European Commission

ISSN 2443-8022 (online)

Replacing Judgment by Statistics: Constructing Consumer Confidence Indicators on the Basis of Data-driven Techniques

Christian Gayer, Alessandro Girardi
and Andreas Reuter

DISCUSSION PAPER 034 | JULY 2016

EUROPEAN ECONOMY

Economic and
Financial Affairs

European Economy Discussion Papers are written by the staff of the European Commission's Directorate-General for Economic and Financial Affairs, or by experts working in association with them, to inform discussion on economic policy and to stimulate debate.

The views expressed in this document are solely those of the author(s) and do not necessarily represent the official views of the European Commission.

Authorised for publication by José E. Leandro, Director for Policy, Strategy, Coordination and Communication.

This paper exists in English only and can be downloaded from http://ec.europa.eu/economy_finance/publications/.

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on http://europa.eu.

Luxembourg: Publications Office of the European Union, 2016

KC-BD-16-034-EN-N (online)
ISBN 978-92-79-54440-8 (online)
doi:10.2765/151008 (online)

KC-BD-16-034-EN-C (print)
ISBN 978-92-79-54441-5 (print)
doi:10.2765/223095 (print)

European Commission

Directorate-General for Economic and Financial Affairs

Replacing Judgment by Statistics:

Constructing Consumer Confidence Indicators on the Basis of Data-driven Techniques

Christian Gayer, Alessandro Girardi and Andreas Reuter

Abstract

This article compares the properties of the European Commission's Consumer Confidence Indicator (CCI) for the euro area with three alternative indices which differ from the former in that they (i) consider a richer set of survey questions and (ii) are the result of data-driven statistical techniques, rather than the simple arithmetic mean of the input series. The alternative indicators are shown to perform only slightly better than the CCI in tracking real private consumption growth and to fail to produce significantly better forecasts of expansions and contractions in private consumption, once information from relevant, timely available hard data is controlled for. The conclusions change, however, if the analysis is re-conducted on well-defined subsets of survey questions. Concretely, the application of the alternative construction techniques to a data set which is limited to questions about consumers' personal finances produces an indicator which, combined with relevant macro-economic time series, yields significant improvements in forecasting expansions and contractions in private consumption.

JEL Classification: C22, C53, E37.

Keywords: consumer surveys, composite indicators, euro area, principal components analysis, partial least squares, ridge regression, macroeconomic forecasting.

Contact: Andreas Reuter, European Commission, Directorate General for Economic and Financial Affairs, R. de la Loi 170, B-1049 Brussels, Belgium, andreas.reuter@ec.europa.eu. Christian Gayer, European Commission, Directorate General for Economic and Financial Affairs. Alessandro Girardi, Italian National Institute of Statistics and LUISS Lab of European Economics.

CONTENTS

LIST OF TABLES

LIST OF GRAPHS

1. INTRODUCTION

Consumer confidence indicators are closely monitored by economic and financial analysts, as well as policy-makers to inform their judgment about the future evolution of private consumption. While there is a raft of other indicators with a proven bearing on consumption (e.g. income, wealth, interest rates), the view that consumer confidence indicators can provide useful, complementary information has received growing support, as evidenced by the increasing amount of consumer surveys across the globe.

From an academic perspective, the role of consumer confidence is still subject to debate with three major schools of thought: The first one, building on a rational expectations framework with frictionless capital markets, rejects the possibility of consumer confidence having an impact on private consumption, once all other relevant variables have been controlled for (Hall, 1978). The second one, epitomised by Acemoglou and Scott (1994), does concede the ability of consumer confidence to signal changes in future consumption levels, but only because it (partially) reflects income expectations. Assuming that rising income expectations do not translate into higher consumption immediately, since consumers face credit constraints obliging them to delay the adaptation of their consumption levels to the time their income factually increases, consumer confidence should, indeed, help predicting future consumption levels. Similarly, Barsky and Sims (2011) diagnose a significant effect of consumer confidence on future economic activity, but do not interpret their finding as an illustration of a causal link between sentiment and economic outcomes, but rather of the ability of consumer confidence to act as a noisy gauge of (changes in) expected long-run productivity growth. A third approach (Eppright et al., 1998) assumes that consumption behaviour does not only depend on economic, but also on psychological aspects ('animal sprits') ([1]), such as the degree of optimism, uncertainty, etc., which are summarised in the concept of consumer confidence and give the latter an independent effect on private consumption going beyond that of economic fundamentals. According to this view, changes in beliefs which are unrelated to economic fundamentals may have a causal effect on the business cycle, as in Angeletos and La'O (2013).

The diverging theoretical views on the value added of consumer confidence find repercussion in the empirical literature, which aims to determine whether and to which extent consumer confidence has a bearing on private consumption, controlling for economic fundamentals. While several authors show consumer confidence indicators to reduce forecast errors in private consumption models (e.g. Ludvigson, 2004 or, lately, Bruno, 2013), others diagnose only weak effects (e.g. Al-Eyd et al., 2008). More recently, a number of works have also documented the relevance of a “confidence channel” in the international transmission of economic shocks, with agents’ confidence in a given country affecting the level of confidence abroad – in particular in significantly smaller economies (Fei, 2011; Dées and Soares Brinca, 2013).

This article neither aims to provide a definite answer to the theoretical, nor the empirical debate about the usefulness of consumer confidence, but it focusses on an aspect which helps to prepare the ground for finding such an answer, notably on how best to construct consumer confidence indicators. After all, the quality of the consumer confidence measure is likely to have a decisive impact on whether an independent confidence effect on private consumption can be distilled or not.

The most common way of measuring consumer confidence, inter alia applied for the construction of the Michigan Index of (US) Consumer Sentiment and the European Commission (EC) Consumer Confidence Indicator (CCI), is to average the readings of a number of consumer survey questions deemed particularly relevant for gauging confidence. While the simplicity of this approach facilitates an easy communication of the indicators' construction method and results to the wide array of users and ensures a relatively easy interpretability, with up- or downswings clearly attributable to developments in the individual underlying survey questions, the approach is arguably of an "ad-hoc" nature and lacks a genuine statistical background. The potential shortcomings of the aggregation technique appear particularly relevant in the light of a number of contributions which show how some individual consumer survey questions display a

([1]) This notion was originally coined by J. M. Keynes, see Keynes, J. M., 1936. For a contemporary interpretation see Akerlof G. A. and R. J. Shiller, 2009.

higher correlation with private consumption than the corresponding aggregate consumer confidence indicator (see e.g. Jonsson and Linden, 2009, as well as ECB, 2015). Given the existence of a multitude of alternative, statistical techniques for the condensation of several variables into a single one (e.g. principal components analysis), a careful examination of their possible advantages for the purpose of the construction of consumer confidence measures appears warranted.

The present work aims to provide such an analysis. Using data from the European Commission's (EC) consumer survey, several new indicators are constructed and their performance is compared to that of the EC's official CCI for the euro area (cci). The new measures differ in respect of (i) the applied computation method (principal components regression - PCR, partial least squares - PLS and ridge regression - RR), as well as (ii) the type of survey questions which are allowed to feed into the indicator. The latter is an attempt to account for the possibility that certain types of variables might capture consumer confidence particularly well, thereby avoiding the ad-hoc selection of questions (as practised in the context of the cci construction) and instead picking questions based on objective characteristics (forward-looking vs. backward-looking questions, etc.) and clear a-priori assumptions about their relative merits in gauging confidence (e.g. forward-looking questions are more likely to capture confidence than backward-looking ones, which should be answered more on factual grounds).

The chosen approach is inspired by and directly complements a number of related publications, inter alia Gelper and Crux (2010), who apply similar data reduction methods we use to the EC's Economic Sentiment Indicator, as well as Slacalek (2005), who also focusses on alternative aggregation methods for the construction of consumer confidence measures, but resorts to a smaller set of techniques in the case of the Michigan Index of (US) Consumer Sentiment.

A comparison of the alternative indicators to the cci in terms of their behaviour in a vector error correction framework (investigating possible leading or lagging behaviour compared to the cci), as well as their directional accuracy when tracking private consumption growth shows that the alternative methods, applied to the full data-set, produce only slight improvements in the measurement of consumer confidence. Furthermore, once information from relevant, timely available hard data is controlled for, the new indicators fail to produce significantly better forecasts of (expansions and contractions in) private consumption. This conclusion is consistent with the evidence in Gelper and Croux (2010) who show that statistically-based confidence measures perform quite similarly to ad-hoc indicators when forecasting the reference series (EU industrial production in their case). The picture greatly changes though, when applying the alternative techniques to (objectively defined) subsets of data. Concretely, indicators generated exclusively on the basis of survey questions about micro-economic concepts (like households' financial situation, saving and purchasing intentions) are shown to provide a higher degree of complementarity to the information in timely hard data series, thus facilitating improvements in forecasting recessions in private consumption.

The structure of the article is as follows: Section 2 starts off with a presentation of the data, followed by a brief description of the different aggregation techniques, as well as the construction methods of the alternative indicators in Section 3. Section 4 presents some preliminary evidence on the different indicators' ability to track private consumption growth. Section 5 turns to the out-of-sample properties of the indicators, notably in terms of predicting recessions in private consumption growth, both in combination with or void of macroeconomic control variables, differentiating between indicators constructed from all survey questions vs. measures relying only on subsets thereof. Section 6 concludes.

2. DATA

Our analysis taps the wealth of data generated by the EC's Joint Harmonised EU Programme of Business and Consumer Surveys (BCS), which provides monthly business and consumer survey data for each Member State, as well as the five candidate countries, according to a common methodology. (2) The EC uses these data to calculate (national and EU/euro-area wide) indicators summarising the level of confidence in a given economic sector (industry, services, retail trade or construction), as well as among consumers. Our analysis uses the data produced by the EC's consumer survey, thereby focussing exclusively on time series aggregated at euro-area level. (3)

The EC's consumer survey aims to capture information about households' spending and savings intentions, as well as their assessment of variables with a likely impact on these plans. To this end, the survey questions are organised in two main blocs: (i) household-specific questions, which cover respondents' financial situation (Q1, Q2, Q12), their savings (Q10, Q11) and purchasing intentions (Q8, Q9), as well as (ii) questions concerning the economy as a whole, notably the general economic situation (Q3, Q4), price changes (Q5, Q6) and future unemployment developments (Q7). As regards the inquired time-horizon, the questions either refer to current/past developments (Q1, Q3, Q5, Q8, Q9, Q10, Q12) or developments over the next 12 months (Q2, Q4, Q6, Q7, Q11). (4) The answers to a given survey question are summarised in the form of so-called balance series, which display the difference between the percentages of respondents giving positive and negative replies over time. (5) The EC calculates its consumer confidence indicator (cci) as the arithmetic mean of four (seasonally adjusted) balance series, whose underlying survey questions are deemed particularly useful for capturing consumer confidence. These four questions inquire developments over the next 12 months, notably households' financial position (Q2) and savings (Q11), as well as their views on the general economic situation in the country (Q4) and the level of unemployment (Q7).

Since the purpose of this article is to test whether consumer confidence indicators based on statistical data reduction methods outperform an indicator based on a judgmental selection of input series (i.e. the cci), the alternative indicators we propose in this paper may resort to the entirety of the questions inquired by the EC's consumer survey (11 in total). (6) Moreover, in order to further broaden the choice of potentially useful input variables, country-specific balance series are used, rather than the corresponding euro-area aggregates. Against the backdrop of varying vintage lengths across countries, it has been decided to only include time series stretching back as far as 1985. This limitation restricts the analysis to data from 10 countries, namely Austria (AT), Belgium (BE), Germany (DE), Greece (EL), Spain (ES), Finland (FI), France (FR), Italy (IT), the Netherlands (NL), and Portugal (PT), which, however, account for some 97% of euro-area real private consumption over the period 1985-2014. In total, our analysis thus includes 110 balance series.

The reference series we consider throughout the paper is euro-area year-on-year (y-o-y) private consumption growth, as retrieved from the Eurostat-database and reconstructed backward by means of the growth rates from Fagan et al. (2001, 2005) so as to ensure data-availability from 1985 onwards.

---

(2) The surveys are conducted according to a common methodology, which consists essentially of harmonised questionnaires and a common timetable. For more details see the methodological user guide of the BCS Programme: http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/bcs_user_guide_en.pdf

(3) National data are aggregated at euro-area level by applying weights which reflect a given country's share in euro-area private final consumption expenditure (at constant prices).

(4) See Annex 1 for a description of the questions inquired by the EU's consumer survey.

(5) In the consumer survey, respondents can usually choose among six options ("got/get a lot better" (PP), "got/get a little better" (P), "stayed/stay the same" (E), "got/get a little worse" (M), "got/get a lot worse" (MM), don't know (N)), (with PP+P+E+M+MM+N=100). Balances are calculated as B=(PP+½P)-(½M+MM), so that their values range from −100, when all respondents choose the (most) negative option to +100, when all respondents choose the (most) positive option.

(6) Country-specific Q10's were excluded from the analysis due to imperfect harmonisation across countries.

3. THE ALTERNATIVE INDICATORS

3.1. AN OVERVIEW OF THE PROPOSED STATISTICAL METHODS

Arguably, when aiming to summarise the information contained in a number of potentially relevant variables into a single indicator, without embarking on any, necessarily subjective, pre-selection based on experience or intuition, the most straight-forward way would be to include all variables in an ordinary least squares (OLS) regression and let the algorithm determine each variable's weight in the aggregate. The downside of this approach is that, (i) as soon as the number of explanatory variables gets relatively large in comparison to the sample size, the OLS estimator fails and (ii) when predictors are (near) collinear (which is the likelier, the more input variables are used), the variance of the estimated parameters is inflated, giving inaccurate estimates. Since we want to resort to the entirety of the balance-series generated by the EC's consumer survey, we have to rely on data-reduction or regularisation methods, notably principal component regression (PCR), partial least squares (PLS) and ridge regression (RR).

The first two techniques share a common logic, which is to summarise the relevant information in the data set in a limited number of latent variables or 'factors', which are computed such that they are uncorrelated with each other (mutual orthogonality).([7]) Each of the resulting factors thus represents a tendency, which is shared by several (or all) variables in the data set and supposedly constitutes a specific phenomenon. Generally, the first factor summarises the highest share of the variables' co-movement, followed by the second, etc. In keeping with Gelper and Croux (2010), who assume there is just a single force influencing all economic sentiment components, our subsequent analysis will consider the first factor extracted by PCR/PLS as the consumer confidence component, rather than a combination of the first x factors.

Although similar, PCR and PLS differ in one essential aspect: While the former extracts factors exclusively from the pool of consumer survey questions, the latter also incorporates information on the target variable, which is, in our case, private consumption growth. PLS thus describes as much as possible of the co-variance between the dependent variable and the regressors.

The third technique applied in this work, RR, is a special case of a Gaussian generalized linear model (Friedman et al., 2010) which seeks to impose a threshold on the values taken by the coefficients. RR is thus a form of regularised (i.e. constrained) regression. Its main advantage is that it works properly even when the number of predictors exceeds the number of available observations; moreover, although biased, the resulting estimators have lower variance than the standard OLS ones. ([8])

([7]) OLS, PCR and PLS have been tied together by Stone and Brooks (1990) in the context of Continuum Regression (CR), a stepwise procedure, where a generalized criteria is maximized in each step. This criteria depends on a parameter α, where $0 \leq \alpha \leq 1$. As discussed in Helland (2001), when $\alpha=0$ CR gives OLS; if $\alpha=0.5$ then CR is equivalent to PLS, while $\alpha=1$ gives PCR.

([8]) See Annex 2 for a technical description of the PCR, PLS and RR algorithms.

3.2. DERIVING THE ALTERNATIVE INDICATORS IN PRACTICE: AGGREGATION ISSUES AND PUBLICATION LAG

The starting point for the construction of the new confidence indicators is a data set featuring all selected consumer survey balance series, as well as, in the case of the PLS- and RR-based indicator, quarterly (y-o-y) private consumption growth. The inclusion of a quarterly measure poses some intricacies, as the other input variables are of monthly frequency. To align the frequencies of all input variables, the monthly balance series have to be transformed into quarterly ones. At the same time, since the ultimate confidence indicator derived from the data shall be monthly, the quarterly input variables must maintain a monthly interpretation such that the level of the confidence indicator in month 2 of a quarter only reflects the level of the underlying input series in month 2 and not their average readings over the first two or all three months of the quarter. This complexity is resolved by splitting every monthly input series, say A, into three quarterly ones, in a way similar to the "blocking approach" recently applied, among others, by Carriero et al. (2012) and Bec and Mogliani (2015) in the context of economic forecasting and borrowed from the engineering literature of signal processing (Chen et al. 2012). The first quarterly series (A-m1) collects all observations of A referring to the first month of a quarter (i.e. January, April, July and October). The value of A-m1 in quarter 1 is thus represented by the value the originally monthly series was featuring in January, the quarter-2 reading is represented by the April value, etc. By the same token, the quarterly variable A-m2 collects observations from the second months (i.e. February, May, August and November), while the last one (A-m3) assembles the observations from the third months (i.e. March, June, September and December).

Since preliminary analyses suggest the consumer survey balance series to be non-stationary in levels, they are transformed into differences before being used for the indicator construction. The quarter-1 value of the variable A-m1 thus does not feature the January-value of the monthly variable A anymore, but instead the difference between the January-value of A and its reading in October, while the quarter-1 value of A-m2 corresponds to A's reading in February minus that in November, etc. ([9]) As regards the only genuinely quarterly input variable (private consumption growth (y-o-y)), the series is also differenced in order to achieve stationarity.

Having properly aggregated the input series, one can turn to the actual indicator construction. The confidence indicators presented in this paper are all computed in a (pseudo) real-time setup, meaning that, to determine the value of a given indicator for January, only information released up to January may be used. The approach is warranted to get a realistic idea of how the different indicators perform when created under normal data-availability conditions.

Table 3.2.1: **Release calendar of survey (svy) and real private consumption growth (rpg) data**

				Reference quarter								
				Q(t-2)			Q(t-1)			Q(t)		
				m1	m2	m3	m1	m2	m3	m1	m2	m3
Calendar quarter Q(t)	[A]	end month 1 (m1)	svy									
			rpg									
	[B]	end month 2 (m2)	svy									
			rpg									
	[C]	end month 3 (m3)	svy									
			rpg									

Dark grey cells represent the availability of data at a given forecasting date (by rows).
Source: European Commission (DG ECFIN, EUROSTAT)

In practice, for each quarter Q(t), the first computation is conducted at the end of month 1 of that quarter (case [A] in Table 3.2.1). At that point in time, the readings of the survey data referring to month 1 have just been released. Accordingly, the m1-version of each survey variable can be used for the analysis. At

([9]) To render the confidence indicators based on the different techniques comparable, the transformation of monthly input series into quarterly (differences) is not only applied to the PLS- and RR-, but also to the PCR-approach. The latter would, strictly speaking, not require such a transformation, since it does not include quarterly private consumption among the input variables.

the same time, private consumption growth, due to its late release (in general, 65 days after the reference quarter), is only available until quarter Q(t-2).

Applying the PLS- and RR-approach to this tailored data-set, one gets two weighting schemes which determine the relative importance (practically a coefficient) that should be attributed to each of the survey variables to get a good measure of consumer confidence. In a subsequent step, the readings of the survey variables for quarter Q(t) are plugged into the weighting scheme. The resulting value is a summary measure of consumer confidence in month 1 of Q(t). Since the value is derived from input series which have been differenced to render them stationary, it has to be transformed so as to regain a level-interpretation. Practically, one simply adds to it all preceding month-1 readings of the confidence indicator. The entire procedure is repeated several times until one has determined the confidence level for (the first month of) every quarter which shall be covered by the indicator.

To generate the values of the confidence indicators in months 2 and 3 of a given quarter, the identical procedure is conducted, but with modifications to the input variables: The confidence level in the second month of the quarter is calculated on the basis of the m2-version of the survey variables (rather than the m1-version). The third month relies on the m3-verison of the surveys and, additionally, allows for one more observation of private consumption growth (namely Q(t-1)) to be included in the estimation of the weighting scheme (see the data availability for case [C] in Table 3.2.1).

Turning to the PCR-based indicator, the computation mechanism is simpler. While it uses, in every month, the same versions of the survey variables as the PLS- and and RR-approaches, it excludes private consumption growth, which is released with a significant delay. As a consequence, the statistical analysis used to determine the weighting scheme of the variables does not have to end in Q(t-2) or Q(t-1), but reaches until the actual quarter of interest (Q(t)). Practically, this means that the first factor reading extracted by the PCR-approach features in Q(t) represents the relevant confidence value. Depending on whether the m1-, m2- or m3-version of the input variables have been used, it refers to the first, second or third month of quarter Q(t). Same as in the case of the PLS- and RR-approach, the monthly confidence measure is, in a last step, transformed so as to give it a level interpretation.

Irrespective of the type of aggregation technique, we construct the confidence indicators for the period 1995 to 2015. In order to make sure that the quality of the indicators does not vary over time, we choose a rolling in-sample window of 36 quarters for the calculation of the indicators. This means that every value of the confidence indicators, no matter if relating to 1995 or 2015, is based on a statistical analysis whose input data feature the same amount of observations (namely 36). ([10])

([10]) Since the data start in 1985 and we use the first four observations to compute y-o-y growth rates, the use of an estimation window of 36 observations implies considering an out-of-sample horizon of 82 quarters (spanning from 1995q1 to 2015q2). Accordingly, we have tested for the existence of a unit root for all 110 survey questions for each month of each quarter of our forecast sample. The total number of ADF tests which have been conducted is thus (110×3×82=) 27060. The ADF results based on the Schwarz Bayesian Criterion (SBC) for lag selection (with the maximum lag length set equal to four) as well as those from the DF-GLS tests (Elliot et al., 1996) indicate that the unit root null hypothesis cannot be rejected at conventional significance levels for the vast majority of cases (98.5% and 97.1% at the 1% and 5% level of significance, respectively). The KPSS (Kwiatkowski et al., 1992) stationarity tests largely corroborate these conclusions. On the other hand, differencing the series appears to induce stationarity in all cases.

4. SOME PRELIMINARY EVIDENCE

4.1. A LOOK AT THE WEIGHTING SCHEME: CCI VS. ITS ALTERNATIVES

To get a first idea of which forces actually drive the newly constructed indicators and whether these differ markedly from the established cci, we categorise the various input series along three criteria: (i) the country to which they refer, (ii) the survey question they represent, and (iii) whether the underlying survey question refers to future or past/current developments. Subsequently, for each of the new measures, the average weight of the different types of input variables in the final indicator is calculated over the horizon 1995 to 2015.

Panel A of Table 4.1.1 reports the average weights of survey questions by country to which they refer. Since the new indicators depart from the cci's approach of using a country's share in (real) private consumption as its weight in the euro-area aggregate, differences between the new indicators and the cci can be expected. Indeed, the results show that all alternative indices assign a comparatively low weight to the largest euro-area countries (DE, FR, IT), while the opposite holds true for the remaining countries (in particular FI and PT). Recalling that the PCR-/PLS-/RR-approaches allocate weights to the input variables based on (i) the degree to which they co-move with the central, overall tendency followed by the variable set and (ii) (in the PLS- and RR-cases) the degree of co-movement with private consumption, the reason for the relative marginalisation of large countries might be that the trajectory of their balance series is characterised by a high degree of idiosyncrasy. Turning to a comparison of the results across the new indicators, there is a high degree of consistency among them in respect of the weights they attach to the different countries. That finding is in line with the evidence reported in Frank and Friedman (1993).

As regards the average weight accorded to the different survey questions, a comparison between the cci and the alternative indices is not possible, since the former only uses four of the survey questions. When focussing on the three new indices, there seems to be, again, broad consistency among them. ([11]) All attach relatively high weights to a cluster of five questions, namely Q1, Q2, Q3, Q4 and Q7. It is worth highlighting that three of them are also included in the established cci. A look at the content of the five questions shows that they do not have obvious commonalities: They inquire macro-economic concepts (general economic situation, unemployment), as well as micro-economic ones (household's financial situation) and refer to both future and past developments.

In line with the latter observation, the relative weights of questions referring to future (fwd) vs. current/past (bwd) developments are almost even, irrespective of which of the alternative techniques has been applied. The same goes for the distinction between micro- (mic) and macro-economic (mac) questions (see Panel C. of Table 4.1.1).

([11]) The finding of broad consistency among the three techniques both in respect of country-specific, as well as question-specific weights contrasts with the evidence reported in Gelper and Croux (2010) who find that PLS and PCR deliver a quite different weighting scheme. The absence of striking differences among the competing methods in this article can be rationalized in the light of the consistency problems of the PLS algorithm, which requires a much larger number of degrees of freedom than PCR-based estimation methods (Chun and Keles, 2010; Kraemer and Sugiyama, 2011; Cubadda and Guardabascio, 2012; Girardi et al., 2016).

Table 4.1.1: **Weighting schemes by country, question and type of question**

Panel A.	By country			
	cci	*cci* $_{PCR}$	*cci* $_{PLS}$	*cci* $_{RR}$
AT	0.029	0.085	0.094	0.108
BE	0.034	0.124	0.116	0.108
DE	0.290	0.147	0.166	0.147
EL	0.026	0.053	0.050	0.060
ES	0.106	0.116	0.107	0.101
FI	0.017	0.079	0.095	0.108
FR	0.209	0.098	0.101	0.099
IT	0.180	0.117	0.084	0.070
NL	0.053	0.088	0.101	0.097
PT	0.021	0.093	0.086	0.102
sum	*0.967*	*1.000*	*1.000*	*1.000*

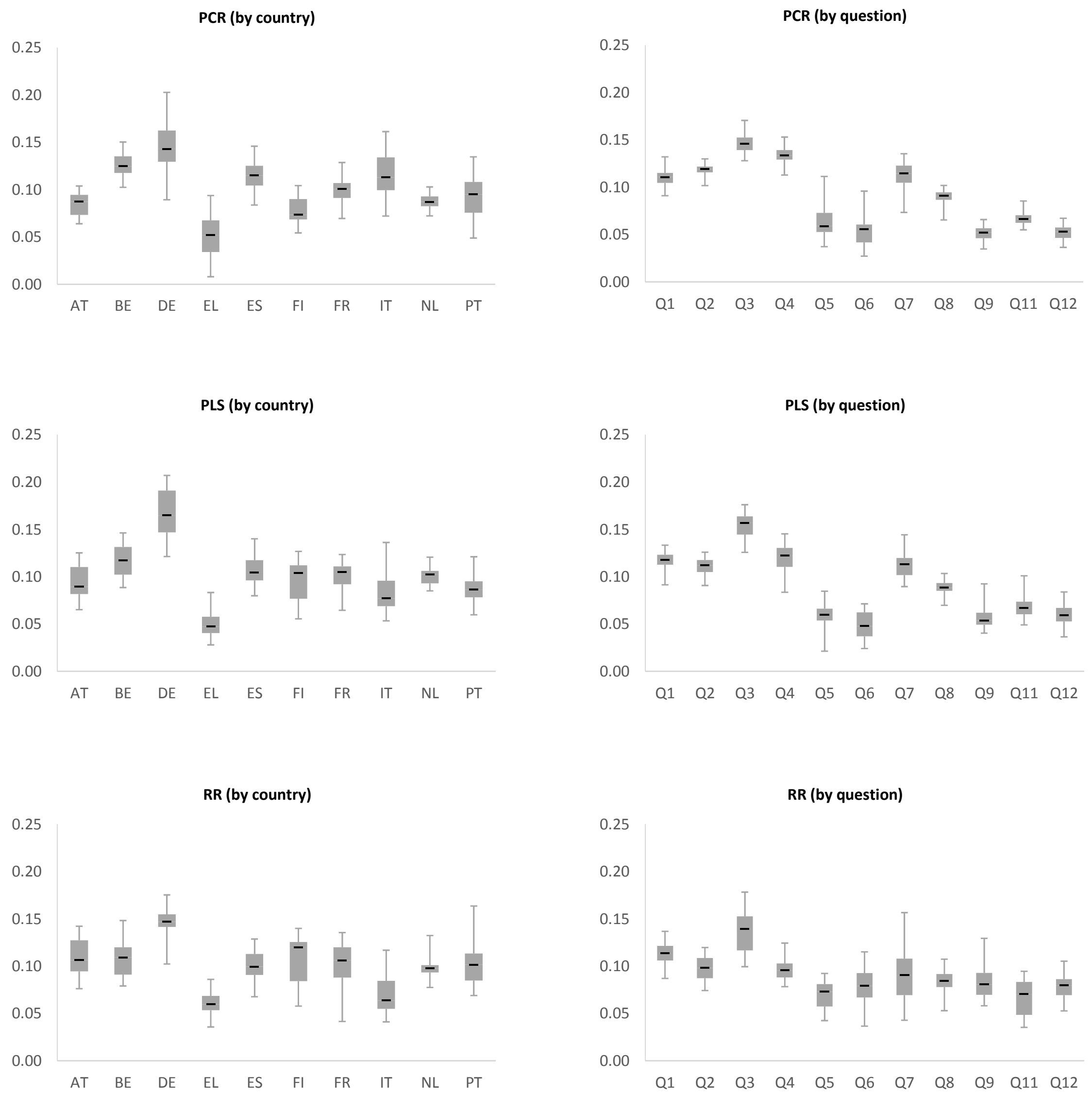
Panel B.	By question			
	cci	*cci* $_{PCR}$	*cci* $_{PLS}$	*cci* $_{RR}$
Q1	0.000	0.111	0.118	0.113
Q2	0.250	0.119	0.110	0.098
Q3	0.000	0.146	0.154	0.137
Q4	0.250	0.134	0.120	0.096
Q5	0.000	0.064	0.058	0.070
Q6	0.000	0.053	0.049	0.079
Q7	0.250	0.113	0.114	0.091
Q8	0.000	0.090	0.089	0.084
Q9	0.000	0.051	0.060	0.085
Q11	0.250	0.067	0.068	0.068
Q12	0.000	0.052	0.060	0.079
sum	*1.000*	*1.000*	*1.000*	*1.000*

Panel C.	By type of question			
	cci	*cci* $_{PCR}$	*cci* $_{PLS}$	*cci* $_{RR}$
bwd	0.0%	51.4%	53.9%	56.8%
fwd	100.0%	48.6%	46.1%	43.2%
mac	50.0%	51.0%	49.5%	47.3%
mic	50.0%	49.0%	50.5%	52.7%

Panel A., B. and C. report the average weights of the input series by the country to which they refer, the survey question they represent and whether the underlying survey question refers to future or past/current developments, respectively.
Source: European Commission

Figure 4.1.1 summarises the estimated country- and question-weights for each of the three alternative techniques in a graphical way, whereby they are expressed in terms of their interquartile range (the height of the grey boxes), min-max range (the distance of the extremes of the vertical lines) and median (the thin black horizontal lines), rather than in terms of their means. It clearly emerges from the graph that the results are robust to the particular way in which the central tendencies are calculated.

Graph 4.1.1: **Weighting scheme by country and by question resulting from PCR, PLS and RR**



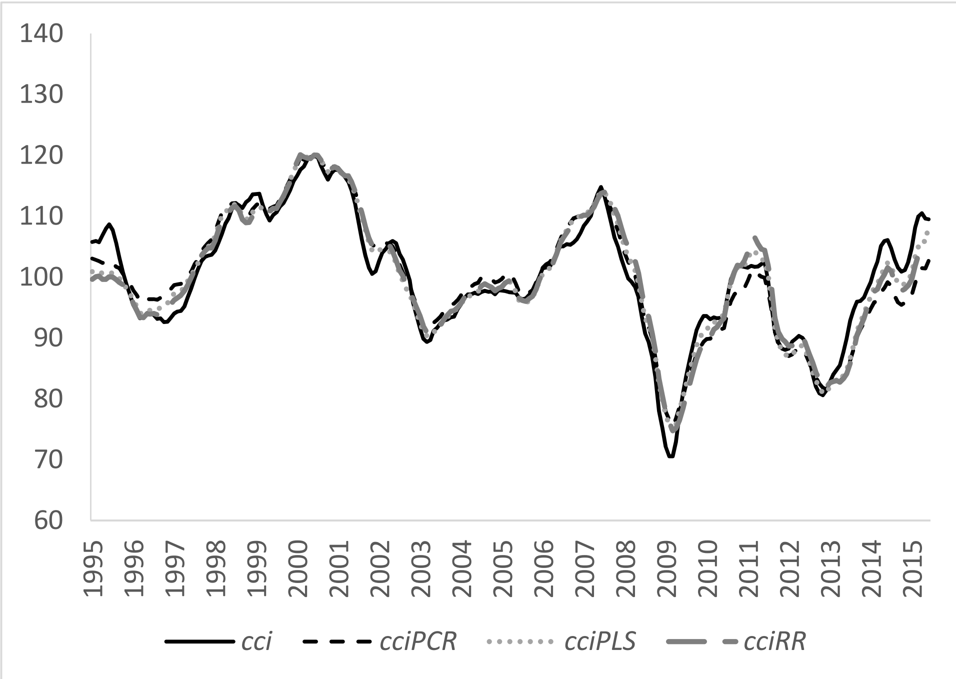
The height of the grey boxes indicates interquartile ranges; the distance of the extremes of the vertical lines are the min-max ranges, while the thin black lines are median values.
Source: European Commission

To get a visual impression of the alternative indicators, they are plotted in Figure 4.1.2. Being rescaled so as to have a mean of 100 and a standard deviation of 10 ([12]), values above 110/below 90 indicate

([12]) This is achieved by (i) subtracting from a given confidence indicator its mean, (ii) dividing the resulting amount by its standard deviation and then (iii) multiplying the measure by 10 and (iv) adding 100.

extremely positive/negative values, when compared to the indicators' usual readings. By and large, the alternative confidence indices move closely together with the established cci, although the latter is based on only four, rather than eleven, survey questions. The major differences relate to 2001/02 and 2008/09, where the competing indices display a relatively less pronounced cyclical behaviour (in terms of amplitude) than the cci. Moreover, the period of the mid and late nineties shows the alternative indicators slightly leading the up- and downswing movements of the cci. Reversely, a slight leading behaviour of the established cci can be detected towards the end of the sample span, where some signs of divergence among the indicators also emerge.

Graph 4.1.2: **Established cci and its three proposed alternatives (monthly values)**



Source: European Commission

In order to better qualify the dynamic relationship between the established cci and its alternatives, a Vector Error Correction (VEC) framework has been applied. Intuitively, VEC models examine whether two (or more) individually non-stationary time-series can be linearly combined in such a way that the resulting series is stationary. The combinations which allow for such stationarity are summarised in so-called co-integration vectors and effectively constitute the steady-state configuration which the model tends to revert to in the long-run, once the effect of transitory shocks fades away. In our case, three bivariate VEC models (cci vis-à-vis $cci_{alt} = cci_{PCR}, cci_{PLS}, cci_{RR}$) of the form

$$\begin{bmatrix} \Delta cci_t \\ \Delta cci_{alt,t} \end{bmatrix} = \sum_{j=1}^{p-1} \Gamma_j \begin{bmatrix} \Delta cci_{t-j} \\ \Delta cci_{alt,t-j} \end{bmatrix} + \Pi \begin{bmatrix} cci_{t-1} \\ cci_{alt,t-1} \end{bmatrix} + u_t, \quad u_t \sim N(0, \Sigma_u) \tag{1}$$

have been specified. Estimating model (1) requires two steps. First, the lag length p is chosen such that the estimated residuals match the multi-normal distribution as closely as possible, this being an essential requirement for a correct statistical inference. Second, the long-term component of the model is identified. The number of cointegration vectors is equal to the (reduced) rank of the matrix Π and is determined on the basis of two tests: the trace test and the maximum eigenvalue test (Johansen 1995). Being of reduced rank, matrix Π can be partitioned as $\alpha\beta'$, where matrix β contains the cointegration vector, while matrix α contains the feedback coefficients (loadings): $\begin{bmatrix} \alpha_{cci} \\ \alpha_{cci_{alt}} \end{bmatrix} [1 \quad -\beta_1]$.

The lag-length has been chosen according to the usual optimal lag criteria, with the maximum tested lag set equal to 12. While the Akaike Information Criterion (AIC) suggests choosing five lags for all models, the Schwarz Bayesian Criterion (SBC) hints at two to four. Faced with that alternative, we prefer to allow for a richer system specification (i.e. five lags). Misspecification tests (available on request) indicate that the estimated residuals match the multi-normal distribution in a satisfactory way both at single equation and system level; moreover, Chow tests indicate the presence of no residual instability in the model, thus suggesting that the estimated pair-wise relationships do not vary over time due to structural breaks.

The results of the VEC models are presented in Table 4.2.1. As Panel A shows, standard cointegration tests indicate the existence of a long-run relationship between the cci and the respective alternative confidence indicator at the 5% (or even a lower) significance level. The central part of the Table contains the specification of the cointegration space, normalized on cci. As evidenced by the standard errors in parentheses, the coefficient of the alternative indicator, which is very close to unity in absolute terms, is statistically significant. Testing for the restriction $\beta_1 = -1$ corroborates this finding in all three cases (see the last row of Panel B.), implying that deviations between the two series are merely erratic.

Panel C reports, for each of the three models, the estimated loading coefficients, which indicate how much a given variable adjusts to deviations from the equilibrium in order to reinstall it. The following considerations are due. Firstly, all models exhibit paths of adjustment to the long-run in a way consistent with an error correction mechanism. Secondly, the feedback coefficients for Δcci are statistically significantly different from zero, while the ones associated with the respective alternative indicator are not (see the p-values in the lowest section of the Panel), suggesting that the adjustment takes place via Δcci with the alternative indicators acting as a sort of short-run exogenous (forcing) variables. Overall, our findings hint at some leading tendency of the new confidence indicators vis-à-vis the cci.

Table 4.2.1: **Bivariate VEC models: long-run structure and adjustment towards the steady-state**

Panel A.	Cointegration tests			
	Trace test		Max eigen. test	
rank	stat	pval	stat	pval
cci				
cci_{PCR}	22.02	[0.03]	18.60	[0.02]
	3.43	[0.50]	3.43	[0.50]
cci_{PLS}	38.07	[0.00]	31.01	[0.00]
	7.06	[0.12]	7.06	[0.12]
cci_{RR}	38.71	[0.00]	31.20	[0.00]
	7.48	[0.11]	7.48	[0.11]

Panel B.	Cointegration space			
	cci	cci_{PCR}	cci_{PLS}	cci_{RR}
	1	-0.917	.	.
		(0.069)		
	1	.	-0.944	.
			(0.040)	
	1	.	.	-1.044
				(0.052)
H_0: $ß_1 = -1$		[0.06]	[0.21]	[0.49]

Panel C.	Short-run adjustment			
	cci	cci_{PCR}	cci_{PLS}	cci_{RR}
	-0.161	0.051	.	.
	(0.042)	(0.034)		
	-0.200	.	0.013	.
	(0.060)		(0.050)	
	-0.142	.	.	0.062
	(0.053)			(0.040)
H_0: $\alpha_{cci} = 0$		[0.00]	[0.00]	[0.00]
H_0: $\alpha_{cci\ alt} = 0$		[0.17]	[0.81]	[0.17]

Source: European Commission

4.3. EVIDENCE FROM DIRECTIONAL ACCURACY TESTS

While the findings of the previous section show that all alternative indicators share a long-term relation with the cci, it is still conceivable that they behave very differently in the short-run. Considering that confidence indicators are usually consulted to get information about short-term developments in private consumption, it is warranted to examine whether one or several of the alternative indicators perform particularly well in tracking short-term developments in private consumption.

Arguably, a good tracking performance has two components: (i) the indicator should provide a rough idea of the expected level of the (later released) reference series, e.g. whether the latter's reading will be above or below its long-term average (test 1); (ii) the indicator should move in the right direction with respect to the series being tracked and thus allow getting an idea of whether the reference series will strengthen or weaken (test 2). To shed light on the relative performance of the confidence indicators against these criteria, we conduct two directional accuracy tests, which determine the percentage of times that a given confidence indicator features the same value as the reference series. For test 1, both the confidence indicators and the reference series are transformed into dummies taking the value 1 if their level (expressed in quarterly y-o-y changes) is above their long-term average. In the context of test 2, the confidence indicators and the reference series are expressed as dummy variables which take the value 1 if the change in the y-o-y series from one quarter to another is positive.

To get a realistic idea of the indicators' performance, the exercise is conducted in (pseudo) real-time, meaning that, for instance, a confidence indicator's quarterly reading in month 1 of quarter 1 is only based on survey releases up to (including) January, while the quarterly reading in month 2 of quarter 1 can resort to survey data up to (including) February, etc. Furthermore, the real-time approach implies that the long-term average used for the calculation of the dummies in test 1 changes over time.

The results of the directional accuracy tests are summarised in a contingency table, where the two columns refer to the reference series (ref^+, ref^-) and the two rows are associated with the four confidence indicators (cof^+, cof^-) that have been considered (cci, cci_{PCR}, cci_{PLS}, cci_{RR}):

$$\begin{array}{ccc} & ref^- & ref^+ \\ cof^- & n_1 & n_2 \\ cof^+ & n_4 & n_3 \end{array} \tag{2}$$

According to condition (2), three directional accuracy rates can be computed: $\%^{all} = (n_1 + n_3)/n$, $\%^{pos} = n_3/(n_2 + n_3)$, $\%^{neg} = n_1/(n_1 + n_4)$, where n indicates the total number of observations (243 months from 1995q1 to 2015q1). When the number of cases in the diagonal (n_1 *and* n_3) is sufficiently large compared to n, the forecasts can be considered to be directionally accurate. To test this feature, we run a χ^2 independence test, as devised in Carnot et al. (2005).

Table 4.3.1 reports these metrics computed for both test 1 (Panel A.) and test 2 (Panel B.). Overall, each of the alternative indicators (in levels) provides a good reflection of year-on-year private consumption growth rates (Panel A. - $\%^{all}$). The percentage of cases where confidence indices indicate correctly whether consumption growth is above or below average is reasonably high, ranging between 69 (for cci and cci_{RR}) and 78% (for cci_{PCR}). Looking at the directional accuracy rates by distinguishing between above- and below-average consumption growth phases ($\%^{pos}$ and $\%^{neg}$, respectively), the share of correct cases ranges between 60 and 87%. As evidenced by Panel B., the cci and the three proposed alternatives also perform satisfactorily in signalling whether private consumption growth accelerates or decelerates, i.e. whether the change in consumption growth rates is positive or negative (Panel B. - $\%^{all}$). The share of correctly identified accelerations or decelerations is around 60% for all alternative confidence indicators. The numbers remain largely unchanged when looking separately at the percentage of correctly identified accelerations and decelerations (Panel B. - $\%^{pos}$ *and* $\%^{neg}$ *respect*ively). The fact that the performance of the indicators is weaker when tracking differences (rather than levels) of the reference series is no particularity of the confidence indicators proposed in this article, but a well-known characteristic of all survey-based indicators.

The last (comforting) conclusion emanating from Table 4.3.1 is that all findings are statistically significant, as illustrated by the rejection of the null hypothesis of the χ^2-based independence test in the last two columns of Table 4.3.1. Hence, it seems that all four confidence measures provide added value when aiming to get an idea of the level as well as the direction of change the reference series is likely to display in the current quarter.

Table 4.3.1: **Directional accuracy tests**

Panel A. - Levels

	Frequencies				Directional accuracy			χ^2 test	
	Correct		*Incorrect*						
	n_1	n_3	n_2	n_4	$\%^{all}$	$\%^{pos}$	$\%^{neg}$	stat	pval
cci	73	94	29	47	68.7%	76.4%	60.8%	33.1	0.000
cci_{PCR}	86	104	16	37	78.2%	86.7%	69.9%	77.5	0.000
cci_{PLS}	79	92	23	49	70.4%	80.0%	61.7%	41.6	0.000
cci_{RR}	77	90	25	51	68.7%	78.3%	60.2%	35.1	0.000

Panel B. - First differences

	Frequencies				Directional accuracy			χ^2 test	
	Correct		*Incorrect*						
	n_1	n_3	n_2	n_4	$\%^{all}$	$\%^{pos}$	$\%^{neg}$	stat	pval
cci	59	85	61	38	59.3%	58.2%	60.8%	7.7	0.006
cci_{PCR}	64	81	56	42	59.7%	59.1%	60.4%	8.3	0.004
cci_{PLS}	65	85	55	38	61.7%	60.7%	63.1%	12.5	0.000
cci_{RR}	59	81	61	42	57.6%	57.0%	58.4%	5.0	0.025

Source: European Commission

5. DETECTING PRIVATE CONSUMPTION EXPANSIONS AND RECESSIONS IN A PSEUDO-REAL TIME CONTEXT

5.1. THE EMPIRICAL SETUP

The evidence hitherto discussed has documented a satisfactory ability of survey-based indicators in tracking real private consumption dynamics. Nonetheless, it is well-known that good in-sample results do not guarantee good out-of-sample properties. To better assess the usefulness of survey-based confidence measures in a forecasting environment, we run a pseudo real-time exercise, simulating the performance of the different indicators in forecasting expansions and contractions in private consumption. ([13]) The questions to be answered are twofold: (i) Do the alternative consumer confidence measures convey additional information to predict recessionary phases in real private consumption compared to the one embedded in the established cci? (ii) If this is the case, is the supplementary information complementary to the one contained in a set of relevant, timely released hard data series or broadly identical?

A common approach to predicting recessions is the use of a probit model (see Estrella and Hardouvelis, 1991, Estrella and Mishkin, 1996, 1998), which allows mapping a set of continuous explanatory variables (in our case, confidence indices) into a binary dependent variable, y_t. *Let* y_t^* be an unobserved dependent variable that determines the occurrence of the event in a way that $y_t = 1$ if $y_t^* > 0$ and $y_t = 0$ otherwise. Let $X_t = [1, x_{1t}, \dots, x_{kt}]'$ be a vector containing timely available, relevant predictors, as well as a constant. The following probit model is fitted to the data:

$$y_t^* = \beta' X_t + \varepsilon_t \tag{3}$$

where ε_t *is* distributed normally. The fitted values of the model represent the probability which the predictors attach to the occurrence of a recession. Mathematically, the probability is expressed by the cumulative normal distribution function Φ, that is: $Pr(y_t = 1 \mid X_t) = \Phi(\beta' X_t)$, where β is obtained by maximizing the log-likelihood function $lnL(\beta) = \sum_{t=1}^{T} y_t ln\Phi(\beta' X_t) + (1 - y_t) ln[1 - \Phi(\beta' X_t)]$, *with* T indicating the length of the estimation span.

Once the different models have been run and generated forecasts of the recession probabilities, the analysis proceeds to a comparison of their forecasting performance. The existing literature has proposed a number of evaluation measures for the probit model case, ranging from the scoring system of Moore and Shiskin (1967) to the pseudo R-squared of Estrella and Mishkin (1996, 1998) or the quadratic and log probability scores of Diebold and Rudebusch (1989). However, all of these metrics focus on model fit and not specifically on the model's ability to correctly determine the presence or absence of a given regime.

As pointed out by Liu and Moench (2014), a formal comparison of the ability of alternative probit specifications to predict the occurrence of recessions is quite problematic since the probability of a recession implied by the models is rarely exactly zero or one. Thus, a cut-off (e.g. 0.50) is usually adopted such that a predicted probability above the cut-off is classified, in our case, as a recession. Obviously, the choice of the cut-off can have a significant bearing on which model performs best.

A possible way out, applied in a number of contributions (Khandani et al., 2010; Jordà and Taylor, 2011, 2012), is to construct for every model a receiver operating characteristic (ROC) curve. The idea is to plot the rate of false positives (x-axis) against the rate of true positives (y-axis) for different cut-off values (from 0 to 1). The intuition is that a good model will always be above a virtual 45 degrees line, separating the x- and y-axis, since it will produce a higher true positive than false positive rate. In essence, the curve thus enables the researcher to (graphically) evaluate the categorization ability of the model over an entire spectrum of different cut-offs, instead of evaluating predictive power at only one (arbitrary) threshold. The visual inspection of the ROC curves can be formalised by integrating the area under the ROC curve (resulting in the $AUROC$) and using statistical tests to determine which of two competing models

([13]) Practically, the dependent variable is a dummy featuring the sequence of expansions (=0) and contractions (=1) of real private consumption. The expansions and recessions are obtained by means of the procedure devised by Harding and Pagan (2002) applied to the cycle extracted through the method by Christiano and Fitzgerald (2003), filtering out fluctuations shorter than 6 quarters and longer than 32 quarters.

produces a larger $AUROC$. Generally, the larger the $AUROC$, the better is the model. By the same token, a value above 0.50 indicates that the model works better than a random guess model.

Since our assessment of the confidence indicators' forecasting performance resorts to both a visual inspection of ROC curves, as well as a statistical comparison of the $AUROC$s, it is worth describing their construction in some more detail.

The ROC curves are constructed in four steps: *Step I:* Once the probabilities of recessions, given by the probit model, $y_t^f \in [0,1]$, have been computed, k evenly spaced cut-offs c_i, $i = 1, \dots, k$ over the range $[0,1]$ are determined. *Step II:* For each cut-off c_i, the model's prediction of the presence or absence of a recession, $\hat{y}_t$, is recorded by setting $\hat{y}_t = 1$ if $y_t^f \geq c_i$ and $\hat{y}_t = 0$ if $y_t^f < c_i$. *Step III:* The comparison between the true y_t and the predicted categorizations $\hat{y}_t$ allows computing the percentage of true positives $\%T^+ = \frac{1}{h_1}\sum_{h=1}^{H} I_h^{t+}$, where $I_h^{t+} = 1$ if $y_t = \hat{y}_t = 1$, 0 otherwise, and the percentage of false positives $\%F^+ = \frac{1}{h_0}\sum_{h=1}^{H} I_h^{f+}$, where $I_h^{f+} = 1$ if $y_t = 1$ and $\hat{y}_t = 0$, 0 otherwise. The number h_1 (h_0) reflects the number of times the true y_t is in a contraction (expansion) phase over the forecasting horizon of length $H = h_1 + h_0$. *Step IV*: The ROC curve is plotted by connecting the coordinates ($\%F_i^+, \%T_i^+$) across all thresholds c_i where $\%F^+$'s are on the x-axis and $\%T^+$'s are on the y-axis.

Based on the calculations of the ROC curves and in line with Jordà and Taylor (2011), the $AUROCs$ are given by:

$$AUROC = \frac{1}{h_0 h_1}\sum_{j=1}^{h_0}\sum_{k=1}^{h_1}[I(z_i > y_k) + 0.5 \times I(z_i > y_k)] \qquad (4)$$

where $I(.)$ is the indicator function, z's are the observations classified to be an expansionary period, while y's, h_0 and h_1 are defined above.

The statistical comparison of the $AUROC$s of competing models (e.g. model 1 and model 2) follows the approach of Hanley and McNeil (1983), who propose the following t-statistic:

$$t_{HM} = \frac{AUROC_1 - AUROC_2}{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{0.5}} \qquad (5)$$

where $AUROC_i$, $i = 1,2$, is the area under the ROC curve for the i-th model under investigation, while σ_i^2 *deno*te its variance:

$$\sigma_i^2 = \frac{1}{h_{0,i}h_{1,i}}[AUROC_i(1 - AUROC_i) + (h_{1,i} - 1)(\Omega - AUROC_i^2) + (h_{0,i} - 1)(\Psi - AUROC_i^2)]^{0.5}$$

with $\Omega \equiv \frac{AUROC_i}{(2 - AUROC_i)}$ and $\Psi \equiv \frac{2 \times AUROC_i^2}{(1 + AUROC_i)}$. The parameter ρ is the correlation between $AUROC_1$ *and* $AUROC_2$. *To obtain* ρ, we estimate the average of the (Kendall-τ rank) correlations for the expansionary observations (z's) and recessionary observations (y's), respectively, across the two models.

Having presented the general set-up of the forecasting simulation, a few pieces of information still have to be provided to ensure a full understanding of the exercise conducted: First of all, our pseudo real-time simulation replicates, for every forecast quarter, the historic data-availability conditions assuming that the forecast is produced at the end of month 3 of the quarter. At that point in time, each of the confidence indicators features three monthly readings for the quarter to be forecast (remember A-m1, A-m2, A-m3, as presented in Section 3.2.), which are averaged across quarters to produce quarterly predictor variables. By contrast, the latest reading of the variable to be forecast refers to the preceding quarter. In total, for every model, 40 forecasts are conducted, corresponding to one forecast per quarter over the out-of-sample period 2005q2 to 2015q1. The in-sample window is a rolling one containing 36 quarterly observations.

Turning to the different specifications of the forecasting model which are tested, there are basically two types: In a first specification (Model A), it is assumed that the set of predictor variables (X_t in condition (3)) contains only a constant term and a survey-based measure of consumer confidence cof (corresponding, alternatively, to cci, cci_{PCR}, cci_{PLS} or cci_{RR}). A comparison of the ROC curves and the $AUROC$s enables us to answer the question whether the alternative consumer confidence measures convey additional information to predict recessionary phases in real private consumption compared to the one embedded in the established cci. A second probit specification (Model B) is an extension of Model A, where the set of predictors X_t is augmented so as to exploit information from timely available hard-data series which have been proven to be relevant in forecasting private consumption. Following Dées and Soares Brinca (2013), among others, the additional predictors consist of the short-term interest rate (3-month euribor, str), the (European) stock market index (Euro Stoxx 50, stk), as well as the (euro-area harmonised) index of consumer prices (cpi). All of them are expressed in quarterly averages, whereby the former remains in levels, while the latter two variables are log-transformed before computing the first difference of quarterly y-o-y changes. ([14]) The ROC curves and $AUROC$s allow testing the extent to which the forecast-relevant information contained in the confidence indicators is complementary to the one contained in the hard data.

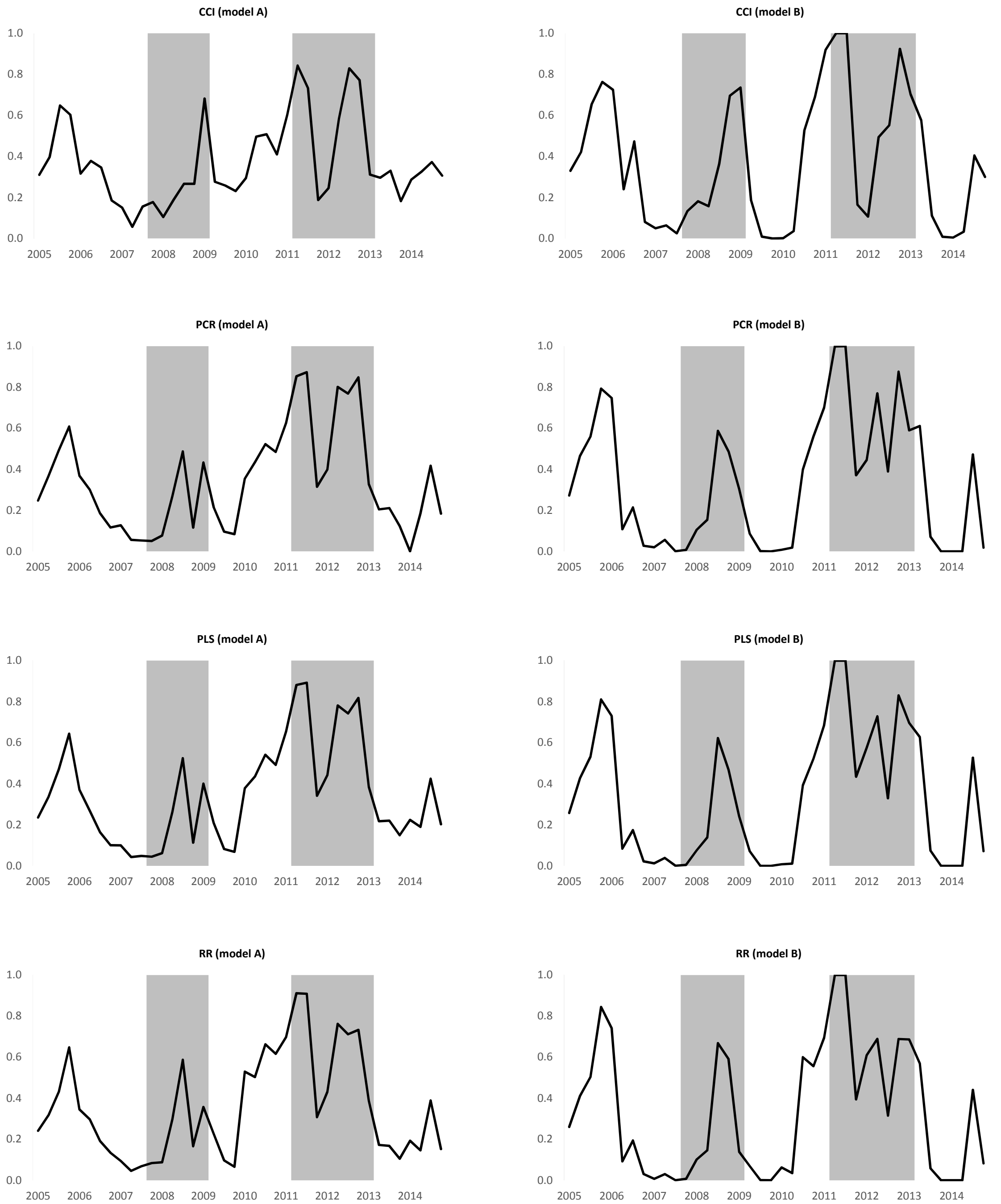
([14]) The publication calendar for the chosen hard-data series is such that we have full information for both str and stk at the end of a calendar quarter. The cpi variable is an exception. Due to its delayed publication, only the readings of the first two months of a given quarter are available by the end of that quarter. Accordingly, in our forecasting exercise, the average of the first two months (rather than all three months) of a given quarter are taken into account when constructing the quarter-on-quarter differences of cpi.

Figure 5.2.1 displays the predicted recession probabilities (solid lines) alongside the actual recession phases (shaded areas). There are two graphs per confidence indicator, with the left one displaying the results of Model A (only the respective confidence indicator and a constant are used) and the right one reporting the outcomes of Model B (where the respective confidence indicator is combined with both a constant and hard-data).

A focus on the grey areas shows that the forecasts cover a period characterised by only two major contractions: the Great Recession of 2008q1 to 2009q2 and the subsequent debt crisis (2011q3-2013q2). By and large, the visual inspection of the forecasted probabilities shows the models to adequately predict the sequence of expansions and contractions in consumption growth, suggesting that confidence indicators are relevant in predicting periods of strong fluctuations in the economy. That result confirms the evidence reported in Garner (1991) and Howrey (2001), among others. Comparing the performances of Models A and B, the major advantage of the hard-data augmented models (irrespective of the confidence indicator used) is their clear identification of the 2008/09 Great Recession, which contrasts with the comparatively low recession probabilities (around 0.50), which the non-augmented models attach to the period 2008/09. The finding can be explained when considering that private households, whose confidence levels are the only predictors in the non-augmented models, have arguably been less affected by the 2008/2009 crisis than, a few years later, by the sovereign debt crisis, which forced states to rein in their spending and implement significant tax hikes.

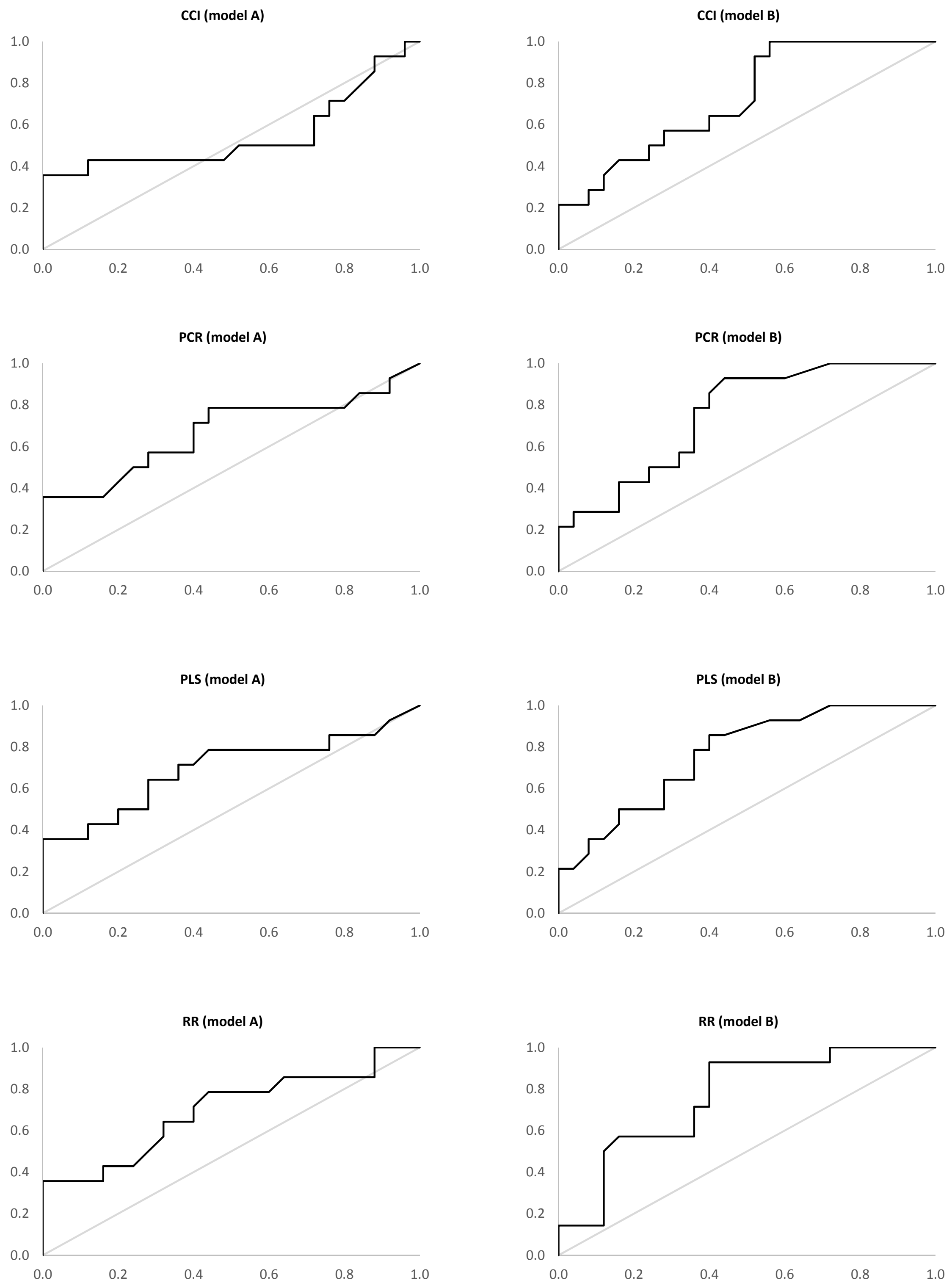
Figure 5.2.2 reports the ROC curves, again by type of Model (A or B) and the respective confidence indicator used. In general, the curve (black line) stands comfortably above the main diagonal (grey line). This suggests that, no matter which confidence indicator is used and irrespective of the cut-off applied for the categorisation of periods into recessions and expansions, all models perform better than the naïve benchmark (i.e. a random guess model).

Graph 5.2.1: **Probit models: out-of-sample contraction probabilities**



Predicted recession probabilities are reported as solid lines, while shaded areas indicate recession phases.
Source: European Commission

Graph 5.2.2: **Probit models: ROC curves**



Source: European Commission

In a next step, we formalise our observations by comparing the different models' AUROCs. The results (see Table 5.2.1) are displayed separately for forecasts based on confidence indicators only (Model A) or confidence indicators in combination with hard data (Model B), whereby each row of the table represents the use of a different confidence indicator.

Table 5.2.1: **Forecast accuracy**

	Model A	Model B
	Out-of-sample summary of models' AUROC	
cci	0.551	0.723
cci_{PCR}	0.666	0.743
cci_{PLS}	0.691	0.760
cci_{RR}	0.697	0.757
	Pair-wise differences wrt *cci*	
cci / cci_{PCR}	3.481 (0.000)	0.596 (0.276)
cci / cci_{PLS}	4.013 (0.000)	1.059 (0.145)
cci / cci_{RR}	3.852 (0.000)	0.869 (0.192)

p-values in parentheses.
Source: European Commission

Our first observation, related to the simple Model A, is that, no matter which of the four confidence indicators is used, all of them help producing forecasts which are better than random guesses (see values above 0.5 in the column labelled "Model A"). The established cci appears to carry the least forecast-relevant information, which is confirmed in the last three rows of Table 5.2.1, where the low p-values indicate that the difference between each of the individual new confidence indicators and the cci is statistically significant. Turning to the results for Model B, the addition of macro-economic variables improves the forecasts, with the corresponding AUROCs generally scoring above 0.7. At the same time, the differences between the models relying on the new indicators and the cci get statistically insignificant (see the last three rows of the second column). This suggests that the information advantage that the new indicators have over the cci is already largely covered by the information contained in the macro-economic variables.

Taken together, the analysis of the AUROCs suggests that indicators resorting to the entire set of consumer survey questions (rather than relying on just four of them) squeeze (a bit) more forecast-relevant information out of the survey data. However, this advantage only seems to be relevant in the (not very relevant) scenario where the forecast is exclusively based on survey data. In this respect, our findings are consistent with the conclusions in Gelper and Croux (2010) who show that statistically-based confidence measures hardly outperform ad-hoc indicators when forecasting the target series (EU industrial production in their case).

A possible explanation why the new indicators do not seem to produce superior forecasts in a realistic forecast scenario (i.e. including hard data), might lie in the number of survey indicators used for their construction. As documented by Boivin and Ng (2006), when too many series conveying a small amount of relevant information in explaining the target variable enter the set of predictors, the statistical efficiency can significantly deteriorate. To remedy this potential shortcoming of our approach, we reconstruct the alternative confidence indicators on the basis of a more limited amount of input variables. To avoid an ad-hoc selection of survey questions feeding into the indicator (as practised in the context of the cci construction), we choose questions based on objectively verifiable characteristics which we assume have a bearing on the degree to which they capture forecast-relevant information complementary to that contained in available hard data. The first criterion we apply is the time period to which the survey questions refer. Questions inquiring consumers' expectations (for the next 12 months) arguably measure a dimension which is, if at all, only partially reflected in the macro-economic series included in our model. At the same time, expectation questions can be assumed to be particularly beneficial for the purpose of forecasting. Our assumption is thus that the inclusion of confidence indicators extracted only from forward-looking questions will yield models performing better than the previous ones where confidence indicators included were derived from questions about both the future, as well as the present and recent past. In particular, we reckon that the forecast-enhancing effect will persist even if the macro-economic control variables are included in the model.

The second criterion we apply is whether the survey questions inquire household-specific (micro) questions, such as households' financial situation, their investment plans, etc., or focus on general economic conditions (unemployment levels, etc.). While we do not have an a priori assumption as to which of the two question types will produce more forecast-relevant confidence indicators, we consider the indicators derived from micro-questions as more likely to remain relevant in the presence of hard-data. After all, the micro dimension can be assumed to be largely absent from the available hard data and thus offer a higher degree of complementarity with the latter.

Table 5.3.1 summarises the results. Against a commonly held view, there does not seem to be any difference between current/backward-questions and those with a forward-looking nature as regards the complementarity of their forecast-relevant information with that contained in the hard data: The alternative confidence indicators do produce forecasts superior to a naïve benchmark (see values larger 0.5 in the upper left parts of Panel A. and Panel B.) and statistically better than a model relying on the established cci (as shown by the significant p-values in the lower left sections of Panels A. and B.). However, as soon as the three macro-economic hard data series are included (Model B), there is no statistically significant difference any more between the model based on the cci and the ones based on the alternative indicators (see the p-values in the lower right section of Panels A. and B.).

Turning to the difference between survey questions inquiring general economic and those focussing on household-specific concepts, our assumptions are proven right. The proposed confidence indicators, derived from questions about the general economic situation, do not produce better forecasts than the cci when included in models featuring the macro-economic control variables (see the p-values in the lower right part of Panel C.). By contrast, the combination of hard data and new indicators derived solely from the realm of household-specific questions is associated with a significantly better forecasting performance than a combination of hard data and the cci (as shown by the significant p-values in the lower right section of Panel D.). At the same time it is interesting to note that, when included in models featuring no hard data, indicators based solely on household-specific questions do not perform significantly better than the cci (see the p-values in the lower left part of Panel D.), unlike all previously considered alternative indicators. This suggests that while focusing on household-specific series does not lead to a better consumer confidence indicator as such, the specific information contained in these questions offers the highest degree of complementarity to the information in hard data series and thus helps improve forecasts of expansions and contractions in consumption.

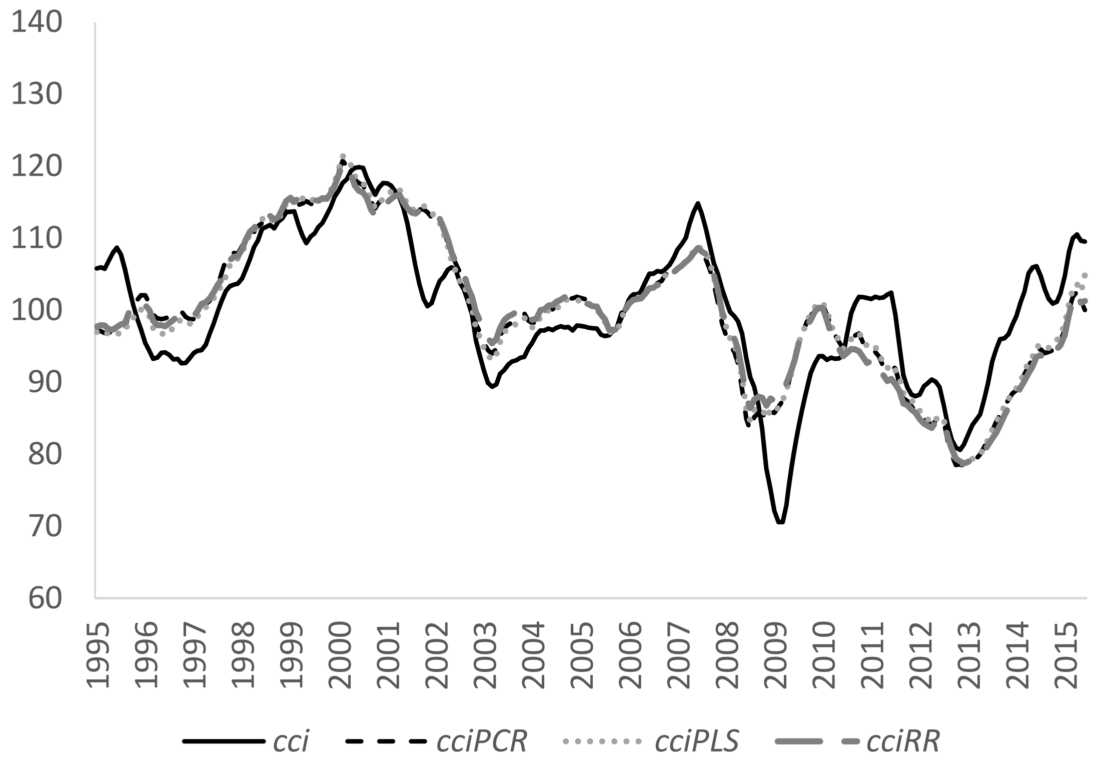
Table 5.3.1: **Forecast accuracy: subsets of questions from the consumer survey**

Panel A.	Current/backward-looking (bwd)	
	Model A	Model B
	Out-of-sample summary of models' AUROC	
cci	0.551	0.723
cci_{PCR}	0.731	0.754
cci_{PLS}	0.720	0.789
cci_{RR}	0.723	0.777
	Pair-wise differences wrt cci	
cci / cci_{PCR}	3.511 (0.000)	0.624 (0.266)
cci / cci_{PLS}	3.444 (0.000)	1.248 (0.106)
cci / cci_{RR}	3.212 (0.001)	1.013 (0.155)
Panel B.	**Forward-looking (fwd)**	
	Model A	Model B
	Out-of-sample summary of models' AUROC	
cci	0.551	0.723
cci_{PCR}	0.640	0.731
cci_{PLS}	0.657	0.729
cci_{RR}	0.660	0.743
	Pair-wise differences wrt cci	
cci / cci_{PCR}	2.972 (0.001)	0.362 (0.359)
cci / cci_{PLS}	3.382 (0.000)	0.220 (0.413)
cci / cci_{RR}	3.103 (0.001)	0.650 (0.258)
Panel C.	**General economy (mac)**	
	Model A	Model B
	Out-of-sample summary of models' AUROC	
cci	0.551	0.723
cci_{PCR}	0.671	0.734
cci_{PLS}	0.686	0.746
cci_{RR}	0.723	0.754
	Pair-wise differences wrt cci	
cci / cci_{PCR}	4.186 (0.000)	0.423 (0.336)
cci / cci_{PLS}	4.385 (0.000)	0.720 (0.236)
cci / cci_{RR}	4.037 (0.000)	0.789 (0.215)
Panel D.	**Household-specific (mic)**	
	Model A	Model B
	Out-of-sample summary of models' AUROC	
cci	0.551	0.723
cci_{PCR}	0.623	0.803
cci_{PLS}	0.623	0.809
cci_{RR}	0.639	0.789
	Pair-wise differences wrt cci	
cci / cci_{PCR}	0.958 (0.169)	1.803 (0.036)
cci / cci_{PLS}	0.945 (0.172)	1.787 (0.037)
cci / cci_{RR}	1.178 (0.119)	1.298 (0.097)

p-values in parentheses
Source: European Commission

To get a better understanding of the consumer indicators derived from household-specific questions, we plot them alongside the current established cci (see Figure 5.3.1). ([15]) Up to the financial crisis of 2008/09, both types of indicators seem to go broadly in lockstep. Subsequently, they clearly diverge from each other: While the cci reaches its lowest ever level at the peak of the financial crisis, in 2009, the alternative indicators show a more profound drop in the economic downturn of 2013. The observed pattern appears convincing when recalling the economic policies of the last years: 2009 saw significant increases in government spending in order to fend off the negative consequences of the crisis, contrasting with the subsequent sovereign debt crisis, which forced states to rein in their spending and implement significant tax hikes, which arguably had a more immediate effect on households' revenue position. Since the alternative confidence indicators rely solely on household-specific questions, it is reasonable that they show a larger response to the sovereign debt crisis.

Graph 5.3.1: **Established cci and its three alternatives based on household-specific questions (monthly values)**



Source: European Commission

([15]) Both indicators have been rescaled as detailed in Section 4.1 to ease their comparability.

6. CONCLUDING REMARKS

This work has provided a comparative assessment of the European Commission (EC) Consumer Confidence Indicator (cci) for the euro area against three alternative consumer confidence indicators which differ from the former in that they take into account a richer data set (10 countries instead of one euro-area aggregate and 11 instead of 4 different consumer survey series), and have been built by means of formal, data-driven statistical techniques (principal component regression, partial least squares and ridge regression methods), rather than an ad-hoc aggregation approach.

The evaluation has been carried out along several dimensions spanning from the weight structure of the alternative aggregation schemes over a comparison of the indicators' evolution over time to the degree of directional accuracy in tracking real private consumption growth and their ability to forecast the occurrence of recessions in consumption. Overall, we find that, despite its simple and ad-hoc aggregation scheme, its limited data input (four series only) and the fact that it is not tailored to its target series by design, the EC's cci performs similarly to the proposed alternative confidence measures. Only slight improvements in tracking private consumption dynamics can be achieved when applying the more complicated construction techniques of the alternative indicators. By the same token, when focussing on the forecasting abilities of the measures, the alternative indicators fare only slightly better than the cci. As soon as timely available hard data are included in the forecasting equation, any competitive edge of the alternative indicators fades. The conclusions change when re-constructing the new indicators solely on the basis of household-specific (micro) questions (households' financial situation, their investment plans, etc.), rather than combining them with questions about general economic conditions (unemployment levels, etc.). The modified indicators are shown to provide the highest degree of complementarity to the information in timely hard data series, thus facilitating improvements in forecasting recessions in private consumption.

On a more abstract level, the findings in this paper suggest two conclusions: (i) The wide-spread approach of constructing consumer confidence indicators as the simple average of a careful selection of (a few) consumer survey series appears to be an appropriate technique. After all, the improvements generated by more complex construction methods must be weighed against their costs: statistically-based aggregation techniques are significantly more difficult to communicate to end-users; moreover, up- or downswings in the indicator cannot be clearly attributed to developments in individual underlying survey questions. (ii) A second finding of the present work is that researchers interested in forecasting private consumption dynamics stand to gain potential windfalls from further exploring the added value of survey questions about micro-economic concepts like households' financial situation, saving and purchasing intentions.

The results presented in this work are particularly promising when considering that the analysis does not include any variables which are highly country-specific, so that it can easily be adapted to the case of other economies. Given the status of the EU BCS programme as international best practice, the same or similar survey questions can also be found in a number of extra-EU survey programmes, rendering the extension to other countries/regions straightforward. In this respect, exploiting the forecast-enhancing information conveyed by household-specific survey questions should be particularly valuable to study not only the predictive content of survey-based confidence measures to forecast macroeconomic aggregates but also the relevance of (consumer) confidence in explaining the international transmission of shocks across economies.

REFERENCES

Acemoglu, D, Scott, A. (1994). Consumer Confidence and Rational Expectations: Are Agents Beliefs Consistent With the Theory? The Economic Journal 104: 1.19.

Akerlof G. A. and Shiller, R. J. (2009). Animal Spirits. How human psychology drives the economy and why it matters for global capitalism. Princeton: Princeton University Press, 2009.

Al-Eyd, A., Barrell, R. and Philip, E. (2008). Consumer Confidence Indices and Short-term Forecasting of Consumption. Manchester School 77: 96-111.

Angeletos, G.-M. and La'O, J. (2013). Sentiments. Econometrica 81: 739-779.

Barsky, R.B. and Sims, E.R. (2011). News Shocks and Business Cycles. Journal of Monetary Economics 58: 273-289.

Bec, F. and Mogliani, M. (2015). Nowcasting French GDP in Real-Time with Survey and “Blocked” Regressions: Combining Forecasts or pooling information? International Journal of Forecasting 31: 1021-1042.

Boivin, J. and Ng, S. (2006). Are More Data Always Better for Factor Analysis? Journal of Econometrics 132: 169-194.

Carnot, N., Koen, V. and Tissot, B. (2005). Economic Forecasting, Palgrave MacMillan, p. 240.

Carriero, A., Clark, T. and Marcellino, M. (2012). Real-time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. Federal Reserve Bank of Cleveland Working Paper, No. 1227.

Chen, W., Anderson, B.D.O., Deistler, M. and Filler, A. (2012). Properties of Blocked Linear Systems. Automatica 48: 2520-2525.

Christiano, L.J. and Fitzgerald, T.J. (2003). The Band Pass Filter. International Economic Review 44: 435-465.

Chun, H. and Keles, S. (2010). Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection. Journal of the Royal Statistical Society, Series B (Statistical Methodology) 72: 3-25.

Cubadda, G. and Guardabascio, B. (2012). A Medium-N Approach to Macroeconomic Forecasting. Economic Modeling 29: 1099-1105.

Dées, S., Soares Brinca, P. (2013). Consumer Confidence as a Predictor of Consumption Spending: Evidence for the United States and the Euro Area. International Economics 134: 1-14.

Diebold, F. and Rudebusch, G. (1989). Scoring the Leading Indicators. Journal of Business and Economic Statistics 62: 369-391.

ECB (2013). Monthly Bulletin, January: 45-58.

Elliott G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit Root. Econometrica, 64: 813-836.

Eppright, D.W., Argues, N.M. and Huth, W.L. (1998). Aggregate Consumer Expectation Indexes as Indicators of Future Consumer Expenditures. Journal of Economic Psychology 19: 215-235.

Estrella, A. and Hardouvelis, G.A. (1991). The Term Structure as a Predictor of Real Economic Activity. The Journal of Finance 46: 555-576.

Estrella, A. and Mishkin, F.S. (1996). The Yield Curve as a Predictor of U.S. Recessions. Current Issues in Economics and Finance 2: 1-6.

Estrella, A. and Mishkin, F.S. (1998). Predicting U.S. Recessions: Financial Variables as Leading Indicators. Review of Economic and Statistics 80: 45-61.

Fagan, G., Henry, J. and Mestre, R. (2001). An Area-wide Model (AWM) for the Euro Area, ECB Working Paper Series, No. 42.

Fagan, G., Henry, J. and Mestre, R. (2005). An Area-wide Model for the Euro Area, Economic Modelling 22: 39-59.

Fei, S. (2011). The Confidence Channel for the Transmission of Shocks. Banque de France Working Paper, No. 314.

Frank, I.E. and Friedman, J.H. (1993). A Statistical View of Some Chemometrics Regression Tools (with Discussion). Technometrics 35: 109-148.

Garner, C.A. (1991). Forecasting Consumer Spending: Should Economists Pay Attention to Consumer Confidence Surveys? Economic Review: 57-71.

Gelper, S. and Croux, C. (2010). On the Construction of the European Economic Sentiment Indicator. Oxford Bulletin of Economics and Statistics 72: 47–62.

Girardi, A., Guardabascio, B. and Ventura, M. (2016). Factor-Augmented Bridge Models (FABM) and Soft Indicators to Forecast Italian Industrial Production. Journal of Forecasting, in press.

Hall, R.E. (1978). Stochastic Implications of the Life-Cycle/Permanent Income Hypothesis: Theory and Evidence. Journal of Political Economy 96: 971-987.

Hanley, J.A., and McNeil, B.J. (1982). The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. Radiology 143: 29-36.

Hanley, J.A., and McNeil, B.J. (1983). A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. Radiology 148: 839-843.

Harding, D. and Pagan, A. (2002). Dissecting the Cycle: A Methodological Investigation. Journal of Monetary Economics 49: 365-381

Hoerl, A.E. and Kennard, R.W. (1970). Ridge-regression: Biased Estimation for Nonorthogonal Problems. Technometrics 8: 27-51.

Helland, I.S. (2001). Some Theoretical Aspects of Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems 58: 97-107.

Helland, I.S. (2006). Partial Least Squares Regression. In Kotz S., Read B., Balakrishnan N., Vidakovic B. (eds), Encyclopedia of Statistical Sciences, John Wiley & Sons: 5957–5962.

Howrey, E.P. (2001). The Predictive Power of the Index of Consumer Sentiment. Brookings Papers on Economic Activity 32: 175-207.

Johansen, S. (1995). Likelihood□Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press.

Jonsson, A., Lindén S. (2009). The Quest for the Best Consumer Confidence Indicator. European Economy Economic Papers, No. 372.

Jordà, O., Taylor, A.M. (2011). Performance Evaluation of Zero Net-investment Strategies. NBER Working Papers.

Jordà, O., Taylor, A.M. (2012). The Carry Trade and Fundamentals: Nothing to fear but FEER Itself. Journal of International Economics 88: 74-90.

Keynes, J. M. (1936). The general theory of employment, interest, and money. London: Macmillan, 1936, 161-62.

Khandani, A.E., Kim, A.J. and Lo, A.W. (2010). Consumer Credit Risk Models via Machine-learning Algorithms. Journal of Banking and Finance 34: 2767-2787.

Kraemer, N. and Sugiyama, M. (2011). The Degrees of Freedom of Partial Least Squares Regression. Journal of the American Statistical Association 106: 697-705.

Kwiatkowski D., Phillips, P.C.B., Schmidt, P. and Shin Y. (1992). Testing the Null of Stationarity against the Alternative of a Unit Root: How Sure Are we that Economic Time Series Have a Unit Root? Journal of Econometrics 54: 159-178.

Liu, W. and Moench, E. (2014). What Predicts U.S. Recessions?, Federal Reserve Bank of New York Staff Reports, No. 691.

Ludvigson, S.C. (2004) Consumer confidence and consumer spending. Journal of Economic Perspectives 18:29-50.

Moore, G.H. and Shiskin, J. (1967). Indicators of Business Expansions and Contractions, NBER Books. National Bureau of Economic Research.

Slacalek, J. (2005). Analysis of Indexes of Consumer Sentiment, mimeo, German Insititute for Economic Research.

Stone, M. and Brooks, R.J. (1990). Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares and Principal Components Regression. Journal of the Royal Statistical Society, Series B (Statistical Methodological) 52: 237-269.

Tibshirani R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J. and Tibshirani, R.J. (2010). Strong Rules for Discarding Predictors in Lasso-type Problems. Journal of the Royal Statistical Society, Series B (Statistical Methodology) 74: 245-266.

Wold, H. (2006). Partial Least Squares. In Kotz S., Read B., Balakrishnan N., Vidakovic B. (eds), Encyclopedia of Statistical Sciences, John Wiley & Sons: 5948–5957.

ANNEX 1

The EU's consumer questionnaire

Q1: How has the financial situation of your household changed over the last 12 months?

It has: (++) got a lot better; (+) got a little better; (=) stayed the same; (−) got a little worse; (−−) got a lot worse; (N) don't know.

Q2: How do you expect the financial position of your household to change over the next 12 months?

It will: (++) get a lot better; (+) get a little better; (=) stay the same; (−) get a little worse; (−−) get a lot worse; (N) don't know.

Q3: How do you think the general economic situation in the country has changed over the past 12 months?

It has: (++) got a lot better; (+) got a little better; (=) stayed the same; (−) got a little worse; (−−) got a lot worse; (N) don't know.

Q4: How do you expect the general economic situation in this country to develop over the next 12 months?

It will: (++) get a lot better; (+) get a little better; (=) stay the same; (−) get a little worse; (−−) get a lot worse; (N) don't know.

Q5: How do you think that consumer prices have developed over the last 12 months?

They have: (++) risen a lot; (+) risen moderately; (=) risen slightly; (−) stayed about the same; (−−) fallen; (N) don't know.

Q6: By comparison with the past 12 months, how do you expect that consumer prices will develop in the next 12 months?

They will: (++) increase more rapidly; (+) increase at the same rate; (=) increase at a slower rate; (−) stay about the same; (−−) fall; (N) don't know.

Q7: How do you expect the number of people unemployed in this country to change over the next 12 months?

The number will: (++) increase sharply; (+) increase slightly; (=) remain the same; (−) fall slightly; (−−) fall sharply; (N) don't know.

Q8: In view of the general economic situation, do you think that now it is the right moment for people to make major purchases such as furniture, electrical/electronic devices, etc.?

(+ +) yes, it is the right moment now; (=) it is neither the right moment nor the wrong moment; (−−) no, it is not the right moment now; (N) don't know.

Q9: Compared to the past 12 months, do you expect to spend more or less money on major purchases (furniture, electrical/electronic devices, etc.) over the next 12 months?

I will spend: (++) much more; (+) a little more; (=) about the same; (−) a little less; (−−) much less; (N) don't know.

Q10:In view of the general economic situation, do you think that now is...?

(++) a very good moment to save; (+) a fairly good moment to save; (−) not a good moment to save; (−−) a very bad moment to save; (N) don't know.

Q11: Over the next 12 months, how likely is it that you save any money?

(++) very likely; (+) fairly likely; (−) not likely; (−−) not at all likely; (N) don't know.

Q12: Which of these statements best describes the current financial situation of your household?

(++) we are saving a lot; (+) we are saving a little; (=) we are just managing to make ends meet on our income; (−) we are having to draw on our savings; (−−) we are running into debt; (N) don't know.

ANNEX 2

PCR, PLS and RR algorithms

This Appendix describes the basic features of the three algorithms that have been used to construct our alternative indicators of consumer confidence. For the sake of simplicity, let us assume that our model of reference can be represented in the form of a multiple linear regression:

$$Y = XB + \varepsilon \tag{A.1}$$

which admits the following least-squares solution:

$$B = (X'X)^{-1}X'Y \tag{A.2}$$

In our context, where the number of variables (columns) in $\boldsymbol{X}$ exceeds the number of observations (rows), we have that $X'X$ is singular. Both PCR and PLS circumvent this problem by decomposing $\boldsymbol{X}$ into orthogonal scores G and loadings P:

$$X = GP \tag{A.3}$$

and regressing Y not on X itself but on the first k columns of the scores G.

In PCR, the X matrix is approximated by the first k principal components (PC) that have been obtained from the singular value decomposition (SVD):

$$X = \tilde{X}_{(k)} + \varepsilon_X = \left(U_{(k)}D_{(k)}\right)V'_{(k)} + \varepsilon_X = G_{(k)}P'_{(k)} + \varepsilon_X$$

Regressing Y on the scores leads to regression coefficients

$$B_{PCR} = P_{(k)}(G'_{(k)}G_{(k)})^{-1}G'_{(k)}Y \tag{A.4}$$

As for PLS, the components are obtained iteratively. One starts with the SVD of the cross-product matrix $= X'Y$, thereby including information on variation in both X and Y, and on the correlation between them. The first left and right singular vectors, w and q, are used as weight vectors for X and Y, respectively, to obtain scores $t = Xw = Ew$ and $u = Yq = Fq$ where E and F are initialised as $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Next, $\boldsymbol{X}$ and $\boldsymbol{Y}$ loadings are obtained by regressing against the same (normalised) vector $\tilde{t} = t/\sqrt{t't}$, *so that* $p = E'\tilde{t}$ *and* $q = F'\tilde{t}$. *Finally, the data matrices are "deflated" by subtracting the information related to the outer products* tp' *and* q': $E_{+1} = E - tp'$ *and* $F_{+1} = F - tq'$, *respectively. The estimation of the next component then can start from the SVD of the* cross-product matrix $E'_{+1}F_{+1}$. *Af*ter k iteration, vectors w, t, p and q are saved as columns in matrices $W_{(k)}$, $T_{(k)}$, $P_{(k)}$ *and* $Q_{(k)}$, *respect*ively. A convenient way to relate weights to the original X matrix is given by the following condition:

$$R_{(k)} = W_{(k)}\left(P'_{(k)}W_{(k)}\right)^{-1} \tag{A.5}$$

since the scores $T_{(k)} = XR_{(k)}$ *can be use*d to calculate the regression coefficients, and later convert these back to the domain of the original variables by pre-multiplying with matrix R:

$$B_{PLS} = R_{(k)}(T'_{(k)}T_{(k)})^{-1}Y \tag{A.6}$$

where the optimal k has to be determined, usually by cross-validation. As pointed out in Section 3, in our context, however, k is set equal to 1, as in Gelper and Croux (2010). A detailed discussion of the PLS method can be found in Wold (2006) and Helland (2006).

Regarding RR (Hoerl and Kennard, 1970), its basic idea is adding a constant ϵ to the diagonal elements of $X'X$:

$$B_{RR} = (X'X + \epsilon I)^{-1}X'Y \tag{A.7}$$

where can be also viewed as a Lagrange multiplier, which amounts to the minimization of $\|XB - Y\|_2^2 + \epsilon\|B\|_2^2$.

A convenient way to perform the required computations is to apply an orthogonal transformation of condition (A.1) by calculating $h = U'Y$ where, U is extracted from the SVD $X = UDV'$. Letting $X'B = \alpha$ the model (A.1) becomes

$$h = D\alpha \tag{A.8}$$

with the estimator of α, being $\hat{\alpha} = (D'D + \epsilon I)^{-1}D'h$ which then is converted back using

$$B_{RR} = V\hat{\alpha} \tag{A.9}$$

EUROPEAN ECONOMY DISCUSSION PAPERS

European Economy Discussion Papers can be accessed and downloaded free of charge from the following address:
http://ec.europa.eu/economy_finance/publications/eedp/index_en.htm

Titles published before July 2015 under the Economic Papers series can be accessed and downloaded free of charge from:
http://ec.europa.eu/economy_finance/publications/economic_paper/index_en.htm

Alternatively, hard copies may be ordered via the “Print-on-demand” service offered by the EU Bookshop: http://bookshop.europa.eu.

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);

- more than one copy or posters/maps:
  - from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  - from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  - by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

KC-BD-16-034-EN-N

Publications Office

ISBN 978-92-79-54440-8