Chapter Thirteen

Peer group and descriptive statistics

Introduction

13.1. Sector balance sheets and income and expense data can disguise important trends. For example, the sector-wide capital to asset ratio for deposit-takers is essentially the average capital to asset ratio for the system (derived by the summation of all institutions capital divided by all institutions assets), and, if symmetrically distributed, would convey information about the middle capital asset ratio (the median) as well as the most frequently observed capital asset ratio (the mode). However, the ratio does not indicate whether the individual institutions' capital ratios are clustered in a narrow range around the average value, or are spread over a wide range. Moreover, if one highly capitalized deposit-taker offsets several other undercapitalized deposit-takers, the aggregate ratio may appear robust, masking significant vulnerabilities from weak deposit-takers whose failures could lead to contagion throughout the system. So in IMF discussions with both users and compilers of FSIs, the need for peer group analysis, and dispersion analysis has been highlighted.

13.2. A wide variety of meaningful peer groups can be created for comparison purposes, and to examine the dispersion and concentration of the institutions in the peer group or sector, descriptive statistics can be compiled. Such information can significantly affect the conclusions about vulnerabilities or strengths of the financial system. This chapter describes some types of peer groups that can be created, and discusses measures of concentration and of dispersion. Issues to address in developing these data are set out, such as the approach to weighting the contribution of the individual institutions, and some explanation of how to analyze the results is also provided. While going beyond the requirements of the agreed FSIs, some recommendations for peer groups and concentration measures to compile are provided. However, the chapter primarily sets out options and ideas for use by compilers and analysts. Indeed, the IMF staff would be interested in learning about country experience in using peer group and dispersion analysis.

Peer group analysis

13.3. A peer group is a statistical set of individual institutions that are grouped on the basis of specific analytically interesting criteria. Peer groups can be used to compare FSI ratios for (1) individual deposit-takers for which data are publicly available against the ratios for similar institutions, (2) peer groups with other domestic peer groups, or (3) peer groups across countries. Peer group analysis can be undertaken using either cross-border or domestically consolidated data.

Types of peer groups

13.4. Depending upon analytical needs, different types of peer groups may be constructed. Some might be on an ad-hoc basis. For example, recent entrants into the market, deposit-takers with low or high capital ratios, with low or high return on equity, with high levels of nonperforming loans, and/or deposit-takers that concentrate lending to particular types of borrowers. Other peer groups might be of a more permanent nature. For example, groups of similarly sized deposit-takers based on their total assets. Flexibility in approach across countries is likely.

13.5. By way of example, peer group data could be constructed for the following major groupings of deposit-takers:

- **Size of assets or revenues**. The size of institutions might affect market competitiveness or market power. Moreover, the condition of the peer group comprised of the largest deposit-takers is often important for understanding overall stability—such as the three or five largest deposit-takers, based on total assets—because these deposit-takers are the most likely to be systemically important and may exercise the greatest market power. Such a group has a small enough number of institutions that it can be constructed for most economies, and can facilitate international comparison.

- **Line of business**, such as distinguishing regular retail banks from mortgage banks.

- **By type of ownership**, such as distinguishing public controlled from private controlled deposit-takers.
- **Offshore deposit-takers** that can only transact with nonresidents.
- Deposit-takers by **region of the economy**.

13.6. From the above list, the *Guide* encourages, as a minimum, the compilation of core FSIs for peer groups based on the relative size of assets, such as percentiles (see paragraphs 13.37 to 13.38 ahead) or groupings of specific numbers of deposit-takers based on size. While peer group analysis may be less useful if large percentiles are chosen, the *Guide* discourages the dissemination of peer group data that might reveal information about specific institutions, unless the country normally requires deposit-takers to publicly disclose individual institution information.

Compilation of peer group data

13.7. A key choice in constructing peer group data is determining how data are to be compiled. Regardless of approach taken, constructing peer groups depends critically on the cost of compiling data and the ease of reorganizing the data according to analytical needs. To allow construction of peer groups, the *Guide* encourages compilers to maintain individual institution data in a database system that allows quick, low-cost data aggregation. Under such an approach, data series can potentially be compiled using the same principles and frameworks as the sector-level data. So, for instance, intra-group income and expense items, and possibly depending on data availability, intra-group equity holdings can be eliminated.

13.8. However, a decision is required as to whether data should be compiled on the basis that the peer group is a sub-group of the total population—that is, the data are the peer group’s contribution to the total for the population—or compiled on a standalone basis—that is, the peer group is self-contained, with all institutions outside the group treated as entirely external to the group. There are advantages for adopting either approach but data compilation considerations may be decisive, particularly if ad-hoc groups are created.

13.9. In this regard, the standalone approach is likely to require less additional data than the sub-group approach. For instance, when aggregating data for all institutions in a peer group, intra-peer group interest income and expense will be eliminated in the net interest income line. But to also eliminate interest income and expense with institutions in the sector but not in the peer group, additional data will be required.

13.10. However, even the standalone approach will require additional data items if they are to be compiled in line with the sector-level approach. Some of this information might be obtainable from the data reported in Table 11.2 or 11.4 depending upon the consolidation approach adopted. For instance, intra-peer group holdings of equity could be eliminated to the extent that deposit-takers identify their holdings of equity on an individual deposit-taker basis. Nonetheless, particularly for ad-hoc groups, peer group data might well be compiled on an approximate best practice basis so allowing the identification of trends but, depending upon the degree of approximation and the scope of analysis, potentially masking relevant inter-relationships. It is encouraged that in such circumstances any relevant potential limitations of the data be identified for the user, such as capital and reserves not being fully adjusted for intra-peer group holdings.

Descriptive Statistics

13.11. In many ways, the use of concentration and dispersion measures is a research and analytical activity in which specific techniques are used based on the nature of the issue under review; the types of data available and the ease of using them; and the sensitivity of the data and limitations on revealing information on specific institutions. Although there are several common elements that will be discussed below, flexibility in selecting techniques should be maintained. This section provides a menu of diverse techniques that are useful in a variety of situations. However, in disseminating information to the public, some types of descriptive statistics may prove more useful, because they can describe concentration and dispersion without revealing information on individual institutions.

Measures of concentration

13.12. The **Gini index** estimates a numeric value for concentration (see the example ahead). It captures the information shown in a Lorenz curve, which is the difference between actual concentration and the hypothetical state in which no concentration exists. With no concentration, every unit has the same endowments (income, market share, volume of market trading, etc.), which generates a Gini index of zero. If only one unit is endowed with all income, assets, etc, and no other unit has any, there is perfect concentration and the Gini index is one. Commonly, Gini indices for personal income fall between 0.20 and 0.45. Gini indices are especially useful to track changes in concentration over time.

13.13. For example, for N deposit-takers, sorted from smallest to largest total assets.

$$\text{Gini} = \sum_{i=1}^{N} 2\left(X_i - Y_i\right)\Delta X_i$$

$$\text{where: } X_i = \frac{i}{N} * 100$$

Y_i = cumulative percentage share

$X\;_i = X_i - X_{i-1}$

Gini Index (Sorted smallest to largest)

Deposit-taker$_i$	Assets	Percent Share	Cumulative Actual Share Y_i	Cumulative Equal Share X_i	Difference X_i-Y_i	Difference*2 $(X_i$-$Y_i)$*2	(Difference*2) * .091[1] $((X_i$-$Y_i)$*2)*$(X_i$ - $X_{i-1})$
11	20	2	2	9.1	7.1	14.2	1.291
10	20	2	4	18.2	14.2	28.4	2.583
9	20	2	6	27.3	21.3	42.6	3.875
8	40	4	10	36.4	26.4	52.8	4.803
7	50	5	15	45.5	30.5	61.0	5.549
6	50	5	20	54.6	34.6	69.2	6.296
5	80	8	28	63.7	35.7	71.4	6.496
4	90	9	37	72.8	35.8	71.6	6.514
3	130	13	50	81.9	31.9	63.8	5.804
2	200	20	70	91.0	21.0	42.0	3.820
1	300	30	100	100.0	0.0	0.0	0.000
							47.030[2/] **Gini Index**

[1] The “equal share” percentage of the total. [2] This index is scaled by a factor of 100.

13.14. The **Herfindahl Index**, H, is the sum of squares of the market shares of all firms in a sector (see the example ahead). Higher values indicate greater concentration. Assuming for simplicity, in the no concentration situation that 100 firms exist, and each has an identical 1 percent of the market, the value of H = 100. In contrast, with perfect concentration, in which one firm has 100 percent market share, H = 10,000. (That is, the contribution of the single monopoly firm is 100 x 100 = 10,000). A rule of thumb sometimes used is that H below 1,000 is considered relatively limited concentration, and H above 1,800 indicates significant concentration.

$$H = \sum_{i=1}^{N} \left(share_i \right)^2$$

13.15. As noted in Chapter 6, the *Guide* encourages dissemination of the Herfindahl Index. For ease of compilation, it is also possible to compile partial Herfindahl indices, such as one based on the shares of the total sector assets of the largest five deposit-takers.

Herfindahl Index

Deposit-taker	**Assets**	**Percent Share**	**Share^2**
1	300	30	*900.0*
2	200	20	*400.0*
3	130	13	*169.0*
4	90	9	*81.0*
5	80	8	*64.0*
6	50	5	25.0
7	50	5	25.0
8	40	4	16.0
9	20	2	4.0
10	20	2	4.0
11	20	2	4.0
Total	1000	100	**1692** **Herfindahl Index** *(Top 5 = 1614)*

Measures of Dispersion

13.16. Descriptive statistics on data dispersion provide measures of average values for groups of institutions, and the size and direction of asymmetry in the distribution of the observations. The four main categories of these statistics are measures of (1) central tendency, (2) variability, (3) skewness, and (4) kurtosis. They can be useful for data analysis, for comparing multiple data sets, and for reporting final results of a survey.[207] In disseminating information, graphical presentations, such as a simple scatter diagram, can be useful to provide users with information on the dispersion of data around the mean.

13.17. **Measures of central tendency** include:

- Mean (first moment of the distribution), or $\overline{X} = \sum_{i=1}^{N} \frac{x_i}{N} => \sum \left(x_i \bullet \frac{n_i}{N} \right)$. This is the arithmetic average of the data. Generalizing $\overline{X} = \sum (x_i \bullet weight_i)$.

Where,

$x_i =$ value of observation i

$n_i =$ number of observations with value x_i

N = total number of observations

$\frac{n_i}{N}$ = weight

$\overline{X} =$ population mean

13.18. **Other measures of central tendency** include:

- *Median* is the middle observation in a data set. It is often used when a data set is not symmetrical, or when there are outlying observations.
- *Mode* is the value around which the greatest number of observation are concentrated, or the most common observation.

[207] An issue arises as to whether dispersion analysis should be undertaken on a standalone basis or on a subgroup basis. As noted in the chapter, there are advantages with both approaches, although the standalone data may be more readily available, but to help understanding of any data disseminated, it is important to know the approach taken. For instance, the mean and variance for FSI ratios for peer groups can vary depending upon the basis on which the data are compiled.

13.19. **Measures of variability** describe the dispersion (or spread) of the data set:

- *Range* is the difference between the largest and the smallest observations in the data set. It has limitations because it depends on only two observations in the data set.

- *Variance* (second moment of the distribution, or $\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}$) => $\sum \left[\frac{n_i}{N} \bullet \left(x_i - \bar{X}\right)^2 \right]$ measures the dispersion of the data around the mean, taking into account all data points. Generalizing, $\sigma^2 = \sum \left[\left(x_i - \bar{X}\right)^2 \bullet weight_i \right]$

Where,

$\sigma =$ population standard deviation

- *Standard Deviation* (or $\sigma = \sqrt{\sigma^2}$) is the positive square root of the variance, and is the most common measure of variability. Standard deviation indicates how close observations are to the mean.

13.20. **Skewness** (third moment of the distribution, or μ_3) indicates the extent to which data are asymmetrically distributed about the mean: Positive skewness indicates a longer right hand-side (tail) of the distribution; negative skewness a longer left tail. One measure of skewness is based on the difference between the mean and the median, standardized by dividing by the standard deviation:

$$\text{Skewness} = \frac{\sum \left[\left(x_i - \bar{X}\right)^3 \bullet weight_i \right]}{\sigma^3}$$

13.21. **Kurtosis** (fourth moment of the distribution, or μ_4) indicates whether the data are more or less concentrated toward the center; that is, the degree of flatness of the distribution near its center. The kurtosis of a normal distribution equals 3, so it is common to subtract 3, as above, to estimate “excess kurtosis” to evaluate whether the distribution has a greater or lesser peak than the normal distribution. Positive excess kurtosis indicates that the distribution is more peaked than the normal distribution; negative excess kurtosis indicates a relatively flat distribution.

$$\text{Kurtosis} = \left(\frac{\sum \left[\left(x_i - \overline{X} \right)^4 \bullet weight_i \right]}{\sigma^4} \right) - 3$$

Weighting options

13.22. In compiling dispersion data, an issue to address is whether data should be compiled on the basis that each observation has the same weight (equal weight approach) or is weighted by its relative contribution to the numerator and denominator (weighted-by-contribution approach). As noted above, the *Guide*'s approach at the sector level is in effect to weight-by-contribution.

13.23. In dispersion analysis, the equal-weight approach facilitates identification of whether weaknesses are concentrated in one or two deposit-takers or spread across a larger number of institutions and helps identify emerging weaknesses regardless of the size of the institution.

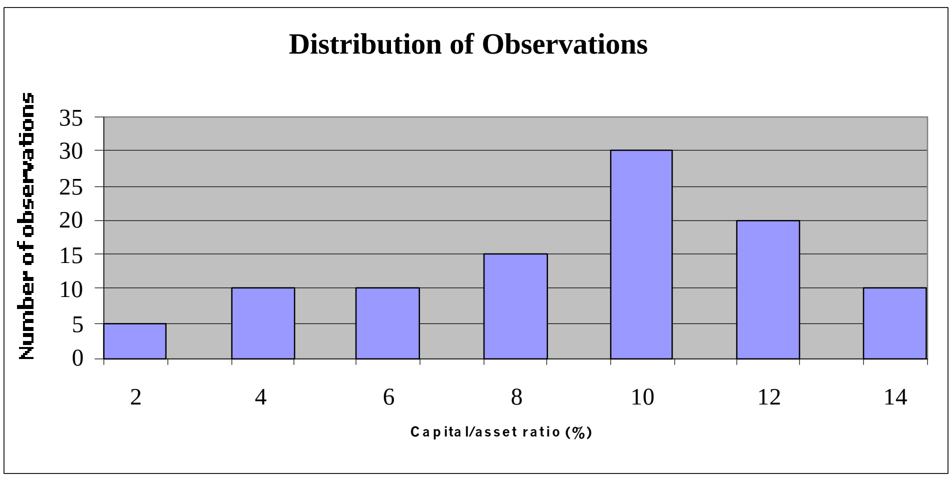
13.24. Nonetheless, variance, skewness, and kurtosis can be calculated using the weight of the contribution from each observation; for the variance the distance of each observation to the mean should be scaled by its weight in the overall average; and the skewness and kurtosis should measure the distribution of the weighted contribution of each observation to the mean, relative to a normal distribution. Compilation (and dissemination) of descriptive statistics on a weighted-by-contribution basis might reveal whether outliers are small or large from a sector perspective.

13.25. Because of their analytical usefulness, dispersion statistics could be disseminated on both bases, with any preferred approach based on data availability. However, if the equal-weight approach is adopted users should be made aware that the mean under this approach might well be different from the FSI itself. Any such difference could be useful information in its own right.

Interpretation of descriptive statistics

13.26. How are dispersion statistics data to be interpreted? Set out in Figure 1 is an example of an economy that has 100 deposit-takers with capital asset ratios that are distributed as shown in the Figure. Table 13.1 provides dispersion statistics data on an equal-weight basis and Table 13.2 on a weighted-by-contribution basis.

Figure 1



Equal weight approach

Table 13.1

Mean	Median	Mode	Variance	Standard deviation	Skewness	Kurtosis
9.1	10.0	10.0	10.7	3.3	-0.5	-0.5

13.27. These statistics could be interpreted as follows: As the value of the mean is less than both the median and mode, this indicates that the distribution is asymmetric, with a leftward skew (i.e. a longer tail toward smaller values). This is confirmed by the negative value for the measure of skewness. Further, the standard deviation indicates some significant dispersion around the mean. This is confirmed by the negative kurtosis, indicating a flat distribution (relative to a normal distribution).[208]

Weighted-by-contribution approach

Table 13.2

Weighted Mean	Standard deviation	Median	Mode	Skewness	Kurtosis
7.4	4.7	10.0	10.0	0.17	-1.51

13.28. The weighted-by-contribution approach produces different results from that of the equal-weight approach. As seen in Table 13.2, the mean is lower and standard deviation higher on a weighted-by-contribution basis. This is due to the big weights at the end of the tails, and large negative Kurtosis reflecting low peakedness (fat tails).

13.29. Figures 2 and 3 add to this analysis. The height of the columns in Figure 2 show the distribution of the individual institutions ratios by weight, that is the contribution of those deposit-takers to the sector-level FSI. The weights are presented in percentage terms and sum to unity. Figure 3 indicates both the weight—through the size of the bubble—and the number of institutions at each ratio—through the bubble's height. These figures show that the outliers in the equal-weighted distribution take on increased significance in the weighted-by-contribution distribution. In this example, of the 100 deposit-takers in the system there are only 5 deposit-takers with ratios of 2 percent and 10 deposit-takers at 14 per cent but between them they account for half the weight—in other words, the outliers are relatively important.

[208] The standard deviation for the population can be used to estimate the percentage of the population members that lie within a specified distance of the mean. Tcehbychev's rule is commonly used for forming such estimates.

Figure 2

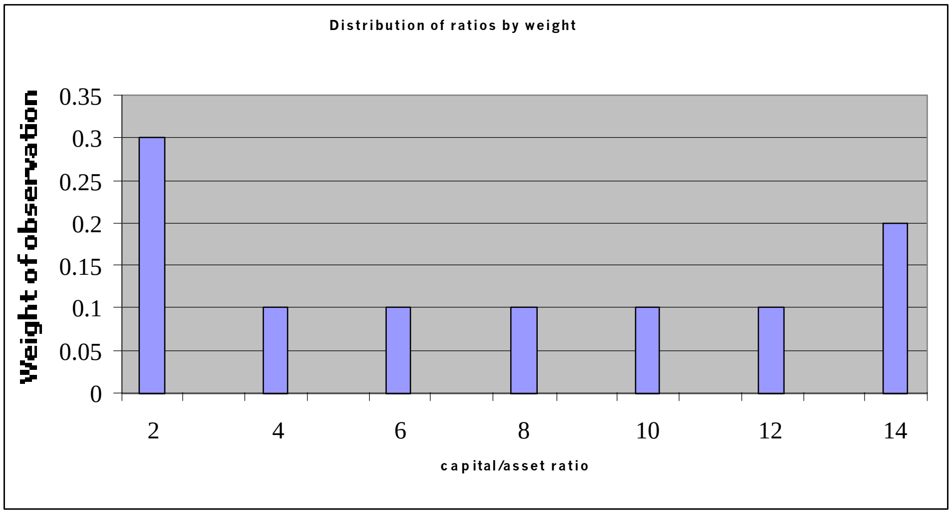
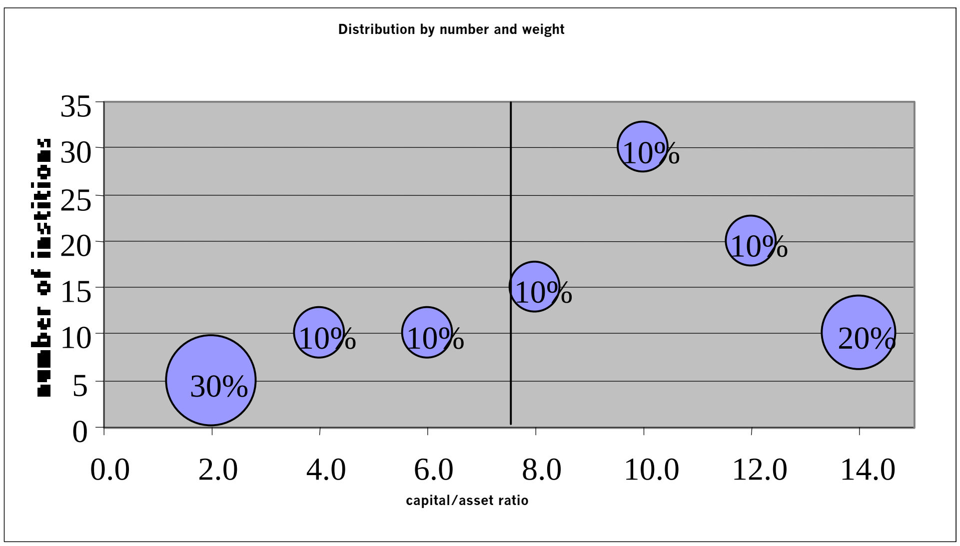


Figure 3



13.30. Another approach to weight-based analysis is to compare individual deposit-takers' (or peer groups) weight-by-contribution to specific FSIs with their relative size in terms of their contribution to sector assets. For instance, a deposit-taker generating large income flows through transactions in the financial market could make a significantly bigger contribution to the sector's income-based FSIs than its asset size would suggest. Such divergence over a period of time might raise the question as to whether the deposit-taker was taking large risks to generate large income flows. Such a comparison might also be used as a tool to check the reliability of data submitted.

13.31. Divergence between the relative balance sheet size of a deposit-taker and its contribution by weight to specific FSIs can be identified by constructing the following comparison ratio:

$$Comparison\,ratio_{i,j} = \frac{Weight - by - contribution_{i,j}}{Weight - by - asset\,size_j}$$

Where, Weight by asset size = $\dfrac{Asset\,size_j}{\sum_{j=1}^{N} Asset\,size_j}$,

i is the ith FSI, j is the jth reporting institution, and N is the total number of reporting institutions.

13.32. A comparison ratio for deposit-taker$_i$ and FSI$_j$ of more (less) than unity indicates that, compared with the rest of the deposit-taking sector, deposit-taker$_i$ has a larger (smaller) weight-by-contribution to the specific FSI than its balance sheet size suggests. A summary matrix of comparison ratios (deposit-taker$_i$ and FSI$_j$) can be constructed.

Extensions of dispersion measures

13.33. Although the set of (core) descriptive statistics provides a useful overview of the distribution, they do not fully illuminate financially weak (strong) spots—that is the left (right) tail of the distribution.[209] In other words, how many deposit-takers populate the left (right) tail and how are they distributed therein? Some possible extensions to the descriptive statistics in the *Guide* are explored below. The examples are provided on an equal-weight basis.

[209] The terms "weak" and "strong" are relative concepts in this context. That is, they are used to convey weakness or strength relative to the mean, which itself may be weak or strong vis-à-vis a predetermined norm or benchmark (such as 8 percent for the capital adequacy ratio).

Option 1: Right and Left Tail Attributes

13.34. The measures of central tendency and variance set out in the *Guide* can be applied to the left and right tails of the distribution, as shown in Table 13.3 below. This provides some additional insight into the size of the skewness, especially if the size of the standard deviation for the left and right tails relative to their respective means are compared; the relatively large standard deviation for the left tail reveals there are a number of institutions with ratios significantly below 5.8. Nevertheless, further disaggregation of the data is needed to get at how many institutions are involved and how far to the left the distribution is skewed.

Table 13.3

	Mean	Median	Mode	Variance	Standard deviation	Skewness	Kurtosis
Total	9.1	10.0	10.0	10.7	3.3	-0.5	-0.5
Left tail	5.8	6.0	8.0	4.6	2.1		
Right tail	11.3	11.0	10.0	2.3	1.5		

Option 2: Ranges

13.35. One way of conveying additional information on the distribution is to show the number of institutions falling within specified ranges (or intervals), for example the number of institutions with FSI ratios between 2 and 4 (see Table 13.4). This can be supplemented with mean and variance information for each interval. While providing additional insight into the shape of the distribution, the usefulness of this approach is dependent upon the size of the intervals. Moreover, cross-country and cross-FSI comparisons can be complicated with this approach because interval size will likely differ across countries and FSIs.

Table 13.4

Range	2-4	5-8	9-11	12-14
Number	15	25	45	30
Mean	3.3	7.2	10.0	12.7
Standard deviation	1.0	1.0	0.0	1.0

13.36. Nevertheless, this approach might be well suited to indicators that have an accepted norm or benchmark, such as the Basel Capital Adequacy Ratio, for which the analysis could focus on the distribution of ratios to the left of the benchmark. This option may become more widely applicable as countries gain experience with FSIs and the calibration of benchmarks to local circumstances.

Option 3: Percentiles

13.37. The percentile distribution of individual deposit-takers' ratios goes some way to address concerns about cross-FSI country comparison of ranges. Percentile analysis involves arranging observations in ascending order and dividing the data into groups of equal number of observations. The values that serve as the dividing lines between groups are called percentiles. For example, Table 13.5 below shows that the 10th percentile corresponds to an observation of 4, and that the 20th percentile corresponds to an observation of 6.[210]

13.38. Combined with the mean and standard deviation for each percentile range (e.g. 0-10%, 10%-20%, 20%-30% etc.), these statistics can reveal areas of financial weakness.[211] For instance, from Table 13.5, the extended left tail is clearly reflected in the

[210] It is important to note this is does not say all deposit-takers with ratios of 4% are in the bottom percentile; some deposit-takers with ratios of 4% may also populate the next percentile. Also, if the percentile value is not a multiple of 1/(n-1), where n is the number of observations arranged in ascending order, the value at the kth percentile is determined by interpolation.

[211] The mean and standard deviation can also be calculated for each percentile range on a cumulative basis (e.g. 0-10%, 0-20%, 0-30% etc), in which case the mean and standard deviation for the population will equal the mean and standard deviation for the entire percentile range.

spread of ratios across the first four percentile ranges. Moreover, the large standard deviation relative to the mean for the bottom percentile indicates that the tail extends below 4% for a number of institutions. By contrast, the standard deviation of zero for other percentile ranges indicates that all observations in each range are equal to the mean for that range.

Table 13.5

Percentile	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
FSI ratio	4	6	8	9.2	10	10	10.6	12	12.2	14
Mean for percentile range	3.0	5.0	7.0	8.0	10.0	10.0	10.0	12.0	12.0	14.0
Standard deviation for percentile range	1.1	1.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

13.39. Nevertheless, as with any system that involves decomposition of aggregated data, the choice of approach can be constrained by confidentially issues, such as not disclosing information that contains less than 3 institutions. Also the usefulness of this approach depends on the number of percentiles used.

Further extensions of dispersion measures

13.40. To extend the data analysis, both the variation in the distribution of FSI ratios and the persistence of individual deposit takers' FSI ratios through time can be observed.

Variation in the distribution [212]

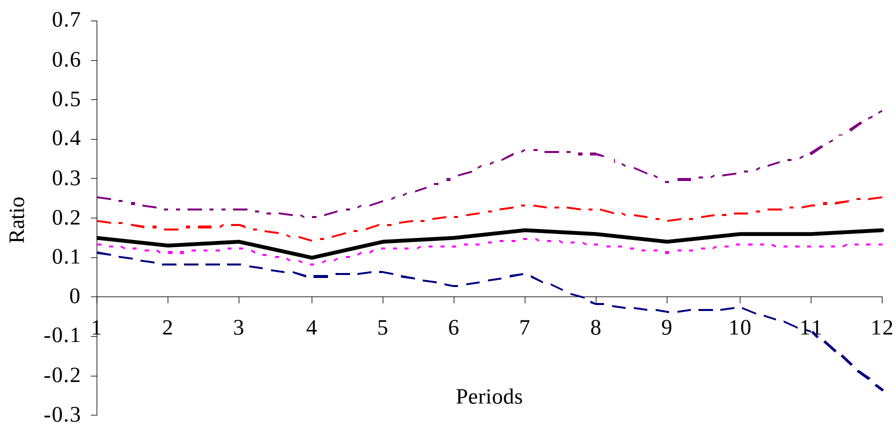
13.41. At different percentiles, the variation in the distribution of deposit-takers' rates of return over time can be observed, so facilitating an understanding of trends within the sector-level data.

13.42. Chart 1 provides an example using profitability data. An interpretation of the chart might be as follows: Until period 4, the rates of return at all percentiles tended to move in the same direction, but thereafter there was a noticeable variation in the distribution. While

[212] Based on "Sytlised facts on UK corporate financial health: evidence from micro-data," by Andrew Benito and Dgertjan Vlieghe, *Financial Stability Review*, June 2000, Bank of England.

the path of profitability of the median deposit-taker (i.e. the return on equity at the 50th percentile) was broadly unchanged, deposit-takers in the top percentile recorded an increasing rate of return (notably from 31 percent in period 10 to 47 per cent in period 12), while those in the bottom percentile recorded falling profitability (notably from -3.0 percent in period 10 to -24.9 per cent in period 12).

Chart 1: Percentiles of distribution of return on equity[(a)]



(a) Percentiles are, from top to bottom, 90th, 75th, 50th (median), 25th, 10th.

Persistence

13.43. Inspection of particular percentiles is not informative about the “persistence” of an individual deposit-takers’ performance from one year to the next. One way of capturing this information is to construct a transition matrix (see Table 13.6) that shows the proportion of deposit-takers that move from one percentile to another over a period of time, for example one year or averaged over a number of years.

13.44. The principal diagonal (top left to bottom right) in a transition matrix gives the proportion of deposit-takers that persist in the same percentile over time. For example, Table 13.6 shows that 65.2% of the deposit-takers that populated the top percentile in period 1 also populated the top percentile in period 2. The remaining 34.8% of deposit-takers that populated the first percentile in period 1 now populate lower percentiles in period 2.

Table 13.6: Transition Matrix for one-year transitions between percentiles of the distribution of return on capital

%	Percentile $1_{t=2}$	Percentile $2_{t=2}$	Percentile $3_{t=2}$	Percentile $4_{t=2}$	Percentile$5_{t=2}$
Percentile $1_{t=1}$	65.2	21.1	6.4	3.1	4.2
Percentile $2_{t=1}$	20.0	50.5	22.6	5.4	1.5
Percentile $3_{t=1}$	7.9	21.6	46.9	20.7	2.8
Percentile $4_{t=1}$	4.1	7.4	21.7	52.3	14.5
Percentile $5_{t=1}$	4.7	2.5	3.9	18.7	70.1

13.45. An interpretation of Table 13.6 might be as follows. There is a relatively high degree of persistence, with typically more than half of the deposit-takers in a particular percentile remaining in that percentile the following period. Moreover, persistence among the very profitable deposit-takers (in the top percentile) and very unprofitable deposit-takers (in the bottom percentile) is greater than that for the three middle percentiles. Mobility from one percentile to the neighboring percentiles is greater than to the more distant percentiles.

Explaining the distribution of financial performance

13.46. Whereas describing the patterns observed in measures of financial health is relatively straightforward, explaining the patterns can be more difficult. Nevertheless, some insights can be provided by examining the characteristics of those entities in the tails of the distribution of these indicators, in effect, by combining peer group and percentile analysis.

13.47. For example, Table 13.7 considers the industrial composition of those nonfinancial companies that in the current period have the lowest level of profitability and highest levels of capital gearing (debt to equity ratio). For illustrative purposes, low profitability refers to levels below the 10^{th} percentile while high capital gearing refers to a level above the 90^{th} percentile for the sector as a whole. The table, based on the number of firms in each industry group expressed as a percentage of the total number of firms, compares the industrial distribution at the tails (rows 2 and 3) with that of the whole sector population (row 1). An interpretation of the data in Table 13.7 might be as follows: While firms with lowest profitability are to be found in each of the broad industry groups, the extraction and transport and communications industries are over-represented relative to their presence in the sector as a whole. Among the companies with high capital gearing, again the transport and communication industry is over-represented.

Table 13.7: Analysis of tails of the distribution by industry classification (per cent).

Industry Group	1	2	3	4	5	6	7	8	Total
1. All firms in sample	5	6	15	12	10	18	20	14	100
2. Firms with low profitability (ROE)	2	16	10	10	4	9	37	13	100
3. Firms with high capital gearing	3	6	8	16	7	11	34	15	100

Note: Industry groups are one-digit non-financial, Standard Industrial Classification (SIC-1980) groups.

1. Energy & water supplies; 2. Extraction of minerals and ores other than fuels; manufacture of metals, mineral products & chemicals; 3. Metal goods, engineering & vehicles industry; 4. Other manufacturing; 5. Construction; 6. Distribution, hotels & catering; 7. Transportation and communication; 8. Other services

Interactions between indicators of financial health

13.48. From a financial soundness perspective, it may matter whether, for example, the companies with high debt levels are also making losses and/or have low liquidity. The overlaps between indicators can therefore be important to the analysis, not least because the interaction between indicators can amplify vulnerability to shocks. Chart 2 provides a stylized example of the overlaps between indicators for companies. One third of the companies (i.e. 32 percent) with the highest gearing also had the lowest profitability. In addition, nearly on third of companies (ie., 29 percent) with the highest gearing (although not the exact same population of those with low profitability) had the lowest liquidity.

Chart 2: Coincidence of Financial Soundness Indicators