Ottawa, Ontario.
19. Stephens, M. 2003. "3rd of the month": Do Social Security Recipients Smooth Consumption Between Checks? *American Economic Review* 93 (1): 406-422.
20. Sudman, S. and R. Ferber. 1971. Experiments in Obtaining Consumer Expenditures by Diary Method. *Journal of the American Statistical Association*, 66(336): 725-735.
21. Tincello, D. G., K. S. Williams, M. Joshi, R. P. Assassa, and K. R. Abrams. 2007. Urinary Diaries: a Comparison of Data Collected for Three Says versus Seven Days. *Obstetrics and Gynecology* 109 (2): 227-280.
22. Tullar, W., M. Pressley, and D. Gentry. 1979. Towards a Theoretical Framework for Mail Survey Response. *Proceedings, Third Annual Conference of the Academy of Marketing Science* 1997: 243-247.
23. Turner, R. 1961. Inter-Week Variations in Expenditure Recorded During a Two-Week Survey of Family Expenditure. *Journal of the Royal Statistical Society – Series C* 10(3): 136-146.