

Schenker, Rolf

**Working Paper**

## Comparing quantitative and qualitative survey data

KOF Working Papers, No. 169

**Provided in Cooperation with:**

KOF Swiss Economic Institute, ETH Zurich

*Suggested Citation:* Schenker, Rolf (2007) : Comparing quantitative and qualitative survey data, KOF Working Papers, No. 169, ETH Zurich, KOF Swiss Economic Institute, Zurich

This Version is available at:

<https://hdl.handle.net/10419/50382>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

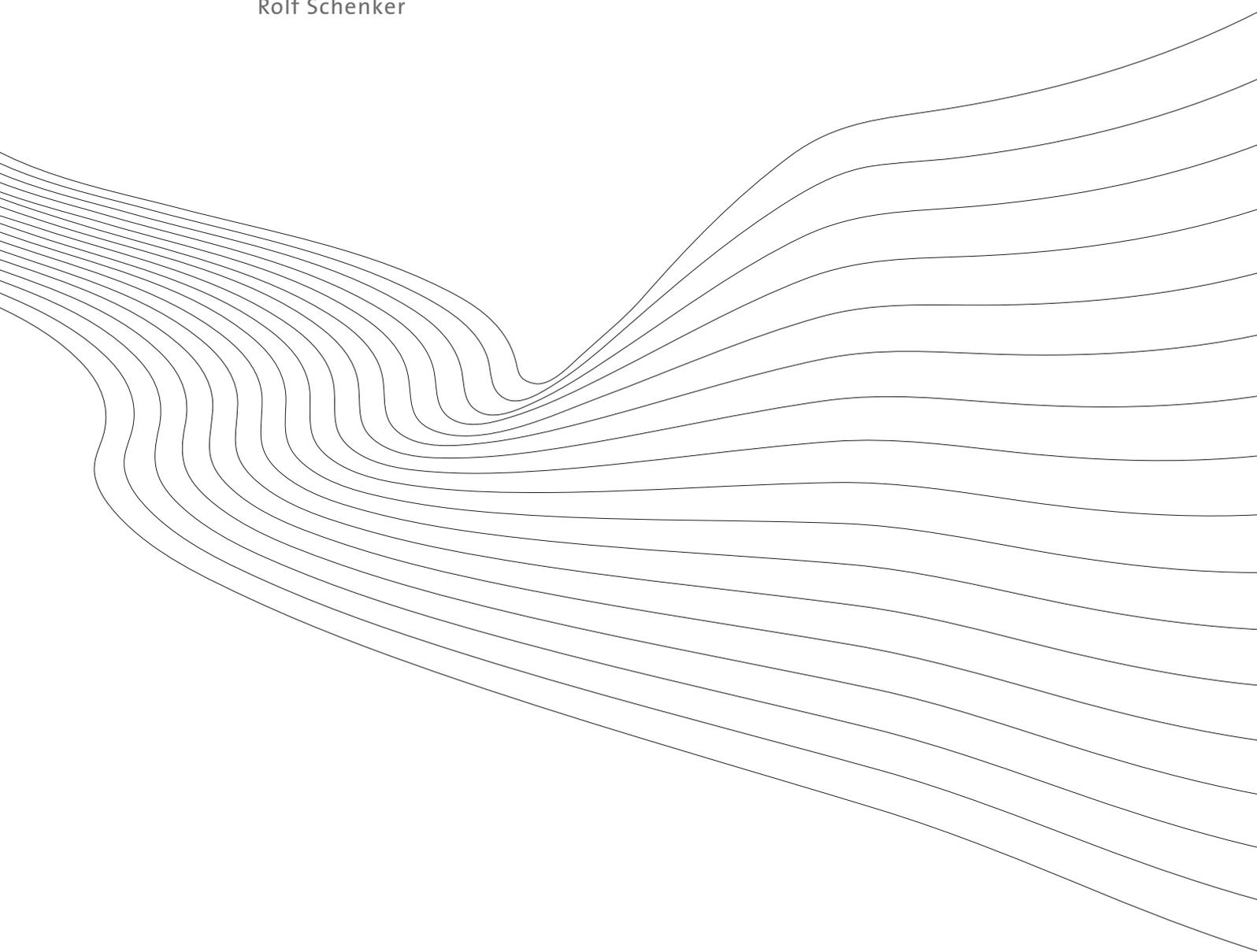
*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# KOF Working Papers

Comparing Quantitative and Qualitative Survey Data

Rolf Schenker



# KOF

ETH Zurich  
KOF Swiss Economic Institute  
WEH D 4  
Weinbergstrasse 35  
8092 Zurich  
Switzerland

Phone +41 44 632 42 39  
Fax +41 44 632 12 18  
[www.kof.ethz.ch](http://www.kof.ethz.ch)  
[kof@kof.ethz.ch](mailto:kof@kof.ethz.ch)

# Comparing qualitative and quantitative survey data

Rolf Schenker\*

June 28, 2007

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The data</b>	<b>2</b>
<b>3</b>	<b>Procedure</b>	<b>4</b>
<b>4</b>	<b>Results</b>	<b>4</b>
4.1	Contingency Tables . . . . .	4
4.2	Distribution and Density Functions . . . . .	5
4.3	Response Functions . . . . .	8
<b>5</b>	<b>Conclusions</b>	<b>9</b>

## Abstract

This paper compares quantitative and qualitative data on firm level. The data is taken from two Swiss investment surveys. This has not yet been done in the literature.

We will see that the mean change in investment of firms planning to increase (decrease) investments is positive (negative). In contrast, the mean change in investment of firms indicating “no change” is indeed virtually zero. Carlson & Parkin (1975) assume the quantitative observations to follow a normal distribution. Other research (e.g. Dasgupta & Lahiri 1992) has been done assuming other distributions. In this paper we show that the micro data does not follow a normal, logistic or exponential distribution.

Furthermore, we adopt the response functions presented by Ronning (1984) to the investment data. They help us to determine the share of firms giving the different qualitative statement for every instance of the quantitative data. We will show that with larger (smaller) quantitative changes, more firms give positive (negative) qualitative statements.

*Keywords:* Response Functions, Investment survey, Qualitative response, Contingency Table

*JEL classification:* C5, E22, C42

---

\*ETH Zurich, KOF Swiss Economic Institute, 8092 Zürich

# 1 Introduction

The basic idea of business surveys is to gather from the development of individual variables (micro variables) to the development of economy-wide variables (macro variables, i.e. from the investment of interviewed firms to the investment in the whole economy). When surveys pose qualitative questions, it is assumed that firms have quantitative assessments of a variable and transform them to qualitative statements. Most qualitative questions propose three answers (e.g. “increase”, “no change”, “decrease”). To be included in econometric models, these qualitative answers have to be quantified in some way. All quantification methods investigate the share of interviewees giving each of the three answers. We will call these shares  $A_t$  (for “increase”),  $B_t$  (for “no change”), and  $C_t$  (for “decrease”).

A widely used quantification method is the one presented by Carlson & Parkin (1975). Assuming the quantitative (change of the) micro variables to follow a normal distribution, they calculate the mean ( $\hat{\mu}_t$ ) and standard-deviation ( $\hat{\sigma}_t$ ) of these distributions from the response shares. Other research (e.g. Dasgupta & Lahiri 1992) has been done assuming normal,  $t$ - or logistic distributions. Dasgupta & Lahiri (1992) showed that the assumed distribution does not have much influence on the estimation results. Section 4.2 gives an explanation, why.

This paper presents Swiss investment data, taken from a quantitative and a qualitative survey. We will be able to compare the data on firm level. Like that it will be possible to test the above assumptions about the distribution of the micro data. This has not yet been done in literature.

The paper is organized as follows: section 2 describes the data, and section 3 highlights the procedure that was adapted to the quantitative data before comparing it to the qualitative data. Section 4 is dedicated to the presentation of the estimation results: It begins with contingency tables (section 4.1), giving a first impression of the relation between the qualitative and the quantitative answers. Section 4.2 presents the distribution and density functions of the qualitative answers. In section 4.3 the answer patterns for the qualitative questions subject to the quantitative changes will be analyzed by reanimating the response functions introduced by Ronning (1984).

## 2 The data

KOF conducts two annual investment surveys. The quantitative survey (IVN) is made in collaboration with the Swiss federal statistical office. It is conducted in spring with about 12'000 firms. The qualitative survey (IVL) is conducted in fall. Its sample comprises 6800 firms. Both samples cover firms from the manufacturing, the construction and the service sector. They partly overlap one another, allowing us to compare the answers on micro (firm) level. This has not yet been done in the literature.

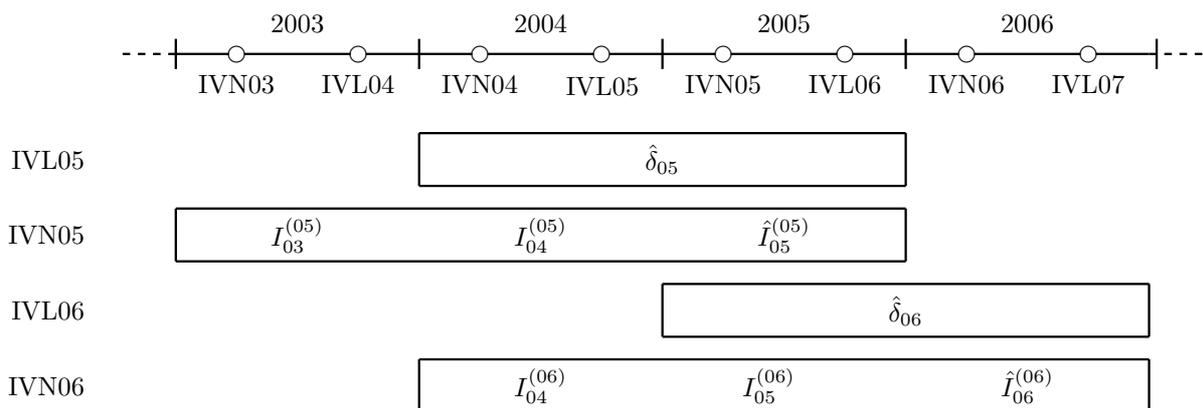


Figure 1: Time pattern

In the quantitative survey (IVN), firms are asked about their investments as monetary amounts. The survey covers investment in the year of interview and the two preceding years. Additionally, firms are asked to indicate the number of employees. In the qualitative survey (IVL), firms are asked about the planned change in investment from the actual to the next year. Firms indicate whether they plan investments to increase, stay the same, or decrease. Additionally, firms are asked about their certainty in their own investment plans. The answers proposed are “very sure”, “rather sure”, “quite unsure”, and “not sure”.

Figure 1 may help to clarify the workflows: Every year KOF conducts two investment surveys: the quantitative survey in spring and the qualitative in fall. The qualitative survey 2005, shortly IVL05, is conducted in fall 2004. Firms are asked about the planned change in their investment sums from 2004 to 2005. The quantitative survey 2005, IVN05, is conducted in spring 2005. There, firms indicate their investment sums in 2003, 2004, and 2005. We use the following general notation: Data collected by IVN in year  $t$  are  $I_{t-2}^{(t)}$ ,  $I_{t-1}^{(t)}$ , and  $\hat{I}_t^{(t)}$  (e.g.  $I_{03}^{(05)}$ ,  $I_{04}^{(05)}$ , and  $\hat{I}_{05}^{(05)}$ ), while data collected by IVL of year  $t$  (conducted in fall of year  $t - 1$ ) are called  $\hat{\delta}_t$  (e.g.  $\hat{\delta}_{05}$ ).<sup>1</sup>

Using data from IVL05 and IVN05, we can compare qualitative and quantitative investment plans. In October-November 2004, when answering IVL05, firms have data about their business in three quarters of 2004. Based on this information, they make their investment plans for 2005. The IVN questionnaire is completed half a year later. This means that firms have business data until the first quarter 2005. When data from IVL and IVN diverge, this does not necessarily mean that firms themselves are inconsistent. It may also happen that new information influences the investment plans of the firms.

In both surveys firms are asked about their investment plans in equipment and software, in constructions and total fixed investment. Each firm that completes all relevant parts of the questionnaires thus generates three observations. For the following analyses we only look at the data for total fixed investment and merge the observations of all years.

<sup>1</sup>Subscripts indicate the time of investment, while superscripts indicate the survey which provides the information.  $\hat{I}_t^{(t)}$  and  $\hat{\delta}_t$  are plans rather than realizations.

### 3 Procedure

As we have seen in section 2, the qualitative survey asks for the changes in investment, while the quantitative survey asks for investment sums themselves. To compare qualitative and quantitative investment statements, the quantitative data has to be transformed into a measure for the “changes in investment”. Whether a certain change in investment is considered as a “constant investment” presumably depends on the average investment sums of a firm. That’s why the absolute change in investment ( $\hat{I}_t^{(t)} - I_{t-1}^{(t)}$ ) is not usable. Furthermore, the measure should be able to distinguish between a slight sliding and an abrupt drop-out of investment. The relative change in investment  $\frac{\hat{I}_t^{(t)} - I_{t-1}^{(t)}}{I_{t-1}^{(t)}}$  is not able to distinguish these variations. If  $\hat{I}_t^{(t)} - I_{t-1}^{(t)}$  is divided by the average investment sum, even this requirement is fulfilled. In what follows, the quantitative change in investment is defined as

$$\hat{\psi}_t \equiv \frac{\hat{I}_t^{(t)} - I_{t-1}^{(t)}}{I_\mu^{(t)}}$$

with  $I_\mu^{(t)} \equiv \frac{\hat{I}_t^{(t)} + I_{t-1}^{(t)} + I_{t-2}^{(t)}}{3}$ . With this definition, positive (negative) values of  $\hat{\psi}_t$  indicate increasing (decreasing) investments in  $t$  compared to  $t - 1$ . The larger the absolute value of  $\hat{\psi}_t$ , the larger is the change in investment relative to the mean investment of the respective firm.  $\hat{\psi}_t = 2$  means that the change in investment from  $t - 1$  to  $t$  amounts twice the mean investment in  $t - 2$ ,  $t - 1$  and  $t$ . If for a certain firm  $I_{t-2}^{(t)} = I_{t-1}^{(t)} = 0$  and  $\hat{I}_t^{(t)} \neq 0$ , then  $|\hat{\psi}_t| = 3$ .<sup>2</sup>

## 4 Results

### 4.1 Contingency Tables

To get a general idea of the results, one should have a look at the contingency table (table 1). It shows the frequency of all combinations of qualitative answers  $\hat{\delta}_t$  and quantitative changes  $\hat{\psi}_t$ . The rows represent different values of  $\hat{\delta}_t$ . A  $-1$  stands for “decrease”, a  $0$  for “stay the same” and a  $1$  for “increase”. In the columns of table 1, different values of  $\hat{\psi}_t$  are shown: a  $-1$  stands for negative, a  $0$  for invariable, and a  $1$  for positive values. The upper left and lower right fields of the table represent consistent combinations of  $\hat{\delta}_t$  and  $\hat{\psi}_t$ , while the upper right and lower left fields stand for inconsistent combinations. The share of consistent answers (38%) is higher than the share of inconsistent answers (20%). Additionally, we can see that the mean of  $\hat{\psi}_t$  is positive (negative) for  $\hat{\delta}_t = 1$  ( $\hat{\delta}_t = -1$ ), i.e. the mean investment of firms with  $\hat{\delta}_t = 1$  ( $\hat{\delta}_t = -1$ ) is in fact larger (smaller) than in the preceding year. The differences are not significant, though. Another interesting result of the contingency table is that for firms with  $\hat{\delta}_t = 0$ , the mean of  $\hat{\psi}_t$  is actually virtually zero.<sup>3</sup>

<sup>2</sup>To avoid a contamination of the estimation results by extreme values, all data with  $|\hat{\psi}_t| > 2$  are omitted. With the guillaotine at  $|\hat{\psi}_t| > 2$ , 380 of the 4830 observations drop out.

<sup>3</sup>Virtually all of the 406 observations with  $\hat{\psi}_t = 0$  are firms with  $\hat{I}_t^{(t)} = I_{t-1}^{(t)} = 0$ .

Mean of $\widehat{\psi}_t$		sign( $\widehat{\psi}_t$ )			
Std. dev. of $\widehat{\psi}_t$		-1	0	1	all
observations					
-1		-0.730	0.000	0.459	-0.349
		0.538	0.000	0.437	0.723
		728	85	307	1120
0		-0.625	0.000	0.511	-0.070
		0.557	0.000	0.462	0.717
		749	204	691	1644
1		-0.648	0.000	0.679	0.216
		0.557	0.000	0.504	0.790
		529	117	1040	1686
all		-0.669	0.000	0.589	-0.032
		0.552	0.000	0.489	0.779
		2006	406	2038	4450

Table 1: Contingency table

## 4.2 Distribution and Density Functions

Contingency tables can only assess the sign of  $\widehat{\psi}_t$ . Density and distribution functions for the different qualitative statements however allow us to get more quantitative insights. We calculate the empirical distribution and estimate the density functions for  $\widehat{\delta}_t = i \forall i = -1, \dots, 1$  and for the entire sample as a whole.

Plotting the empirical distribution function (cf. fig. 2(a)), we see that the distribution functions for  $\widehat{\delta}_t = i \forall i = -1, \dots, 1$  show a jump at  $\widehat{\psi}_t = 0$ . These jumps are caused by the observations standing in the column  $\widehat{\psi}_t = 0$  of the contingency table. When estimating the density functions, this concentration leads to a bunch in the region of  $\widehat{\psi}_t = 0$  (cf. fig. 2(b)). One could argue that each distribution is a combination of two distributions: a point-distribution with  $\widehat{\psi}_t = 0$  and a (maybe parametric) distribution. Figure 3 shows the distribution and density functions of the data when all observations with  $\widehat{\psi}_t = 0$  are excluded.

An observation that does not change when shifting from fig. 2(b) to fig. 3(b) are the large intersections of the density functions. The areas under the density functions in the left (right) of  $\widehat{\psi}_t = 0$  are the observations of the left (right) column in the contingency table. The area under the density function for  $\widehat{\delta}_t = 1$  on the left of  $\widehat{\psi}_t = 0$  represent the inconsistent observations in the lower left field of the contingency table.

Figure 4 shows stylized density functions. It will help us to develop a measure for the intersection of the density functions. We first determine  $\mathcal{A}$ , the area of all density functions together (the shaded area in fig. 4(a)). In the case of 3 density functions,  $\mathcal{A}$  will be at most 3 (when the density functions are disjoint). Then we determine  $\mathcal{B}$ , the area lying under more than one of the densities (cf. fig. 4(b)). In the case of disjoint densities,  $\mathcal{B} = 0$ .  $\mathcal{B}/\mathcal{A}$  is a measure of intersection.

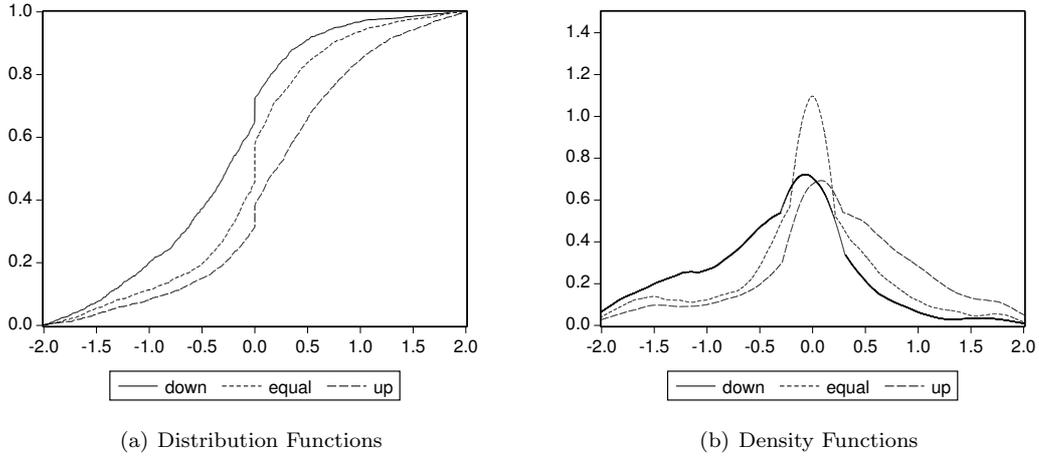


Figure 2: Distribution of  $\hat{\delta}_t = i \forall i = -1 \dots 1$  for total fixed investment. Nonparametric estimation of the density functions using Epanechnikov.

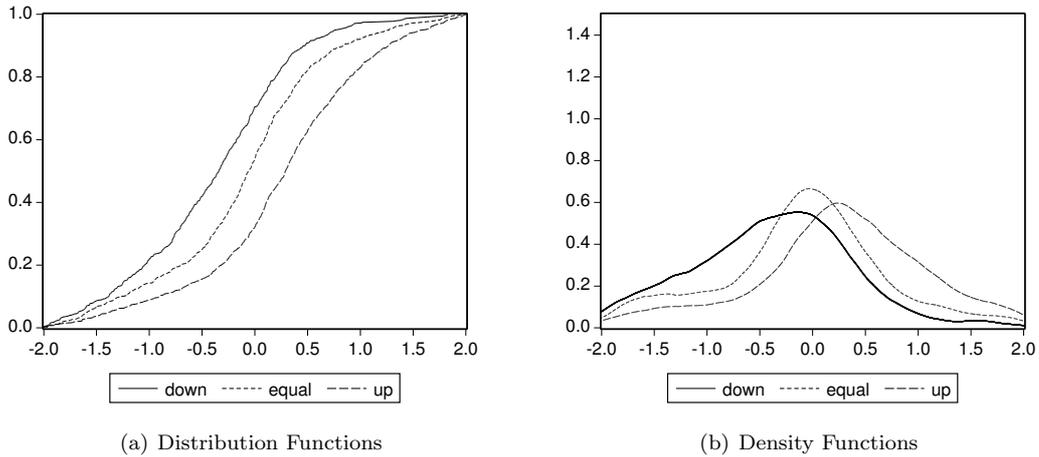


Figure 3: Distribution of  $\hat{\delta}_t = i \forall i = -1 \dots 1$  for total fixed investment. Nonparametric estimation of the density functions using Epanechnikov. All observations with  $\hat{\psi}_t = 0$  are excluded.

It is bounded to the interval  $[0, 1]$ . If  $B/A = 1$  then the density functions are congruent. In fig. 2(b),  $\mathcal{A} = 1.395$  and  $\mathcal{B} = 0.918$ , thus  $\mathcal{B}/\mathcal{A} = 65.8\%$ . In fig. 3(b), the intersection is slightly larger:  $\mathcal{A} = 1.374$  and  $\mathcal{B} = 0.953$ , thus  $\mathcal{B}/\mathcal{A} = 69.3\%$ .

Now, we are going to test whether the distributions functions follow some parametric distributions. First, we test the distributions of  $\hat{\psi}_t$  conditioned on  $\hat{\delta}_t = i \forall i = -1, \dots, 1$ . The distributions we test against are the normal, the logistic and the exponential. The measure to test the empirical distributions for parametric distributions is the Watson (U2)-statistics. For details on the computation of the test statistics and the p-values see Durbin (1973) or Stephens (1986). As we can see in the first part of table 2, the three tested empirical distributions are significantly different from

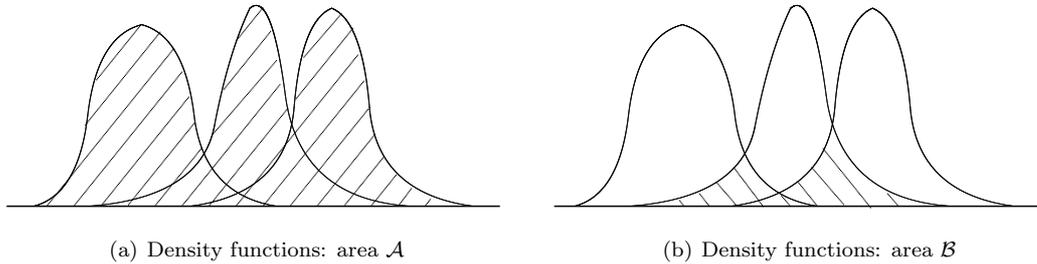


Figure 4: Overlap

Sample	Density	normal		logistic		exponential	
		Watson	p-value	Watson	p-value	Watson	p-value
total	$\delta_t = -1$	0.556	0.000	0.495	<0.005	11.873	0.000
	$\delta_t = 0$	2.387	0.000	1.484	<0.005	21.178	0.000
	$\delta_t = 1$	0.880	0.000	0.380	<0.005	25.382	0.000
truncated	$\delta_t = -1$	0.164	0.010	0.173	<0.005	9.885	0.000
	$\delta_t = 0$	0.858	0.000	0.425	<0.005	15.895	0.000
	$\delta_t = 1$	0.615	0.000	0.192	<0.005	22.792	0.000

Table 2: Empirical Distribution Tests

the parametric distribution functions.<sup>4</sup> The second part of table 2 shows the test results for the truncated sample (observations with  $\hat{\psi}_t = 0$  are excluded). In the truncated sample, the values for  $\hat{\psi}_t$  do not follow a normal, logistical, or exponential distribution, either.

After examining the distribution functions individually, we are now going to compare them. Looking at the distributions for the three statements  $\hat{\delta}_t = i \forall i = -1, \dots, 1$ , it seems clear that they are different. To compare distributions, different test are known in the literature. A widely used measure to compare the means of empirical distributions is the Anova F-statistics. To compare the medians, we use the Kruskal-Wallis-Statistics. This is a generalization of the Mann-Whitney test. The idea is to rank the series from smallest value to largest, and to compare the sum of the ranks from of the different subgroups. For details, see Sheskin (2000). To compare the variances of different distributions, we use the Brown-Forsythe test. It is based on an analysis of variance (ANOVA) of the absolute difference from the median. For details, see Brown & Forsythe (1974). The results of these tests show that the means, medians and variances of these distributions are significantly different (cf. table 3).

Now, we will analyze the distribution of the  $\hat{\psi}_t$  regardless of  $\hat{\delta}_t$ . Carlson & Parkin (1975) assume it to be normal. Some research (e.g. Dasgupta & Lahiri 1992) has been done assuming the distribution to be logistic or exponential. They showed that the assumed distribution does not have much influence on  $\hat{\mu}_t$  and  $\hat{\sigma}_t$ . Figure 5 shows the distributions for all  $\hat{\psi}_t$ .

As we can see in table 4, the distribution of  $\hat{\psi}_t$  for all  $\hat{\delta}_t$  does not follow any of the three

<sup>4</sup>These results hold too if the observations with  $|\hat{\psi}_t| > 2$  are not excluded.

equality of	test	value of test-statistics	p-value
means	Anova F-statistic	195.658	0.000
medians	Kruskal-Wallis (tie-adj.)	420.848	0.000
variances	Brown-Forsythe	16.844	0.000

Table 3: Comparing the density functions

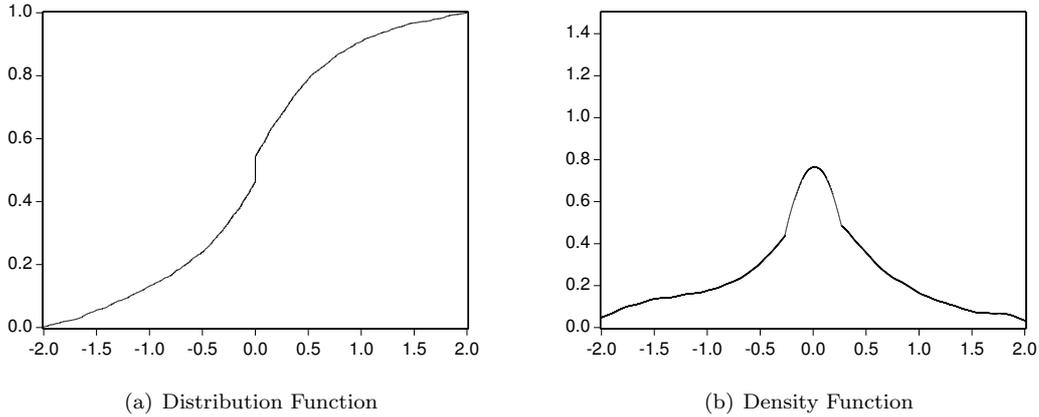


Figure 5: Distribution of  $\hat{\psi}_t$  for total fixed investment. Nonparametric estimation of the density function using Epanechnikov.

parametric functions, either. That means that when applying the Carlson-Parkin - method with any of these distributions, one of the key assumptions is violated. This may be the cause why the results are somewhat independent of the distribution assumed (cf. Carlson 1975, Lahiri & Teigland 1987, Batchelor & Orr 1988, Dasgupta & Lahiri 1992).

### 4.3 Response Functions

The density functions give an impression of the distribution of  $\hat{\psi}_t$  for a specific qualitative statement  $\hat{\delta}_t = i$ . But they can not be used to identify the importance of the qualitative statements  $\hat{\delta}_t$  for different values of  $\hat{\psi}_t$ .

Ronning (1984) presented response functions indicating the probability of the qualitative statements ( $\hat{\delta}_t = i \forall i = -1, \dots, 1$ ) when a firm is confronted with a specific quantitative value ( $\hat{\psi}_t$ ). Response functions can be considered as advancements of the response shares: They indicate the share of firms giving different qualitative statements for different values of  $\hat{\psi}_t$ .

Sample	normal		logistic		exponential	
	Watson	p-value	Watson	p-value	Watson	p-value
total	3.379	0.000	2.029	<0.005	59.191	0.000
truncated	1.116	0.000	0.548	<0.005	48.040	0.000

Table 4: Empirical Distribution Tests

Ronning expects that the probability of a “completely wrong qualitative statement” ( $\hat{\delta}_t = -1$  when  $\hat{\psi}_t > 0$  or  $\hat{\delta}_t = 1$  when  $\hat{\psi}_t < 0$ ) decreases with larger values of  $|\hat{\psi}_t|$ . Furthermore he assumes an indifference interval in which the probability of  $\hat{\delta}_t = 0$  is larger than 0. He assumes this interval to include  $\hat{\psi}_t = 0$ .

We adopt this concept on the investment data and calculate response functions:

$$\gamma_{\hat{\psi}_t}^{(i)} = \frac{N_{\hat{\psi}_t}^{(i)}}{\sum_{k=-1}^1 N_{\hat{\psi}_t}^{(k)}} = \frac{\beta_{\hat{\psi}_t}^{(i)} \cdot N^{(i)}}{\sum_{k=-1}^1 N_{\hat{\psi}_t}^{(k)}}$$

with  $\gamma_{\hat{\psi}_t}^{(i)}$ : value of the response function for  $\hat{\delta}_t = i$  and  $\hat{\psi}_t$ ,  $\beta_{\hat{\psi}_t}^{(i)}$ : Value of the density function for  $\hat{\delta}_t = i$  at  $\hat{\psi}_t$ ,  $N_{\hat{\psi}_t}^{(i)}$ : Quantity of firms with  $\hat{\psi}_t$  and  $\hat{\delta}_t = i$ ,  $N^{(i)}$ : quantity of firms with  $\hat{\delta}_t = i$ .

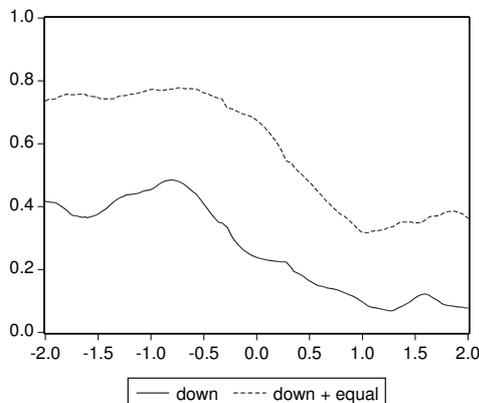


Figure 6: Response functions for total fixed investment

Figure 6 shows the response functions. In the negative interval of  $\hat{\psi}_t$ , the share of firms with  $\hat{\delta}_t = -1$  amounts 40% and decreases with augmenting  $\hat{\psi}_t$ . Contrariwise, the share of firms with  $\hat{\delta}_t = 1$  increases with augmenting  $\hat{\psi}_t$ . The probability for  $\hat{\delta}_t = 0$  is larger than zero and quite the same on the whole range  $|\hat{\psi}_t| \leq 2$ . This means that the indifference interval, as defined by Ronning, is very large. All these observations do not change much if the answers are weighted by the corresponding subjective certainty indicated in IVL. This is because the share of firms indicating to be “not sure” and “quite unsure” is only 7% and these answers are distributed virtually uniformly on  $|\hat{\psi}_t| \leq 2$ .

## 5 Conclusions

In this article we compared quantitative and qualitative investment data on firm level. It showed that firm’s answers in these two surveys are in most cases congruent. The density functions show large intersections, though. Further research should try to investigate the determinants of these intersections. Possible factors of influence could be the firm size and the sector of the firm.

Another interesting issue would be to develop a quantification method that is based on the response functions rather than the response shares ( $A_t$ ,  $B_t$ , and  $C_t$ ) only. As the response functions investigate the response shares for different values of  $\hat{\psi}_t$ , a quantification method based on the response functions should show better performance than the methods based on the response shares.

## References

- R. A. Batchelor & A. B. Orr (1988). 'Inflation expectations revisited'. *Economica* **55**(219):317–31.
- M. Brown & A. Forsythe (1974). 'Robust Tests for the Equality of Variances'. *Journal of the American Statistical Association* **69**:364–367.
- J.-A. Carlson (1975). 'Are Price Expectations Normally Distributed?'. *Journal of the American Statistical Association* **70**(352):749–54.
- J.-A. Carlson & J. M. Parkin (1975). 'Inflation Expectations'. *Economica* **42**(166):123–38.
- S. Dasgupta & K. Lahiri (1992). 'A Comparative Study of Alternative Methods of Quantifying Qualitative Survey Responses Using NAPM Data'. *Journal of Business and Economic Statistics* **10**(4 (October)):391–400.
- J. Durbin (1973). *Distribution theory for tests based on the sample distribution function*. Society for industrial and applied mathematics, Philadelphia - Pa.
- K. Lahiri & C. Teigland (1987). 'On the Normality of Subjective Probability Distribution of Forecasts'. *International Journal of Forecasting* **3**(2):269–279.
- G. Ronning (1984). 'Welche Informationen enthält die Antwortkategorie 'unverändert' in Tendenzbefragungen?'. *Zeitschrift für empirische Wirtschaftsforschung (Ifo-Studien)* **18**:261–72.
- D. J. Sheskin (2000). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC Press, Boca Raton, 2nd edn.
- M. A. Stephens (1986). 'Tests Based on EDF Statistics'. In R. B. D'Agostino & M. A. Stephens (eds.), *Goodness-of-Fit Techniques*. Dekker, New York Basel.