DNB Working Paper

No. 499 / February 2016

Does banknote quality affect counterfeit detection? Experimental evidence from Germany and the Netherlands

Frank van der Horst, Martina Eschelbach,
Susann Sieber and Jelle Miedema

DeNederlandscheBank

EUROSYSTEEM

Does banknote quality affect counterfeit detection? Experimental evidence from Germany and the Netherlands

(A joint study by De Nederlandsche Bank and Deutsche Bundesbank)

Frank van der Horst, Martina Eschelbach, Susann Sieber and Jelle Miedema *

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

Working Paper No. 499

February 2016

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

Does Banknote Quality Affect Counterfeit Detection? Experimental evidence from Germany and the Netherlands

Frank van der Horst[a], Martina Eschelbach[b], Susann Sieber[b] and Jelle Miedema[a]

[a] *De Nederlandsche Bank (DNB)*
[b] *Deutsche Bundesbank (DBB)*

9 February 2016

Abstract

Counterfeit prevention is a major task for central banks, as it helps to maintain public confidence in the currency. It is often maintained that a high quality of the banknotes in circulation helps the public detect counterfeits. However, there has not been any scientific evidence in support of this assertion so far. The present study is a first attempt to fill this research gap.

To investigate whether banknote quality affects counterfeit detection, De Nederlandsche Bank (DNB) and the Deutsche Bundesbank (DBB) conducted a field study in 2014 and 2015 amongst 250 consumers and 261 cashiers in the Netherlands and Germany. Participants received a set of 200 banknotes with either a high or a low average soil level, based on the actual circulation in two different countries. Real-life circulation in both Germany and the Netherlands is in between these values. Each set contained 20 counterfeit notes, which testees were asked to detect.

On average, untrained consumers detect 79% of the counterfeits, whereas retail cashiers detect 88%. Cashiers are found to detect more counterfeits when the set is clean, even after controlling for a wide range of personal characteristics in a regression. The estimated effect of cleanliness on the cashiers' detection rate is an additional 0.87 out of 20 counterfeits (4.4%) For consumers, the quality of the sets does not change the hit rate in a statistically significant way.

While a high detection rate is crucial, a low false alarm rate, i.e. the share of banknotes that is incorrectly classified as counterfeit, is important as well. A high detection rate discourages counterfeiters, while a low false alarm rate reduces the financial and non-financial costs for the public and retailers for checking banknotes. Both contribute to maintaining trust in the currency, which is a prerequisite for the functioning of central banks' policies. Consumers have a false alarm rate of 8.3% (14.9 out of 180 banknotes), whereas cashiers incorrectly declare 4.8% of the genuine banknotes (8.6 out of 180) to be counterfeits. In the clean sets, the false alarm rates are on average 1.5 percentage points higher than in the less clean sets.

The ability to tell apart genuine notes from counterfeits, also called sensitivity, combines detection and false alarm rates. It appears that cleanliness makes neither consumers nor cashiers more sensitive when differentiating between genuine banknotes and counterfeits. At the same time, cleanliness increases bias, i.e. people are more suspicious if the average quality of banknotes is high. This results in higher detection rates for cashiers, as mentioned, but also in more false alarms.

A cleaner circulation proves to help cashiers check banknote authenticity by hand, but central banks' costs for the replacement of soiled banknotes are considerable. A less clean circulation in a country reduces societal costs considerably, but makes it harder for cashiers to authenticate banknotes by hand. This implies that even more attention should be given to training and the use of automatic detection aids. Central banks may well assign different priorities to this. In addition, it must be remembered that banknote quality has implications beyond counterfeit detection and costs. For example, a high level of banknote cleanliness contributes to the smooth functioning of banknote equipment machinery.

The experience people gain during the test proves to matter to a great extent. At the start of the test, it appears that clean banknote circulation facilitates detection. Later on, this positive effect of cleanliness disappears, while the detection rates rise considerably. The study also shows that the more security features of the banknotes are reported as having been checked, the more counterfeits are detected. Elderly people perform significantly worse on the task. Furthermore, the study shows that genuine banknotes that look brand new are very often mistaken for counterfeits (14.7% false alarm rate versus 6.5% overall).

Our findings further suggest that central banks should continue their efforts aimed at informing the public and cashiers about banknote security features. It might be a worthwhile strategy to target their communication campaigns specifically at the elderly, since they perform worse and use cash more often than younger people.

Keywords: banknotes, counterfeits, banknote quality, signal detection theory.
JEL classifications: E40, E41, E50, E58.

1. INTRODUCTION

The national central banks (NCBs) of the Eurosystem aim to ensure a consistently high quality of euro banknotes in circulation. There are several reasons for this. First and foremost, providing fit banknotes to the public helps to maintain confidence in the euro, which is at the core of central banks' policies. Furthermore, banknotes must be acceptable to retailers and be capable of being used in banknote equipment, such as automated teller machines (ATMs) and vending machines. Another much-cited reason for ensuring a good quality of banknotes is that it helps recognise security features.

According to the public, central banks are successful in keeping the quality of the banknotes in circulation high. For the euro area, the results of the ECB online survey on the quality of euro banknotes in circulation (European Central Bank, 2015) show that citizens consider the physical condition of the Euro 50 notes in circulation to be very high: 69% consider them to be in good or in excellent condition, and 20% believe they are in acceptable condition. Several studies show that both the Dutch and the German public are even more satisfied with the present banknote quality than the average European citizen. According to a representative survey of the Dutch population, 83% of the Dutch people state that euro banknotes in general are fairly clean or very clean. With regard to the Euro 50 banknote, even 88% of respondents are of this opinion (Randsdorp and Zondervan, 2015, p.13). A representative survey of the German population (Deutsche Bundesbank, 2015) draws a similar picture. 87% of German respondents state that they are satisfied or fairly satisfied with the quality of euro banknotes (e.g. their cleanliness and intactness). The same study also finds that a majority of the German population (58%) would not agree to a reduction in banknote quality, even if this were to reduce the costs to the public.[1]

In order to assure the high quality of banknotes in circulation, the national central banks of the Eurosystem agreed upon minimum standards for sorting. In practice, NCBs differ in their sorting policy, which is one of the reasons why the quality of circulation differs between the euro area countries. Some NCBs increase the minimum requirements for the cleanliness of banknotes by adding a 'counterfeiting factor'. They argue that counterfeits are easier to detect in a clean circulation. This argument is also implicitly put forward in ECB Decision 2010/14:

"To (…) enable a proper detection of counterfeits, euro banknotes in circulation must be maintained in good condition to ensure that they can be easily and reliably checked for genuineness, and therefore, euro banknotes must be checked for fitness."

However, we are not aware of any research that provides evidence in support of the assumption that more counterfeits are detected when the quality of banknotes in circulation is high. To our knowledge, the only study touching on this issue was performed by the Dalhousie University and commissioned by the Bank of Canada (Klein et al., 2004). In this research, no evidence is found that quality of circulation has an impact on the detection rate of counterfeits, while the performance of detecting low quality genuine notes proves less accurate in a high quality notes context (more false alarms). Unfortunately, this study does not focus predominantly on this issue. Furthermore, the Canadian setting might differ from the European one.

From a central bank policy perspective, a combination of a high detection rate and a low false alarm rate is important. A high detection rate discourages counterfeiters, while a low false alarm rate reduces the financial and non-financial costs of checking banknotes for the public and retailers. Central banks give different priorities to one aspect or the other, but both contribute to maintaining trust in the currency. In addition, if the quality of the banknote circulation did not affect counterfeit detection, it would not be necessary to destroy soiled banknotes solely for this reason. In that case central banks could adjust their sorting policies by abandoning the counterfeiting factor. As a result, fewer sorted banknotes would need to be replaced, which would save taxpayers' money.

[1] 36% of the respondents agree to a reduction in banknote quality, while 6% refused to answer the question or did not have an opinion.

The main objective of the present study, therefore, is to determine whether the quality of the banknote circulation has an impact on the ability to discriminate between counterfeits and genuine banknotes.

The study was performed jointly by De Nederlandsche Bank and the Deutsche Bundesbank, in cooperation with the VU University Amsterdam, which provided advice on set-up and analysis. It was performed identically and simultaneously in both countries, which had the key benefits of increased power of expression and sharing the workload.

The structure of this report is as follows. Section 2 formulates our research questions and review existing literature. Section 3 describes the data collection. Section4 presents our hypotheses and the empirical strategy. Section 5 describes our sample and reports summary statistics, before our main results will be presented in Section 6. Further regressions using a different observation unit can be found in Section 7, while Section 8 contains some additional descriptive results. Finally, Section 9 sets out our conclusions.

2. RESEARCH QUESTIONS AND LITERATURE REVIEW

The main research question in this study is whether the cleanliness of banknotes has an effect on the public's ability to distinguish between genuine banknotes and counterfeits. The results of Jonker et al. (2006) show that, overall, the Dutch public is quite capable of recognising a counterfeit note: without having practised or received any feedback, members of the general public correctly identify 83% of the counterfeit notes they have been given to examine. Klein et al. (2004) find that Canadian consumers detect 69% of counterfeits they examine in a test.[2] Detection rates of genuine banknotes are also very high. While Canadian consumers correctly classify 84% of the genuine notes (a higher percentage than with the counterfeits), Dutch consumers sort 82% of the genuine notes correctly.

While Jonker et al. (2006) completely disregard note quality, Klein et al. (2004) look at both the quality of the individual note and the quality of the surrounding notes in the test sets that their participants examine. For individual notes, they find that counterfeits are best detected when their quality is high, while genuine banknotes of medium quality produce the lowest number of false alarms. The note quality in the test sets has no impact on counterfeit detection, but clean sets produce more false alarms.

Besides consumers, other stakeholders with regard to counterfeit detection are retailers or cashiers in a shop. It is assumed that counterfeit distributors try to spend the fake banknotes quickly and with minimum risk. They usually do so by spending the counterfeits to get change back. Retailers receive 120 banknotes a day on average, while members of the general public receive just one or two (De Heij, 2010). This makes cashiers the first line of defence in the fight against counterfeits. Around twenty percent of the retailers in the Netherlands who claim to authenticate banknotes do not make use of authentication devices (De Nederlandsche Bank, 2015), so for this group it is even more important that they are able to distinguish counterfeits from genuine notes. But are they able to do so? According to Jonker et al. (2006), Dutch cashiers without detection aids are better at identifying counterfeits than Dutch consumers. The same is true for professional cash handlers in Canada. But will the cashiers be affected by the quality of the circulation in the same way as consumers? To answer that question, in this research we distinguish between cashiers and consumers.

The rich data set which we have gathered also allows us to study some additional research questions. We knew in advance that some people declare a note to be a counterfeit only when they are very sure, which means they also fail to detect some counterfeits. However, other people declare a note to be a counterfeit much sooner, and more often incorrectly. The general tendency to declare a note a counterfeit or a genuine banknote is called bias. We investigate whether bias differences between people are influenced by the quality of banknote circulation.

Our participants differ widely in terms of their socio-demographic characteristics. We therefore ask whether personal characteristics, such as age, are relevant for their counterfeit detection performance. The results can help to identify vulnerable groups of the public on which education campaigns on counterfeit detection should focus.

Besides personal characteristics and the overall banknote quality, the properties of a specific banknote might also have an influence on whether it is perceived as a counterfeit or a genuine note. Important characteristics of banknotes are the denomination, the soil level and – in the case of counterfeits – the class of the counterfeit. We seek to analyse which characteristics of counterfeits make them less or more difficult to detect. The results of Klein et al. (2004, p.4) imply that the soil level might be important. In their study, both very new banknotes and notes of low quality are often incorrectly classified as counterfeit. Jonker et al. (2006, p. 20-22) report that there are fewer false alarms for Euro 20 banknotes compared with Euro 50 banknotes in their tests.

In addition, we are interested to find out how the performance of our participants evolves throughout the tests. Jonker et al. (2006) and Klein et al. (2004) divide the banknotes they use in their tests into several stacks of equal size, and we did the same. In both papers, the authors report

[2] Some participants received training on counterfeit detection in the middle of the tests.

considerable learning effects from stack to stack – an increase in detection rates and a reduction in false alarm rates. As our main regressions are on the level of the individual, it is not possible to include banknote or stack characteristics. We therefore run a set of further regressions that allow us to analyse these aspects in detail.

The tests we conduct also allow us to study the process of banknote authentication. The following research questions guide our analysis: What are the main features and characteristics that the testees use to authenticate the banknotes? What is the average time needed to check a banknote for authenticity?

When we use the term 'banknote quality' in this study, we solely refer to the soil level (cleanliness) of the banknotes and not to other, less frequently encountered defects such as limpness or dog-ears. This is because, in practice, the soil level is the main reason why central banks destroy unfit banknotes. In addition, varying only one defect in the test limits the variables that may influence the measurements.

3. DATA COLLECTION

3.1 Test sets

We created eight test sets with equal characteristics, except for the level of cleanliness. All sets consisted of 200 notes, 20 of which were counterfeits. DBB and DNB were both supplied with two clean sets and two less clean sets. See Table 1 for an overview of set characteristics.

Table 1 Test set overview

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
Average level of cleanliness	Clean	Clean	Clean	Clean	Less clean	Less clean	Less clean	Less clean
Testing NCB	DNB	DNB	DBB	DBB	DNB	DNB	DBB	DBB
No. of genuine notes	180	180	180	180	180	180	180	180
No. of counterfeit notes	20	20	20	20	20	20	20	20

We chose to have sets of 200 banknotes to facilitate a relatively small (10%) counterfeit presence while making it possible to investigate several classes of counterfeits. It was aimed at reflecting the low likelihood in real life of encountering a counterfeit and lowering the a priori chance of selecting counterfeits. A relatively high percentage of counterfeits in the test set was assumed to trigger participants more than required. We considered that testing more than 200 notes might lead to disrupted results due to testees getting tired or bored.

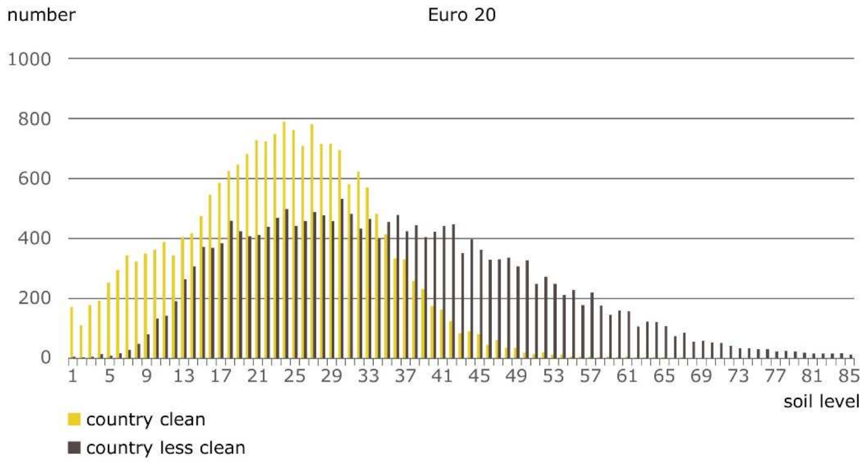
Each set consisted of equal numbers of Euro 20 and Euro 50 notes – both genuine and counterfeit. These denominations were chosen because - according to an ECB press release on 17 July 2015 - together they account for 86% of all counterfeits found in circulation in the Eurosystem.

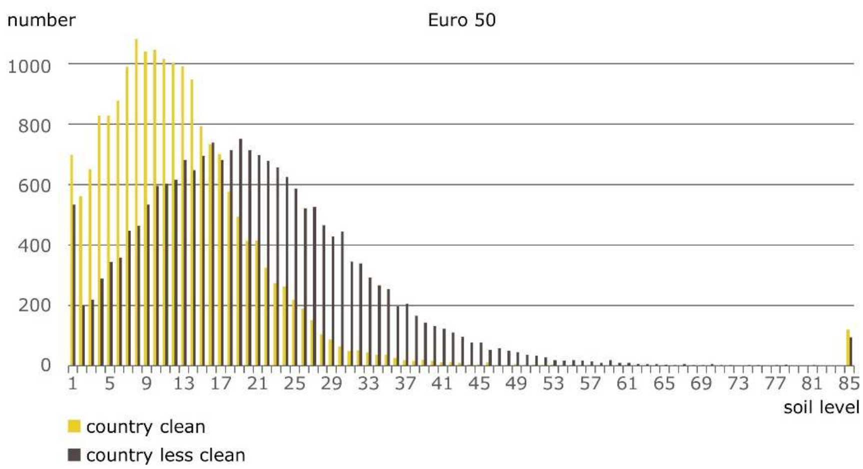
All counterfeit notes we used in the test sets are frequently found in the Eurosystem, and they were retrieved from circulation. They varied in terms of professionalism, but overall, they were rather deceptive. Also, all of the selected counterfeit notes are part of the Eurosystem test set for checking banknote handling machines and therefore represent the majority of counterfeits. We aimed to have an equal average soil level of the counterfeit notes in both the clean and less clean sets, because we assumed that counterfeiters do not deliberately deteriorate their product to mimic the quality of the circulation. The 20 counterfeit notes were randomly mixed with the 180 genuine notes.

3.2 Soil distribution

To establish a meaningful difference between the clean and less clean sets, we looked at a Eurosystem internal quality survey from 2013. This survey measures the quality of notes in circulation per country (sample size: 20,000), expressed in a scale of soil levels from 0 to 100. Typically, the distribution of the soil level is skewed: there are a large number of relatively clean notes, with a tail of dirty ones. Figure 1 shows the actual difference between a country with a relatively clean circulation and a country with a less clean circulation.

Figure 1 Distribution of soil levels of Euro 20 and Euro 50 banknotes in two countries





To prepare the test sets, we considered the difference between a country C with a relatively clean circulation and a country LC with a relatively less clean circulation as a basis for the difference between the clean and less clean sets. In practice, the clean test set was based on the real-life average of the distributions of Euro 20 and Euro 50 in country C, whereas the less clean test set was based on the distribution of the Euro 20 in country LC – which is the more soiled denomination. In this way, we achieved a mean less clean soil level (33.8) which was somewhat more than twice the mean of the clean distribution (15.9). The actual circulation in both Germany and the Netherlands is in between these extreme values. The soil level distribution of counterfeits was similar in all eight sets, in such a way that their soil level is not obviously different from genuine notes. Figure 2 and Figure 3 show the soil level distribution inside the test sets on a scale from 1 to 16, which is linearly related to the scale of 1 to 100 from the quality survey.

Figure 2 Soil level distribution of genuine banknotes in test sets
(per set)

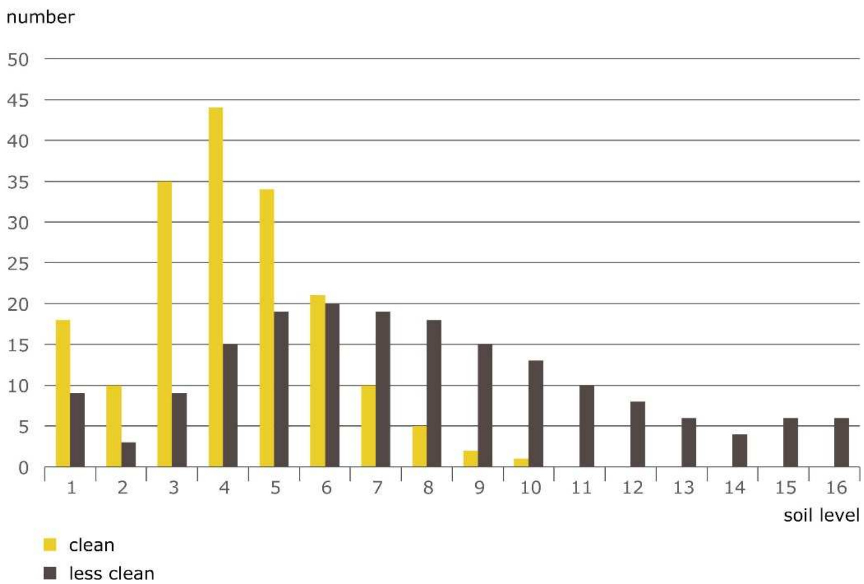
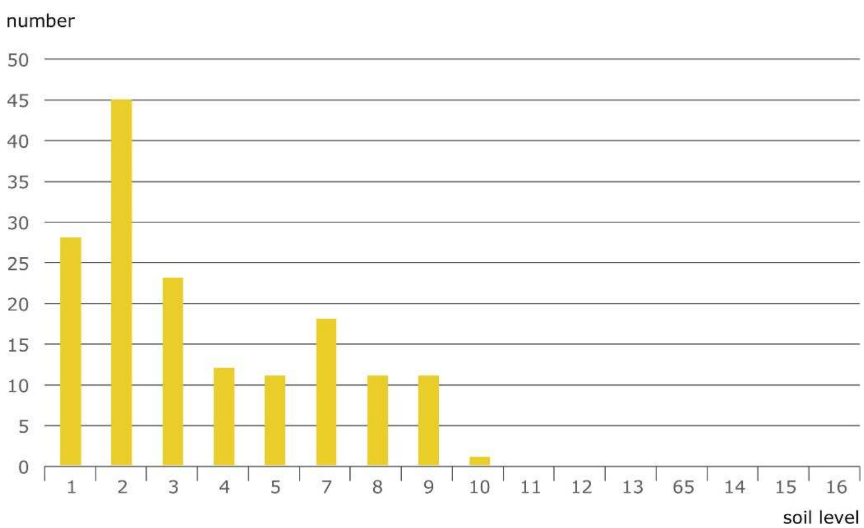


Figure 3 Soil level distribution of counterfeits in test sets
(total of all 8 sets)



To monitor the level of cleanliness before, during and after the tests, some test sets were analysed with the *Brain²* technology (Balke et al., 2012). This technology is a quick and reliable means of fitness scoring for banknotes, and it enabled us to keep track of the average fitness level of test sets. The assumption was that the banknotes decrease in fitness level simply by being used for testing, but the average fitness difference between the clean and less clean sets will remain constant, as they were tested an equal number of times. This assumption appeared to be fairly accurate. Based on measurements of a sample, the *Brain²* fitness score of less clean notes appeared to deteriorate slightly more (15%) than the score of clean notes (12%). The percentage distance between the clean and less clean sets remained very clear before the testing (15%) and after the testing (19%).

Figure 4 and Figure 5 show examples of the clean and less clean banknotes that were used during the tests.

Figure 4 Examples of a clean Euro 20 and Euro 50 note

Figure 5 Examples of a less clean Euro 20 and Euro 50 note

3.3 Participants

After a pilot with 180 participants, we could establish the required number of consumers and cashiers. For statistically significant results, a minimum of 244 consumers and 244 cashiers was needed (see Table 2). This was based on an anticipated effect (Cohen's d) of 0.32, a desired statistical power level of 0.8 and a probability level of 0.05.

Table 2 Required minimum number of participants

	DNB	DBB	Total
Consumers	122	122	244
Cashiers	122	122	244
Total	244	244	488

We managed to have more than the necessary minimum number of participants. Our results are based on data gathered from 511 participants – 250 consumers and 261 cashiers.

To obtain a representative sample of consumers we attended very diverse locations, such as a city town hall, open days at both NCBs, visitor centres, shopping streets, community centres, various courses and speech events, etc.

Participating cashiers mostly worked at the check-out of large supermarkets, but some of them were shopkeepers in shopping streets. The test took place during their working hours. We assured participating cashiers that their individual results or the results of their store would not be passed on to local management.

We made it clear to all participants that the tests were part of a scientific study and that their participation was voluntary. Participants did not receive any remuneration. After finishing the test, participants were given a small gift, usually a USB stick. In case they were interested, we also educated them after the test on security features of euro banknotes and handed out information materials.

3.4 Test setting

The participants were gathered around a table and were asked to separate 200 banknotes in genuine ones and counterfeit ones, see Figure 6. They were told that their sets consisted of mostly genuine ones, but that there was at least one counterfeit in each stack. The banknotes were offered in five stacks of 40 notes to each participant in the same random order. The reason for not presenting all 200 notes at once was to avoid boredom, and that participants refreshed a bit after every stack. This was also achieved by using a short period for noting down the counterfeits and the time used. Despite the fact that the time was recorded, we emphasised to the participants that they should take the time they needed, as it was not a competition. Time was measured with a stopwatch. After the test participants were asked to fill in a questionnaire, amongst others on socio-demographic factors (see Appendix A.1).

The tests were performed mostly on the locations of the participants, especially with cashiers, who do not usually have time to travel for a test. Test conditions, e.g. light, were fairly the same, but could not be controlled for pragmatic reasons.

Figure 6 Pictures of actual test setting

4. EMPIRICAL STRATEGY

4.1 Hypotheses

Our main research question formulated in Section 2 is whether the quality of the banknotes in circulation has an influence on counterfeit detection. Following the argumentation of the ECB and many national central banks in the euro area, we pose the following hypothesis:

H1: Counterfeits are easier to detect in a clean banknote circulation than in a less clean banknote circulation.

In view of the results of Klein et al. (2004, p. 9) we assume that a clean banknote circulation makes people more suspicious of the genuineness of banknotes, which in turn affects bias. We therefore formulate the following hypothesis:

H2: In a clean banknote circulation, respondents are more suspicious and tend to declare more banknotes to be counterfeits.

Furthermore, as explained in Section 2, we investigate the influence of stack and banknote characteristics on participants' results in the counterfeit detection experiment. Furthermore, we are interested in the process of banknote authentication, in particular the security features used and the time needed to check a banknote. The research questions guiding this part of the analysis are set out in Section 2. However, we do not provide any hypotheses relating to these additional research questions.

4.2 Measures of respondents' performance in counterfeit detection

When a respondent examines a banknote in the test set, four outcomes are possible. These are displayed in Table 3.

Table 3 Possible outcomes during the test

		Respondent's classification of banknote	
		Counterfeit	Genuine
Actual quality of banknote	Counterfeit	Hit	Miss
	Genuine	False Alarm	Correct Reject

If the respondent examines a counterfeit banknote, he or she can either

- correctly classify the banknote as counterfeit ('hit'), or
- incorrectly classify the banknote as genuine ('miss').

If the respondent examines a genuine banknote, he or she can either

- correctly classify the banknote as genuine ('correct reject') or
- incorrectly classify the banknote as counterfeit ('false alarm').

From the number of hits, one can calculate the so-called 'hit rate' (or 'detection rate'), which is defined as the number of hits divided by the total number counterfeits in the test set (20). Dividing the number of false alarms by the total number of genuine banknotes (180) gives the 'false alarm rate'. Hit rate and false alarm rate jointly quantify the performance of a respondent in the experiment (Stanislaw and Todorov, 1999).

The (scarce) previous research on counterfeit detection that we are aware of analyses hit rates and false alarm rates. In order to be able to compare our results with previous studies, we also start by looking at these measures. However, a single measure, such as the hit rate, does not suffice to measure performance. In an extreme case, a respondent could classify all banknotes as counterfeits. Although the hit rate would be at its maximum (100%), one could hardly say that the person was successful in telling apart genuine banknotes and counterfeits. A high counterfeit detection rate is certainly beneficial, however, if it comes at the price of being suspicious and extensively examining almost every banknote (a high false alarm rate), the benefits of a further increase in the detection rate have to be weighed against the pecuniary and non-pecuniary costs of extensive checks for authenticity.

Therefore researchers started to use other measures of participants' performance in yes/no tasks such as the detection of counterfeits. An alternative theory is called 'Signal detection theory' (SDT) and has been applied in the field of psychology for decades (see also Stanislaw and Todorov, 1999). The two basic elements of the theory are 'response bias' and 'sensitivity'. The response bias describes the general tendency to answer yes or no in a yes/no task. Applied to our experiment, this is the tendency to declare a banknote as counterfeit. Sensitivity is the ability to tell apart counterfeits and genuine banknotes.

Different measures of sensitivity and response bias have been suggested in the literature. The two prevalent measures for sensitivity are known in the literature as *A′* and *d′*. Response bias is operationalised with the measures β and *c*. All of these measures are functions of the hit rate and the false alarm rate as described above. The actual values of these four measures do not have a simple intuitive interpretation.[3] Three of the four measures are parametric,[4] i.e. they are based on the assumption that both genuine and counterfeit banknotes are normally distributed with the same variance but different means along a dimension which one could call 'counterfeit resemblance' or 'counterfeit obviousness'. The location parameters are specific to the individual.

If a person cannot differentiate at all between genuine and counterfeit banknotes, the two distributions have the same mean – in other words, they overlap completely. For a person who does not make any mistakes (no misses or false alarms) there is no overlap in the distributions, which is equivalent to saying that their means are very far apart. Following these considerations, the sensitivity measure *d′* gives the distance between the means of the two distributions. It ranges from zero (no ability to differentiate between the two types of banknotes) to infinity.

An overlap between the two distributions means that there are a number of cases where the individual is not sure whether a banknote is genuine or counterfeit. In such cases, some people tend to say that the banknote in question is counterfeit, which results in a high hit rate, but also a large number of false alarms. Others tend to say that the banknote is genuine, which maximizes the number of correct rejects, but also produces several misses. This tendency is called bias. The former type of person is said to have a liberal criterion, the latter type has a conservative criterion. β an *c* both measure the position of this criterion with respect to a neutral point. *c* gives the distance of the criterion to the neutral point in standard deviation units, β is a likelihood ratio.

A′ is a non-parametric measure of sensitivity, i.e. one does not make assumptions about the distribution of genuine and counterfeit banknotes along the 'counterfeit resemblance' scale. The main driver of *A′* is the distance between the hit rate and the false alarm rate. The larger it is, the larger is *A′*. When an individual is not able to differentiate counterfeits and genuine notes, the hit rate and the false alarm rate will be the same. One example is a person who declares all banknotes counterfeits. This person has a hit rate and a false alarm rate of one. The difference between the two rates is zero, thus indicating that the person is not able to discriminate between the two types of banknotes. In that case, *A′* takes on its lowest possible value. *A′* is at its maximum when the hit rate is one and the false alarm rate is zero.

We use *A′* as our primary measure of sensitivity. In the appendix we show results using the sensitivity measure *d′*. Response bias is captured using *c*. The appendix provides alternative estimations using the measure β. All measures are calculated on the level of the individual, i.e. we

---

[3] For an extensive discussion of these measures and the formulae to calculate them, see Stanislaw and Todorov, 1999, pp. 139-142.
[4] An exception is *A′*, which is a non-parametric measure of sensitivity.

end up with one value for each of the four measures and for each participant. The three parametric measures (*d′*, *c*, and *β*) are not defined when the hit rate and/or the false alarm rate are zero or one. Therefore we set the maximum hit rate at 0.975 and the minimum false alarm rate at 0.025.[5,6]

4.3 Regression analysis

After calculating the measures of performance according to SDT, we estimate several linear regression models with the individual (*i*) as the level of observation. Each model is of the form:

$$y_i = \alpha_0 + \alpha_1 clean_i + \gamma' X + \varepsilon_i$$

In order to test hypothesis H1, the dependent variable y is, first, the hit rate and second, a measure for sensitivity (*A′* or *d′*). When testing hypothesis H2, y is instead a measure for response bias (*c* or *β*).

The main explanatory variable *clean* is an indicator variable which assumes the value one if the respondent has a clean set and the value zero if the participant has a less clean set during the test. The matrix *X* contains several individual-specific variables: a country indicator (Dutch/German), age, gender, education, the preferred mode of payment (cash, cashless or both), whether the respondent was impaired by any visual handicaps during the test and whether the respondent had checked banknotes for authenticity in the last six months. Some of these variables are control variables, while others help to identify possible socio-demographic differences in counterfeit detection performance. α_0 is a constant, α_1 is a coefficient, γ is a vector of coefficients and ε_i is an error term. We run separate regressions for consumers and cashiers. In the regression testing H1, we expect $\alpha_1$to be positive. If H2 holds, α_1 will be negative in these regressions.

As these regressions on the level of the individual participant do not allow for including banknote or stack characteristics, we later report further regressions on the level of the banknote.

---

[5] The caps must not be closer to 1 resp. 0 in order not to violate the normality assumption underlying the various signal detection theory measures. 45.4% of our respondents who have 4 or fewer false alarms and thus a false alarm rate of below 0.025 are affected by the cap. The cap on the hit rate only affects respondents who have a hit rate of exactly one (26.6% of respondents).

[6] Hit rates of zero (i.e. no counterfeit banknote is detected) and false alarm rates of one (i.e. all genuine banknotes are classified as counterfeits) do not occur in the dataset.

5. SAMPLE AND DESCRIPTIVE STATISTICS

5.1 Data quality

The overall quality of the data derived from the tests is very good. Missing or ambiguous answers were clarified during the interview. For the few inconsistencies remaining we make simple imputations as described in further detail in Appendix A.2.

Yet, some of the variables must be interpreted carefully:

- **Education:** Many cashiers in the retail business are young people studying at university, who have not yet reached the final level of education they aim for.
- **Having checked banknotes for authenticity in the last 6 months:** The questionnaire did not specify whether checking is done manually or with banknote authentication devices. From discussions with participants after the test it became clear that at least some persons who answered 'yes' only relied on technical devices and had never tested banknotes manually.
- **Security features:** Some answers concerning the security features tested by respondents are ambiguous. Sometimes respondents write 'tilting', which leaves open whether they checked the hologram or the colour-changing number. Wherever possible, 'tilting' is translated into a security feature, usually the hologram. Where this is not possible, 'tilting' is counted as 'other' security features. 'Other' security features also include 'feel', 'serial number', 'banknote number', 'image' and 'picture'. 'Signature' is included as a security feature although in fact it is not. However, the signature is named as a security feature by a considerable part of the respondents. Printing quality and colour are combined to one feature, as are 'hologram' and 'hologram stripe'. Furthermore, comparing own observations during the tests with respondents' answers sometimes revealed that not all checked features were mentioned, in particular paper quality or printing.

5.2 Descriptive statistics of consumer and cashier sample

Table 4 Descriptive statistics of the consumer sample

Variables	Mean	Standard deviation
Hit rate (Detection rate)	0.79	0.20
False alarm rate	0.08	0.10
Share of banknotes selected	0.15	0.09
Clean set	0.50	
Dutch	0.50	
Age	47.74	16.98
Female	0.45	
No secondary education	0.03	
Secondary education or education missing	0.28	
Higher secondary education	0.24	
University	0.46	
Visual handicaps during the test	0.07	
Prefers cash as payment instrument	0.24	
Prefers cards as payment instrument	0.37	
No payment preference or missing	0.38	
Has checked banknotes in the last 6 months	0.14	
A'	0.92	0.07
c	0.27	0.43
d'	2.51	0.92
β	2.76	1.86
Number of security features checked	2.43	1.05
Average sorting time per banknote (seconds)	6.29	2.69
Number of observations	250	

For the study, 250 consumers are interviewed. Descriptive statistics for the consumer sample can be found in Table 4. Consumers perform rather well in the tests. They detect 79% of the counterfeits, or in other words 15.8 of the 20 counterfeits they encountered during the test. Their average false alarm rate is 8%, i.e. almost 15 (of 180) genuine banknotes are incorrectly declared as counterfeits. Compared to the study by Jonker et al. (2006) – who test the counterfeit detection performance of Dutch consumers and cashiers using euro banknotes – our consumers have both a lower hit rate and a lower false alarm rate. Jonker et al. (2006) report a hit rate of 92% and a false alarm rate of 14%.[7] However, the test settings differ significantly. First, half of the respondents in the previous study receive training on counterfeit detection before the test. Second, they have more time (up to 15 seconds) to examine the banknotes. Third, the test is split into three rounds and participants are informed about the percentage of correct answers after each round. Non-trained consumers identify 83% of counterfeits in the first test round, which is close to our own results.

The sensitivity and bias measures *A′*, *c*, *d′* and *β*, which we report in Table 4, do not have a simple intuitive interpretation. However, some general observations are possible. *A′* usually lies between

[7] Although the authors do not directly report it, having a higher hit rate and a higher false alarm rate is equivalent to having a more liberal bias. A common finding in Signal Detection Theory is that the bias is more liberal when the share of signal trials (in our case, the percentage of counterfeits) is higher. This is certainly the case here, as the share of counterfeits in the test sets of Jonker et al. (2006) is one third. In our sets, it is only one tenth.

0.5 and 1, where 1 indicates a perfect performance. The average *A′* of consumers is above 0.9, which implies that respondents do rather well in telling apart counterfeits and genuine banknotes. Similarly, values of *d′* larger than zero indicate respondents' ability to distinguish between counterfeits and genuines. The bias measure *c* may be both positive and negative, with zero as the neutral point. A positive value implies a tendency to answer 'no', i.e. not to declare a banknote a counterfeit. In other words, respondents apply a conservative criterion and only call a banknote a counterfeit when they are rather confident about it. In a similar vein, values of β larger than one point towards a conservative criterion.

The average time used per banknote is 6.29 seconds, which is close to the 6 seconds we indicated as the time that should suffice to recognise a counterfeit. However, the time needed per banknote varies widely, from 1.9 to 17.0 seconds, and it markedly decreases from stack to stack. Concerning socio-demographics, our sample of consumers has more males and more persons with an above-average level of education compared to the general population in Germany and the Netherlands. There are an equal number of German and Dutch participants. Only a small minority of participants (14%) has recently checked banknotes for authenticity. In the tests, they check on average 2.43 security features. This is in accordance with to the number of security features spontaneously mentioned by Dutch respondents (between 1.9 and 2.6) in the biennial studies about knowledge and appreciation of euro banknotes (Randsdorp and Zondervan, 2015). In theory this should be sufficient to follow the ECB's advice to always check multiple security features.[8]

[8] For more information how to check banknotes rapidly see http://www.bundesbank.de/Navigation/DE/Aufgaben/Bargeld/Falschgeld/Falschgelderkennung/falschgelderkennung.html and http://www.dnb.nl/en/payments/euro-banknotes-and-coins/euro-banknotes/genuine-or-counterfeit/security-features/

Table 5 Descriptive statistics of the cashier sample

Variables	Mean	Standard deviation
Hit rate (Detection rate)	0.88	0.14
False alarm rate	0.05	0.06
Share of banknotes selected	0.13	0.06
Clean set	0.50	
Dutch	0.41	
Age	37.97	13.72
Female	0.77	
No secondary education	0.02	
Secondary education or education missing	0.52	
Higher secondary education	0.31	
University	0.15	
Visual handicaps during the test	0.05	
Prefers cash as payment instrument	0.23	
Prefers cards as payment instrument	0.29	
No payment preference or missing	0.48	
Has checked banknotes in the last 6 months	0.81	
A'	0.96	0.05
c	0.17	0.37
d'	3.06	0.72
β	2.78	1.62
Number of security features checked	2.65	1.14
Average sorting time per banknote (seconds)	5.04	2.10
Number of observations	261	

In total, 261 cashiers are interviewed in our study – 106 in the Netherlands and 155 in Germany, see Table 5. They are quicker and better at detecting counterfeits than consumers. The hit rate is 88%, which corresponds to 17.6 out of 20 counterfeits correctly identified. They also declare 8.6 genuine banknotes as counterfeits (a false alarm rate of 5%). The average of both sensitivity measures *A′* and *d′* is higher for cashiers than for consumers, which implies that they are better at telling apart counterfeits and genuines. The bias measures *c* and *β* indicate that in case of doubt cashiers are more prone to call a banknote a genuine banknote rather than a counterfeit, just as consumers.

In comparison with consumers, cashiers are younger, the female share is higher and their average (completed) educational level is lower. The preference for cash as a payment instrument is about the same as for consumers, but 37% of consumers and only 29% of cashiers prefer cards.

Most cashiers have checked banknotes in the last 6 months. However, some of them have not checked them manually, but only with banknote authentication devices. The average number of security features checked is slightly higher for cashiers than for consumers.

5.3 Descriptive statistics on the set level

Tables A1 and A2 in the Appendix show some summary statistics for the clean and the less clean sets. Both the hit rate and the false alarm rate are higher in the clean set, i.e. in the clean sets

more counterfeits are detected, but more genuine banknotes are incorrectly declared as counterfeit. The average sorting time is similar in both types of sets.
The socio-demographic profile of participants sorting the clean and less clean stacks is also rather similar, which is in line with the set-up of our experiment.

Table 6 Performance measures by set

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
Average level of cleanliness	Clean	Clean	Clean	Clean	Less clean	Less clean	Less clean	Less clean
Testing NCB	DNB	DNB	DBB	DBB	DNB	DNB	DBB	DBB
Hit rate	0.830	0.834	0.892	0.868	0.782	0.825	0.839	0.818
False alarm rate	0.079	0.068	0.065	0.078	0.071	0.071	0.034	0.059
A′	0.927	0.932	0.952	0.943	0.918	0.929	0.946	0.933
c	0.174	0.222	0.097	0.106	0.302	0.250	0.311	0.290
Observations	56	59	70	71	58	59	69	69

Table 6 gives an overview of the performance measures by set and provides some interesting insights. Hit rates vary between 78% and 89%. A comparison of the individual pairs of sets with the same characteristics (same level of cleanliness and same country) reveals that they are usually rather similar. Exceptions are sets 5 and 6 with respect to the hit rate and sets 7 and 8 with respect to the false alarm rate.

Detection rates are higher in the clean than in the less clean sets in both the Netherlands and Germany, however, this difference is roughly 3 percentage points in the Netherlands and more than 5 percentage points in Germany. False alarm rates are also higher in the clean sets. While the gap between clean and less clean sets is only 0.2 percentage points in the Netherlands, in Germany the difference is on average 2.5 percentage points.

German participants detect on average more counterfeits than Dutch participants. The pattern is particularly pronounced in the clean sets, but less distinct with respect to the less clean sets, as the hit rate of the Dutch respondents varies strongly between sets 5 and 6. False alarm rates are similar in both countries when only the clean sets are examined, but much lower in Germany when the less clean sets enter the focus.

A′ is above 0.9 in each set, which implies that in general respondents do rather well in telling apart counterfeits and genuine banknotes. The sensitivity measure is slightly higher for the German sets than for the Dutch sets. Differences between clean and less clean sets are small or non-existent. It can clearly be seen from Table 6 that the average value of *c* is lower in the clean sets, which points towards a tendency to declare more banknotes counterfeits when the quality of the circulation is high.

6. REGRESSION RESULTS

As explained in Section 3, interpreting hit rates is the standard approach in the (scarce) scientific literature on counterfeit detection. In addition, hit rates allow for a more intuitive interpretation of results. We therefore start with the analysis of hit rates (see Table 7 and Table 8) before we apply signal detection theory measures in the following sections. All regressions are run separately for consumers and cashiers.

6.1 Hit rate

Table 7 Results from a linear regression model with the hit rate as dependent variable (consumers only)

Variables	Coefficient	Robust standard error
Clean set	0.0280	0.0180
Dutch	-0.0815***	0.0128
Age	-0.0045***	0.0005
Female	-0.0460*	0.0236
No secondary education	0.0156	0.0293
Secondary education or education missing	Ref.	
Higher secondary education	0.0284	0.0303
University	0.0176	0.0324
Visual handicaps during the test	-0.1117*	0.0516
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.0102	0.0293
No payment preference or missing	0.0169	0.0190
Has checked banknotes in the last 6 months	0.0525	0.0318
Constant	1.0270***	0.0443
Number of observations	250	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the hit rate as dependent variable. Standard errors are clustered on the set level.
Note 2: Hit rates take on values between and including zero and one.
Note 3: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table 8 Results from a linear regression model with the hit rate as dependent variable (cashiers only)

Variables	Coefficient	Robust standard error
Clean set	0.0435**	0.0158
Dutch	-0.0096	0.0201
Age	-0.0025***	0.0005
Female	-0.0250	0.0259
No secondary education	-0.1297*	0.0624
Secondary education or education missing	Ref.	
Higher secondary education	0.0022	0.0197
University	0.0383	0.0230
Visual handicaps during the test	0.0610***	0.0142
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.0667***	0.0171
No payment preference or missing	0.0156	0.0268
Has checked banknotes in the last 6 months	0.0735**	0.0262
Constant	0.8842***	0.0389
Number of observations	261	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the hit rate as dependent variable. Standard errors are clustered on the set level.
Note 2: Hit rates take on values between and including zero and one.
Note 3: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

For cashiers, the hit rate is higher in clean sets than in less clean sets. The coefficients are significant at the 5% level. For consumers, the coefficient is also positive, but insignificant. The size of the coefficient for cashiers indicates that they find on average 0.87 (out of 20) counterfeits more in the clean sets. Considering that cashiers miss on average 2.4 counterfeits, detecting 0.87 counterfeits more is a considerable improvement. The results are in line with the descriptive statistics discussed in Section 4.

Our results imply that German consumers detect more counterfeit banknotes than Dutch consumers. Various explanations are possible. Germans use cash much more often than the Dutch. A recent study by Bagnall et al. (2014), which compares the results of several international studies on consumers' payment behaviour at the point of sale, finds that cash accounts for 52% of all transactions and 34% of the turnover in the Netherlands, while the cash share in Germany is 82% in terms of volume and 53% in terms of value. Nevertheless, our estimation results on preferred method of payment shed some doubts on this explanation. Another possible reasoning is that participants in Germany and the Netherlands were recruited in different ways. In Germany, most of the consumers were interviewed during events of the Deutsche Bundesbank, such as on the 'Open Day' or during Bundesbank presentations targeted at the general public. As a result, German participants might have an above average level of interest in central banking topics.

Both regressions show that the age of the respondent has a significant effect on the hit rate. The older the person, the less counterfeits he or she detects. The out-of-sample prediction is that at age 25 consumers will find 89.3% (17.9 out of 20) of all counterfeits, while at age 65 they will find only 71.5% (14.3 out of 20). For cashiers, the gap is smaller, but still pronounced and highly significant. We estimate that a 25-year old cashier will find 91.4% (18.3 out of 20) of all counterfeits, while at age 65 cashiers will only find 81.6% (16.3 out of 20).

With respect to education, our results are mixed: while there is no effect of educational achievement on the hit rate for consumers, there appears to be a moderate positive relationship between education and hit rates for cashiers.

Results for visual handicaps are as expected when it comes to consumers. Visual impairments significantly reduce the hit rate. However, cashiers with visual handicaps perform significantly better. We do not have an explanation for this result, but it should be interpreted cautiously as only 5% of the cashier participants say they could not see properly.

Surprisingly, using cash as one’s preferred payment instrument does not help in detecting counterfeits. While payment behaviour has no statistically significant effect for consumers, cashiers who prefer to pay with cashless means even have a higher hit rate than those who predominantly use cash. Consequently, payment behaviour does not help to explain differences between Dutch and German participants. Having checked banknotes in the last six months increases the hit rate for cashiers, but not for consumers. However, only a small share of consumers has recently checked banknotes for authenticity.

Our results here are based on a linear regression model with standard errors clustered at the set level. Since the hit rate is a fraction that can take on values between (and including) zero and one, fractional response models might also be appropriate. This is particularly true as we have a high share of respondents with a hit rate of exactly one. Fractional logit and probit models (results not shown) produce virtually the same results as the linear model with respect to the sizes of the effects and their significance for both consumers and cashiers.

6.2 Sensitivity

Table 9 Results from a linear regression model with sensitivity measure A' as dependent variable (consumers only)

Variables	Coefficient	Robust standard error
Clean set	0.0030	0.0053
Dutch	-0.0342***	0.0051
Age	-0.0013***	0.0002
Female	-0.0198	0.0115
No secondary education	-0.0200	0.0146
Secondary education or education missing	Ref.	
Higher secondary education	0.0015	0.0119
University	-0.0052	0.0130
Visual handicaps during the test	-0.0367*	0.0176
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.0131	0.0157
No payment preference or missing	0.0085	0.0130
Has checked banknotes in the last 6 months	0.0178	0.0152
Constant	0.9986***	0.0147
Number of observations:	250	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sensitivity measure *A'* as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

In a next step, we investigate whether banknote quality affects participants' sensitivity. In the regressions of Table 9 and Table 10 the coefficient of cleanliness is not significant. Thus, in contrast to the previous results on the hit rate for cashiers, having a clean set does not help in telling apart genuine banknotes and counterfeits. The result holds for both consumers and cashiers. Tables A3 and A4 in the Appendix show the same regression with the alternative sensitivity measure *d'* as dependent variable. Regression results for consumers and cashiers using *d'* are broadly in line with results reported in Table 9 and Table 10 as regards both sign and significance of the estimated coefficients.

Interestingly, almost all explanatory variables besides cleanliness have the same effect on sensitivity which they have on the hit rate. Dutch consumers and those with a visual impairment are less sensitive. Sensitivity also decreases with age. Cashiers show a better performance if they have recently checked banknotes. In addition, we find that cashiers who prefer payment cards over cash are more sensitive. There also appears to be a moderate positive effect of education on sensitivity in the case of cashiers, while for consumers payment behaviour and education are insignificant.

Table 10 Results from a linear regression model with sensitivity measure A′ as dependent variable (cashiers only)

Variables	Coefficient	Robust standard error
Clean set	0.0080	0.0073
Dutch	-0.0072	0.0074
Age	-0.0006**	0.0002
Female	-0.0132	0.0081
No secondary education	-0.0399	0.0342
Secondary education or education missing	Ref.	
Higher secondary education	0.0067	0.0045
University	0.0158**	0.0065
Visual handicaps during the test	0.0232***	0.0054
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.0234***	0.0049
No payment preference or missing	0.0076	0.0073
Has checked banknotes in the last 6 months	0.0309**	0.0099
Constant	0.9465***	0.0167
Number of observations	261	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sensitivity measure *A′* as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

6.3 Bias

Table 11 Results from a linear regression model with bias measure c as dependent variable (consumers only)

Variables	Coefficient	Robust standard error
Clean set	-0.1038*	0.0529
Dutch	0.1120*	0.0513
Age	0.0084***	0.0010
Female	0.0458	0.0538
No secondary education	-0.1189	0.1196
Secondary education or education missing	Ref.	
Higher secondary education	-0.0991	0.0573
University	-0.0870	0.0469
Visual handicaps during the test	0.0965	0.0854
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.0639	0.0674
No payment preference or missing	0.0358	0.0594
Has checked banknotes in the last 6 months	-0.0774	0.0679
Constant	-0.1258	0.0861
Number of observations	250	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the bias measure *c* as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table 11 and Table 12 show regression results with the bias measure *c* as dependent variable. As can be seen from the coefficient of the clean set indicator, clean banknotes reduce participants' bias measure *c*. In other words, they become more suspicious when the set is clean and are more prone to declare a banknote a counterfeit in case of doubt. The effect is moderately significant for consumers and highly significant for cashiers. A possible explanation is that in a circulation with many unusually dirty banknotes, no banknotes stick out in particular. Consequently, a minor deviation from the other banknotes might not be recognised and therefore not give rise to suspicion.

Similar to what we saw earlier, using the alternative bias measure β instead yields substantially the same results. Socio-demographic variables besides age are insignificant for both groups of participants. Older participants apply a more conservative criterion. Elderly adults are known to very much dislike making errors (Starns and Ratcliff, 2010). What that means depends on what they perceive as the most important error. Since there were far fewer counterfeits than genuine notes in the test sets, they may have considered a false alarm as the most embarrassing error, which could have led to the conservative bias for elderly adults.

The results of the effect of cleanliness on bias might also help to shed some light on the puzzling fact that the hit rate is higher in clean sets, but at the same time the cleanliness of banknotes does not make people more sensitive to distinguish between genuine banknotes and counterfeits. Taking all the evidence together, in clean sets it is easier to recognise small deviations in banknotes, which makes people more suspicious about all banknotes. As a result, they sort out more banknotes, both the actual counterfeits and genuine banknotes. This produces a higher hit rate, but at the same times causes the false alarm rate to rise. Consequently, a high hit rate does not necessarily result in a high sensitivity.

Table 12 Results from a linear regression model with bias measure c as dependent variable (cashiers only)

Variables	Coefficient	Robust standard error
Clean set	-0.1684***	0.0251
Dutch	-0.0315	0.0314
Age	0.0065***	0.0013
Female	-0.0054	0.0533
No secondary education	0.2925	0.1730
Secondary education or education missing	Ref.	
Higher secondary education	0.0553	0.0511
University	-0.0398	0.0755
Visual handicaps during the test	-0.0344	0.0471
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	-0.0910	0.0516
No payment preference or missing	0.0168	0.0627
Has checked banknotes in the last 6 months	-0.0593	0.0557
Constant	0.0751	0.0848
Number of observations	261	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the bias measure *c* as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

7. THE INFLUENCE OF BANKNOTE AND STACK CHARACTERISTICS

In the previous sections, we analysed the effect of the cleanliness of the set and of the individuals' personal characteristics on various measures of their counterfeit detection performance. For each individual we calculated one performance measure, based on the overall number of counterfeits detected (and, if applicable, the number of false alarms). However, such an approach does not allow us to determine the influence of stack characteristics and properties of individual banknotes on the performance. For this reason, we run regressions in which we treat individual banknotes (instead of respondents) as the unit of observation. We first analyse the sample of counterfeits and then the sample of genuine notes. The dependent variable is binary, indicating whether the respective banknote is selected by the individual as a counterfeit. This gives us an idea what determines the hit rate and the false alarm rate. Before we proceed with the regression estimation, we present some bivariate analysis of set or banknote characteristics and hit rates or false alarm rates.

7.1 Banknote characteristics

7.1.1 Class of counterfeit

All counterfeits seized by authorities are analysed by national analysis centres within the euro area and categorised into classes. A class of counterfeits is usually produced by the same (group of) counterfeiters and has the same characteristics, e.g. production technique and quality of imitation of security features. Table 13 provides a basic breakdown of the counterfeits in our test sets by denomination and production technique (print or copy).

Table 13 Counterfeit classes breakdown by denomination, technique and detection rates

Class-ID	Denomination	Technique	Overall detection rate	Detection rate among younger people (<=60 y.)	Detection rate among elderly people (>60 y.)
1	20	Copy	96.5%	97.7%	89.3%
2	20	Copy	94.8%	96.3%	86.0%
3	50	Copy	86.9%	89.1%	74.0%
4	50	Print	92.0%	93.6%	82.7%
5	20	Print	82.9%	84.6%	72.7%
6	20	Print	81.9%	84.1%	69.3%
7	50	Print	78.9%	81.9%	61.3%
8	50	Print	85.1%	87.6%	70.7%
9	50	Print	73.5%	78.0%	47.6%
10	20	Print	85.1%	87.0%	74.0%

The results show very clearly that there is a large variation in detection rates between different classes of counterfeits. The counterfeits which are easiest to detect are the copied 20 euro banknotes. The lowest detection rates (below 80%) were measured for two different 50 euro counterfeits using print. We also investigate whether detection rates differ between younger and elderly people. Our earlier estimates imply that the average detection rates of elderly people would be lower. We find this pattern confirmed with respect to every single class of counterfeits in our test sets. Participants above the age of 60 generally show detection rates which are 10% to 15% lower than those of younger people. For the two most difficult to recognise counterfeit classes in our tests (IDs 7 and 9), the gap is even larger. They appear to be particularly hard to recognise for the elderly.

A part of the age gap can be explained by the fact that the majority of cashiers are younger than 60 and therefore push the detection rates in this group upwards. However, the age pattern remains the same if we only look at consumers. A further explanation is that elderly people tend to have limited knowledge of banknote security features (Randsdorp and Zondervan, 2015, p. 16). In our tests they report to check fewer security features on average than younger participants. With respect to every single security feature, the share of elderly participants who test it is the same or lower than the share of younger participants. The only exception is the signature, which is tested by an exceptionally high share of people aged 60 and above. We can only speculate here, but maybe it is more difficult for elderly people to keep up with new developments in banknote security features. Thus they rely on features which have already been in use for a long time. Lastly, it is possible that the counterfeits that are the most difficult to recognise for the elderly require good vision and feeling, which are known to decline with age.

7.1.2 Soil level and denomination

Both the genuine banknotes and the counterfeits used in our test sets differed in terms of soil levels. Descriptive statistics of the relationship between hit rates and the cleanliness of counterfeit notes are difficult to interpret as the soil level is also correlated with the class of the counterfeit. Nevertheless, we include the soil level of the banknote as an explanatory variable in the regression for the detection probability of counterfeits.

Another interesting question is whether the soil level of genuine banknotes has an effect on false alarms. Our primary expectation was that dirty banknotes would cause more false alarms because they might look unfamiliar and security features could be more difficult to test. However, the previous regressions on bias have already indicated that false alarm rates might actually be higher in clean sets.

Figure 7 Average false alarm rates at different soil levels of genuine banknotes

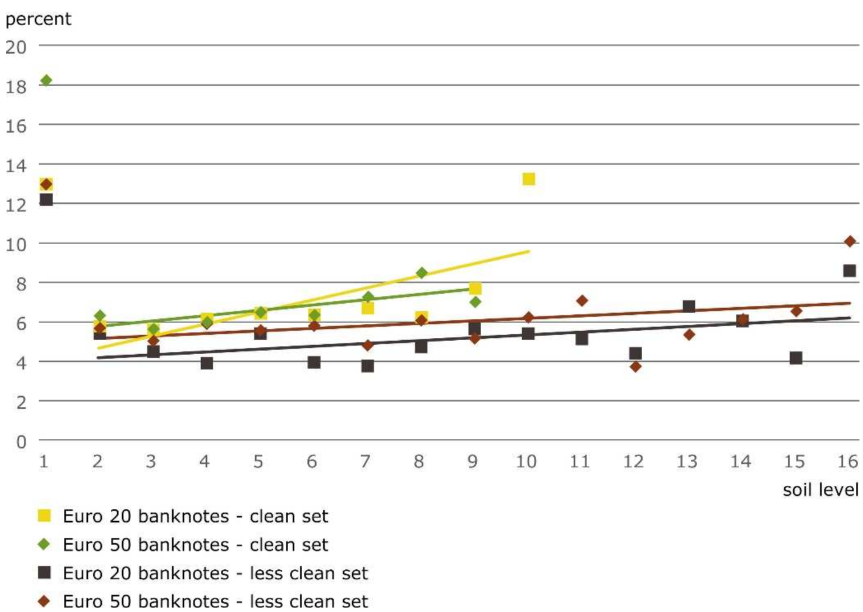


Figure 7 shows the average false alarm rates at different soil levels, for clean and less clean sets and for Euro 20 and Euro 50 banknotes separately.

The first thing to note is that by far the highest false alarm rates are recorded for banknotes of soil level 1, i.e. brand new banknotes. They range between 12% and more than 18%. Obviously participants consider these banknotes very suspicious. If one looks at the individual banknotes with the highest false alarm rates, almost all have soil level 1. For very clean banknotes of soil level 2 to very dirty banknotes of soil level 16 we find average false alarm rates of roughly 4% to10%. The lines in Figure 7 show a fitted line of a linear regression of false alarm rates on soil level when banknotes of soil level 1 are <u>not</u> considered. All four regression lines have a moderately positive slope, indicating that false alarm rates rise with soil level.[9] It can also be seen that the regression lines for the less clean sets (dotted lines) are below the regression lines for the clean sets (solid lines). This indicates that there are more false alarms in the clean sets.[10] Finally, Figure 7 shows that – at least in the less clean sets – the false alarm rates are higher for the Euro 50 banknotes.

In order to control for the effect of soil level on the false alarm rates, we include 'soil level' as a continuous variable and use an additional dummy variable for soil level 1. In addition, we include a dummy variable for the denomination of the banknote.

[9] If banknotes of soil level 1 were included in the regression, all slopes would be negative.
[10] The regression lines for the clean sets are shorter because these sets did not include any banknotes with very high soil levels.

7.2 Probability of detecting a counterfeit

Table 14 Results from a linear regression model of the probability to detect a counterfeit

Variables		Coefficient	Robust standard error
Set characteristics			
Clean set (ref.: stack 1)		0.0428**	0.0181
Dutch		-0.0490***	0.0079
Stack characteristics			
Stack	1	Ref.	
	2	0.0804***	0.0208
	3	0.1038***	0.0217
	4	0.1558***	0.0222
	5	0.1893***	0.0225
Interaction Stack*Clean set	1	Ref.	
	2	-0.0112	0.0241
	3	-0.0300	0.0246
	4	-0.0455*	0.0236
	5	-0.0265	0.0253
Running number of counterfeit in stack		0.0044*	0.0024
Number of counterfeits in stack		-0.0060**	0.0026
Banknote characteristics			
Class of counterfeit	1	Ref.	
	2	0.0312**	0.0123
	3	-0.0802***	0.0147
	4	0.0060	0.0155
	5	-0.1166***	0.0144
	6	-0.0931***	0.0130
	7	-0.1578***	0.0143
	8	-0.0699***	0.0181
	9	-0.2015***	0.0171
	10	-0.0919***	0.0165
Soil level		0.0014	0.0024
Individual characteristics			
Cashier		0.0200**	0.0101
Age		-0.0038***	0.0002
Female		-0.0326***	0.0077
No secondary education or education missing		-0.0423	0.0262
Secondary education		Ref.	
Higher secondary education		0.0112	0.0089
University		0.0223**	0.0096
Visual handicaps during the test		-0.0389**	0.0167
Prefers cash as payment instrument		Ref.	
Prefers cards as payment instrument		0.0338***	0.0103
No payment preference or missing		0.0198**	0.0096
Has checked banknotes in the last 6 months		0.1001***	0.0171
Interaction Stack*Checked banknotes	1	Ref.	
	2	-0.0248	0.0223
	3	-0.0385*	0.0225
	4	-0.0408*	0.0212
	5	-0.0738***	0.0204
Constant		0.9757***	0.0292
Number of observations		10,220	
F(26; 10,183)		34.40 [0.0000]	
R^2		0.1133	

Note 1: The table presents estimated coefficients and standard errors of a linear regression model.
Note 2: The level of observation is the participant-counterfeit combination, resulting in 20 observations per participant.
Note 3: The dependent variable is binary, taking on the value one when a participant correctly classified a specific counterfeit banknote as counterfeit, and zero otherwise.
Note 4: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table 14 shows the results of a linear probability model where the dependent variable assumes the value of one if a given counterfeit is detected by the testee. We use robust standard errors because error terms are correlated across individuals and banknotes.[11]

Regressors are split up in four groups: set characteristics, stack characteristics, banknote characteristics, and individual characteristics. The first and the last type were already included in the regressions on the hit rate, sensitivity and bias. For them, the present analysis can serve as a robustness test. In addition, estimation on the level of banknotes allows controlling for stack and banknote characteristics.

The main research question in this paper is whether cleanliness has an effect on counterfeit detection. Cleanliness is a set characteristic, but we assume that its effect might vary between stacks. We therefore interact cleanliness and stack. The results in Table 14 need to be interpreted in the following way: the coefficient of 'Clean set' is the effect of cleanliness in the reference category for 'Stack', which is stack 1. In the first stack, the probability to detect a counterfeit in a clean set is roughly 4.3 percentage points higher than in a less clean set. In order to see the effect of cleanliness in the remaining stacks, one has to look at the interaction term 'Stack*Clean set'. All interaction terms are negative, which implies that the difference in detection rates between clean and less clean sets is smaller in those stacks compared to stack 1. At the same time, detection rates increase from stack to stack. This can be seen from the positive coefficients of the variable 'Stack'. In the fifth stack, detection rates are 18.9 percentage points higher than in the first stack. The results are summarised in Figure 8.

Figure 8 Predicted detection rates and their 95% confidence intervals by stack and cleanliness

percent

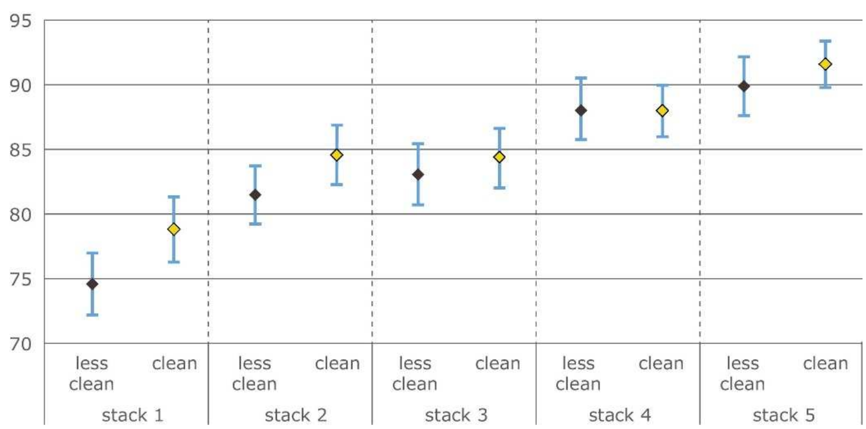


Figure 8 shows the predicted detection rates (red dots) and their 95% confidence intervals (black lines) in each stack of the clean and less clean sets. When the predicted detection rate of the one cleanliness level is outside the range of the 95% confidence interval of the other cleanliness level, the difference in the probability to detect a counterfeit between the two types of sets is significant at the 95% level. This is the case for sets 1 and 2, but not for the remaining sets.

It can also be seen that in general detection rates increase from stack to stack, however, the increase appears to be largest from the first to the second stack. We interpret this as a learning

[11] Although the dependent variable is binary, we opt for a linear specification instead of the straightforward nonlinear models such as logit or probit. The reason for this is that in nonlinear models, error terms which are not identlically and independently distributed can result in inconsistent coefficient estimates (e.g. Wooldridge, 2010). Autocorrelated error terms, as in our case, require the use of special panel data models such as the random effects probit model. Yet, in these models, obtaining marginal effects, which are necessary for an intuitive interpretation of the results, is not trivial because it heavily relies on assumptions (e.g. that the individual specific or the banknote specific error terms are zero). The estimation becomes even more complicated when the model contains interaction effects (Ai and Norton, 2003).

effect. Similar learning effects are found by Jonker et al. (2006). They use test sets with three stacks, each consisting of 72 banknotes, and find that consumers (cashiers) are able to increase their detection rates from 83% (96%) in the first stack to 93% (99%) in the third stack.[12] Klein et al. (2004) also find significant learning effects in a banknote detection test in Canada. Unfortunately, they do not distinguish between hits and false alarms.

Our findings suggest that when individuals encounter counterfeits for the first time, the quality of the banknote circulation does actually play a significant role. More counterfeits are detected in the clean sets. As experience and routine increase, the quality of the circulation becomes of minor importance. However, our results show that most people have very limited knowledge of banknote security features. Hence, when they encounter a counterfeit in real life, the situation will rather resemble that in stack 1 than that in stack 5.

In order to understand the learning effect even better, we also include the running number of the counterfeit within the stack in the regression. The coefficient is positive and significant, indicating that even within each stack, learning takes place.

The distribution of counterfeits within each stack and between stacks of the same set was randomly determined for each set. The number of counterfeits per stack varied between one and nine. We are thus interested in finding out whether the number of counterfeits in each stack has an effect on detection rates. Participants quickly form expectations on how many counterfeits there are in a stack. When the number is above their expectations, they might reduce their efforts to detect even more, or they might fall victim to 'gambler's fallacy', i.e. they expect to receive one (or several) genuine banknotes after a counterfeit, even though the distribution of counterfeits in the sets is random. This is particularly interesting as in two of the four less clean sets there were a large number of counterfeits already in the first stack (9 out of a total of 20), while in the clean sets the maximum number of counterfeits in the first stack was 5. The results confirm that a high number of counterfeits in a stack significantly depresses detection rates. Nevertheless, the effect appears not to be strong enough to be the main driver in the differences between clean and less clean sets.

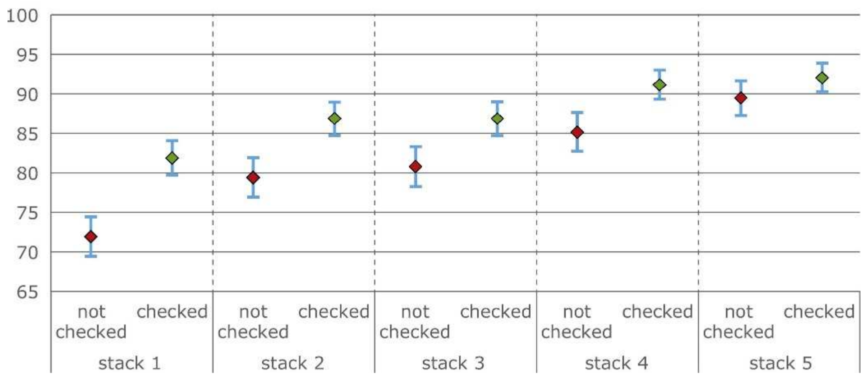
Turning to banknote characteristics, the results of the descriptive analysis in Section 6.1.1 are confirmed: the class of the counterfeit has a considerable and highly significant effect on the probability of detection. Colour copies of Euro 20 banknotes are easy to spot, while some of the print counterfeits of Euro 50 banknotes are much more difficult to detect. The soil level of the counterfeit has no significant impact.

The results regarding personal characteristics are generally in line with the findings of the regression on the hit rate. Fewer counterfeits are found in the Dutch sets. The highly significant negative effect of age on the probability of detection is in line with earlier results for both consumers and cashiers. A higher detection rate of participants who use cards as their preferred means of payment matches the earlier findings for cashiers. Women show a moderately worse performance than men, while detection rates increased with education. Being a cashiers and having checked banknotes in the last 6 months (either manually or with detection aids) both have positive and significant effects on the hit rate.

In order to understand the effect of having previous experience in checking banknotes even better, we interact it with stack (see Figure 9). Having checked banknotes for authenticity recently strongly improves the performance of respondents. In the first stack, persons with experience have a probability of detecting a counterfeit that is about 10 percentage points higher than the probability of persons without experience. The gap between the two groups narrows from stack to stack. This can be interpreted as evidence that learning about security features of banknotes takes place rather fast and that seeing a limited number of counterfeit banknotes is sufficient to achieve a good level of knowledge.

[12] The detection rates and learning effects reported in Jonker et al. (2006) should not be directly compared to our results. First, their tests take place in a laboratory setting. Second, they have a much higher share of counterfeits (one third) in their test sets. Third, participants receive feedback on the percentage of correct answers after each stack.

Figure 9 Predicted detection rates and their 95% confidence intervals by stack and previous experience
(having checked banknotes in the last 6 months)
percent



7.3 Probability of a false alarm

Table 15 Results from a linear regression model of the probability to cause a false alarm

Variables		Coefficient	Robust standard error
Set characteristics			
Clean set (ref.: stack 1)		0.0125***	0.0043
Dutch		0.0173***	0.0018
Stack characteristics			
Stack	1	Ref.	
	2	-0.0148***	0.0038
	3	-0.0279***	0.0037
	4	-0.0253***	0.0038
	5	-0.0391***	0.0037
Interaction Stack*Clean set	1	Ref.	
	2	0.0067	0.0056
	3	0.0075	0.0054
	4	-0.0004	0.0054
	5	0.0120**	0.0055
Running number of genuine banknote in stack		0.0000	0.0001
Number of counterfeits in stack		-0.0016***	0.0005
Banknote characteristics			
Soil level		0.0019***	0.0003
Soil level 1		0.0947***	0.0045
Denomination: Euro 50		0.0087***	0.0016
Individual characteristics			
Cashier		-0.0191***	0.0023
Age		0.0001	0.0001
Female		0.0153***	0.0018
No secondary education or education missing		0.0085	0.0061
Secondary education		Ref.	
Higher secondary education		-0.0043**	0.0020
University		0.0102***	0.0023
Visual handicaps during the test		0.0054	0.0037
Prefers cash as payment instrument		Ref.	
Prefers cards as payment instrument		-0.0229***	0.0025
No payment preference or missing		-0.0181***	0.0022
Has checked banknotes in the last 6 months		-0.0218***	0.0022
Constant		0.0740***	0.0060
Number of observations		91,980	
F(25; 91,954)		57.93 [0.0000]	
R^2		0.0208	

Note 1: The table presents estimate ed coefficients and standard errors of a linear regression model.
Note 2: The level of observation is the participant-banknote combination, resulting in 180 observations per participant.
Note 3: The dependent variable is binary, taking on the value one when a participant incorrectly classified a specific genuine banknote as counterfeit, and zero otherwise.
Note 4: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table 15 presents the results of a linear regression model of the probability to incorrectly select a genuine banknote in the test (a 'false alarm'). As the structure of that data is the same as in the previous section, we use the same estimation method and presentation of results as before. The dependent variable is again binary, now indicating that the respondent suspects a specific genuine banknote to be a counterfeit. As before, we interact cleanliness as the most important set characteristic and the number of the stack. The coefficient of the variable 'Clean set' gives the difference in false alarm rates between clean and less clean sets in the reference category for 'Stack', which is stack 1. False alarm rates are around 1.3 percentage points higher in the first stacks of the clean sets. The interaction variable 'Stack*Clean set' is positive for all stacks except stack 4, which means that in stacks 2, 3 and 5, the difference in false alarm rates between clean and less clean sets tends to be even larger than in stack 1. Simultaneously, false alarm rates decrease from stack to stack. This can be seen from the negative and significant coefficient of the dummy variables for 'Stack'.

Figure 10 Predicted false alarm rates and their 95% confidence intervals by stack and cleanliness
percent

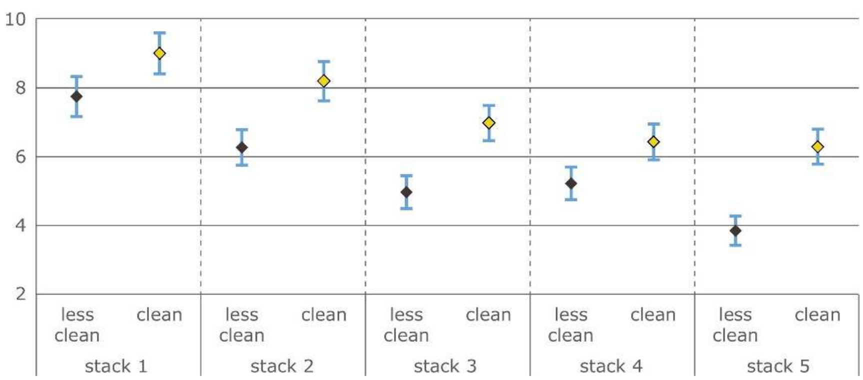


Figure 10 summarises the results. The learning effect – which manifests itself by falling false alarm rates - is clearly visible for both types of sets. Irrespective of stack, false alarm rates in the clean sets are significantly higher than in the less clean sets. These findings are in line with the research by Klein et al. (2004, p. 9) who report that in a clean circulation there are more false alarms. In their study, genuine banknotes of low quality are incorrectly classified as counterfeit in the clean circulation.

With the exception of stack 4, for which false alarm rates in the less clean sets are unexpectedly high, the gap in false alarm rates between the two types of sets appears to widen slightly from stack to stack. Interestingly, this is the opposite development to the one that we observe for the detection rates. We can only speculate about the causes of this result. One possible explanation is that respondents' sensitivity increases during the test, resulting in higher hit rates and lower false alarm rates. However, bias might generally be higher when the average quality of banknotes in circulation is good, which leads to a sustained high level of false alarms in the clean sets. As the experience of most consumers and even cashiers with detecting counterfeits is very limited, they can be expected to be in a situation similar to the one in stack 1 when they encounter a suspicious banknote in circulation. From our results we can conclude that in this situation the cleanliness of the circulation has a significant yet moderate effect on the incidence of false alarms.

Our findings accord with the results of Jonker et al. (2006). They also find a considerable reduction in false alarm rates of consumers (cashiers), from roughly 18% (14%) in the first stack to 9% (9%) in the second stack. While in their study the performance of cashiers improves even further in the third stack – although at a slower pace – the false alarm rates of consumers increase slightly.

Interestingly, the number of counterfeit banknotes in the stack has a significant effect on the false alarm rate. The more counterfeits there are, the lower the false alarm rate is. In the regression on the hit rate we saw a similar result: the more counterfeits there were, the lower the hit rate was. As participants form their expectations concerning the likelihood of encountering a counterfeit banknote based on experiences gathered during the tests, they might feel insecure if there suddenly is a much higher or lower number of counterfeits than in previous stacks. It is possible that participants suffer from 'gambler's fallacy', i.e. they incorrectly assume that one or several genuine banknotes follow a counterfeit or, the other way around, they assume that after a certain number of genuine banknotes, a counterfeit follows. Thus, when they do not find a counterfeit for a while they tend to declare the next best banknote that is somehow 'suspicious' a counterfeit. In general, our results show that the design of the test sets (ratio of genuine banknotes and counterfeits; distribution of counterfeit banknotes within the sets and stacks) has an effect on key outcomes. In further research, it might be advisable to properly control for these effects.

Banknote characteristics also have a highly significant and sizable effect on the false alarm rate. The findings of the descriptive analysis in Section 6.1 are confirmed by the regression results. Brand new banknotes are highly suspicious: They have a false alarm rate that is more than 9 percentage points higher than that of still clean banknotes of soil level 2. When we control for the effect of these very new banknotes, false alarm rates significantly increase with soil level. Our results thus confirm the findings of Klein et al. (2004, p. 4) in the Canadian context. They also report that both very new banknotes and notes of low quality are often incorrectly classified as counterfeit.

Keeping soil level constant, false alarm rates of Euro 50 banknotes are almost 0.9 percentage points higher compared to Euro 20 banknotes in our tests. This finding is in line with previous research by Jonker et al. (2006, p. 20-22) who also report fewer false alarms for Euro 20 banknotes compared to Euro 50 banknotes.

Some of the characteristics of the individual also prove to be important for the determination of false alarm rates. As expected, previous experience results in fewer false alarms. Being a cashier and having checked banknotes for authenticity in the previous 6 months each reduce the false alarm rate by roughly two percentage points. Dutch and female participants sort out more genuine banknotes, as do people who prefer to use cash. The effect of education is not linear and hard to interpret. However, we find that participants with a university degree have a higher false alarm rate. In the previous regression they were also found to have a higher hit rate. Both aspects together imply that they have a more conservative criterion, which is in line with the results of the regressions on bias. Age, which had some effect on detection rates, is insignificant with respect to false alarms.

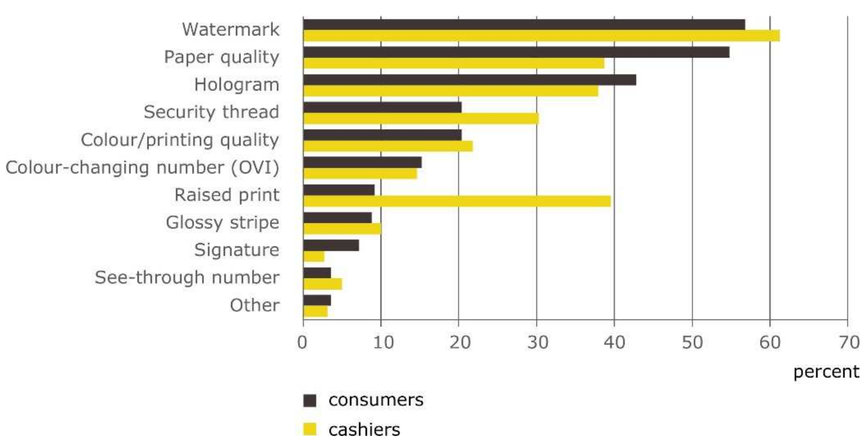
8. FURTHER RESULTS

8.1 Checking security features

After the test, participants were asked which security features they had tested. Although the test supervisors offered their help in finding the correct name of the features, some answers are ambiguous. Sometimes respondents write 'tilting', which leaves open whether they checked the hologram or the colour-changing number. Wherever possible, 'tilting' is translated into a security feature, usually the hologram. Where this is not possible, 'tilting' is counted as 'other' security features. 'Other' security features also include 'feel', 'serial number', 'banknote number', 'image' and 'picture'. 'Signature' is included as a security feature although in fact it is not. However, the signature is named as a security feature by a considerable number of respondents. Printing quality and colour are combined to one feature, as are 'hologram' and 'hologram stripe'.

Figure 11 gives an overview of the frequency of testing specific security features by consumers and cashiers. More than 50% of consumers report they checked the watermark and the paper quality, respectively, and more than 4 out of ten have a look at the hologram. All of the remaining security features are only inspected by a minority of respondents. One in five tests the security thread and the colour/printing quality, respectively. The two most prominent security features on the back side of the notes – the colour-changing number and the glossy stripe – as well as the raised print and the see-through number are almost unknown to consumers. Around 7% of the consumers try to verify the genuineness of a banknote with the help of the signature on the front. This is interesting because the signature is not a security feature. Experiences from conducting the test show that respondents who checked this feature often became highly confused when they realised that there were different signatures.[13]

Figure 11 Percentage of participants who test the various security features of euro banknotes



Watermark, hologram and paper quality are also the security features which cashiers test most often. Interestingly, fewer cashiers than consumers test the paper quality, although it is likely that handling cash every day is particularly helpful in recognising differences in substrate. Somewhat more cashiers than consumers inspect the security thread. With respect to colour and printing quality, the colour-changing number, the glossy stripe and the see-through number, the habits of cashiers are similar to those of consumers. What is striking is that cashiers rely on the raised print much more frequently than consumers. It turns out that this security feature has been promoted among cashiers. In the interviews with the cashiers in both Germany and the Netherlands it became apparent that some of the retail chains which took part in the study had explicitly trained their employees in checking the raised print.

---

[13] Euro banknotes bear the signature of the president of the European Central Bank. Our test sets included banknotes with the signatures of all three presidents who have been in office to date: Wim Duisenberg, Jean-Claude Trichet and Mario Draghi.

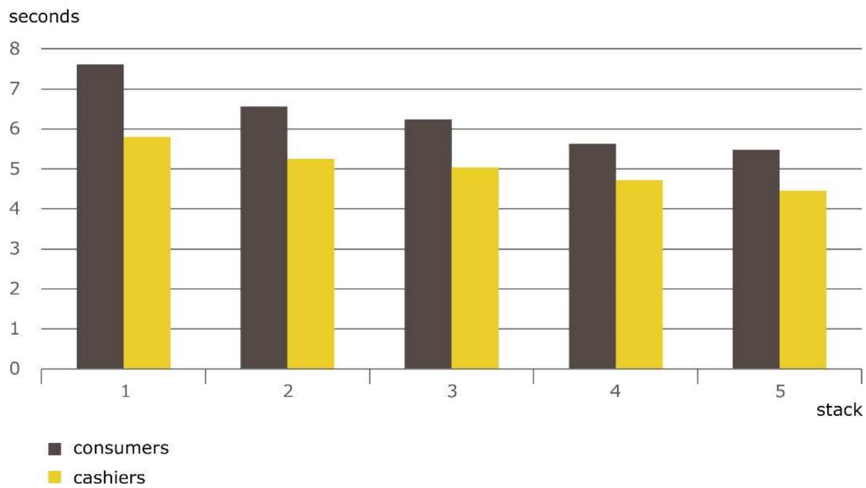
The Eurosystem advises to test more than one security feature of banknotes before accepting them, since there are high-quality imitations of each individual security feature, but they are usually not combined on the same counterfeit banknote. This was certainly the case with the banknotes we used in our tests. Consequently, checking several security features should have increased detection rates. From Table 16 one can see that this is actually the case. Most of our respondents (39.3%) test three security features, 46.8% test less than three, and 13.9% test more than three. The average hit rate increases substantially and continuously in the number of security features checked, from less than 80% to almost 100%. This clearly shows that knowledge about banknote security features is very helpful in detecting counterfeits. However, one should keep in mind that good knowledge of security features can also be a sign of experience with or an interest in checking banknotes for authenticity.

Table 16 Distribution of the number of security features checked and average hit rate

Number of security features checked	Share among respondents	Average hit rate
0 or 1	17.4%	76.0%
2	29.4%	81.0%
3	39.3%	86.1%
4	10.0%	91.8%
5 or more	3.9%	96.8%

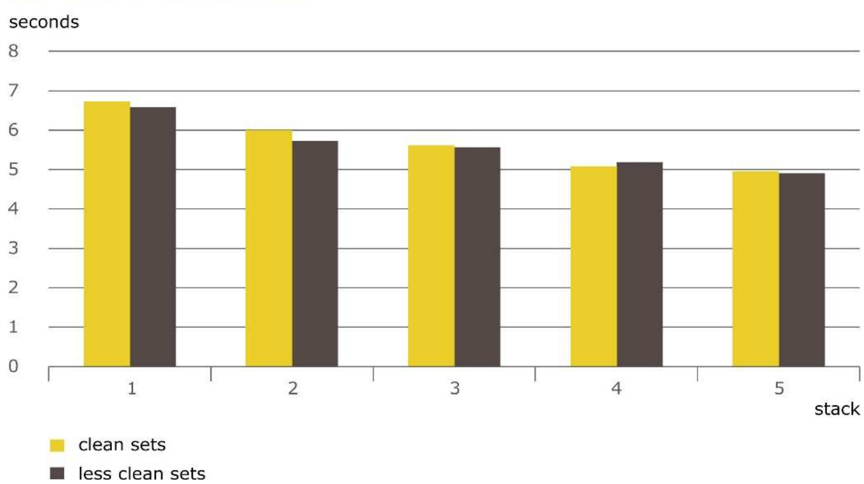
8.2 Sorting time

Figure 12 Average sorting time per banknote in seconds for consumers and cashiers



The results in Section 6 have already provided some indication that learning takes place during the tests. From stack to stack, the probability of detecting a counterfeit goes up and the probability of having a false alarm goes down. In addition, learning during the test decreases the average time needed. As can be seen from Figure 12, for both consumers and cashiers the average sorting time per banknote decreases steadily from stack to stack.[14] Cashiers are considerably faster than consumers. Nevertheless, even for them we find learning effects, although checking banknotes for genuineness should already be part of their work routine.

Figure 13 Average sorting time per banknote in seconds in the clean sets and the less clean sets

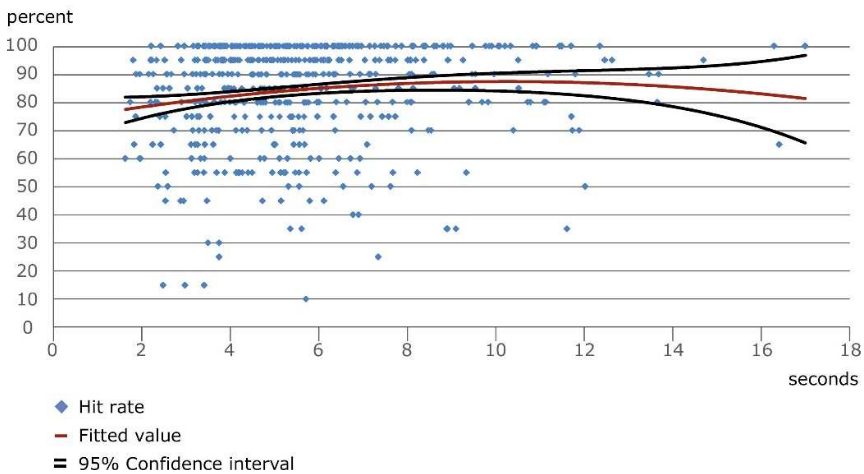


At best, the average cleanliness of the banknotes has a small influence on sorting time, and existing differences seem to disappear with experience (see Figure 13). Interestingly, the average time to check a banknote is 0.15 seconds higher in the clean set than in the less clean set, but only in stacks one and two.[15] However, we can still conclude that a high level of cleanliness of the banknote circulation does not help in speeding up the checking for genuineness.

[14] We cannot rule out that participants became bored with the task after a while and speeded up checking in order to finish as quickly as possible.
[15] We do not have a definite explanation for this observation. Our assumption is that the large number of highly 'suspicious' brand new banknotes drives the result.

This is confirmed by the regression results reported in Tables A7 and A8 in the Appendix. In those regressions, sorting time per banknote is the dependent variable. The explanatory variables are the same as in the regressions for hit rate, sensitivity and bias, which are reported in Chapter 5. While cleanliness of the set is not significant in the regression for consumers, it even results in higher sorting times for cashiers. This is in line with the previous findings: consumers' test performance appears to be almost unaffected by cleanliness, while cashiers are more suspicious when the average banknote quality is high. They take more time, find more counterfeits, but also produce more false alarms. Socio-demographic variables are of little help in explaining sorting time.

Figure 14 Relationship between average sorting time per banknote and hit rate



Apart from the question what determines the sorting time, one might also ask whether sorting time has an effect on the counterfeit detection performance. Figure 14 shows the relationship between the average sorting time per banknote and the hit rate. The bold line is the predicted hit rate from a regression of the hit rate on time and time squared, while the shaded area is the confidence interval. It can be seen that the predicted hit rate somewhat increases in time up to a sorting time of around ten seconds per banknote. There is no evidence that taking even more time than ten seconds increases the detection rate any further.

9. CONCLUSIONS AND RECOMMENDATIONS

The main purpose of this study is to improve knowledge about the effect of cleanliness of the banknote circulation on counterfeit detection. For this purpose, we test 250 consumers and 261 cashiers from the Netherlands and Germany, using 8 different test sets of 2 different cleanliness levels, each consisting of 20 counterfeits and 180 genuine banknotes. In order to describe the performance of participants, we use different measures: the hit rate (the share of counterfeits correctly selected), the false alarm rate (the share of genuine notes incorrectly selected), sensitivity (ability to distinguish genuine and counterfeit banknotes) and bias (tendency to be more or less suspicious about genuineness in general).

The following hypotheses were tested;
H1: Counterfeits are easier to detect in a clean banknote circulation than in a less clean banknote circulation.
H2: In a clean banknote circulation, respondents are more suspicious and tend to declare more banknotes to be counterfeits.

Our research draws a differentiated picture of the relationship between cleanliness and counterfeit detection.

The average hit rate for consumers is 79%, and for cashiers it is 88%. Our results show that cashiers detect significantly more counterfeits in the clean sets than in the less clean sets. The size of the effect (0.87 out of 20 counterfeits detected additionally) is quite substantial, given the overall high average detection rate. This means that H1 holds for this group. For consumers, however, cleanliness is not statistically significant in the regression on the hit rate and H1 is rejected.

The advantages of a high detection rate are reduced if it is accompanied by a large number of false alarms. For this reason, we also study sensitivity. Our results suggest that cleanliness does not help to increase sensitivity, i.e. the ability to differentiate between genuine banknotes and counterfeits. In accordance with H2, cleanliness rather drives up bias; people are generally more suspicious when the average quality of banknotes is good. The higher hit rate for cashiers in the clean sets is therefore accompanied by more false alarms. This is best illustrated by looking at the average false alarm rates of consumers and cashiers in the two types of sets. Cashiers incorrectly select 9.8 (out of 180) genuine banknotes in the clean set, but only 7.5 in the less clean set. For consumers, the average additional false alarms in the clean sets are 3.1 out of 180, with an overall average of 14.9 false alarms.

In general, we find that cashiers' performance is affected by cleanliness to a greater extent than consumers' performance. While cleanliness is highly significant in the regressions for both the hit rate and bias on the sample of cashiers, it is only marginally significant for bias and insignificant for the hit rate in the case of consumers.

A cleaner circulation proves to help cashiers check banknote authenticity by hand, but central banks' costs for the replacement of soiled banknotes are considerable. A less clean circulation in a country reduces societal costs, but makes it harder for cashiers to authenticate banknotes by hand. This implies that even more attention should be given to training and the use of automatic detection aids. Central banks may well assign different priorities to this. In addition, it must be remembered that banknote quality has implications beyond counterfeit detection and costs. For example, a high level of banknote cleanliness contributes to the smooth functioning of banknote equipment machinery.

Additional regressions on the level of individual banknotes reveal further interesting insights. There are substantial learning effects, which cause hit rates to go up and false alarm rates to go down from stack to stack. In the first two stacks, the probability to detect a counterfeit is higher in the clean sets than in the less clean sets. False alarm rates are significantly, but only moderately higher in the first stacks of the clean sets. This may suggest that when people have very little experience in the task – which probably applies to the majority of citizens in the Netherlands and Germany – cleanliness of the banknote circulation helps in detecting counterfeits. In the final stacks, counterfeit detection rates are the same, irrespective of the cleanliness of the banknote circulation. False alarm rates are increasingly reduced in the course of the test, but this effect is less strong in clean sets. A question that remains open is which stack is the best representation of real-life conditions. One could argue that the first stack best represents real life because participants have no previous experience with counterfeits and learning has not yet taken place. In contrast, one could claim that in real life people have *a priori* expectations of what the share of counterfeits in a banknote circulation is. As the test sets are unfamiliar to the participants, they need to form expectations on the counterfeit rate during the tests and should have a rather accurate estimate in the last stacks. In addition, during the test people get more at ease while performing their task of selecting counterfeits - just as in everyday life, when people are not focused on selecting counterfeits.

However, for those who have some routine and are already relatively good at spotting counterfeit banknotes, like cashiers, cleanliness has a rather detrimental effect because it seems to increase bias and thus cause more false alarms. It is worthwhile to study which other factors, apart from cleanliness, have an effect on bias and in which situations people become insecure about the genuineness of banknotes. After all, a very high false alarm rate might in theory indicate a low level of trust in the euro banknotes. In general, bias is highly subjective. It can change from day to day and from situation to situation. Our regressions do not provide many clues. Apart from age, socio-demographic variables are insignificant. With respect to banknote characteristics, we find that brand new banknotes cause a large number of false alarms. According to our observations during the tests, these are mainly caused by the feel of the banknote. A possible recommendation from our study is therefore to thoroughly consider the feel of the banknote and the changes in tactile features of the banknotes when decisions about banknote substrate and coating are made. Besides brand new banknotes, false alarm rates increase with soil level. Nevertheless, further research on the determinants of bias in counterfeit detection is needed.

Our rich data allows for a multitude of additional conclusions and recommendations beyond the optimal level of banknote quality.

Age of the participant is found to be significant in most of our analyses. Elderly people detect fewer counterfeits and are less sensitive. They test fewer security features on average than younger participants. At the same time, various studies on payment behaviour show that they use cash more often (see e.g. Deutsche Bundesbank, 2015, pp. 32-33). We therefore recommend to develop information material targeted at the elderly population and to consider the needs of the elderly when developing new banknote designs and security features. However, we could not identify specific security features that are particularly relevant for elderly people. Additional research that allows a closer look at the relationship between respondent characteristics, classes of counterfeits and detection rates, could fill this knowledge gap.

In accordance with previous research, we find sizable learning effects. This is remarkable because our participants only received a total of 20 counterfeit banknotes. A possible recommendation is to extend central banks' training activities for cashiers and the public. It appears to be particularly helpful to use counterfeits seized from circulation in these training sessions. Even within a limited amount of time and with only a few examples of counterfeit banknotes, the training effect can be substantial. Retailers should also be encouraged to train their employees regularly, as having checked banknotes in the past six months improves cashiers' performance significantly.

Finally, our results confirm some of the Eurosystem's basic assumptions with regard to authenticity checking. First, a few seconds are enough to recognise a counterfeit. Those who use more than 10 seconds do not perform better than others who are faster. Second, it is indeed advisable to check more than one security feature, as detection rates are much higher when several features are checked. When doing this, one might also want to stress that the signature is not a security feature.

REFERENCES

Ai, C. and E. Norton (2003), Interaction terms in logit and probit models, *Economics Letters* 80(1), 123-129.

Bagnall, J., D. Bounie, K. P. Huynh, A. Kosse, T. Schmidt, S. D. Schuh and H. Stix (2014), Consumer Cash Usage: A Cross-Country Comparison with Payment Diary Survey Data, *International Journal of Central Banking,* forthcoming.

Balke, P., J.-M. Geusebroek and P. Markus (2012), BRAIN2 - Machine Learning to Measure Banknote Fitness, De Nederlandsche Bank, Amsterdam, http://www.dnb.nl/binaries/Paper%20ODS_tcm46-267424.pdf.

De Heij, H. (2010), Banknote design for retailers and public, *DNB Occasional Studies 8(4)*, De Nederlandsche Bank, Amsterdam.

De Nederlandsche Bank (2015), Factsheet – Cash: behaviour and perception by retailers, De Nederlandsche Bank, Amsterdam, http://www.dnb.nl/en/binaries/712585_factsheet_retailersonderzoek_EN_tcm47-321626.pdf.

Deutsche Bundesbank, Falschgelderkennung (*Counterfeit detection*), http://www.bundesbank.de/Navigation/DE/Aufgaben/Bargeld/Falschgeld/Falschgelderkennung / falschgelderkennung.html.

Deutsche Bundesbank (2015), Payment behaviour in Germany in 2014: Third study of the utilisation of cash and cashless payment instruments, Frankfurt am Main.

European Central Bank, Decision of the European Central Bank of 16 September 2010 on the authenticity and fitness checking and recirculation of euro banknotes (ECB/2010/14).

European Central Bank, Results of the 2014 survey on the quality of euro banknotes in circulation (ESCB internal document).

European Central Bank, Biannual information on euro banknote counterfeiting, Press release dated 17 July 2015, http://www.ecb.europa.eu/press/pr/date/2015/html/pr150717.en.html.

Jonker N., B. Scholten, M. Wind, M. Van Emmerik and M. van der Hoeven (2006), Counterfeit or genuine: can you tell the difference? Quantitative research on the ability of the general public and cash handlers to distinguish counterfeit from genuine euro banknotes, *DNB Working Paper* No. 121, De Nederlandsche Bank, Amsterdam.

Klein M., S. Gadbois and J.J. Christie (2004), Perception and Detection of counterfeit currency in Canada: note quality, training and security features. *Proceedings of SPIE, Optical Security and Counterfeit Deterrence Techniques V*, R.L. van Renesse (ed.), Vol. 5310, 1-12.

Randsdorp, Y. and I. Zondervan (2015), A biennial study about knowledge and appreciation of euro banknotes among the Dutch, Study prepared for De Nederlandsche Bank, Amsterdam, http://www.dnb.nl/en/binaries/Rapport%202015%20-%20Awareness%20and%20appreciation%20of%20euro%20banknotes_tcm47-322785.PDF.

Stanislaw, H. and N. Todorov (1999), Calculation of signal detection theory measures, *Behavior Research Methods, Instruments, & Computers* 31(1), 137-149.

Starns J. J. and R. Ratcliff (2010), The effects of aging on the speed–accuracy compromise: boundary optimality in the diffusion model. *Psychology and Aging* 25(2), 377–390.

Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, London: MIT Press.

ACKNOWLEDGEMENTS

We would like to thank Professor Martijn Meeter of VU University Amsterdam for his useful advice. We also thank Ger Otten for programming the software and Hans Broeders, Stephan Raspoort and Reiner Gruhle for their help in producing and analysing the test sets. We are grateful to the co-testers Jan-Mark Geusebroek, Esther van den Kommer, Gerard Konincks, Nikolaus Bartzsch, Michael Bastius, Dagmar Boy, Johana Cabinakova, Erwin Gladisch, Ben Groth, Reiner Gruhle, Tobias Kronberger, Jens Riechmann, Tanja Roth, Dirk Storm, Rüdiger Walter, Silvia Welander, Andreas Wenzel, Steffi Winkler, and Heike Wörlen.
Furthermore, we extend our gratitude to Paul de la Combé, Marcel Swart and Susan Bun for supporting our access to a large number of cashiers.
Most of all, of course, we thank all volunteering participants, both the testees from the public category and the employees from the following companies.

From Germany:
- real; SB Warenhaus, in particular Jürgen Sottong;
- EDEKA-Georg;
- dm-drogerie markt, in particular Michael Geißler; and
- Several shops in Frankfurt and the Rhine-Main-area: TeeGschwendner Wiesbaden, Inhaber Wolfgang Michel; Donys Lieblingsschuhe; Buchhandlung Kerst + Schweitzer; Buchhandlung an der Paulskirche; Fachbuchhandlung Gebicke am Gericht; Wissenschaftliche Buchhandlung Theo Hector; Buchhandlung Carolus; Buchhandlung Bollinger.

From the Netherlands:
Supermarkets:
- Albert Heijn (Almere, Amsterdam, Amstelveen, Culemborg, Honselersdijk, Naaldwijk and Putten);
- Plus (Naarden, Dommelen, Benschop, Almere, Duivendrecht and Zeewolde);
- Nettorama (Zeist);
- Hoca (Nunspeet)

Several shops in Amsterdam (café Oosterling, A la carte, stomerij Libas, Foto Mignon, Handyman, Centre-neuf, Rodolfo's, Hans Struijk fietsen, Gloria verlichting, Seventy-five, Me gorgeous, Zwartjes van 1883, Marije Buffing edelsmid, Discovery games, Handyman, Elzinga wijnen, drogisterij Oorbeek, Studio Bazar and the Amsterdam Experience)
- Monarch (Diemen).

And, last but not least, we wish to thank all our colleagues for their valuable input and remarks.

APPENDICES

Appendix A.1: Survey

Questions	**Answers**
1. What is your gender?	1. Male 2. Female
2. What is your age?	.. year
3. What is your highest level of education?	1. No educational degree 2. Secondary education 3. Higher secondary education 4. University
4. When doing this test, were you impaired by any visual handicaps, such as colour blindness or because you have forgotten your glasses?	1. Yes 2. No
5. Is handling banknotes part of your daily work routine?	1. Yes 2. No
6. Assuming that you are offered various payment options when shopping, how would you pay for your purchases?	1. Predominantly in cash 2. Predominantly by non-cash payment instrument 3. In cash but also by non-cash payment instrument
7. What were the most important features that helped you detect counterfeit banknotes during this test?	………………………………. ………………………………. ……………………………….
8. Have you checked any banknotes for authenticity in the last six months?	1. Yes 2. No

Appendix A.2: Data cleaning

- In the German dataset, there were two respondents who said they handled banknotes in their professional life, but they were of retirement age (75 and 80). Since these respondents most likely referred to their jobs before retirement, their answers were recoded and they were classified as consumers. Two more persons (aged 76 and 70) did not answer the question about the professional handling of banknotes. For the sake of consistency, these persons were also classified as consumers.
- One person did not answer the question about handicaps. He was classified as having no handicaps.
- Three persons did not state whether they had checked banknotes in the previous six months. They were classified as not having checked banknotes.
- One consumer did not provide information about age. The average age of consumers (48) was imputed in this single case.
- A few people did not provide information about their educational degree and/or their preferred payment instrument. As these missing answers are too few to justify creating a 'missing' category in the regressions, respondents with missing information on education were grouped together with 'secondary education' (the largest group of respondents). Respondents who did not answer the question on their preferred payment instrument were grouped together with those who had no preference.

Appendix A.3: Additional results

Table A1 Descriptive statistics of the clean sets

Variables	Mean	Standard deviation
Hit rate (Detection rate)	0.86	0.17
False alarm rate	0.07	0.09
Share of banknotes selected	0.15	0.08
Cashier	0.51	
Dutch	0.45	
Age	42.40	16.86
Female	0.61	
No secondary education	0.04	
Secondary education or education missing	0.38	
Higher secondary education	0.30	
University	0.27	
Visual handicaps during the test	0.08	
Prefers cash as payment instrument	0.23	
Prefers cards as payment instrument	0.32	
No payment preference or missing	0.45	
Has checked banknotes in the last 6 months	0.50	
A'	0.94	0.06
c	0.14	0.40
d'	2.84	0.87
β	2.49	1.64
Number of security features checked	2.53	1.09
Average sorting time per banknote (seconds)	5.69	2.47
Number of observations	256	

Table A2 Descriptive statistics of the less clean sets

Variables	Mean	Standard deviation
Hit rate (Detection rate)	0.82	0.18
False alarm rate	0.06	0.08
Share of banknotes selected	0.13	0.07
Cashier	0.51	
Dutch	0.46	
Age	43.10	15.41
Female	0.62	
No secondary education	0.01	
Secondary education or education missing	0.42	
Higher secondary education	0.25	
University	0.33	
Visual handicaps during the test	0.04	
Prefers cash as payment instrument	0.24	
Prefers cards as payment instrument	0.34	
No payment preference or missing	0.42	
Has checked banknotes in the last 6 months	0.47	
A'	0.93	0.06
c	0.29	0.39
d'	2.74	0.86
β	3.05	1.79
Number of security features checked	2.55	1.12
Average sorting time per banknote (seconds)	5.61	2.50
Number of observations	255	

Table A3 Results from a linear regression with sensitivity measure d' as a dependent variable (consumers only)

Variables	Coefficient	Robust standard error
Clean set	-0.0014	0.0888
Dutch	-0.4660***	0.1051
Age	-0.0178***	0.0030
Female	-0.2212	0.1508
No secondary education	0.0875*	0.0441
Secondary education or education missing	Ref.	
Higher secondary education	0.1215	0.1368
University	-0.0080	0.1795
Visual handicaps during the test	-0.5490*	0.2475
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.1749	0.1927
No payment preference or missing	0.1666	0.1394
Has checked banknotes in the last 6 months	0.3607	0.1981
Constant	3.5190***	0.1755
Number of observations	250	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sensitivity measure *d'* as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table A4 Results from a linear regression with sensitivity measure d' as a dependent variable (cashiers only)

Variables	Coefficient	Robust standard error
Clean set	0.1407	0.1228
Dutch	-0.0861	0.1091
Age	-0.0102***	0.0028
Female	-0.1748	0.1248
No secondary education	-0.3437	0.3757
Secondary education or education missing	Ref.	
Higher secondary education	0.0900	0.0730
University	0.2351*	0.1080
Visual handicaps during the test	0.3548**	0.1260
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.3382***	0.0655
No payment preference or missing	0.0920	0.1078
Has checked banknotes in the last 6 months	0.4766***	0.1218
Constant	2.9494***	0.2307
Number of observations	261	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sensitivity measure *d'* as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table A5 Results from a linear regression with bias measure β as a dependent variable (consumers only)

Variables	Coefficient	Robust standard error
Clean set	-0.4557*	0.2370
Dutch	0.1332	0.2842
Age	0.0290***	0.0049
Female	0.1671	0.2666
No secondary education	-0.4533	0.5306
Secondary education or education missing	Ref.	
Higher secondary education	0.0288	0.2775
University	-0.0266	0.2674
Visual handicaps during the test	0.0637	0.3439
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.2759	0.3312
No payment preference or missing	-0.0090	0.2826
Has checked banknotes in the last 6 months	0.0585	0.3185
Constant	1.3630***	0.3627
Number of observations	250	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sensitivity measure β as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table A6 Results from a linear regression with bias measure β as a dependent variable (cashiers only)

Variables	Coefficient	Robust standard error
Clean set	-0.6383***	0.1337
Dutch	-0.0814	0.1024
Age	0.0216***	0.0049
Female	-0.0224	0.1866
No secondary education	1.3944	0.8821
Secondary education or education missing	Ref.	
Higher secondary education	0.3675	0.2983
University	-0.0473	0.2649
Visual handicaps during the test	-0.0844	0.3198
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	-0.3703	0.2142
No payment preference or missing	-0.0168	0.2341
Has checked banknotes in the last 6 months	0.0224	0.1841
Constant	2.2890***	0.3230
Number of observations	261	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sensitivity measure β as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table A7 Results from a linear regression with average sorting time per banknote (in seconds) as a dependent variable (consumers only)

Variables	Coefficient	Robust standard error
Clean set	-0.0881	0.3143
Dutch	0.6021	0.3658
Age	0.0073	0.0122
Female	-0.3485	0.3152
No secondary education	-0.9553	0.6238
Secondary education or education missing	Ref.	
Higher secondary education	0.1490	0.4809
University	-0.2456	0.4143
Visual handicaps during the test	-0.1009	0.6425
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.0188	0.4957
No payment preference or missing	0.1930	0.4131
Has checked banknotes in the last 6 months	1.1660	0.8057
Constant	5.7002***	0.7799
Number of observations	250	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sorting time per banknote in seconds as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table A8 Results from a linear regression with average sorting time per banknote (in seconds) as a dependent variable (cashiers only)

Variables	Coefficient	Robust standard error
Clean set	0.2679**	0.1124
Dutch	0.4149*	0.1765
Age	0.0125	0.0090
Female	-0.1831	0.3213
No secondary education	-1.0704*	0.5056
Secondary education or education missing	Ref.	
Higher secondary education	-0.2607	0.2606
University	0.1740	0.4002
Visual handicaps during the test	-0.1881	0.4542
Prefers cash as payment instrument	Ref.	
Prefers cards as payment instrument	0.3957	0.2248
No payment preference or missing	0.3184	0.1678
Has checked banknotes in the last 6 months	-0.5734	0.4266
Constant	4.6938***	0.5533
Number of observations	261	

Note 1: The table presents estimated coefficients and robust standard errors of a linear regression model with the sorting time per banknote in seconds as dependent variable. Standard errors are clustered on the set level.
Note 2: Statistical inference is based on two-sided t-tests. *, **, and *** indicate statistical significance at the ten, five, and one percent level, respectively.

Previous DNB Working Papers in 2016

No. 493 **Jacob Bikker, Dirk Gerritsen and Steffie Schwillens,** Competing for savings: how important is creditworthiness during the crisis?
No. 494 **Jon Danielsson and Chen Zhou**, Why risk is so hard to measure
No. 495 **Gabriele Galati, Irma Hindrayanto, Siem Jan Koopman and Marente Vlekke**, Measuring financial cycles with a model-based filter: Empirical evidence for the United States and the euro area
No. 496 **Dimitris Christelis, Dimitris Georgarakos, Tullio Jappelli and Maarten van Rooij**, Consumption uncertainty and precautionary saving
No. 497 **Marco Hoeberichts and Ad Stokman**, Price level convergence within the euro area: How Europe caught up with the US and lost terrain again
No. 498 **Janko Cizel, Jon Frost, Aerdt Houben and Peter Wierts**, Effective macroprudential policy: Cross-sector substitution from price and quantity measures

DeNederlandscheBank

EUROSYSTEEM

De Nederlandsche Bank N.V.
Postbus 98, 1000 AB Amsterdam
020 524 91 11
dnb.nl