The Expected Error Method: A new measure of overprecision applied to a representative sample *

Ciril Bosch-Rosa[1], Bernhard Kassner[2], and Steffen Ahrens[1]

[1]Chair of Macroeconomics, Technische Universität Berlin
[2]Seminar for Economic Policy, Ludwig Maximilians Universität München

July 13, 2020

Abstract

[*Preliminary, please do not cite or circulate*] In this paper we introduce the Expected Error Method, a new way of eliciting overprecision. Overprecision is a type of overconfidence that results from an excess of confidence in one's own judgment which may lead to buy less insurance than optimal, the under-diversification of portfolios, or to increase asset price volatility. Unlike previous approaches to measure overprecision our method is simple to understand, quick to implement, and can be carried out by paper and pen. We study the predictive power of our new method by including it in a nationally representative survey. In line with the theoretical predictions, our measure shows that a higher degree of overprecision results in lower portfolio diversification, larger stock price forecasting errors, more extreme political views, but also a higher likelihood of not voting.

Keywords Overconfidence, SOEP, Survey

JEL Classification C83 · D91 · G41

*Corresponding author: cirilbosch@gmail.com. We acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG) though the CRC TRR 190 "Rationality and Competition." The order of the authors is not alphabetical as the first two authors contributed relatively more to the paper.

1 Introduction

Overprecision is a type of overconfidence that results from an excess of confidence in one's own judgment (Moore et al., 2015). Some examples of overprecision would be to claim that a newly-minted transatlantic boat is "unsinkable," to launch an invasion on Imperial Russia in summer, or to not vaccinate your children against measles. In all of the previous cases, decision-makers underestimated the noisiness of the data they had at hand and assumed that their judgments were less subject to error than they really were.

From an economic point of view, overprecision may lead consumers to buy less insurance than they should (Grubb, 2015) or to large distortions in corporate investment decisions (Ben-David et al., 2013; Moore et al., 2015). In finance, overprecision has been shown to be responsible for under-diversification of portfolios as well as asset price volatility (Goetzmann and Kumar, 2008; Scheinkman and Xiong, 2003). However, even if overprecision has such negative consequences, we understand very little of it (Mannes and Moore, 2013).

One of the reasons for our lack of understanding of overprecision is that it is hard to measure. The most common way was introduced by Alpert and Raiffa (1982) and consists in asking respondents for confidence intervals (CI) for a series of numerical questions (e.g., how long is the river Nile).[1] However, the literature has shown that this method creates implausibly high measures of overprecision resulting from participants not fully understanding how to use CI's (Moore et al., 2015). This is best shown in Teigen and Jørgensen (2005) where participants give 98% CI's that are practically identical to those resulting from asking for 50% CI's. Other alternatives to measure overprecision are the the two-alternative forced-choice (2AFC) (Griffin and Brenner, 2004) or the Subjective Probability Interval Estimate (SPIES) (Haran et al., 2010) which are either not fully suited to elicit individual level measures of overprecision or are too time-consuming (see Section 2.1 for a more detailed discussion on all methods).

In this paper we contribute to the literature by introducing a new way of eliciting overprecision using the "Expected Error Method." This method consists of a two-step

[1]The idea behind this method is that a perfectly calibrated respondent should get nine out of ten correct answers within the confidence intervals. An overprecise subject would get less than ten out of ten within her CI while an underprecise respondent would get all ten questions within the CI's. In their seminal paper, the 98% CI's of their participants (MBA students) contain the correct answer only 60% of the time.

procedure where we first ask participants a numerical question (e.g., what year did the US Army capture Saddam Hussein?) and then ask them to estimate how "far away" from the correct answer is their response to the first question. In other words, in the second step we ask participants to estimate their expected error. By comparing the true "realized error" to their expected error, we can measure the overprecision of participants in a simple and direct way that is easy to implement and understand.

To test our new method, we bring it to the 2018 Innovation Sample of the German Socio-Economic Panel (SOEP-IS). The SOEP-IS is a representative sample of German households. The richness of this data set allows us not only to study the determinants of overprecision, but also to test whether overprecision affects respondent's life choices (e.g., voting behavior, portfolio diversification or career choices) as predicted by the theory.

The results show that overprecision is negatively correlated with age, years of education, and gross income. Yet, overprecision does not differ across genders. More importantly, we find that overprecision has strong predictive power for several theoretical conjectures. For example, our measure predicts larger forecast errors in the prediction of stock prices and lower portfolio diversification as suggested by Odean (1998) and Barber and Odean (2000). Regarding subjects' political views and behavior, our new measure of overprecision predicts a tendency to hold extreme political ideologies as suggested by Ortoleva and Snowberg (2015b). Yet, in contrast to Ortoleva and Snowberg (2015b), our measure of overprecision is associated with voting absenteeism rather than with an increased likelihood to vote. However, given the different incentives that arise for partisan voters from the two-party in the US and the multiple party system in Germany, this result is to be expected.

The literature on overprecision is rich (see Moore et al. (2015) for a brief overview). Yet only a small number of papers study the effects of overprecision in a representative sample.[2] Ortoleva and Snowberg (2015a,b) estimate a measure of individual overprecision using a representative sample of the US adult population and study the influence of overpresision on ideological extremeness, partisan identification, and voter turnout. Using the same data set, Stone (2019) studies partisan hostility. However, these papers focus on political preferences and voting behavior, while we go beyond this by including data on

[2]Friehe and Pannenberg (2019) study the overconfidence using a representative sample of the German population (using the 2014 wave of the SOEP-IS). However, they study *overplacement*, which, as shown in Moore and Healy (2008), is quite different from *overprecision*.

financial behavior and career choices. Moreover, while Ortoleva and Snowberg (2015a,b) and Stone (2019) have to estimate the individual measure of overprecision of respondents, our new method allows us to directly elicit the overprecision of respondents.

Another important related paper is Enke and Graeber (2019). While not trying to measure overprecision, they use a method very similar to the proposed Expected Error Method to elicit what they call "cognitive uncertainty." In Enke and Graeber (2019), what the authors measure is the *subjective uncertainty* of experimental subjects around subjects' selected *optimal actions*. To do so, they allow subjects to provide a symmetric interval of "uncertainty" around the answers provided to a battery of questions across different domains of (economic) decision-making. Their results show that with this method they can explain different well-known biases such as probability weighting functions in choices under risk (Kahneman and Tversky, 1979) or the tendency to be ambiguity averse for likely events, but ambiguity seeking for unlikely outcomes (Trautmann and Van De Kuilen, 2015). Overall, while the setup of Enke and Graeber (2019) is not designed to measure overprecision and is limited to an online experimental population, it supports our new Expected Error Method as a robust tool to elicit the uncertainty of respondents around a given answer.

To summarize, our paper contributes to the existing literature on overprecision in three dimensions: first, we introduce a novel technique, which we call the Expected Error Method, which gives a direct measure of overprecision, is easy to understand, and can be quickly implemented in surveys. Second, we show that in line with the theoretical predictions, our measure shows that a higher degree of overprecision results in lower portfolio diversification and larger stock price forecasting errors as well as ideological extremism. Third, while most of the existing literature on overprecision uses university students (e.g., Alpert and Raiffa, 1982) or special pools of subjects (e.g, Glaser and Weber (2007) use finance professionals and McKenzie et al. (2008) use IT professionals), we show the validity of our new method using a representative sample.

Our paper proceeds as follows: Section 2 discusses the notion of overprecision, introduces our new measure for overprecision, using the Expected Error Method, and presents the SOEP IS dataset. In Section 3 we correlate overprecision with various personal characteristics. In Section 4 we use our measure of overprecision to predict various outcomes in the domains of prediction errors, portfolio diversification or voting behavior as predicted by the theory. The last section concludes.

2 Expected Error, Overprecision, and Data Details

2.1 Overprecision

Overconfidence is considered to be one of the most pervasive and potent biases in human judgment (Mannes and Moore, 2013; Kahneman, 2013). It leads to fighting wars (Johnson, 2009), to excessive entry into markets (Camerer and Lovallo, 1999), or (less critically) for 80% of the people to think that they are above median drivers (Svenson, 1981). However, overconfidence is a general term that encompasses three different phenomena which are: overestimation, overplacement, and overprecision (Moore and Healy, 2008; Moore and Schatz, 2017). The first one has to do with absolute values, thinking that you are better than you really are. Overplacement has to do with relative values, thinking that your performance is better than that of others. Finally, overprecision has to do with the degree of certainty with which one judges her own knowledge. In other words, overprecision relates to the second moment of the distribution, in which case a person may hold accurate beliefs on average, but underestimate the variance of the possible outcomes (Malmendier and Taylor, 2015).

Of the three types of overconfidence, overprecision is the most robust and least studied (Moore et al., 2015). While widely used in theoretical finance (e.g., Odean, 1998; Scheinkman and Xiong, 2003; Daniel and Hirshleifer, 2015), the empirical literature usually has to make do with proxies such as gender (Barber and Odean, 2001) or using manager's own portfolio of the company's stock options (Malmendier and Tate, 2015). One of the reasons for this lack of direct measures of overprecision is that overprecision is a hard variable to measure (Moore et al., 2015).

The most common way of measuring overprecision is by asking for 90% confidence intervals (CI) for a series of numerical questions (e.g., how long is the river Nile). Using this paradigm, a perfectly calibrated person would get wrong one question out of every ten, while an overprecise person would have more than one answer falling outside of the stated CI. However, the literature has shown that such method creates implausible high measures of overprecision, with the stated 90% CI's of respondents containing the correct answer only between 30% to 60% of the time (e.g., Russo, 1992; Bazerman and Moore, 2013; Moore et al., 2015). The best explanation for such results is that participants are not used to using CI's and do not fully grasp what they are being asked (Moore et al., 2015). This was made clear by Teigen and Jørgensen (2005) who show that the elicited intervals resulting from asking 90% CI's are practically identical to those resulting from asking for 50% CI's.

While there are some alternatives to CI to measure overprecision, these tend to be either time-consuming or limited in the information they provide. For example, the two-alternative forced-choice has respondents choose between two possible answers to a question and must then indicate how confident they are about their answer being correct. By comparing the number of correct answers to the stated confidence one can measure if, on average, respondents are overconfident. However, this method has several drawbacks, as it cannot distinguish between different types of overconfidence, is not suitable for individual measures of overconfidence, and results in data that has several statistical limitations such as non-continuous probability distributions (see Moore et al. (2015); Griffin and Brenner (2004) for a further discussion). Another approach to measure overprecision is to elicit complete probability distributions from respondents as suggested in Haran et al. (2010). While this method seems to measure overprecision more accurately than CI's Moore et al. (2015), it is time-consuming as it requires respondents to first understand the concept of probability distribution and then to build such a distributions for each question. Additionally, because distributions can only be elicited by partitioning the support into discrete bins, researchers need to make a series of *ad hoc* decisions to implement and define the desired 90% boundaries of the distribution.

In contrast to the methods listed in this section, in the following we introduce a new measure which is both easy to understand and implement.

2.2 New Measure of Overprecision: The Expected Error Method

Our new method to measure overprecision - which we call the Expected Error Method - consists in asking two consecutive questions to respondents. The first question (a) can be on any topic, but needs to have a numerical answer.[3] The second question (b) asks

[3]Some examples are, the result multiplying 385 times 67, the Length of the Nile, or the year of the death of Lady Diana. Some examples of questions that do not work are: the name of the oldest son of Lady Diana, the color of the Batmobile, or the gender of the current prime minister of the United Kingdom.

respondents how far away they expect their answer to question (a) to be from the true answer. In other words, the second question asks respondents to estimate their expected error. For example:

(a) *How long (in kilometers) is the river Nile?*

(b) *How far away (in kilometers) do you think your answer to (a) is from the true answer?*

By comparing the expected error of respondents to their realized error, we get a measure of how over-/underprecise the respondent is about her own knowledge on the topic of question (a). Formally, call the answer of respondent i to question j $a_{i,j}$, her expected error for question j $ee_{i,j}$, and the true answer to the question ta_j. Hence, our measure of overprecision for respondent i for question j is:

$$error_{i,j} = |a_{i,j} - ta_j|, \tag{1}$$

$$overprecision_{i,j} = error_{i,j} - ee_{i,j}, \tag{2}$$

where equation (1) measures the true error ($error_{i,j}$) of respondent i to question j. Notice, that this equation calculates the *absolute error*, that is, we do not care about the direction of the error, but rather the size of the error. In equation (2), we calculate the difference between the expected error ($ee_{i,j}$) and the true error ($error_{i,j}$) of subject i to question j. This time we do care about the direction of the error, as a respondent who underestimated her expected error (i.e., $error_{i,j} > ee_{i,j}$) is considered to be *overprecise*, while a respondent who overestimated the expected error (i.e., $error_{i,j} < ee_{i,j}$) is *underprecise*. Finally, those subjects who guessed their expected error (i.e., $error_{i,j} = ee_{i,j}$) are considered to be perfectly calibrated for that question.

Eliciting overprecision using the Expected Error Method rather than confidence intervals has several advantages. First and foremost, respondents do not need to have any statistical knowledge to answer the questions and the setup is easy to explain. Additionally, questions can be answered quickly, and it can be implemented easily in either computerized or pen and paper surveys. Finally, it is easy to make the Expected Error Method incentive compatible, since one can put on top of each question an incentive compatible payment mechanism such as a quadratic scoring rule (Brier, 1950) or the binarized scoring rule (Hossain and Okui, 2013), and then pay randomly only one of the two outcomes to avoid hedging across questions. This stands in contrast with the more complicated scoring rules necessary to make CI's incentive compatible (e.g., Jose and Winkler, 2009).

2.3 SOEP-IS

The data that we use comes from the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). The Innovation Sample is a subset of the larger SOEP panel (SOEP-Core, which has approximately 30,000 individual respondents) and it is designed to host and test novel survey items (see, Richter et al., 2015). We use the 2018 wave of the SOEP-IS which in total had 4,860 individual respondents distributed across 3,232 different households.

The data we use for the construction of our measure is composed by seven different questions in which we ask respondents two answer two things, (a) the year of a specific historical event that occurred not further away than 100 years (b) the distance (in years) between their answer to (a) and the correct answer to (a).[4] In other words, we ask respondents to answer a general knowledge question and then we ask them to state the error they expect to make; their expected error (see Section 2.2).

We asked seven different questions of events taking place between 1938 and 2003. The questions were thought to vary in difficulty and to cover different decades. The content of the questions ranges from the year in which the first version of the Volkswagen Beetle was introduced (1938) to the year in which Microsoft was founded (1975) or when Saddam Hussein was captured by the US Army (2003) (see Table B.1 in the appendix for all of the question and correct answers). These questions were asked to a subset (902) of the individuals in the SOEP-IS 2018 who joined the panel after 2016. We supplement the data by additional personal-related variables from the survey years 2016-2018. We drop 55 individuals which did not answer any of the overprecision questions, since this is the

[4]The precise formulation was in German. For the example in which we ask about the year of the death of Lady Diana we ask: (a) *In welchem Jahr starb Lady Diana, die erste Frau von Prinz Charles?* and then (b) *Was schätzen Sie, wie viele Jahre Ihre Antwort von der richtigen Antwort entfernt ist?*.

main variable of interest, and 43 individuals with incomplete information. In total we end up with a sample of 804 subjects across 583 different households.[5]

3 Overprecision and its Determinants

In Figure 1 we plot the density of the answer $a_{i,j}$ for each question j. The red vertical line marks the correct answer. It is clear from the dispersion of the densities that some questions were easier for respondents than others. In Figure 2 we plot the true error ($error_{i,j}$) in the vertical axis and expected error ($ee_{i,j}$) in the horizontal axis for each of the seven questions. Additionally, we plot a a 45 degree red line, so that any dot above is a subject that was overprecise ($error_{i,j} > ee_{i,j}$) and any point below corresponds to an underprecise answer ($error_{i,j} < ee_{i,j}$). It is therefore clear that, as expected, respondents are overprecise in their answers across all questions, independent of the difficulty.

Since overprecision is measured across seven different questions, internal consistency across these questions is important. The standard approach, the Cronbach α (Cronbach, 1951), assumes that the outcomes are τ-equivalent, i.e. the outcomes differ by an additive constant and thus have equal factor loadings ($T_i = \mu_i + \lambda F + e_i$, where T_i is the outcome of item i with mean μ_i, e_i is the score error and λ is the factor loading on the common factor F) (Morera and Sotkes, 2016). We use an alternative, more general measure – congeneric reliability (see e.g. Cho, 2016). In contrast to the standard α, this measure does not assume equal factor loadings ($T_i = \mu_i + \lambda_i F + e_i$). To construct the congeneric reliability measure, we estimate the factor loadings, $\hat{\lambda}_i$, of each overprecision measure with respect to a common factor and estimate the congeneric reliability according to the formula $\frac{(\sum \hat{\lambda}_i)^2}{(\sum \hat{\lambda}_i)^2 + \sum \hat{\sigma}^2_{e_i}}$, where $\hat{\sigma}^2_{e_i}$ is the estimated variance of the error. This results in a congeneric reliability of .76.

To combine the overprecision measures across all seven questions into a unique value for each subject (OP_i) in a simple manner, we average the measure of overprecision ($overprecision_{i,j}$) for each subject (i) across all questions (j). We plot the density of OP_i

[5]To test whether our estimation sample is still representative of the German population we compare the means of important characteristics in our sample with the weighted means according to the sampling weights in the larger SOEP Core which is representative for the German population. The results in Table B.3 in the appendix show that our sub-sample is at large still representative of the larger SOEP Core with only some minor differences.

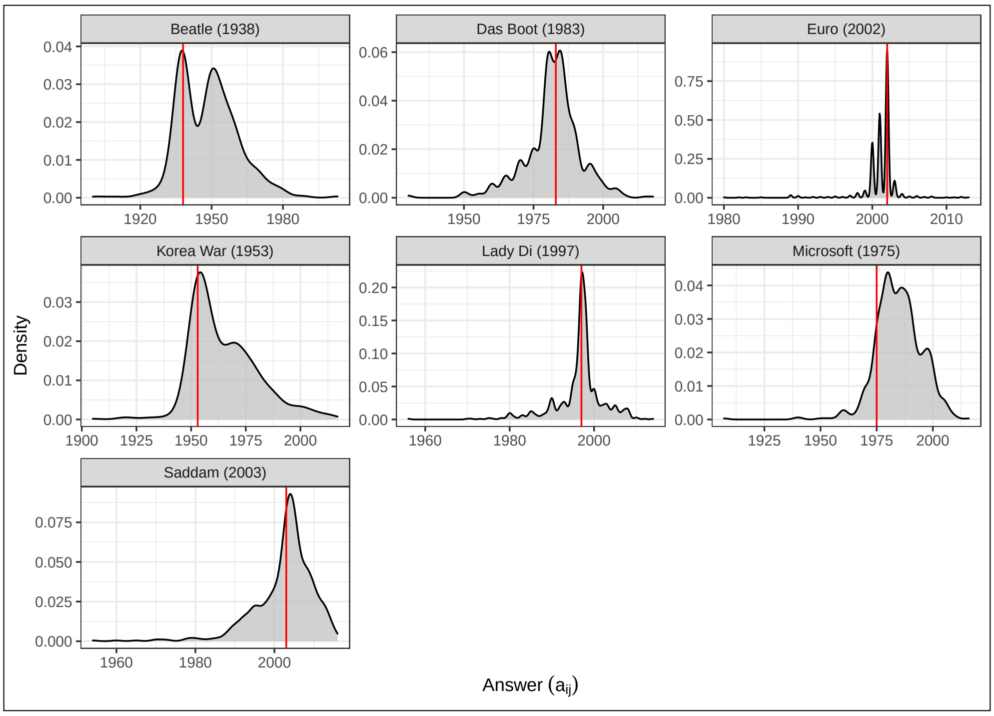


Figure 1: Density of the answers ($a_{i,j}$) for each question. The red vertical line marks the correct answer. Note that the vertical axis is different for each question.

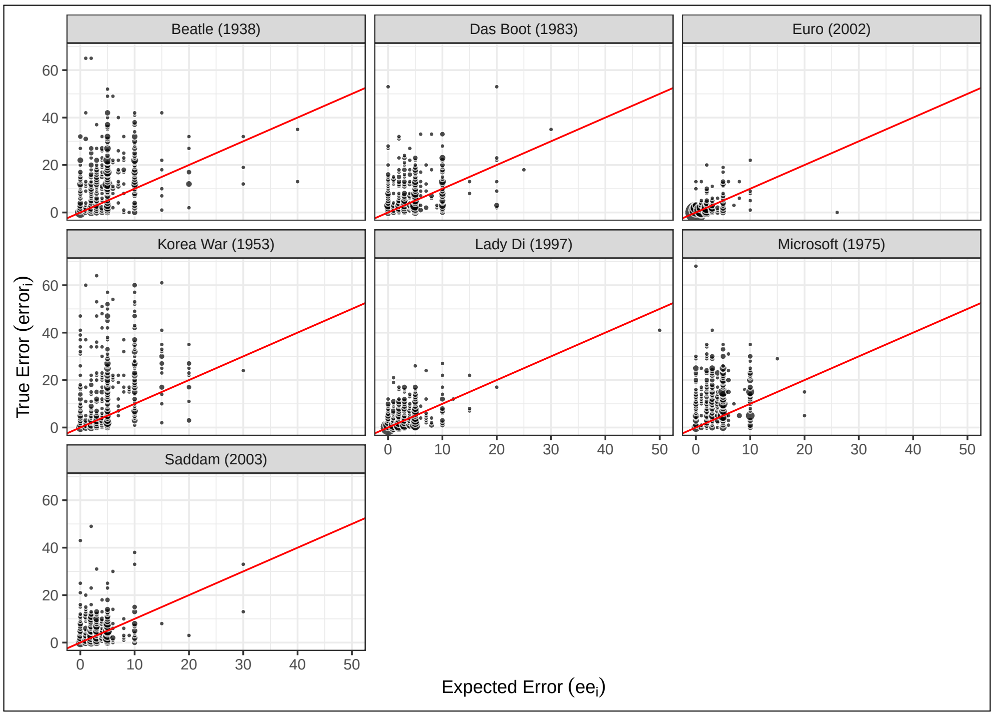
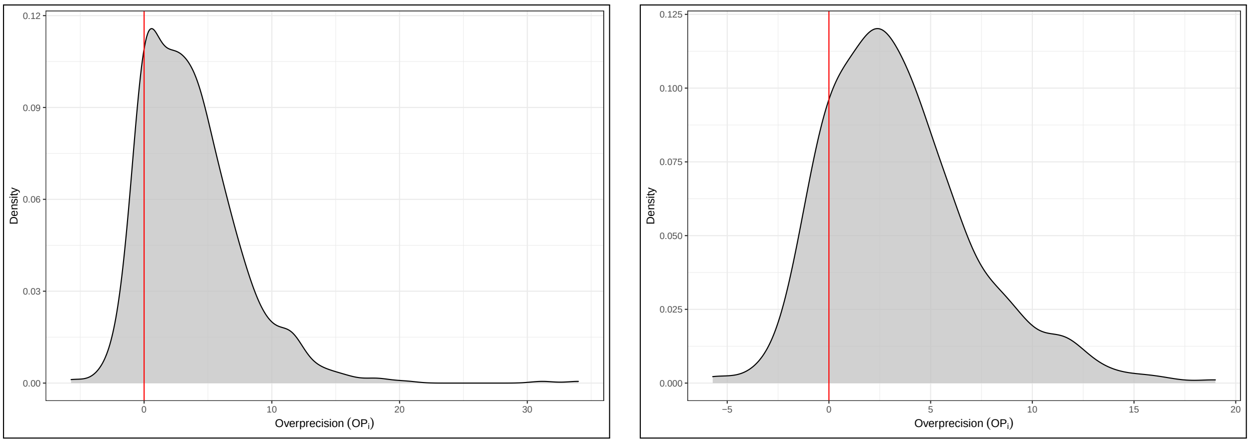


Figure 2: Relation between the true error ($error_{i,j}$) in the vertical axis and the expected error ($ee_{i,j}$) in the horizontal axis. Any dot above (below) the 45 degree red line is an overprecise (underprecise) answer by the respondent.

in Figure 3a.[6] In line with Figure 2, OP_i shows that the large majority of respondents are overprecise. On the other hand, and in contrast with most of the literature using CI's to measure overprecision, we find a relatively large number of subjects that are *underprecise* (approximately 11%). Moreover, 7.5% of respondents seem to be perfectly calibrated (vertical red line in Figure 3a).



(a) Density of OP_i for all respondents (N=804)   (b) Density of OP_i for subset (N=410)

Figure 3: Density of Overprecision (OP_i). In the left panel we plot the density of OP_i, which is the average overprecision for each subject i across all questions j. In the right panel we plot the density of OP_i only for those respondents who answered all questions in the survey.

However, one should take into account that subjects could decide not to answer questions; 50% of the subjects answered all questions with 5% answering only one (see Figure A.2 in the appendix for a detailed breakdown). Of those subjects that are perfectly calibrated 43% answered only one question and only $\sim$ 6% answered all seven. This means that what we see in Figure 3a is an "upper bound" of perfectly calibrated subjects. As can be seen in Figure 3b, once we plot the density function for the subset of subjects that answered *all questions*, then respondents are substantially less calibrated, with the mode of OP_i shifting to the right and leaving only 1% of the subjects being perfectly calibrated, while, at the same time, having and increase in the proportion of underprecise respondents (16%).

Finally, for ease of interpretation, we standardize the aggregate score (OP_i) to be mean zero and standard deviation one and call it Sop_i. In Table 1 we regress the Sop on a series

[6] An alternative would be to construct the composite measure OP_i using a principal component approach as in Ortoleva and Snowberg (2015b). The results of using such approach are very similar to using the average (Spearman $\rho = 0.876$).

of demographic measures using four different OLS models. In all the models we control for age, gender, and years of education. Of these, both age and education are significantly and negatively correlated with overprecision. Given the units, both seem to have a relatively large impact. For example, for every two years of education overprecision is reduced about one tenth of a standard deviation. It is also important to note that the number of questions answered by respondents (*Answered*) is not random, with overprecision increasing as subjects answer more questions (see Figures A.1 and A.2 in the appendix for a graphical overview of these results).

In columns (3) and (4) we add the monthly gross individual income (*Gross Income*) measured in thousands of euros.[7] Additionally, we add dummies for the labor force status (e.g., employed, unemployed, maternity leave, etc.) as well as a dummy for those respondents that were living in East Germany in 1989. The results show a relatively strong effect of income on overprecision, with every two thousand euros reducing overprecision by almost one tenth of a standard deviation. Finally, in column (4) we add federal state (Bundesland) and time of interview fixed effects, which have no qualitative effects on the model of column (3).

The results from Table 1 are in contrast with those of Ortoleva and Snowberg (2015b) who find that neither income nor education are correlated with their measure of overprecision, and that females are significantly less overprecise than males. To analyze the reasons behind these divergent results, in Appendix C we analyze two alternative measures of overprecision. The first one, eliminates the mechanical effect that the age of respondents has on their measured overprecision. This effect comes from the fact that respondents that have lived through an event are better calibrated than those that haven't (see Figure C.1 in the appendix). The second measure is a measure of overprecision that we construct following the estimation method of Ortoleva and Snowberg (2015b) but using the data from our Expected Error Method. As we report in Appendix C, the results replicate those of Ortoleva and Snowberg (2015b), with the surprising exception of *answered*, which has a significant and negative effect of the measure of overprecision defined by Ortoleva and Snowberg (2015b). However, despite these differences in results concerning the determinants of overprecision, the correlation between the two 'robustness'

[7]Since this item is only available for employed individuals, we code missing variables as 0 and include a dummy that is one for missing observations.

Depenedent Variable: *Sop*	(1)	(2)	(3)	(4)
Age	-0.008***	-0.007***	-0.007**	-0.007**
	(0.002)	(0.002)	(0.003)	(0.003)
Female	0.084	0.130*	0.104	0.082
	(0.069)	(0.071)	(0.072)	(0.072)
Years Education	-0.053***	-0.063***	-0.052***	-0.045***
	(0.013)	(0.013)	(0.013)	(0.014)
Answered		0.062***	0.066***	0.069***
		(0.020)	(0.020)	(0.021)
Gross Income			-0.045**	-0.048**
			(0.022)	(0.022)
Constant	1.054***	0.773***	0.550**	0.436
	(0.209)	(0.227)	(0.266)	(0.338)
N	804	804	804	804
adj. R^2	0.035	0.045	0.060	0.084
Fixed Effects	No	No	No	Yes
Employment Status Dummy	No	No	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Determinants of overprecision. In columns (1) - (4) we run an OLS with *Sop* as the dependent variable. In column (3) we include dummies for the labor force status (employed, unemployed, retired, maternity leave, non-working), and whether the respondent was a citizen of the GDR prior to 1989. In column (4) we also control for the federal state (*Bundesland*) where the respondent lives and time at which he/she responded the questionnaire.

measures and the original measure which we use in Table 1 is high. We take these results as a reinforcement of the validity of our new Expected Error Method and proceed to analyze the impact on real life outcomes using our original measure, where any concerns about confounding effects should be mitigated by the inclusion of the determinants as controls.

4 Overprecision and Real Life Outcomes

In this section we exploit the richness of our dataset and examine how overprecision affects real life outcomes in the domains of financial markets and politics. To do so, first, we describe the empirical methodology in section 4.1 and then present the predictions derived from the literature along with our results in section 4.2.

4.1 Methodology

To test the predictions from the theoretical literature on overprecision we use three different procedures. First, we run a regression of each outcome (y_i) on our measure of overprecision and a vector of control variables of the form:

$$y_i = \alpha + \beta Sop_i + \boldsymbol{\gamma' X_i} + \epsilon_i, \tag{3}$$

where Sop_i denotes the standardized overprecision measure, $\boldsymbol{X_i}$ a vector of control variables, and ϵ_i is the random error term.[8] We estimate (3) using OLS and present the point estimate for each coefficient and its p-value respectively in columns (1) and (2) of Table 2. Since we test several different outcomes, we also report the Sidak-Holm p-values adjusted for multiple hypothesis testing in column (3).

Second, we employ a least absolute shrinkage and selection operator (LASSO) to test whether our overprecision measure has predictive power for the outcome variable. LASSO

[8] We include all possible control variables we assume to be be correlated either with the dependent variable or with overprecision. These are: age, gender, financial literacy, years of education, a measure of risk aversion, the monthly gross labor income, dummy variables for the labor force status (employed, unemployed and retired), a measure for impulsivity, patience and narcissism as well as interview date (month and year) and state fixed effects. Additionally we include a dummy for house owners, landlords and for having lived in the GDR in 1989. A test for multicollinearity mitigates concerns of multicollinearity. For the year 2017, we can also construct the Big Five personality traits. However, not all of our subjects were part of the survey in 2017 and by including these personality traits we loose 63 observations. In Table B.4 we include the Big Five personality traits. The results remain robust to including the Big Five personality traits.

is a machine learning application that is frequently applied to improve the predictive power of statistical models. The objective of the LASSO criteria is to choose those variables with the highest predictive power from the set *of all possible control variables*. It does so by estimating a penalized regression by minimizing the sum of squared residuals and a penalty term for the sum of the coefficients.[9] This is implemented via cross-validation, i.e., the estimator partitions the data into different folds of training and testing data and selects the penalty term that minimizes the out-of-sample prediction error in the testing data. If our variable of interest, in our case, Sop, is included in the model, then it has predictive power for the outcome. We report the results of this exercise in column (5) of Table 2.

Finally, we follow Cobb-Clark et al. (2019) and estimate the "R^2 rank" of our variable of interest (Sop). This is obtained doing a step-wise regression in which we sequentially keep adding variables to the model. To do so, we regress the outcome on each variable in our vector of possible controls separately and pick the variable that delivers the highest R^2. We then proceed and regress the outcome on the variable from the previous round and each of the remaining variables separately. This is continued until all variables have been added to the model. The higher the rank of our variable of interest, the more the variable is able to explain the variation in the outcome, i.e. rank 1 delivers the highest R^2.

Using the combination of these approaches allows us to better assess the relevance of overprecision for real life outcomes.

4.2 Prediction Results

The results of our three analytical approaches are summarized in Table 2. The number of observations (Column 7) varies due to missing observations in some outcome variables. Column 1 lists the point estimate of the standardized overprecision measure Sop from the full regression as specified in Section 4.1 along with the unadjusted p-value (Column 2) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column 3) which is slightly less conservative than the Bonferroni adjustment. Column 4 displays the result from the R^2 procedure as specified in Section 4.1 along with the maximum possible vari-

[9] Formally $\min_\beta \frac{1}{2N}\sum_{i=1}^{N}(y_i - \alpha - \sum_j \beta_j x_{ij})^2 + \lambda \sum_j |\beta_j|$ for the linear case, where j are the coefficients which are included in the model and λ is a given tuning parameter. See Tibshirani (1996) for more details.

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	1.046**	0.037	0.232	2/40	yes	0.19	577
opt_dax	0.092***	0.009	0.078	3/40	yes	0.40	577
err_rent	0.332*	0.064	0.282	4/40	yes	0.07	669
err_buy	0.155	0.281	0.733	8/40	no	0.00	643
B Diversification:							
std_divers	-0.126***	0.001	0.010	3/40	yes	0.14	773
C Ideological Positioning:							
std_extreme	0.088**	0.040	0.217	6/41	yes	0.05	715
std_lr	-0.006	0.892	0.892	19/41	no	0.08	715
D Voting behavior:							
non_voter	0.031**	0.012	0.092	3/41	yes	0.12	705

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: This table shows the estimation results of Section 4. The number of observations (Column 7) varies due to missing observations in the outcome variable. Column 1 lists the point estimate of the standardized overprecision measure *Sop* from the full regression as specified in Section 4.1 along with the unadjusted p-value (Column 2) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column 3) which is slightly less conservative than the Bonferroni adjustment. Column 4 displays the result from the R^2 procedure as specified in Section 4.1 along with the maximum possible variables to be included in the model. The political outcomes additionally include political interest. Column 5 specifies the result of the LASSO procedure as specified in Section 4.1 along with the R^2 of the estimated model (Column 6).

ables to be included in the model. The political outcomes additionally include political interest. Column 5 specifies the result of the LASSO procedure as specified in Section 4.1 along with the R^2 of the estimated model (Column 6).

Financial Market Outcomes

We first test predictions regarding financial market outcomes. The first set of predictions concerns forecast errors of asset price predictions in the stock market and in the real estate market. The theory of overprecision in financial markets argues that overprecise investors overweight their private signals when forming expectations, leading investor to hold incorrect beliefs about the future valuation of an asset (e.g., Benos, 1998; Odean, 1998). Moreover, the disagreement regarding the fundamentals of assets driven by overprecision leads traders to contribute to asset price bubbles as they over-optimistically believe that in the future they will find a buyer paying an even higher price (e.g., Scheinkman and Xiong, 2003; Hong et al., 2006). Direct empirical support for this association of overprecision and forecast errors in financial markets is provided by Deaves et al. (2019) who analyse the predictions of German stock market forecasters and correlate them to an accompanying survey-based measures of overprecision. Additionally, Hilary and Menzly (2006) provide evidence consistent with this association for Northern American analysts and Hayunga and Lung (2011) for the US real estate market. The former derive a proxy of overprecision from forecasts, while While the latter proxy overprecision with excessive asset turnover.

Following the logic outlined above, we expect that overprecise subjects exhibit a lower forecast performance and that overprecise subjects systematically make prediction errors into the positive direction, i.e. that they overestimate returns to the stock market and the real estate market. We test the first prediction using the absolute distance of one year ahead predictions of the German Stock Index (DAX), Germany's blue chip stock market index, from the realized value.[10] We test the second prediction using the standardized difference between the one year ahead prediction and the realized value, so that a positive value denotes an overestimation of the stock market realization. Analogously, following Hayunga and Lung (2011), we expect overprecise respondents to systematically make prediction errors regarding the development of real estate markets. Therefore, we test the

[10]Note that the observations from the 2018 waves are almost all within the period before March 2019 and are thus unaffected by the stock market decline caused by the Corona-crisis in March 2020.

predictive power of overprecision on the absolute distance of two year ahead predictions of German housing and rental prices from the realized values.

The results in Table 2 show that our measure of overprecision is an important predictor of forecast errors in asset prices. A one standard deviation increase in overprecision is associated with an increase in the absolute forecast error of 1 percentage point and a .09 standard deviation increase in overestimation of the one year-ahead stock market forecast. Moreover, the results show that a one standard deviation increase in overprecision leads to an increase in the absolute forecast error of rental and housing prices whereby the latter is less strong. The results from the LASSO estimation reveal that overprecision is also a good predictor of these forecast errors since it is selected as an explanatory variable for the models of the stock market forecast and the rental prices, and ranking high (between 2 and 4) in the the R^2 rank approach. We thus conclude, that overprecision is associated with a higher forecast error and an overestimation of future returns both in the financial and real estate markets.

Second, we test the theoretical prediction of Odean (1998) that overprecision is associated with underdiversified portfolios. Intuitively, overprecise investors overweight their private information, thereby trading too frequently while being concentrated on too few favorable assets. Goetzmann and Kumar (2008) provide empirical evidence in support of this prediction for traders in the US and Merkle (2017) for traders in the UK. While the former rely on the asset turnover proxy, the latter elicit overprecision directly through survey questions. We test this hypothesis for the representative German sample using a standardized measure with mean zero and standard deviation of one that captures the degree to which a subject diversifies her portfolio among stocks, real estate, government bonds, savings, and gold.[11]

Our results for the German sample confirm the theoretical prediction that overprecision is associated with underdiversification. The point estimate in column 1 in Table 2 shows that a one standard error increase in overprecision leads to .1 standard deviation decrease in our diversification measure. That means that their optimal portfolio is skewed towards a certain asset category. Moreover, overprecision is among the variables chosen by the LASSO estimation and ranked third in the R^2 rank approach.

[11] For a detailed description of the measure refer to Table B.2 in the appendix.

Political Views and Voting Behavior

For the last set of predictions we study the political views and voting behavior of the respondents. Ortoleva and Snowberg (2015b) define overprecision in the political context as the belief that one's own experiences are more informative about actual politics than they really are. For instance, overprecise people may visit biased media outlets, without fully accounting for this bias, exchange information on social media without realizing that much of the information comes from politically like-minded peers, or inferring from adverse local economic conditions on the general state of the economy opting for drastic policy responses. Against this background, the authors show theoretically and empirically that overprecision in one's own beliefs leads to ideological extremeness and strengthens the identification with political parties, increasing the likelihood to vote. Yet, the literature remains inconclusive whether these associations holds for liberals and conservatives alike. While Moore and Swift (2011) and Ortoleva and Snowberg (2015b) find that conservatives seem more susceptible to overprecision than liberals, Ortoleva and Snowberg (2015a) show that this association only holds in election years.

We use several measures to test the predictive power of overprecision for political views and voting behavior. First, to test whether overprecision predicts ideological extremeness, we use a measure on a scale from 0 to 5 that indicates how far away from the center respondents see themselves in the political spectrum. Second, to test whether overprecision tends to push respondents towards one end of the ideological spectrum , we use respondents answers to the question where they see themselves in the political spectrum with 0 being extreme left and 10 being extreme right. Third, to test whether overprecise respondents a more likely to vote, we use a dummy that equals one if a person indicated to be a non-voter in the opinion poll (Sonntagsfrage) for the 2017 federal elections to the German Bundestag.

In line with Ortoleva and Snowberg (2015b), our results suggest that overprecision is a good predictor of ideological extremeness. Overprecision is among the variables chosen by the LASSO estimation and ranking high (sixth) in the the R^2 rank approach. Confirming Ortoleva and Snowberg (2015a), we do not find evidence that overprecision is associated more strongly with conservatism or liberalism, as overprecision is neither correlated to political ideology nor among the variables chosen by the LASSO estimation. Furthermore, overprecision is ranked quite low (19/41) in the R^2 rank approach. Finally, in contrast to Ortoleva and Snowberg (2015b), we find that overprecision is a strong predictor of voting absenteeism, with overprecision being chosen by the LASSO estimation and ranked fourth in the R^2 rank approach. Hence, as it seems, overprecision increases the likelihood of voting absenteeism rather than increasing the likelihood of voting: with a one standard deviation increase in overprecision resulting in a 3 percentage point increase in the likelihood of not voting.

The last result seems to be in contradiction with the result of Ortoleva and Snowberg (2015b). However, one should be cautious when comparing the voting behavior of overprecise respondents in the US and Europe. In Ortoleva and Snowberg (2015b) partisanship is measured *within* the republican and democratic parties. Because both of these parties have high chances of winning the elections, those more identified with such parties have stronger incentives to vote for them (e.g., Miller and Conover, 2015). As opposed, in Germany, more extreme respondents gravitate to fringe parties (e.g., Die Linke, AfD, NPD)[12] with smaller chances of winning elections,[13] so the incentives to vote are very different than for those in the dataset of Ortoleva and Snowberg (2015b). Hence, the theoretical assumptions underlying Ortoleva and Snowberg (2015b)'s predictions of voter turnout and overprecision are a good description of voting behavior in the two-party system of the US, but are not appropriate for the more disperse German system.

Overall, we believe our results align well with Ortoleva and Snowberg (2015a,b). We confirm their theoretical predictions and empirical findings on how overprecise respondents process information and how this translates into their political views. One would expect this since these results are independent from the electoral system. What the electoral system affects are the incentives to vote, and there we see a difference in the behavior of overprecise respondents.

[12] If we pool all respondents voting for radical parties (AfD, NPD, and Die Linke) and compare it to the voters of the rest of parties, a non-parametric test confirms the tendency of radical party voters to ideological extremeness (Mann-Whitney U p-value<0.001).

[13] Take as an example the explicit (self-imposed) *cordon sanitaire* that all major democratic parties have imposed around the AfD. Angela Merkel's intervention, and the series of resignations, that resulted after the 2019 Thuringian election shows how strongly such *cordon* is enforced.

5 Conclusion

In this paper we introduce the Expected Error Method as a new way to measure the overprecision of survey respondents that is intuitive to respondents and quick to implement. This distinguishes it from all previously used alternatives such as confidence intervals (Alpert and Raiffa, 1982) or probability distributions (Haran et al., 2010), which are either hard to understand for subjects or hard to implement in surveys.

To test the predictive power of the Expected Error Method, we implement it in the 2018 wave of the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). To do so, respondents of the survey are first asked about the year that a relatively known historical event took place (e.g., the year of the introduction of the Euro). They then were asked to estimate “how far away (in years)” their answer to the previous question is from the true answer, in other words, what is their “expected error” in answering the first question. By taking the difference between the “estimated error” and the realized error we get an individual measure of overprecision. We then exploit the rich data of the SOEP-IS to test how well our new method can predict the life outcomes of respondents and their political behavior using both linear regressions and LASSO techniques.

The results show that the Expected Error Method has strong predictive power for several theoretical implications of overprecision. For example, as suggested in Odean (1998) and Barber and Odean (2000), our new measure predicts larger forecast errors in the prediction of blue chip stock prices and lower levels of portfolio diversification. Additionally, we find that, as predicted and shown in Ortoleva and Snowberg (2015a), more overprecise respondents will hold more extreme political ideologies.

As for the determinants of overprecision, as in Ortoleva and Snowberg (2015b) we find that gender has no effect on overprecision. We also find that, contrary to Ortoleva and Snowberg (2015b) and Prims and Moore (2017), on average, the older a respondent, the less overprecise she is. However, this effect seems to be driven by the type of questions we ask. Finally, we also find that both years of education and gross income reduce the overprecision of respondents.

In summary, we exploit the rich data-set of the German Socio-Economic Panel to show that the Expected Error Method renders a robust measure of overprecision. Our measure not only predict real life outcomes of respondents, but it does so for different specifications and techniques. Some advantages of our new method are that it it is intuitive, fast to implement, and requires practically no instructions. Additionally, the Expected Error Method can be implemented both in pen and paper and computerized surveys. Therefore, we believe that our approach is an important step forward in measuring overprecision, a type of overconfidence that can have catastrophic consequences (e.g., Scheinkman and Xiong, 2003; Grubb, 2015), from which we all suffer (Moore et al., 2015), but which we rarely acknowledge.

References

Alpert, M. and H. Raiffa (1982): "A progress report on the training of probability assessors," . Cited on pages 2, 4, and 21.

Barber, B. M. and T. Odean (2000): "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *The journal of Finance*, 55, 773–806. Cited on pages 3 and 21.

——— (2001): "Boys will be boys: Gender, overconfidence, and common stock investment," *The quarterly journal of economics*, 116, 261–292. Cited on page 5.

Bazerman, M. H. and D. A. Moore (2013): *Judgment in managerial decision making*, New York: Wiley, 8th edition ed. Cited on page 5.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013): "Managerial miscalibration," *The Quarterly Journal of Economics*, 128, 1547–1584. Cited on page 2.

Benos, A. V. (1998): "Aggressiveness and survival of overconfident traders," *Journal of Financial Markets*, 1, 353–383. Cited on page 17.

Brier, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly weather review*, 78, 1–3. Cited on page 8.

Camerer, C. and D. Lovallo (1999): "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89, 306–318. Cited on page 5.

Cho, E. (2016): "Making Reliability Reliable:A Systematic Approach toReliability Coefficients," *Organizational Research Methods*, 19, 651–682. Cited on page 9.

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2019): "Self-Control: Determinants, Life Outcomes and Intergenerational Implications," *SOEPpapers on Multidisciplinary Panel Data Research*, 1047. Cited on page 15.

Cronbach, L. J. (1951): "Coefficient alpha and the internal structure of tests," *psychometrika*, 16, 297–334. Cited on page 9.

Daniel, K. and D. Hirshleifer (2015): "Overconfident investors, predictable returns, and excessive trading," *Journal of Economic Perspectives*, 29, 61–88. Cited on page 5.

Deaves, R., J. Lei, and M. Schröder (2019): "Forecaster Overconfidence and Market Survey Performance," *Journal of Behavioral Finance*, 20, 173–194. Cited on page 17.

Enke, B. and T. Graeber (2019): "Cognitive uncertainty," Tech. rep., National Bureau of Economic Research. Cited on page 4.

Friehe, T. and M. Pannenberg (2019): "Overconfidence over the lifespan: Evidence from Germany," *Journal of Economic Psychology*, 74, 102207. Cited on page 3.

Glaser, M. and M. Weber (2007): "Overconfidence and trading volume," *The Geneva Risk and Insurance Review*, 32, 1–36. Cited on page 4.

Goetzmann, W. N. and A. Kumar (2008): "Equity Portfolio Diversification," *Review of Finance*, 12, 433–463. Cited on pages 2 and 18.

Griffin, D. and L. Brenner (2004): "Perspectives on probability judgment calibration," *Blackwell handbook of judgment and decision making*, 177–199. Cited on pages 2 and 6.

Grubb, M. D. (2015): "Overconfident Consumers in the Marketplace," *Journal of Economic Perspectives*, 29, 9–36. Cited on pages 2 and 22.

Haran, U., D. A. Moore, and C. K. Morewedge (2010): "A simple remedy for overprecision in judgment," *Judgment and Decision Making*, 5, 467. Cited on pages 2, 6, and 21.

Hayunga, D. K. and P. P. Lung (2011): "Explaining Asset Mispricing Using the Resale Option and Inflation Illusion," *Real Estate Economics*, 39, 313–344. Cited on page 17.

Hilary, G. and L. Menzly (2006): "Does Past Success Lead Analysts to Become Overconfident?" *Management Science*, 52, 489–500. Cited on page 17.

Hong, H., J. Scheinkman, and W. Xiong (2006): "Asset Float and Speculative Bubbles," *The Journal of Finance*, 61, 1073–1117. Cited on page 17.

Hossain, T. and R. Okui (2013): "The binarized scoring rule," *Review of Economic Studies*, 80, 984–1001. Cited on page 8.

Johnson, D. D. (2009): *Overconfidence and war*, Harvard University Press. Cited on page 5.

Jose, V. R. R. and R. L. Winkler (2009): "Evaluating quantile assessments," *Operations research*, 57, 1287–1297. Cited on page 8.

Kahneman, D. (2013): *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 1st edition ed. Cited on page 5.

Kahneman, D. and A. Tversky (1979): "Prospect theory: an analysis of decision under risk," *Econometrica*, 47, 263–291. Cited on page 4.

Malmendier, U. and G. Tate (2015): "Behavioral CEOs: The role of managerial overconfidence," *Journal of Economic Perspectives*, 29, 37–60. Cited on page 5.

Malmendier, U. and T. Taylor (2015): "On the verges of overconfidence," *Journal of Economic Perspectives*, 29, 3–8. Cited on page 5.

Mannes, A. E. and D. A. Moore (2013): "A Behavioral Demonstration of Overconfidence in Judgment:," *Psychological Science*. Cited on pages 2 and 5.

McKenzie, C. R. M., M. J. Liersch, and I. Yaniv (2008): "Overconfidence in interval estimates: What does expertise buy you?" *Organizational Behavior and Human Decision Processes*, 107, 179–191. Cited on page 4.

Merkle, C. (2017): "Financial overconfidence over time: Foresight, hindsight, and insight of investors," *Journal of Banking & Finance*, 84, 68–87. Cited on page 18.

Miller, P. R. and P. J. Conover (2015): "Red and Blue States of Mind: Partisan Hostility and Voting in the United States," *Political Research Quarterly*, 68, 225–239. Cited on page 20.

Moore, D. A. and P. J. Healy (2008): "The trouble with overconfidence," *Psychological Review*, 115, 502–517. Cited on pages 3 and 5.

Moore, D. A. and D. Schatz (2017): "The three faces of overconfidence," *Social and Personality Psychology Compass*, 11, e12331. Cited on page 5.

Moore, D. A. and S. A. Swift (2011): "The three faces of overconfidence in organizations," in *Organization and management series. Social psychology and organizations*, Routledge/Taylor & Francis Group, 147–184. Cited on page 19.

Moore, D. A., E. R. Tenney, and U. Haran (2015): "Overprecision in Judgment," in *The Wiley Blackwell Handbook of Judgment and Decision Making*, John Wiley & Sons, Ltd, 182–209. Cited on pages 2, 3, 5, 6, and 22.

Morera, O. F. and S. M. Sotkes (2016): "Coefficient α as a Measure of Test Score Reliability: Review of 3 Popular Misconceptions," *American Journal of Public Health*, 106, 458–461. Cited on page 9.

Odean, T. (1998): "Volume, Volatility, Price, and Profit When All Traders Are Above Average," *The journal of finance*, 53, 1887–1934. Cited on pages 3, 5, 17, 18, and 21.

Ortoleva, P. and E. Snowberg (2015a): "Are conservatives overconfident?" *European Journal of Political Economy*, 40, 333–344. Cited on pages 3, 4, 19, 20, 21, and 34.

——— (2015b): "Overconfidence in political behavior," *American Economic Review*, 105, 504–35. Cited on pages 3, 4, 11, 12, 19, 20, 21, 34, 35, and 36.

Prims, J. P. and D. A. Moore (2017): "Overconfidence over the lifespan," *Judgment and decision making*, 12, 29–41. Cited on page 21.

Richter, D., J. Schupp, et al. (2015): "The SOEP Innovation Sample (SOEP IS)," *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 135, 389–400. Cited on page 8.

Russo, J. E., . S. P. J. H. (1992): "Managing overconfidence," *Sloan Management Review*, 33, 7–17. Cited on page 5.

Scheinkman, J. A. and W. Xiong (2003): "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111, 1183–1220. Cited on pages 2, 5, 17, and 22.

Stone, D. F. (2019): ““Unmotivated bias” and partisan hostility: Empirical evidence,” *Journal of Behavioral and Experimental Economics*, 79, 12–26. Cited on pages 3 and 4.

Svenson, O. (1981): “Are we all less risky and more skillful than our fellow drivers?” *Acta Psychologica*, 47, 143–148. Cited on page 5.

Teigen, K. H. and M. Jørgensen (2005): “When 90% confidence intervals are 50% certain: on the credibility of credible intervals,” *Applied Cognitive Psychology*, 19, 455–475. Cited on pages 2 and 6.

Tibshirani, R. (1996): “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288. Cited on page 15.

Trautmann, S. T. and G. Van De Kuilen (2015): “Ambiguity attitudes,” *The Wiley Blackwell handbook of judgment and decision making*, 1, 89–116. Cited on page 4.

A Extra Figures

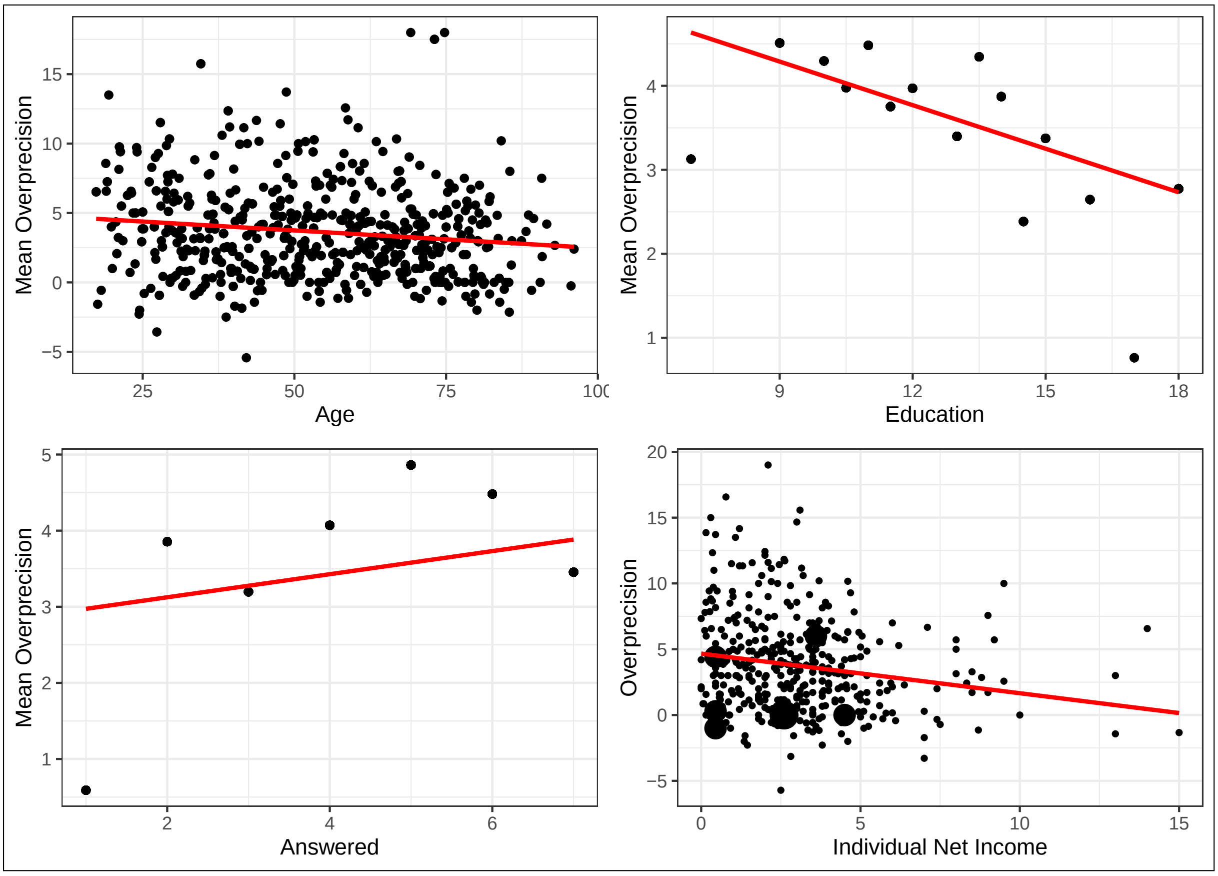


Figure A.1: Correlation of Overprecision. In the vertical axis of the left panel we plot the mean overprecision across all age groups which we plot in the horizontal axis (one outlier with one subject dropped). In the middle axis we plot the mean overprecision across the years of education which we plot in the horizontal axis. In the vertical axis of the right panel we plot the mean overprecision across the number of answered questions, which we plot on the horizontal axis. In all four cases the red line is the fitted linear regression. We dropped one individual outlier in all cases to make the graphs more readable.

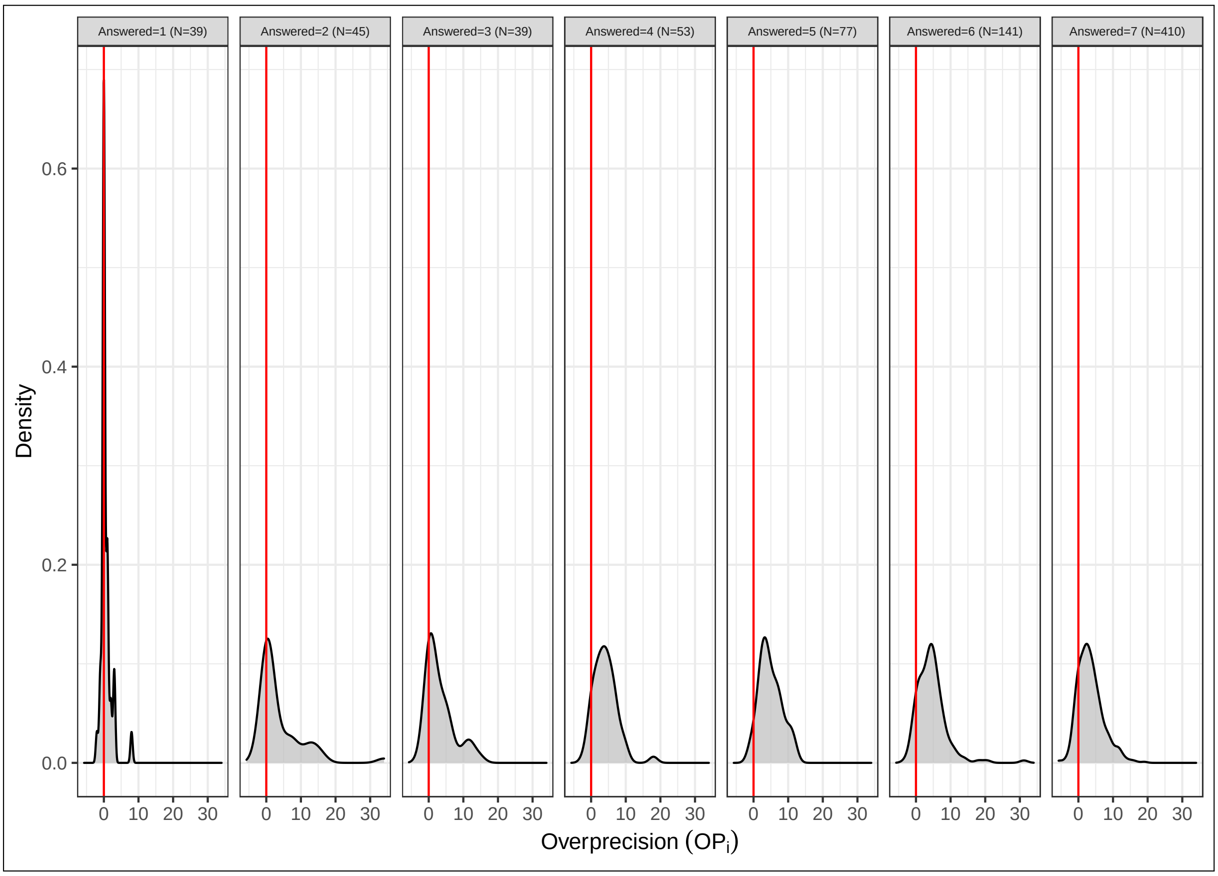


Figure A.2: Density of Overprecision (OP_i) for each of the subsets of questions answered. We plot from left to right the densities of OP_i for those respondents who answered from the minimum number of answers (1) to the maximum number of answers (7). In the title we report the number of respondents for each density.

B Extra Tables

SOEP-IS Code	Question	Answer
Q467 - IGEN02a	In welchem Jahr wurden Euro-Geldscheine und -Münzen eingeführt?	2002
Q470 - IGEN03a	In welchem Jahr wurde das Unternehmen Microsoft (Herausgeber des Betriebssystems Windows) gegründet?	1975
Q473 - IGEN04a	In welchem Jahr kam der Film “Das Boot” (Regie: Wolfgang Petersen) in die deutschen Kinos?	1983
Q476 - IGEN05a	In welchem Jahr wurde Saddam Hussein von der US-Armee gefangen genommen?	2003
Q479 - IGEN06a	In welchem Jahr wurde der erste Volkswagen Typ 1(auch bekannt als “Käfer”) produziert?	1938

Q482 - IGEN07a	In welchem Jahr endete der Korea-Krieg mit einem Waffenstillstand?	1953
Q485 - IGEN08a	In welchem Jahr starb Lady Diana, die erste Frau von Prinz Charles?	1997

Table B.1: Original questions in German language from the SOEP-IS 2018

Variable	Definition
A Prediction error:	
err_dax	Absolute distance between one year-ahead prediction of the DAX realization and the actual realization over the horizon. Data from the first trading day of each month was used depending on the month of the interview. The data does not contain the Corona crash.
opt_dax	Difference between one year-ahead prediction of the DAX realization and the actual realization over the horizon. Positive values indicate an overestimation of the returns. Data from the first trading day of each month was used depending on the month of the interview. The data does not contain the Corona crash.
err_rent	Absolute distance between one year-ahead prediction of rental prices in Germany and the actual realization over the horizon. One year-ahead predictions were linearly derived from two year-ahead predictions. Quarterly data according to the month of the interview was used.
err_buy	Absolute distance between one year-ahead prediction of house prices in Germany and the actual realization over the horizon. One year-ahead predictions were linearly derived from two year-ahead predictions. Quarterly data according to the month of the interview was used.
B Diversification:	
std_divers	Aggregate diversification measure over five asset classes. For each asset class a penalty score is calculated expressing the distance to an equally diversified portfolio. Diversification equals maximum attainable penalty score less actual penalty. The diversification measure is standardized to have mean 0 and standard deviation 1.
C Ideological Positioning:	

Variable	Definition
std_extreme	Absolute distance to center of an ideology scale from 0 (left) to 10 (right). Standardized to have mean 0 and standard deviation 1.
std_lr	Location on an ideology scale from 0 (left) to 10 (right). Standardized to have mean 0 and standard deviation 1.
	D Voting Behavior:
non_voter	= 1 if respondent indicated not to vote in the Sonntagsfrage for the Bundestagswahl 2017.
Controls:	
age	Difference between interview month/year and birth month/year in years.
gender	=1 if female.
east1989	= 1 if living in East Germany in 1989.
std_risk	Location on risk scale from 0 (risk avers) to 10 (risk loving). Standardized to have mean 0 and standard deviation 1.
pgbilzt	Years of education.
pglabgro	Monthly gross labor income in thousands. Missings are coded with zero.
mispglabgro	=1 if missing pglabgro.
finlit	Share of correct answers to 6 questions related to financial knowledge.
owner	=1 if living in own property.
owner_rent	=1 if earning money from renting out property.
std_narcis	Average narcissiscm measure over 6 items on scale from 1 to 6. Standardized to have mean 0 and standard deviation 1.
std_impuls	Location on impulsivity scale from 0 (not impulsive) to 10 (fully impulsive). Standardized to have mean 0 and standard deviation 1.
std_patient	Location on patience scale from 0 (not patient) to 10 (fully patient). Standardized to have mean 0 and standard deviation 1.
empl	=1 if employed.
unempl	=1 if unemployed.
nonwork	=1 if non-working.
matedu	=1 if on maternity, educational or military leave.

Variable	Definition
retire	=1 if retired.
answered	Number of questions answered for overprecision.
retire	Political interest on a scale from 1 (high) to 4 (low). Reversed and standardized to have mean 0 and standard deviation 1.

Table B.2: Overview and definition of the variables used in the analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	SOEP IS		SOEP Core		Difference		
	mean	sd	mean	sd	difference	p-value	N[Core]
Age	53.884	(0.627)	50.535	(0.180)	-3.349	0.000	30,997
Gender	0.509	(0.018)	0.508	(0.005)	0.000	0.984	30,997
German	0.933	(0.009)	0.877	(0.003)	-0.056	0.000	30,997
East (current)	0.174	(0.013)	0.172	(0.003)	-0.002	0.904	30,997
East (1989)	0.187	(0.014)	0.198	(0.004)	0.012	0.413	24,591
Years Education	12.705	(0.097)	17.276	(0.027)	-.429	0.000	28,482
Employed	0.537	(0.018)	0.603	(0.005)	0.066	0.003	30,967
Retired	0.228	(0.015)	0.221	(0.004)	-0.006	0.672	30,967
Unemployed	0.037	(0.007)	0.042	(0.002)	0.005	0.490	30,697
Gross Income	2.943	(0.112)	2.837	(0.029)	-0.106	0.359	17,829
Married	0.567	(0.018)	0.521	(0.005)	-0.047	0.010	30,896
N[SOEP IS]	804						

Table B.3: Representativeness of SOEP IS subsample. This table shows the descriptives of selected personal characteristics of the respondents for the SOEP IS and the SOEP Core. The results for the SOEP IS in comlums (1) and (2) are unweighted whereas the results for the SOEP Core in Columns (3) and (4) are weighted using the sampling weights provided. Columns (5) and (6) show a simple t-test on the difference between the means. Column (7) shows the sample size of the SOEP Core. The sample size varies due to missing observations.

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	1.200**	0.024	0.156	2/45	yes	0.17	536
opt_dax	0.094**	0.011	0.095	3/45	yes	0.41	536
err_rent	0.321*	0.068	0.297	5/45	yes	0.07	623
err_buy	0.139	0.345	0.816	12/45	no	0.00	601
B Diversification:							
std_divers	-0.114***	0.003	0.030	3/45	yes	0.13	718
C Labor market:							
entrepreneur	0.001	0.977	0.977	10/41	no	0.15	393
leader	0.023	0.355	0.732	9/41	no	0.15	393
D Voting behavior:							
std_extreme	0.078*	0.059	0.306	7/46	yes	0.07	715
std_lr	-0.017	0.677	0.896	33/46	no	0.10	715
non_voter	0.029**	0.017	0.128	3/46	yes	0.15	705

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: This table shows the estimation results of Section 4 including the Big Five personality traits. The number of observations (Column 7) varies due to missing observations in the outcome variable. Column 1 lists the point estimate of the standardized overprecision measure *Sop* from the full regression as specified in Section 4.1 along with the unadjusted p-value (Column 2) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column 3) which is slightly less conservative than the Bonferroni adjustment. Column 4 displays the result from the R^2 procedure as specified in Section 4.1 along with the maximum possible variables to be included in the model. The labor market outcomes are estimated only among the working population which decreases the number of observations and the maximum number of variables (labor market dummies are omitted). The political outcomes additionally include political interest. Column 5 specifies the result of the LASSO procedure as specified in Section 4.1 along with the R^2 of the estimated model (Column 6).

C Robustness of Descriptive Results

This initial set of results in Section 3 are in contrast with those of Ortoleva and Snowberg (2015b), who find that neither income nor education are correlated with overprecision, and that females are significantly less overprecise than males.

The negative correlation between age and overprecision in our sample is likely to be driven by the type of questions that was asked in the survey. Since we asked about specific historical events that took place within the last 100 years, individuals who lived during these events might be better calibrated. This becomes obvious in Figure C.1 where, for every question, we divide the density of our overprecision measure $op_{i,j}$ between those respondents born before and after the event. As expected, those subjects born before the event are better calibrated in their answers than those born after the event. As a robustness test we construct, for every respondent, a measure of overconfidence (Sop_i') composed only of the questions asking about events that happened before the individual was born. So, for example, the measure Sop_i' for a person born in 1992 would be composed using only the questions on events posterior to this year, which means using only three of the seven possible answers.[14] Columns 1 to 4 of Table C.1 replicate the regressions from Table 1 using the 'robust' measure (Sop_i'). The results show that, if we exclude the mechanical effect of age, then overprecision and age are positively correlated which is consistent with the earlier results from the literature (e.g., (Ortoleva and Snowberg, 2015a). Otherwise, all of our results remain robust.

As a further robustness test we construct a more complex measure of overprecision similar to the estimation method of Ortoleva and Snowberg (2015b) and run the same OLS models as in Table 1. Ortoleva and Snowberg (2015b) construct their measure of overconfidence by asking respondents about their assessment of the current and one year-ahead inflation rate and unemployment rate as well as their confidence about the respective answers. They proceed by regressing confidence on a fourth-order polynomial of accuracy to isolate the effect of knowledge. The principal component of the four residuals from these regressions is used as measure of overconfidence. To replicate their measure, we regress the expected error on a fourth-order polynomial of the true error and take the principal component of the negative of the residuals across all seven questions to get the

[14]Note that, in doing so, we lose a substantial amount of information and give more weight to events that took place earlier.

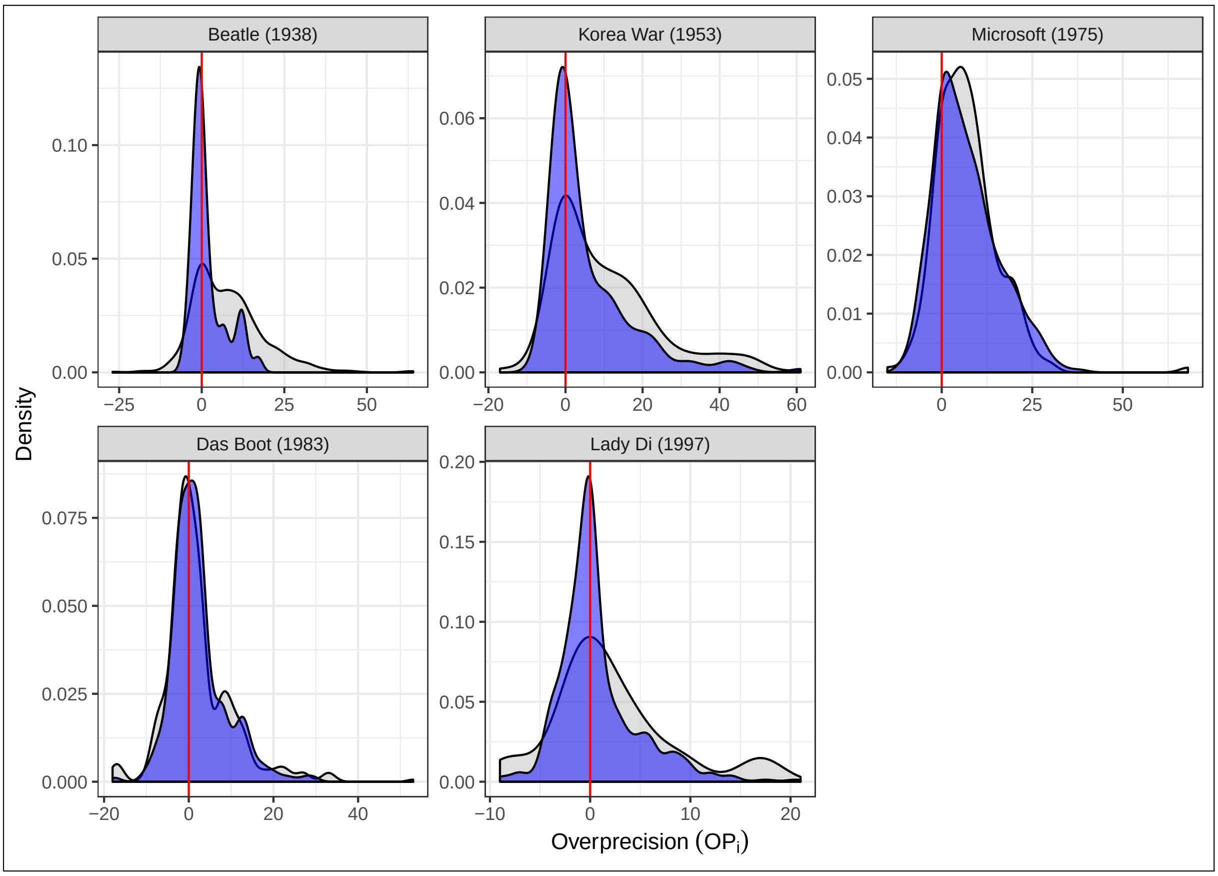


Figure C.1: Density of Overprecision (OP_{ij}) and Age. From left (less recent) to right (more recent) We plot the density of the measured overprecision ($op_{i,j}$) for each question j. In blue we plot the density of all respondents born at the year of the event or before. In grey we plot the density of the measured overprecision for the question ($op_{i,j}$) of those subjects that were born after the event took place. Note that the scale of the vertical axis is different across the five plots. Questions with (correct) answers after 2000 are not included as there were no underage respondents.

new individual measure of overprecision op'_i.[15] This new measure of overprecision (op'_i) also differs from our original measure op_i by taking the relative difference between the expected error and the true error into account.

The results in Table C.1 in the most part replicate the outcomes of Ortoleva and Snowberg (2015b), with females being less overprecise and income and education not showing up as statically relevant. Moreover, age is positively correlated with the estimated overprecision. Surprisingly, the number of answered questions has a negative effect on overprecision. In other words, contrary to observed measure of overprecision, if we estimate overprecision suing the methodology of Ortoleva and Snowberg (2015b), then the more questions a respondent answers, the less overprecise she is.

[15]We use the negative of the residual since both the errors that we use are the inverse of the variables used by Ortoleva and Snowberg (2015b).

Despite the different results concerning the determinants of overprecision, the correlation between the measures is high. Therefore, we proceed the analysis using our baseline measure as described above. The reason is that it is a simple and straightforward approach which can easily be implemented and which does not require the specification of an econometric model, such as the approach of Ortoleva and Snowberg (2015b), and which takes all available information into account, as compared to the robust measure presented in this section. Controlling for the determinants in the subsequent estimations mitigates concerns about confounding effects.

	Robust measure				Residual measure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	0.016***	0.016***	0.017***	0.018***	0.006***	0.006***	0.007**	0.008***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Female	0.014	0.056	0.040	0.020	-0.128*	-0.204***	-0.187**	-0.195***
	(0.068)	(0.069)	(0.072)	(0.071)	(0.070)	(0.071)	(0.073)	(0.072)
Years Education	-0.021*	-0.029**	-0.020	-0.015	-0.017	-0.002	-0.006	0.001
	(0.012)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Answered		0.054***	0.056***	0.064***		-0.102***	-0.102***	-0.088***
		(0.020)	(0.020)	(0.021)		(0.020)	(0.020)	(0.021)
Gross Income			-0.037*	-0.042*			0.014	0.003
			(0.023)	(0.023)			(0.023)	(0.023)
Constant	-0.579***	-0.828***	-0.931***	-0.618*	-0.056	0.403*	0.109	0.039
	(0.205)	(0.225)	(0.309)	(0.371)	(0.211)	(0.227)	(0.314)	(0.374)
N	800	800	800	800	804	804	804	804
adj. R^2	0.080	0.087	0.093	0.123	0.017	0.046	0.047	0.096
Fixed Effects	No	No	No	Yes	No	No	No	Yes
Employment Status	No	No	Yes	Yes	No	No	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.1: Determinants of overprecision using two alternative measures of overprecision. In columns (1) - (4) we run an OLS with the robust measure as the dependent variable. In columnes (5) - (8) we use the approach of Ortoleva and Snowberg (2015b). In column (3)/(7) we include dummies for the labor force status (employed, unemployed, retired, maternity leave, non-working), and whether the respondent was a citizen of the GDR prior to 1989. In column (4)/(8) we also control for the federal state (*Bundesland*) where the respondent lives and time at which he/she responded the questionnaire.