Portfolios of stocks are common constructs for financial researchers and practitioners. Researchers use portfolios to test asset pricing models or to measure risk premiums related to certain characteristics of companies (e.g., risk premiums). Practitioners use portfolios to measure performance of a target or control group of companies or to study the ability of certain investment strategies (styles) to produce abnormal profits. Portfolios can provide useful information about the stock market in a summary way that is helpful, especially when there are many assets in the universe.

Despite several applications for portfolios, there are hardly any studies of issues related to their construction, especially in the case of small or emerging stock markets. These markets share characteristics that affect portfolio construction and make it different from the process in larger markets.

The biggest differences are the thin trading of many companies, small number of companies, and the number of companies with multiple share classes. Many small and emerging markets also have restrictions on foreign ownership, which has often led to the quotation of both restricted and unrestricted shares (see, e.g., Hietala [1989] and Domowitz, Glen, and Madhavan [1997], for studies in Finland and Mexico).

We have two aims in this article. First, we consider the issues related to portfolio construction with an emphasis on the issues that are frequently encountered in data from small or emerging markets. Second, we provide documentation of portfolio construction in practice, using as an example Emerging Markets Data Base (EMDB) data for the Turkish stock market (the Istanbul Stock Exchange). Turkey is a good example for our purposes since it brings up most of the typical problems in portfolio construction.

STOCK RETURN MEASUREMENT

Portfolio construction requires that return time series be available for the stocks selected to be in our portfolios. In many cases, we have price but not returns and we have to calculate return series ourselves. This is not a trivial exercise.

There are a number of issues to consider. First, we have to consider the kind of price information to use as the basis for the calculation. Typically, there are several different price observations available for each stock: best bid and ask offers at the end of the day, and first (opening), last (closing), highest, and lowest transaction prices. If intraday data are not available, researchers have usually preferred to use the average transaction price measured as the arithmetic average of the highest and lowest transaction prices or the closing transaction price. These so-called realization methods have the advantage that they both reflect real...
trading opportunities (see Roll [1984]). The closing price has the advantage that it can reduce the non-synchronous trading bias when compared to market (index) returns. Small and emerging markets frequently exhibit thin trading, and one or more of these price observations may be missing. This is a stock could have no transactions during a day (the closing price is either zero or equal to the closing price for the previous trading day, depending on the database), and sometimes even the bid or the ask price is also missing if illiquidity is really severe.

If we have bid and ask offers available but no transaction price, the alternatives differ, depending on whether we have chosen to use the average transaction price or closing price approach. If we have chosen to use the average of the highest and lowest transaction prices, we typically use an average of the bid and ask offers to calculate the missing price observation. If we have chosen to use closing prices, we can either use the last available transaction price as the closing price until the next transaction takes place, or we can use the average of the bid and ask offers as earlier.

The first alternative has the disadvantage of a greater non-synchronous trading bias than the latter method, as the last transaction price observation can be several days old. Using only transaction prices, on the other hand, reflects real trading opportunities more accurately.

The situation is more complicated if a bid or ask offer (or both) observation is missing. In this situation, one is usually forced to use previous price observation, i.e., the previous closing or average trading price, or the previous average bid–ask offer. Ultimately, the choice between average price method, closing price method, or a combination of them has to be made after considering the benefits and disadvantages. The way the market index is constructed also affects the choice, because totally different methods to construct market index and portfolios can introduce measurement problems.

Once we have constructed the price series, we have to consider the kind of corrections to make to the price series, since price series are typically not appropriate for return calculations as such. Usually we have to correct them at least for cash dividends, stock dividends, issues, and splits. Sometimes this can be hard, not least because one has to track down all the correct dates for these events if they are not provided in the database.

Finally, we have to consider how to measure returns. Traditionally there are two alternatives: percentage returns, and logarithmic returns (i.e., continuously compounded returns). Percentage returns have the advantage that they are frequently used in practice, but log returns (i.e., log-arithmetic relative of the price) have become standard in financial research.

There are several reasons for this. First, multi-period returns are just sums of period returns. Second, log returns reduce the bias in returns induced by bid/ask spread and price discreteness if compared with the proportional percentage returns (see Mucklow [1994]). Third, compound returns often exhibit a higher degree of normality than percentage returns. Fourth, log returns reduce the heteroscedasticity found in most stock return series.

The use of log returns is not totally without problems. The log return on a portfolio is not exactly the value-weighted sum of log returns on assets in the portfolio, but the bias is typically small (see Campbell, Lo, and MacKinlay [1997]).

When we are measuring returns from an international point of view, we also have to take into consideration changes in currency exchange rates. Normally, this is not a problem, but in some emerging countries the currency is not freely convertible or the information is not otherwise available. In some countries, there may even be restrictions on capital outflows. Some countries also tax foreign investors’ dividends differently from their capital gains.

Finally, there are several practical problems. Calculating returns for a last listing period (so-called delisted returns) is always problematic. There can be several different reasons for delisting (e.g., bankruptcy, merger and acquisition, liquidation, or migration to another exchange), which makes return calculations almost impossible in practice when there are a large number of assets. Typically researchers have simply removed delisted returns from the sample, but this delisting bias can have major implications for the empirical results.

In most cases we also want to aggregate daily returns over longer periods (e.g., weekly or monthly). If we are using continuously compounded returns, we can accumulate returns over the desired period. That is, to get monthly returns we just accumulate daily returns over the whole month. If we want to calculate weekly returns, we usually accumulate returns from Wednesday to the next Wednesday to avoid the day-of-the-week effect found in many markets.

**ISSUES IN PORTFOLIO CONSTRUCTION**

**Portfolios Versus Individual Assets in Tests of Asset Pricing Models**

The idea of using portfolios in tests of asset pricing models (APM) is to reduce noise in the individual asset
return series due to non-synchronous trading and other measurement errors, for example. This is especially the case when betas (risk sensitivities) are estimated as independent variables in a two-step cross-sectional estimation.10

In a time series approach, the use of portfolios also makes estimation of the covariance matrix of the residual term easier. This is the case when the number of assets, \( N \), is “too close” to the number of time series observations, \( T \), or when there are many assets. At the same time, the portfolio residual terms are often lower than for individual assets.

There are also other advantages of using portfolios. Namely, many models can be estimated only with a relatively small number of assets (e.g., GARCH-type models). Yet, studies of only marketwide indices may miss important aspects of individual markets. Unlike U.S. studies, most studies on other markets have so far used mostly aggregated market/industry proxies to study asset pricing models; something interesting may thus have gone unnoticed. Finally, grouping only a few test assets on the basis of an interesting characteristic can reveal interesting insights into the data set. The results are also much easier to report and to compute.

There are also problems associated with the use of portfolios. First, the power of the tests against the null hypothesis diminishes as the number of assets is reduced (Campbell, Lo, and MacKinlay [1997]). This is because portfolios reduce heterogeneity in a sample and obscure possible relevant return characteristics in portfolio averages (Roll [1977] Gibbons, Ross, and Shanken [1989]). On the other hand, if a researcher forms portfolios on the basis of characteristics that prior evidence has found to be relevant, the null hypothesis may be rejected too often due to a “data-snooping” bias (see Brennan, Chordia, and Subrahmanyam [1998] Lo and MacKinlay [1990]).

Second, information is lost when assets are grouped together. For example, cross-correlations and other dependencies between assets are removed to a large extent from the data. Furthermore, the characteristics of the asset return distribution are altered with the use of portfolios. Finally, multivariate asset pricing tests are sensitive to the choice and number of test assets (Stambaugh [1982]).

Ultimately, the decision whether to test asset pricing models using a group of portfolios or a full set of individual assets seems to depend on our objectives. There is no definitive answer, and one has to weight the pros and cons against each other every time.

Common Issues

Once a decision to construct portfolios is made, there are still several issues that need careful consideration. Several problems in portfolio construction are common to all markets. Others are typically encountered in small or emerging stock markets. Although the answers depend partly on researchers’ interests or the research setting, some general guidelines can be given in order to minimize the problems and avoid mistakes.

The first question is the kind of coverage we want in our portfolios. Should we include stocks only from one country or from several countries? And if we choose only one country, we may still need to select between multiple stock exchanges (e.g., NYSE versus the Nasdaq). Furthermore, most stock exchanges have several lists (e.g., Main List and OTC List), and sometimes we are forced to select between them because they differ so much (one may be far more liquid than the other) or the data are not available.

The second question is the kind of portfolios to construct. That is, what information or variables should we use to group \( N \) assets into \( P \) (\( P < N \)) portfolio? In asset pricing tests, the test assets (portfolios) should: 1) have a spread in mean returns; 2) have a spread in loadings on the true factors; 3) have low cross-correlations; and 4) be economically interesting and investable (Carhart et al. [1996]). That is, to test factor pricing models (with one or more factors), we should construct the portfolios so that they show strong diversification in relation to the factors. In addition, if conditional pricing models are tested, we have to consider whether our interest is in the time series and cross-sectional behavior of the stocks, since the classification information can often also be used as conditioning (time series) information.

Portfolios can be formed on the basis of some market information (e.g., size or beta) or company-specific information (e.g., industry classification or accounting measures). Of course a combination of the market and company-specific information can also be used (e.g., book-to-market value). Sorting criteria on the basis of market information allows for more frequent reranking of the assets in the portfolios, while company-specific information typically allows for reranking of the portfolios only once a year. Market information is also often easier to use in practice, since it is usually readily available. Portfolios can also be formed using two-way ranking. If this is the case, companies are first ranked into groups on the basis of the first criterion and then within each group again on the basis of the second criterion.
The third question is the number of portfolios to construct. In asset pricing tests, the question is related to the time series observations, T, as it sets the upper limit for the number of portfolios; when N is larger than T, the estimated residual matrix is singular (Korajczyk and Viallet [1990]). The number of portfolios depends partly on the researcher’s interests and measurement problems in the return data; it is also a matter of convenience.

Usually there are not more than twenty portfolios, but the number can be less if there are few assets available. The number of portfolios can sometimes be as low as two, depending on the question of interest. Usually, however, if the number of assets is limited, the number of portfolios should be kept as high as possible while keeping the number of assets in a portfolio above a certain minimum level.

Once the number of portfolios is decided, we have to keep an eye on the number of assets in the portfolios over time so that it does not drop below a certain limit between the revisions. A related problem that the number of companies quoted on the stock market can vary remarkably over the sample period. If this happens, the number of assets in portfolios might be satisfactory for a part of the sample period but not all the time. Sometimes, however, having enough portfolios is more important. So one is faced with a trade-off, and the decision must be made on the basis of the availability of companies over time and the requirements of the tests. As a rule of thumb, one should have at least five or more assets in a portfolio.

The fourth question is whether the portfolio return should be a value-weighted or equal-weighted average of the returns on assets included in the portfolio. Once again, the choice depends on the question. Value-weighted portfolios highlight the role of size (and investibility), while equal-weighted ones highlight the cross-sectional variety of all available assets.

The value-weighted approach has generally been the predominant choice for research purposes. It has the advantage of matching practitioners’ return calculations. Furthermore, value-weighted portfolios imply lower transaction costs than (contrarian) equally weighted portfolios. Empirically, value-weighted returns usually exhibit lower autocorrelation due to thin trading, especially on higher frequencies (see, e.g., Campbell, Lo, and MacKinlay [1997]). Value-weighted portfolios, however, have the disadvantage of requiring more information (i.e., equity capitalization size) and work to construct.

If the value-weighted approach is selected, we have to decide the kind of weighting scheme to use. In other words, when and how often should the weights be updated, such as at the end of the year or some another date and such as once a year or more frequently? Typically, weights are updated either yearly to minimize transaction costs and to increase investibility or using the same frequency as the return data (i.e., daily, weekly, or monthly) to reflect the fact that asset pricing models are usually one-period models. If researchers choose to update weights only once a year, they typically do it either at the end of the year or in the middle of the year.11

The sixth question is when and how often to update portfolio content; i.e., how often the assets should be re-ranked into portfolios. This is closely related to the weighting issue. In some cases, the availability of instrument variables for grouping determines the suitable re-ranking frequency, although with certain variables (like size) we can choose to re-rank the assets almost continual (every period) or less frequently (usually yearly). The choice is ultimately up to the data frequency and the research setting.

Next is what to do when a stock is delisted and when a new one enters the list. Exchange listings are dynamic—new companies are listed, and others disappear. A new company presents a problem when it should be included in the portfolios. Usually, new companies are included as soon as we have their market values are available, i.e., with a one-period lag. Delisting of a company presents a bigger problem. It is very complicated and sometimes even impossible to calculate the return for the final period. Therefore, researchers usually exclude the last-period return from the portfolios.

Finally, are there ways to avoid the survivorship bias (only companies that have survived are included in the sample) and the look-ahead bias (the portfolio construction process uses information that was not available for investors at that time)? Survivorship bias can be a serious issue in portfolio construction. It frequently occurs when the researcher uses a database that is backfilled. That is, historical information is added to the database only for those companies that are listed today.12

Another source of survivorship bias is the selection of companies for the portfolios on the basis of some external information that is available only for companies that have survived through the period. This is very common; most publications provide information only for those companies that are listed at the time of the publication. The good thing about survival bias is that it can be avoided with careful checking of the data once one is aware of it.

A possible source for look-ahead bias in portfolio construction is the information that is used to group the
assets into portfolios. If that information was not available in reality at the time of the grouping, the results are subject to the bias.

The most common example is when accounting information is used to group assets into portfolios at the end of the accounting year. In reality, financial reports are published with a couple of month’s lag. Since it is almost impossible to keep track of when accounting information is released for every firm, researchers usually ignore this problem and assume that all firms make this information available at a certain time (such as by the end of April).

**Issues on Small Thinly Traded Stock Markets**

Small and emerging stock markets share several unique features that are of concern in the portfolio construction. First, the number of listed companies is small. While the New York Stock Exchange has thousands of listed companies, many small stock markets have a hundred listed companies or fewer. This restricts the number of portfolios that can be constructed. Moreover, extra steps have to be taken to ensure that the sample data are used to the fullest. Besides using a complete database, a good practice is not to remove companies from portfolios sooner than one period before their market capitalization values hit zero (i.e., they are removed from the exchange list) or when they are otherwise unavailable. Similarly, companies should be added to the portfolios one period after their market value becomes available.13

Second, companies in many countries issue several listed or unlisted classes of stock (e.g., common and preference shares), while in the US firms typically have only one class of stock.14 There also may be restrictions on (foreign) ownership.15 If this is the case, unrestricted and restricted stock series may be listed separately. A good example of multiple stock series is from the Finnish stock market.16

The problem with multiple stock series is twofold. Typically small stock markets have only a few companies that are actively traded, and listing multiple stock series separately exacerbates the thin trading effect on the return measurement as the available liquidity is divided across a number of stock series. A related issue is that multiple stock series measure the same thing to some degree, so multicollinearity could be a problem in the econometric analysis. To mitigate these problems, one representative stock series is typically selected to represent the company.

Selecting a representative stock series presents additional problems:

1. Which series do we select as the representative series for a company?
2. How do we measure the weight for the representative stock series?
3. What do we do with the unlisted series?

If the number of available companies is small, the best policy is to follow two principles. The first principle is to select the series that has been available the longest during the sample period. This guarantees the maximum number of stock series available at all times. The second principle is aimed at reducing the thin trading effect on the portfolio return measurement. If there are multiple series available for equally long times during the sample period, we should select the most liquid one. Liquidity, can be very hard to define, but minimizing the number of non-trading days is often a rule of thumb.17

A small forward-looking bias is caused if we select one stock series for the whole sample period using liquidity information from the same period. Yet making the decision on a monthly (or yearly) basis and switching series according to past liquidity would change the distribution of the returns unpredictably, since there are various reasons why the liquidity between the share classes may change.18

Representative stock selection has the drawback that the selected series could change once new information is revealed. That is, historical portfolio returns could change if the stocks selected for the portfolios are changed. Ultimately, the choice depends on the research agenda.

When there are several listed stock series available, we should first apply these guidelines to choose between different classes of stock (e.g., between common and preference stocks). If restricted and unrestricted stock series are listed separately, we first have to choose which one of them to use as the representative series for that class of stock. Then we choose between common and preference classes of stock.

Another issue is that in many countries restrictions on foreign ownership have been lifted during the last few decades. If constricted series are suddenly converted to unrestricted, we have to make further choices. In many cases, it is difficult to say whether the unrestricted stock series is an extension of the old restricted stock or an (often) more liquid restricted stock price series. A good practice is to let the liquidity up to the conversion date decide. That is, trading after the conversion date does not affect the choice between unrestricted and restricted series, only the choice between ordinary and preference shares.
Multiple share series also complicates measurement company's weight in the value-weighted portfolios. Usually, the market values are first calculated separately for each series, and then they are added up to represent the company's weight in the portfolio construction. A problematic issue here is unlisted stock series. Basically there are two alternatives: Either ignore them, or value them somehow and add that to the company's market value. Using the latter alternative raises a valuation problem: the price to use for the unlisted stocks. Using the market price of the listed stocks is questionable because the unlisted stocks usually have premium voting rights, as they are used to control the company. In some cases the controlling stocks are later converted to more liquid preference stocks on a 1:1 basis.

Restrictions on foreign ownership can also affect portfolio construction through weights. Namely, if we test asset pricing models from foreign investors' point of view, it seems reasonable that the weights not exceed the restrictions set on foreign ownership. In some cases, some companies may not even be available for foreign investors. A similar problem is caused by government ownership of publicly listed companies, which reduces the so-called free float of the stock. Comparing full weights versus benchmarks that take into consideration only the free-float portion of the company can be problematic.

On some small stock exchanges, only one or a few companies clearly dominate the market. This raises two kinds of considerations. First, in asset pricing tests we might think about removing this company from the portfolios, especially if the number of portfolios is small. Second, if a company grows rapidly in size, using lagged market values as weights in the portfolio construction will bias the return on that particular portfolio downward. This can seriously affect tests of asset pricing models when (for example) monthly returns are used and if the market index is calculated using weights lagged only by one day.

There are often also several minor problems. One occurs when two publicly listed companies merge, and continue under the name of one of the companies. Basically, we have to decide whether we have a new company or not and construct the data accordingly. Similarly, a company can change its name after restructuring. Dual-listing of causes the companies another problem with respect to estimation of market value (number of shares). In most cases we cannot use the total number of shares to calculate the market value if only a small portion of shares were initially sold to a new market and if there are some barriers to the free flow of shares from one market to the other. Typically one uses either the initial number of shares issued to a new market or the market's proportion of total trading as the basis for market value calculations.

Finally, there are bound to be market specific-occasions or characteristics to deal with. Changes in legislation can have huge impacts on the stock market. Governments are typically concerned with the foreign ownership of companies, and they can set additional restrictions on foreign ownership to protect national interests. Companies can also change the characteristics of their stocks themselves. A company could set up new series, change the ownership restriction status of its outstanding series, or combine two or more series.

Common to all these problems is that there are no clear rules on how to proceed. Thus, decisions have to made on a case-by-case basis.

**DATA**

We use monthly data to construct size portfolios. The sample period is January 1987 through November 1999. Data for Turkey are from the Emerging Markets Data Base (EMDB) of Standard & Poor's, which bought the database from the International Finance Corporation, part of The World Bank. The EMDB has become the most widely used database in emerging markets research. It provides comprehensive information for the companies that meet certain stock selection criteria aimed at reflecting foreign investors' real investment opportunities. In practice, the database includes most active stocks covering 60% to 75% of the total market capitalization.

Exhibit 1 shows some statistical figures for the Istanbul Stock Exchange (ISE). The total number of companies included in the database is 71. At the end of the sample period (November 1999), the database covered 53 companies, which represents less than one-fifth of the total number of stocks traded (285) on the ISE at the time. This is because the EMDB targets 60% of the total market capitalization and liquidity of the country.

Although the company coverage in the database is relatively modest, this is not the case when market values are taken into consideration. Namely, the total market capitalization value of the listed companies in the ISE was 33,000,000 billion Turkish lira (US $65 billion), and the companies included in the portfolios represent more than 80% of the total capitalization value. The highest individual market capitalization value in the database is for a company called Erdemir (in December 1989, 40.9% of the total market value).
To calculate portfolio returns, we need daily price series (or month-end values) and the number of stocks outstanding for all listed share series. These are readily available in the data from the EMDB. In addition, we may need some additional background on the market, as discussed earlier. It turns out that there are currently no restrictions on foreign portfolio investors trading in the Turkish securities markets. Decree No. 32, passed in August 1989, removed all restrictions on overseas institutional and individual investment in securities listed on the Istanbul Stock Exchange. Hence, the Turkish stock market is open to foreign investors, without any restrictions on the repatriation of capital and profits.

**Stock Prices and Returns**

Stock returns are calculated for all stocks included in the database. Stock returns are measured as the continuously compounded return on stock price indices adjusted for cash dividends, i.e., the logarithmic relative of the price. This price index is based on the last available closing price provided by the EMDB, corrected for stock dividends, issues, and splits.

We select the last closing price (i.e., so-called realization method) as a basis for return calculations, because it is readily available, and it has the advantage that it reflects real trading opportunities. We measure all returns in local currencies.

---

**EXHIBIT 1**

**Turkish Market and Portfolio Data**

**Panel A: General Information**

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock exchange</td>
<td>Istanbul Stock Exchange (ISE)</td>
</tr>
<tr>
<td>Market capitalization value in November 1999</td>
<td>TL 33e6 bill. (USD 65 bill.)</td>
</tr>
<tr>
<td>Number of companies at end of 1999</td>
<td>285</td>
</tr>
<tr>
<td>Number of listed stock series at end of 1999</td>
<td>285</td>
</tr>
<tr>
<td>Remarks</td>
<td>Stock Exchange established in 1986. All restrictions on foreign ownership abolished in August 1989</td>
</tr>
</tbody>
</table>

**Panel B: EMDB Data**

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample period</td>
<td>January 1988–November 1999</td>
</tr>
<tr>
<td>Number of time series observations</td>
<td>155 months</td>
</tr>
<tr>
<td>Number of stocks available in sample period</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
</tr>
<tr>
<td>At end of sample period</td>
<td>53</td>
</tr>
<tr>
<td>Average</td>
<td>35.3</td>
</tr>
<tr>
<td>Maximum</td>
<td>December 1996–November 1999: 58</td>
</tr>
<tr>
<td>Return series, largest observation:</td>
<td>Rabak, August 1989: 120.4%</td>
</tr>
<tr>
<td>Return series, smallest observation:</td>
<td>Petkim, February 1994: -107.7%</td>
</tr>
<tr>
<td>Maximum relative company market value</td>
<td>Erdemir, December 1989: 40.9%</td>
</tr>
<tr>
<td>Remarks</td>
<td>No multiple classes of stock in database.</td>
</tr>
</tbody>
</table>

**Panel C: Portfolios**

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of companies in portfolios</td>
<td>53</td>
</tr>
<tr>
<td>Average</td>
<td>35.1</td>
</tr>
<tr>
<td>Market value covered at end of sample period</td>
<td></td>
</tr>
<tr>
<td>Of all available companies</td>
<td>100%</td>
</tr>
<tr>
<td>Of total market value</td>
<td>TL 27e6 bill (82%)</td>
</tr>
</tbody>
</table>

---

\(^a\) 33,000,000 billion Turkish lira.

\(^b\) Continuously compounded returns.
Market Values

To calculate the value-weighted average of the returns, we need market values for all companies. The market value of a particular company is the stock price times the number of shares outstanding. The stock price is simply the month-end closing price. The number of stocks outstanding is corrected for splits, subscription and bonus stock issues, tender issues, and directed issues. The database does not include multiple classes of stock for any company. Thus, we do not have to worry about selecting representative stock series.

Weights are then calculated as the total market value of a company’s equity capital at the end of the previous weight revision period divided by the sum of the market values of all companies included in the portfolios at that time. Weights are updated in our case monthly. This means that portfolio return for a time t to t + 1 is the value-weighted average using weights calculated as of t.23

We do this in order to guarantee the investibility of the portfolios and that no forward-looking bias is induced through the weights. As we take the local investor’s point of view, no restriction on the market weights is set.

Portfolio Sorting Information

Size portfolios are constructed by ranking companies on the basis of their total market values as of time t. Sorting occurs every month at the same time the weights are updated. Companies are excluded from the portfolios if they are not available at the end of the next month.

As Exhibit 1 shows, we have on average only 35 companies available and the minimum number of companies is as low as 14, so we choose to construct only 6 size portfolios.

PORTFOLIO STATISTICS

Descriptive statistics for the portfolios are given in Exhibit 2. It shows minimum, mean, and maximum, relative market weights for each portfolio. Relative market weight is the portfolio’s market capitalization value divided by the total market capitalization value. Similar minimum, mean, and maximum, statistics are also given for the number of companies in each portfolio. Averages are calculated as the time series mean of the month-end values.

The largest-size portfolio accounts, on average, for more than 53% of the total market capitalization value; the smallest-size portfolio accounts for less than 3%. If we look at how the weight of the companies in each portfolio has changed on a monthly basis, we can see that there have been huge variations in the relative weights of the portfolios (results not reported). This can be seen in Exhibit 2 in the minimum and maximum relative weight figures. For example, the percentage of total market capitalization included in the database in the smallest portfolio has varied from 0.42% to over 6.84%.

The number of companies in the portfolios has been on average more than 5.5 for all size-portfolios. Smaller portfolios have more stocks on average due to the way the size portfolios are constructed. At the beginning of the sample period, all portfolios have only two or three companies.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Relative Value</th>
<th>Number of Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Mean</td>
</tr>
<tr>
<td>Size portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – Largest</td>
<td>29.33</td>
<td>53.24</td>
</tr>
<tr>
<td>2</td>
<td>14.89</td>
<td>20.95</td>
</tr>
<tr>
<td>3</td>
<td>5.28</td>
<td>11.74</td>
</tr>
<tr>
<td>4</td>
<td>2.21</td>
<td>6.92</td>
</tr>
<tr>
<td>5</td>
<td>1.27</td>
<td>4.88</td>
</tr>
<tr>
<td>6 – Smallest</td>
<td>0.42</td>
<td>2.29</td>
</tr>
</tbody>
</table>


a Time series average of ratio of portfolio’s market value to total market capitalization value at end of each month.

b Time series average of number of assets in a portfolio at end of each month.
**EXHIBIT 3**

Portfolio Returns

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Return Mean</th>
<th>Min</th>
<th>Max</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Excess Kurtosis</th>
<th>Bera-Jarque</th>
<th>Autocorrelationa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Sample Statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market Portfolio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISE100 Index</td>
<td>5.5%</td>
<td>-49.5%</td>
<td>81.9%</td>
<td>0.178</td>
<td>0.553</td>
<td>1.984</td>
<td>&lt;0.001</td>
<td>0.086</td>
</tr>
<tr>
<td>Size Portfolios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – Largest</td>
<td>4.5%</td>
<td>-48.9%</td>
<td>74.5%</td>
<td>0.188</td>
<td>0.622</td>
<td>1.493</td>
<td>&lt;0.001</td>
<td>0.069</td>
</tr>
<tr>
<td>24.9%</td>
<td>-62.7%</td>
<td>47.8%</td>
<td>0.180</td>
<td>-0.309</td>
<td>1.037</td>
<td>0.009</td>
<td>0.089</td>
<td>0.069</td>
</tr>
<tr>
<td>35.5%</td>
<td>-59.6%</td>
<td>57.9%</td>
<td>0.180</td>
<td>0.018</td>
<td>0.551</td>
<td>0.373</td>
<td>0.030</td>
<td>0.090</td>
</tr>
<tr>
<td>45.3%</td>
<td>-58.6%</td>
<td>69.5%</td>
<td>0.185</td>
<td>0.458</td>
<td>1.071</td>
<td>0.002</td>
<td>0.118</td>
<td>-0.004</td>
</tr>
<tr>
<td>55.7%</td>
<td>-49.2%</td>
<td>61.0%</td>
<td>0.175</td>
<td>0.513</td>
<td>0.980</td>
<td>0.002</td>
<td>0.097</td>
<td>0.020</td>
</tr>
<tr>
<td>6 – Smallest</td>
<td>5.3%</td>
<td>-51.5%</td>
<td>63.3%</td>
<td>0.195</td>
<td>0.348</td>
<td>0.239</td>
<td>0.173</td>
<td>0.094</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel B: Cross-Correlation Coefficients</strong></th>
<th>ISE-Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISE100 Market Index</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – Largest</td>
<td>0.885</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.859</td>
<td>0.742</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.847</td>
<td>0.758</td>
<td>0.768</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.877</td>
<td>0.790</td>
<td>0.774</td>
<td>0.844</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.907</td>
<td>0.793</td>
<td>0.768</td>
<td>0.852</td>
<td>0.852</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 – Smallest</td>
<td>0.877</td>
<td>0.776</td>
<td>0.741</td>
<td>0.799</td>
<td>0.821</td>
<td>0.874</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The normality of the return series is tested using Bera-Jarque test (p-value provided in the table). Sample size is 155 monthly observations from January 1987 to November 1999. All returns are measured in local currency.
Exhibit 3 describes the portfolio return series. As a comparison, we also provide similar statistics for the ISE National-100 price index. It measures the development of the largest 100 companies listed on the ISE. As in Harvey [1995], we find the Turkish companies to have high returns that average above 4.5% (54%, annualized). Similarly, all portfolio returns exhibit high volatility. Contrary to U.S. findings, the mean realized return is generally higher for the smaller companies.

Most asset returns also show evidence of non-normality. The Bera-Jarque test for normality rejects the null hypothesis of a normal distribution for four of the six portfolios. Surprisingly, there is no evidence of significant first-order serial correlation in the portfolio or market returns. Most of the first-order autocorrelation coefficients are below 0.10, which is similar to what Bekaert et al. [1998] find for Turkey (0.08) using the IFC’s market index and slightly different sample period.

Panel B of Exhibit 3 shows the cross-correlation matrix of portfolio returns. In general, portfolios exhibit high positive cross-correlations with each other, which is typical for size portfolios. Since the EMDB includes only the largest and most liquid companies, the cross-correlations are even higher than otherwise.

CONCLUSIONS

We have discussed the problems associated with portfolio construction in many small and emerging stock markets. As we have seen, researchers have a surprising number of choices when they are constructing portfolios, especially so in small or emerging stock markets. While I touch on many issues related to portfolio construction, one can see that a variety of questions remain unanswered. Most notably, how do the choices that researchers or practitioners make during portfolio construction affect their empirical results? There is clearly room for further research on these issues.

ENDNOTES

The author thanks Eva Liljeblom, Anders Löflund, and an anonymous referee for their helpful comments and suggestions, and Campbell R. Harvey for providing the emerging markets data for Turkey.

1For example, in many small markets more than half of the companies have at least two (listed) share classes. Only a few companies on the New York Stock Exchange have more than one class of shares (excluding preferred shares, which more resemble fixed-income securities).

2The Turkish stock market is a midsized emerging market. Its market capitalization was more than USD 60 billion at the end of 1999, making it the second-largest emerging market in Europe after Greece and before Russia. The Istanbul Stock Exchange was formally established at the end of 1985. For more information see http://www.ise.org/.

3Exceptions are research databases such as the CRSP database in the United States. In many cases, however, the data are collected directly from the stock exchange or data vendors such as Reuters or Datastream, and we have to calculate returns ourselves.

4In some cases, one may even have enough data to calculate an average (volume-weighted) transaction price using all transactions within a day.

5For more information on correcting for value events, see, e.g. “HEX” [2000].

6For more information on correcting for value events, see, e.g. “HEX” [2000].

7See Shumway and Warther [1999] for more information.

8In practice, weekly return calculations can be a little problematic, since a Wednesday (for example) can be a holiday. This problem can be overcome by constructing an indicator variable (in a spreadsheet) to show when the week changes. After this it is easy to calculate weekly returns.

9It is well-known that the results of the beta pricing and factor models are severely affected by the thin trading (see Dimson [1979]).

10Measurement problems are especially important in small or emerging markets, where the return series show various degrees of autocorrelation and heteroscedasticity, as one of the main ideas behind portfolio construction is to reduce these problems.

11Updating weights at the end of a year can highlight the January effect in returns, yet it requires less information than the alternative. In practice, we need a half a year longer time period to use middle-year weights, and if we have only a short sample, spending half a year’s worth of observations is not worth the benefits.

12See Harvey [1995] on survivorship bias in the emerging markets databases.

13This also reduces survival bias, because, (alternatively) we could select companies every year, and typically only companies that have survived through the year are selected in order to avoid the problems of calculating last-period returns.

14Common (or ordinary) stocks have premium voting rights; preference stocks have voting rights but less than common stocks (preferred stock in the U.S. has no voting rights). Preference shares, often have the first right to the dividend, although often up to a certain limit (e.g., 10% dividend on the nominal value). In some cases, preference shares always have higher dividend rights than common stocks.

Old and newly issued stocks are listed separately if they have different dividend rights (usually newly issued stocks are entitled to the dividend starting next period). On discussion is
applicable to some degree even to this kind of situation.

Restrictions can be sector restrictions, limits on single or aggregate foreign ownership, and limits on foreign investments. To decide whether these restrictions affect portfolio construction, we have to know whether we take the foreign or the domestic investor’s point of view. If these restrictions are binding, we may have to set limits on the company weights in our portfolios.

Prior to 1993, Finnish companies could have a maximum of four listed share series (ordinary and preference classes of stocks, both available as restricted and non-restricted series). Beginning in 1993, all stocks were declared unrestricted, and the maximum number of different series of stock came down to two (ordinary and preference). See Hietala [1989] for more information.

If the trading is so active that there are few days with no trading, a limit (>0) could be used instead. Alternatively, we could select the series with the highest volume accumulated over the whole sample period, although this alternative could pick up series that have been thinly traded most of the time but have several large trades. To mitigate this problem, we could estimate liquidity from less aggregated data (e.g., calculate years with superior liquidity and use that information), but the liquidity is still more volume-, not trading-, oriented.

In a corporate takeover, for example, the series with the voting premium (ordinary shares) is often more liquid, the preference stocks are often more liquid at other times.

In Finland, for example, all stocks of insurance companies have been converted to restricted stocks unless owned by foreigners at the time of the conversion.

Price observations and market values are available from December 1986.

Had there been multiple classes of stocks available, we would have needed the daily trading volumes (number of shares traded) to select among series. This is not the case with Turkey.

See Bekaert and Harvey’s chronology of important financial, economic, and political events in emerging markets for more detailed information on regulation in Turkey (http://www.duke.edu/~charvey/Country_risk/couindex.htm).

To be exact, the weights are calculated at the end of period t - 1, i.e. at the end of the previous month.

REFERENCES


