Performance Attribution for Equity Portfolios☆

Yang Lu[1,*]

SU 2896 Paresky Center, Williams College, Williamstown, MA 01267, United States

David Kane[2,*]

260 Franklin Street, Boston, MA 02109, United States

Abstract

The pa package provides tools for conducting performance attribution for equity portfolios. The package uses two methods: the Brinson-Hood-Beebower model (hereafter referred to as the Brinson model) and a regression-based analysis. The Brinson model takes an ANOVA-type approach and decomposes the active return of any portfolio into asset allocation, stock selection, and interaction effect. The regression-based analysis utilizes estimated coefficients, based on a regression model, to attribute active return to different factors. This paper focuses on attribution models for equity portfolios. After describing the Brinson and regression approaches and demonstrating their use via the pa package, we show that the Brinson model is just a special case of the regression approach.

Keywords: Attribution, Brinson, Regression, Portfolio Performance

1. Introduction

Many portfolio managers measure performance with reference to a benchmark. The difference in return between a portfolio and its benchmark is the

☆We thank colleagues at Acadian Asset Management, LLC, especially Michael Flynn, Matthew Lyberg, Michael Kraus, and Henry Philip, for their helpful comments.

*Corresponding author

Email addresses: yang.lu@williams.edu (Yang Lu), dkane@acadian-asset.com (David Kane)

[1]Williams College

[2]Acadian Asset Management, LLC and Harvard University

active return of the portfolio. Portfolio managers and their clients want to know what caused this active return. Performance attribution decomposes the active return. The two most common approaches are the Brinson-Hood-Beebower (hereafter referred to as the Brinson model) and a regression-based analysis.[3]

Portfolio managers use different variations of the two models to assess the performance of their portfolios. Managers of fixed income portfolios include yield-curve movements in the model [8] while equity managers who focus on the effect of currency movements use variations of the Brinson model to incorporate "local risk premium" [12]. In contrast, in this paper we focus on attribution models for equity portfolios without considering any currency effect.

The `pa` package provides tools for conducting both methods for equity portfolios. The Brinson model takes an ANOVA-type approach and decomposes the active return of any portfolio into asset allocation, stock selection, and interaction effects. The regression-based analysis utilizes estimated coefficients from a linear model to attribute active return to different factors. After describing the Brinson and regression approaches and demonstrating their use via the `pa` package, we show that the Brinson model is just a special case of the regression approach.

2. Data

We demonstrate the use of the `pa` package with a series of examples based on modified versions of the real-world data sets from MSCI Barra's Global Equity Model II (GEM2).[4] According to MSCI Barra:

> *GEM2 is the latest Barra global multi-factor equity model. It provides a foundation for investment decision support tools via a broad range of insightful analytics for developed and emerging market portfolios.*

The original data set contains selected attributes such as industry, size, country, and various style factors for a universe of approximately 48,000 securities on a monthly basis. The three modified versions are `year`, `quarter`, and `jan`, each of which contains 3000 securities. `quarter` is a subset of `year`,

[3] See [2] and [7] for more information.

[4] See www.msci.com and [11] for more information.

containing data from the first quarter. `jan` is a subset of `quarter` with data from January 2010.

```
> data(year)
> names(year)

 [1] "barrid"    "name"      "return"
 [4] "date"      "sector"    "momentum"
 [7] "value"     "size"      "growth"
[10] "cap.usd"   "yield"     "country"
[13] "currency"  "portfolio" "benchmark"
```

The following definitions, besides `portfolio` and `benchmark`, are from MSCI Barra. See [11] for more information.

- `barrid`: Barra security identifier.
- `name`: security name.
- `return`: monthly total return in security's trading currency.
- `date`: first day of the month.
- `sector`: consolidated sector categories based on the Global Industry Classification Standard (GICS).
- `momentum`: includes historical alpha from a 104-week regression performance and relative strength (over trailing six and twelve months) with a one-month lag.
- `value`: utilizes official MSCI data items for Value factor descriptors and describes a major investment style which seeks to identify stocks that are priced low relative to fundamentals. It is standardized on a country-relative basis.
- `size`: logarithm of the total market capitalization of the firm.
- `growth`: a combination of long-term predicted earnings growth, sales growth (trailing five years), and earnings growth (trailing five years).
- `cap.usd`: market capitalization in USD.

- `yield`: annualized dividend-per-share divided by the current price.
- `country`: country in which the security is traded.
- `currency`: currency in which a security is traded.
- `portfolio`: we define the portfolio holdings as the top 200 securities based on `value` scores in January 2010. They are held through December 2010. This is to avoid the complexity of trading in the analyses. The portfolio column refers the weight of a security in the portfolio. Securities that are not in the portfolio have portfolio weights of 0; each of the portfolio holdings has a portfolio weight of 0.5%.
- `benchmark`: we define the benchmark as the top 1000 securities based on `size` each month. The benchmark weight of each security is proportional to its `cap.usd`.

Here is a sample of rows and columns from the data frame `year`:

```
      date                         name country
2010-01-01 BLUE STAR OPPORTUNITIES CORP     USA
2010-01-01 SEADRILL                         NOR
2010-05-01 BUXLY PAINTS (PKR10)             PAK
2010-08-01 CDN IMPERIAL BK OF COMMERCE      CAN
2010-11-01 CDN IMPERIAL BK OF COMMERCE      CAN
    sector  size   return portfolio benchmark
    Energy  0.00  0.00000     0.000  0.000000
    Energy -0.27 -0.07905     0.000  0.000427
 Materials  0.00 -0.01754     0.005  0.000000
Financials  0.52  0.02613     0.005  0.000012
Financials  0.55 -0.00079     0.005  0.000012
```

The portfolio has 200 equal-weighted holdings. Canadian Imperial Bank of Commerce is one of the 200 portfolio holdings with a weight of 0.5% in 2010. Its return was 2.61% in August and almost flat in November.

3. The Brinson Model

3.1. Single-Period Brinson Model

Consider an equity portfolio manager who uses the S&P 500 as the benchmark. In a given month, she outperforms the S&P by 3%. The outperformance has three components - *allocation effect*, *selection effect*, and *interaction effect*. Allocation effect is caused by her allocating more of the portfolio to sectors that perform well. Selection effect is attributed to some of the stocks in the portfolio doing better than their sector as a whole. The difference between the active return and the sum of the allocation and selection effects is the residual, which is also called the interaction effect between allocation and selection effects. The Brinson model provides mathematical definitions for these terms and methods for calculating them.

The example above uses sector as the classification scheme. However, the same approach can work with any other variable which places each security into one, and only one, discrete category: country, industry, and so on. In fact, a similar approach can work with continuous variables that are split into discrete ranges, e.g., the first tercile, the second tercile, and the last tercile. For generality, we will use the term "category" to describe any classification scheme which places each security in one, and only one, category.

Below we outline the Brinson model based on works done by Brinson et. al. (1986, 1991).[5]

Notations:

- n is the number of securities in the benchmark.
- N is the number of categories in the benchmark.
- w_i^B is the weight of security i in the benchmark.
- w_i^P is the weight of security i in the portfolio.
- W_j^B is the weight of category j in the benchmark. $W_j^B = \sum\limits_{i \in j} w_i^B$.
- W_j^P is the weight of a category j in the portfolio. $W_j^P = \sum\limits_{i \in j} w_i^P$.
- R_j^B is the weighted sum of the returns of all stocks in a category j in the benchmark. $R_j^B = \sum\limits_{i \in j} w_i^B r_i$.
- R_j^P is the weighted sum of the returns of all stocks in a category j in the portfolio. $R_j^P = \sum\limits_{i \in j} w_i^P r_i$.

[5]See [2] and [3] for more informaiton.

- $\sum_{i=1}^{n} w_i^B = \sum_{i=1}^{n} w_i^P = \sum_{j=1}^{N} W_j^B = \sum_{j=1}^{N} W_j^P = 1.$

Lower case w and r refer to the weight and the return of an individual security. Upper case W and R refer to those of a collection of stocks within the same category value of a category.

The return of a portfolio, R_P, can be calculated in two ways:

- On an individual security level by summing over n stocks: $R_P = \sum_{i=1}^{n} w_i^P r_i$.

- On a category level by summing over N categories: $R_P = \sum_{j=1}^{N} W_j^P R_j^P$.

Similar definitions apply to the return of the benchmark, R_B,

- $R_B = \sum_{i=1}^{n} w_i^B r_i$.

- $R_B = \sum_{j=1}^{N} W_j^B R_j^B$.

The arithmetic active return of a portfolio, R_{active}, is the portfolio return R_P less the benchmark return R_B:

$$R_{active} = R_P - R_B.$$

One can also define the active return by taking geometric average returns. The `pa` package implements the arithmetic measure of the active return for a single-period Brinson model as it is more intuitive in a single period.

Since the allocation of the portfolio based on a category (e.g., sector) is generally different from that of the benchmark, allocation plays a role in the active return, R_{active}. The same applies to stock selection where assuming that the portfolio has the exact same categorical exposures as the benchmark does, equities within each category are different. This contributes to R_{active} as well.

Allocation effect, $R_{allocation}$, over N categories is defined as:

$$\begin{aligned} R_{allocation} &= \sum_{j=1}^{N} W_j^P R_j^B - \sum_{j=1}^{N} W_j^B R_j^B \\ &= \sum_{j=1}^{N} (W_j^P - W_j^B) R_j^B, \end{aligned}$$

The intuition is that a portfolio would produce different returns with different allocation schemes (W_j^P vs. W_j^B) while having the same stock selection and thus the same return (R_j^B) for each category. The difference between the two returns is caused by the allocation scheme.

Selection effect, $R_{selection}$, over N categories is defined as:

$$\begin{aligned} R_{selection} &= \sum_{j=1}^{N} W_j^B R_j^P - \sum_{j=1}^{N} W_j^B R_j^B \\ &= \sum_{j=1}^{N} (R_j^P - R_j^B) W_j^B, \end{aligned}$$

Two different returns can be produced when two portfolios have the same allocation (W_j^B) yet dissimilar returns due to differences in stock selection within each category (R_j^p vs. R_j^B).

Interaction effect, $R_{interaction}$, is

$$R_{interaction} = \sum_{j=1}^{N} (W_j^P - W_j^B)(R_j^P - R_j^B).$$

It can be thought of as the difference in return caused by the interaction between the allocation scheme and the stock selection within each category. Alternatively, $R_{interaction}$ is the result of subtracting return due to allocation, $R_{allocation}$, and return due to selection, $R_{selection}$, from the active return, R_{active}:

$$R_{interaction} = R_{active} - R_{allocation} - R_{selection}.$$

3.2. Weakness of the Brinson Model

The Brinson model allows portfolio managers to analyze the relative return of a portfolio using any attribute of a security, such as country or sector.

One weakness of the model is to expand the analysis beyond two categories.[6] As the number of categories increases, this procedure is subject to the *curse of dimensionality*.

Suppose an equity portfolio manager wants to find out the contributions of any two categories (for instance, country and sector) to her portfolio based on the Brinson model. She can decompose the active return into three broad terms – $R_{allocation}$, $R_{selection}$, and $R_{interaction}$. The allocation effect can be further split into country allocation effect, sector allocation effect and the product of country and sector allocation effects:

$$R_{allocation} = R_{country\,allocation} + R_{sector\,allocation} + R_{country\,allocation} R_{sector\,allocation}.$$

Specifically, the country allocation effect is the return caused by the difference between the actual country allocation and the benchmark country allocation while assuming the same benchmark return within each level of the category country, that is,

$$R_{country\,allocation} = \sum_{j=1}^{N} {}_{C}W_j^P \, {}_{C}R_j^B - \sum_{j=1}^{N} {}_{C}W_j^B \, {}_{C}R_j^B,$$

where

- ${}_{C}W_j^P$ and ${}_{C}W_j^B$ refer to the weight of each country j (N_C countries in total) in the portfolio and that in the benchmark, respectively.
- ${}_{C}R_j^B$ refers to the benchmark return of any country j.

Similarly, the sector allocation effect is the difference in return between a portfolio's sector allocation and the benchmark's sector allocation while having the same benchmark returns:

$$R_{sector\,allocation} = \sum_{j=1}^{N} {}_{S}W_j^P \, {}_{S}R_j^B - \sum_{j=1}^{N} {}_{S}W_j^B \, {}_{S}R_j^B,$$

${}_{S}W_j^P$ and ${}_{S}W_j^B$ refer to the weight of the sector j in the portfolio and the weight of the sector j in the benchmark, respectively. ${}_{S}R_j^B$ is the benchmark return of any given sector j of all N_S sectors.

[6] Brinson et al. [3] proposed a framework to include two variables in the Brinson analysis.

In the same vein, the return as a result of the selection effect $R_{selection}$ is the sum of country selection effect, sector selection effect, and the product of country and sector selection effects:

$$\begin{aligned}
R_{selection} &= R_{country\,selection} + R_{sector\,selection} + R_{country\,selection} * R_{sector\,selection} \\
&= \sum_{j=1}^{N} {}_CW_j^B\,{}_CR_j^P - \sum_{j=1}^{N} {}_CW_j^B\,{}_CR_j^B \\
&+ \sum_{j=1}^{N} {}_SW_j^B\,{}_SR_j^P - \sum_{j=1}^{N} {}_SW_j^B\,{}_SR_j^B \\
&+ (\sum_{j=1}^{N} {}_CW_j^B\,{}_CR_j^P - \sum_{j=1}^{N} {}_CW_j^B\,{}_CR_j^B) \\
&* (\sum_{j=1}^{N} {}_SW_j^B\,{}_SR_j^P - \sum_{j=1}^{N} {}_SW_j^B\,{}_SR_j^B).
\end{aligned}$$

The interaction effect, $R_{interaction}$, includes the interaction between country allocation and sector selection and that between country selection and sector allocation.

Therefore, in the case of Q categories where $Q > 1$, the Brinson model becomes very complex (assume $Q \geq 3$):

$$\begin{aligned}
R_{allocation} &= \sum_{j=1}^{Q} R_{allocation_j} + \sum_{j=1}^{Q}\sum_{k=1}^{Q} R_{allocation_j} R_{allocation_k} \\
&+ \sum_{j=1}^{Q}\sum_{k=1}^{Q}\sum_{p=1}^{Q} R_{allocation_j} R_{allocation_k} R_{allocation_p} \\
&= \ldots,
\end{aligned}$$

$$\begin{aligned}
R_{selection} &= \sum_{j=1}^{Q} R_{selection_j} + \sum_{j=1}^{Q}\sum_{k=1}^{Q} R_{selection_j} R_{selection_k} \\
&+ \sum_{j=1}^{Q}\sum_{k=1}^{Q}\sum_{p=1}^{Q} R_{selection_j} R_{selection_k} R_{selection_p} \\
&= \ldots,
\end{aligned}$$

where $R_{allocation_j}$ is the allocation effect of any given category j, $j \in Q$ and $R_{selection_j}$ is the selection effect of any given category j, $j \in Q$. i, j, k have different values.

As the number of categories grows, the numbers of terms for the allocation and the selection effects grow exponentially. Q categories will result in $2^Q - 1$ terms for each of the allocation and selection effect.

Due to the interaction between allocation and selection of each of the Q categories (it could be interaction between 2, 3 or even all Q categories), the number of terms included in the interaction effect grows exponentially to take into all the interaction effects among all categories.

$$
\begin{aligned}
R_{interaction} &= \sum_{j=1}^{Q}\sum_{k=1}^{Q} R_{allocation_j} R_{selection_k} \\
&+ \sum_{j=1}^{Q}\sum_{k=1}^{Q}\sum_{p=1}^{Q} R_{allocation_j} R_{selection_k} R_{allocation_p} \\
&+ \ldots.
\end{aligned}
$$

Q categories has $2^{2n} - 2^{n+1} + 1$ terms of interaction effects.

For instance, when there are 3 categories, the allocation effect and the selection effect each have $2^3 - 1 = 7$ terms. The interaction effect has $2^6 - 2^4 + 1 = 49$ terms. When there are 4 categories, $2^4 - 1 = 15$ terms have to be estimated for the allocation effect as well as the selection effect, respectively. $2^8 - 2^5 + 1 = 225$ terms have to be calculated for the interaction effect of 4 categories. This poses a significant computational challenge when a portfolio manager performs a multivariate Brinson analysis.

To some extent, the regression-based model detailed later solves the problem of multivariate attribution.

3.3. Single-Period Brinson Tools

Below shows the implementation of the Brinson model to analyze the return of a single-period portfolio via the `pa` package.

Brinson analysis is run by calling the function `brinson` to produce an object of class `brinson`.

```
> data(jan)
> br.single <- brinson(x = jan, date.var = "date",
+                      cat.var = "sector",
```

```
+                  bench.weight = "benchmark",
+                  portfolio.weight = "portfolio",
+                  ret.var = "return")
>
```

The data frame, `jan`, contains all the information necessary to conduct a single-period Brinson analysis. `date.var`, `cat.var`, and `return` identify the names of the columns containing the date, the factor to be analyzed, and the return variable, respectively. `bench.weight` and `portfolio.weight` specify the name of the benchmark weight column and that of the portfolio weight column in the data frame.

Calling `summary` on the resulting object reports essential information about the input portfolio (including the number of securities in the portfolio and the benchmark as well as sector exposures) and the results of the Brinson analysis.

```
> summary(br.single)

Period:                              2010-01-01
Methodology:                         Brinson
Securities in the portfolio:         200
Securities in the benchmark:         1000

Exposures
            Portfolio Benchmark
Energy          0.085    0.2782
Materials       0.070    0.0277
Industrials     0.045    0.0330
ConDiscre       0.050    0.0188
ConStaples      0.030    0.0148
HealthCare      0.015    0.0608
Financials      0.370    0.2979
InfoTech        0.005    0.0129
TeleSvcs        0.300    0.1921
Utilities       0.030    0.0640

Returns
                   2010-01-01
Allocation Effect    -0.00140
Selection Effect      0.01418
```

```
Interaction Effect    0.00191
Active Return         0.01469
```

The `br.single` summary shows that the active return of the portfolio in January 2010 was 1.47%. This return can be decomposed into allocation effect (-0.14%), selection effect (1.42%), and interaction effect (0.19%).

Figure 1 plots the absolute return of the portfolio and the benchmark sector by sector in January 2010. Utilities performed the best with a gain of more than 5% and Consumer Discretionary, the worst performing sector, lost more than 10% in the portfolio. Utilities was also the sector with the highest active return in the portfolio. Further, almost all active return of the portfolio was due to the selection effect.

3.4. Multi-Period Brinson Model

To obtain Brinson attribution on a multi-period data set, one calculates allocation, selection and interaction within each period and aggregates them across time. There are five methods for this – arithmetic, geometric, optimized linking by Menchero [10], linking by Davies and Laker [4], and linking by Frongello [6]. We focus on the first three methods in this paper. Arithmetic measure calculates relative performance of a portfolio and its benchmark by a difference; geometric measure does so by a ratio. Arithmetic measure is more intuitive but a well-known challenge in arithmetic attribution is that active returns do not add up over multiple periods due to geometric compounding.[7] Geometric is able to circumvent the adding-up problem. Menchero [10] discusses various linking algorithms to connect arithmetic return with geometric return and argues that the *optimized linking algorithm* is the best way to link attribution over time.

`Arithmetic Attribution.` The arithmetic attribution model calculates active return and contributions due to allocation, selection, and interaction in each period and sums them over multiple periods. The arithmetic active return over T periods $R_{arithmetic}$ is expressed as:

$$R_{arithmetic} = \sum_{t=1}^{T} R_t^{active},$$

and R_t^{active} is the active return in a single period t.

[7] See [1] for a complete discussion of the complexity involved.

```
> plot(br.single, var = "sector", type = "return")
```

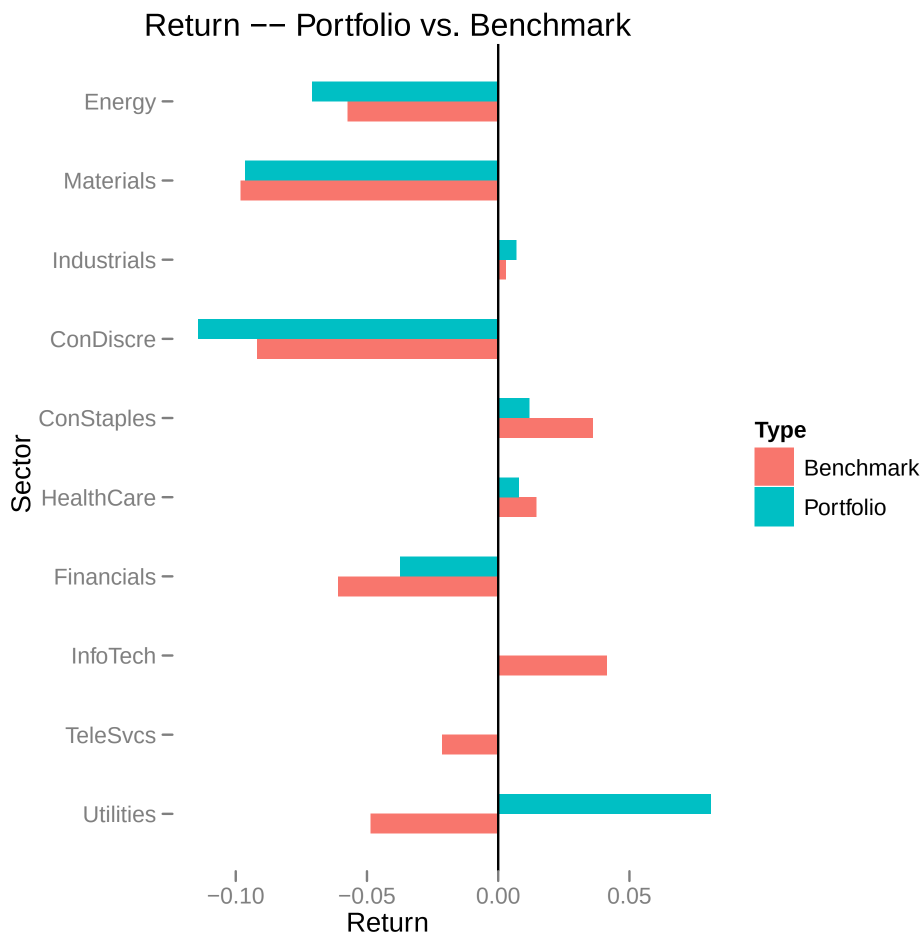

Figure 1: This shows the absolute return of the portfolio and the benchmark sector by sector in January 2010. In the portfolio, Utilities performed the best with a gain of more than 5% and Consumer Discretionary, the worst performing sector, lost more than 10%. Utilities was also the sector with the highest active return in the portfolio. Almost all active return of the portfolio was due to the selection effect.

`Geometric Attribution.` The geometric attribution is to compound various returns over T periods where,

$$1 + R_P = \prod_{t=1}^{T} (1 + R_t^P),$$

$$1 + R_B = \prod_{t=1}^{T} (1 + R_t^B),$$

and R_t^P and R_t^B are portfolio and benchmark returns in a single period t, respectively. Geometric active return, $R_{geometric}$, is thus the difference between R_P and R_B:

$$R_{geometric} = R_P - R_B.$$

`Optimized Linking Algorithm.` The well-known challenge faced in arithmetic attribution is that the actual active return over time is not equal to the arithmetic summation of single-period active returns,

$$R_{geometric} \neq R_{arithmetic},$$

i.e.,

$$R_P - R_B \neq \sum_{t=1}^{T} R_t^{active}.$$

Menchero [10] proposed an optimized linking coefficient b^{opt} to link arithmetic returns of individual periods with geometric returns over time,

$$R_P - R_B = \sum_{t=1}^{T} b_t^{opt} R_t^{active},$$

where b_t^{opt} is the optimized linking coefficient in a single period t.

The optimized linking coefficient b_t^{opt} is the summation of a *natural scaling* A and an *adjustment* a_t specific to a time period t,

$$b_t^{opt} = A + a_t,$$

where A is an coefficient for linking from the single-period to the multi-period return and a_t is an adjustment to eliminate residuals.[8]

Since active return over time, R_{active} where $R_{active} = R_P - R_B$, is a summation of active return in each period adjusted to the optimized linking algorithm, the following is true:

[8]See [9] for more information on the optimized linking coefficients.

$$R_{active} = \sum_{t=1}^{T} b_t^{opt}(R_t^{allocation} + R_t^{selection} + R_t^{interaction}),$$

where $R_t^{allocation}$, $R_t^{selection}$, and $R_t^{interaction}$ represent allocation, selection and interaction in each period t, respectively.

Within each period t, the adjusted attribution is thus expressed as

$$\hat{R}_t^{allocation} = b_t^{opt} R_t^{allocation},$$

$$\hat{R}_t^{selection} = b_t^{opt} R_t^{selection},$$

and

$$\hat{R}_t^{interaction} = b_t^{opt} R_t^{interaction}.$$

Therefore, across T periods, active return, R_{active}, can be written as

$$R_{active} = \sum_{t=1}^{T}(\hat{R}_t^{allocation} + \hat{R}_t^{selection} + \hat{R}_t^{interaction}).$$

3.5. Multi-Period Brinson Tools

In practice, analyzing a single-period portfolio is meaningless as portfolio managers and their clients are more interested in the performance of a portfolio over multiple periods. To apply the Brinson model over time, we can use the function `brinson` and input a multi-period data set (for instance, `quarter`) as shown below.

```
> data(quarter)
> br.multi <- brinson(quarter, date.var = "date",
+                     cat.var = "sector",
+                     bench.weight = "benchmark",
+                     portfolio.weight = "portfolio",
+                     ret.var = "return")
```

The object `br.multi` of class `brinsonMulti` is an example of a multi-period Brinson analysis.

```
> exposure(br.multi, var = "size")
```

```
$Portfolio
     2010-01-01 2010-02-01 2010-03-01
Low       0.140      0.140      0.155
2         0.050      0.070      0.045
3         0.175      0.145      0.155
4         0.235      0.245      0.240
High      0.400      0.400      0.405

$Benchmark
     2010-01-01 2010-02-01 2010-03-01
Low      0.0681     0.0568     0.0628
2        0.0122     0.0225     0.0170
3        0.1260     0.1375     0.1140
4        0.2520     0.2457     0.2506
High     0.5417     0.5374     0.5557
```

The `exposure` method on `br.multi` shows the exposure of the portfolio and the benchmark based on a user-defined category. Here, it shows the exposure on `size`. The portfolio overweights the benchmark in the lowest quintile in `size` and underweights in the highest quintile.

```
> returns(br.multi, type = "linking")

$Raw
              2010-01-01 2010-02-01 2010-03-01
Allocation       -0.0014     0.0064     0.0046
Selection         0.0146     0.0178    -0.0152
Interaction       0.0020    -0.0074    -0.0087
Active Return     0.0151     0.0168    -0.0193

$Aggregate
              2010-01-01, 2010-03-01
Allocation                    0.0095
Selection                     0.0173
Interaction                  -0.0142
Active Return                 0.0127
```

The `returns` method shows the results of the Brinson analysis applied to the data from January 2010 through March 2010. The optimized linking algorithm is applied here by setting the type to *linking*. The first portion of the `returns` output shows the Brinson attribution in individual periods. The second portion shows the aggregate attribution results. The portfolio formed by top 200 value securities in January had an active return of 12.7% over the first quarter of 2010. The allocation and the selection effects contributed 0.95% and 1.73%, respectively; the interaction effect made a loss of 1.42%.

Figure 2 plots the returns of both the portfolio and the benchmark of the allocation effect from January 2010 through March 2010. Utilities performed the best with an absolute gain of more than 5% in January and February 2010 but did not performa well in March 2010.

4. Regression

4.1. Single-Period Regression Model

One advantage of a regression-based approach is that such analysis allows one to define their own attribution model by easily incorporating multiple variables in the regression formula. These variables can be either discrete or continuous.

Suppose a portfolio manager wants to find out how much each of the value, growth, and momentum scores of her holdings contributes to the overall performance of the portfolio. In his paper, Grinold [7] outlines how a regression analysis is conducted conventionally.

Consider the following linear regression without the intercept term based on a single-period portfolio of n securities with k different variables:

$$\mathbf{r}_n = \mathbf{X}_{n,k}\mathbf{f}_k + \mathbf{u}_n$$

where

- $\mathbf{r}_n$ is a column vector of length n. Each element in $\mathbf{r}_n$ represents the return of a security in the portfolio.

- $\mathbf{X}_{n,k}$ is an n by k matrix. Each row represents k attributes of a security. There are n securities in the portfolio.

- $\mathbf{f}_k$ is a column vector of length k. The elements are the estimated coefficients from the regression. Each element represents the *factor return* of an attribute.

- $\mathbf{u}_n$ is a column vector of length n with residuals from the regression.

```
> plot(br.multi, type = "return")
```

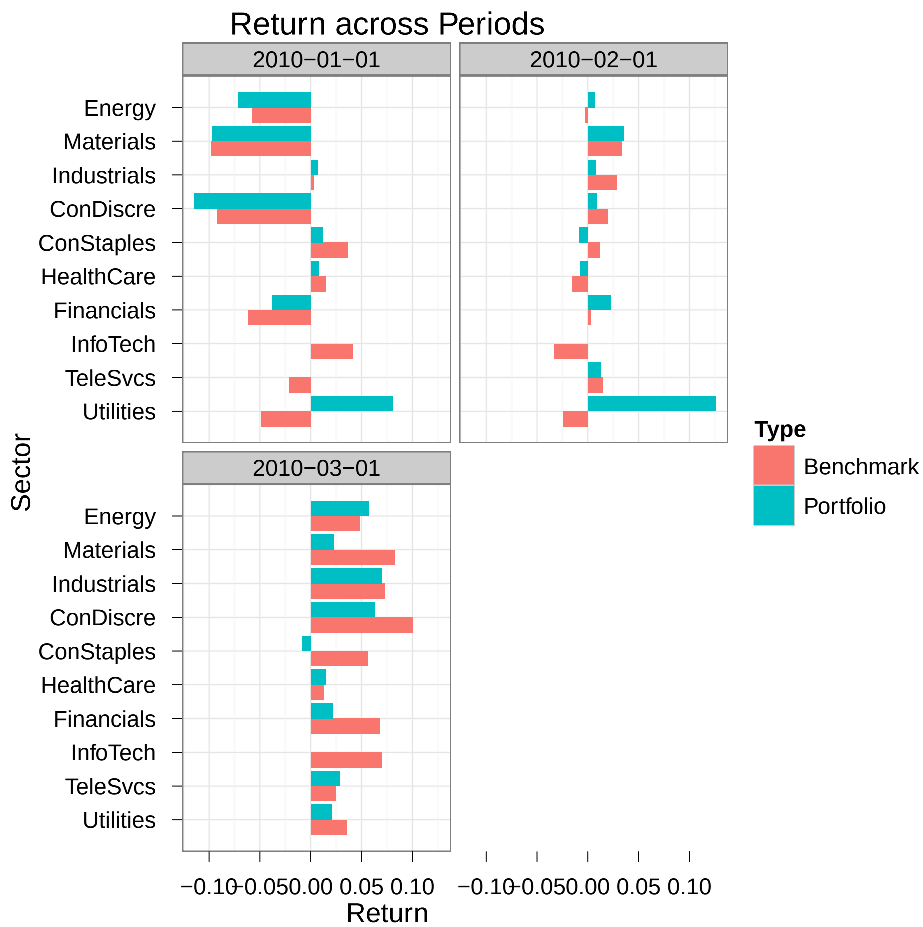

Figure 2: The figure plots the returns of both the portfolio and the benchmark of the allocation effect from January 2010 through March 2010. Utilities performed the best with an absolute gain of more than 5% in January and February 2010 but did not perform well in March 2010.

In the case of this portfolio manager, suppose that she only has three holdings in her portfolio. r_3 is thus a 3 by 1 matrix with returns of all her three holdings. The matrix $\mathbf{X}_{3,3}$ records the score for each of the three factors (value, growth, and momentum) in each row. $\mathbf{f}_3$ contains the estimated coefficients of a regression $\mathbf{r}_3$ on $\mathbf{X}_{3,3}$.

The active exposure of each of the k variables, X_i, $i \in k$, is expressed as

$$X_i = \mathbf{w}_{active}\prime \mathbf{x}_{n,i},$$

where X_i is a constant representing the active exposure of the attribute i in the portfolio, $\mathbf{w}_{active}$ is a column vector of length n containing the active weight of every security in the portfolio, and $\mathbf{x}_{n,i}$ is a column vector of length n with attribute i for all securities in the portfolio. Active weight of a security is defined as the difference between the portfolio weight of the security and its benchmark weight.

Using the example mentioned above, the active exposure of the attribute `value`, X_{value} is the product of $\mathbf{w}_{active}\prime$ (containing active weight of each of the three holdings) and $\mathbf{x}_3$ (containing value scores of the three holdings).

The contribution of a variable i, R_i, is thus the product of the factor returns for the variable i, f_i and the active exposure of the variable i, X_i. That is,

$$R_i = f_i X_i.$$

Continuing the example, the contribution of value is the product of f_{value} (the estimated coefficient for value from the linear regression) and X_{value} (the active exposure of value as shown above).

Therefore, the active return of the portfolio R_{active} is the sum of contributions of all k variables and the residual u.

$$R_{active} = \sum_{i=1}^{k} R_i + u.$$

For instance, a hypothetical portfolio has three holdings (A, B, and C), each of which has two attributes – size and value.

```
  Return Name Size Value Active_Weight
1    0.3    A  1.2   3.0           0.5
2    0.4    B  2.0   2.0           0.1
3    0.5    C  0.8   1.5          -0.6
```

Following the procedure as mentioned, the factor returns for size and value are -0.0313 and -0.1250. The active exposure of size is 0.32 and that of value is 0.80. The active return of the portfolio is -11% which can be decomposed into the contribution of size and that of value based on the regression model. Size contributes -1% of the active return of the portfolio and value causes the portfolio to lose the other 10.0%.

4.2. Single-Period Regression Tools

The pa package provides tools to analyze both single-period and multi-period portfolios based on a regression model.

```
> rb.single <- regress(jan, date.var = "date",
+                  ret.var = "return",
+                  reg.var = c("sector", "growth", "size"),
+                  benchmark.weight = "benchmark",
+                  portfolio.weight = "portfolio")
> exposure(rb.single, var = "growth")
```

```
     Portfolio Benchmark
Low      0.305    0.2032
2        0.395    0.4225
3        0.095    0.1297
4        0.075    0.1664
High     0.130    0.0783
```

`reg.var` specifies the columns containing variables whose contributions are to be analyzed. Calling `exposure` with a specified `var` yields information on the exposure of both the portfolio and the benchmark by that variable. If `var` is a continuous variable, for instance, `growth`, the exposure will be shown in 5 quantiles. Majority of the high `value` securities in the portfolio in January have relatively low `growth` scores.

```
> summary(rb.single)
```

```
Period:                              2010-01-01
Methodology:                         Regression
Securities in the portfolio:         200
Securities in the benchmark:         1000

Returns
```

```
                  2010-01-01
Sector              0.003189
Growth              0.000504
Size                0.002905
Residual            0.008092
Portfolio Return   -0.029064
Benchmark Return   -0.043753
Active Return       0.014689
```

The `summary` method shows the number of securities in the portfolio and the benchmark, and the contribution of each input variable according to the regression-based analysis. In this case, the portfolio lost 2.91% and the benchmark lost 4.38%. Therefore, the portfolio outperformed the benchmark by 1.47%. `Sector`, `growth`, and `size` contributed 0.32%, 0.05%, and 0.29%, respectively.

4.3. Multi-Period Regression Model

The same challenge of linking arithmetic and geometric returns is present in multi-period regression model. We apply the optimized linking algorithm proposed by Menchero [9] in the regression attribution.

Within each period t,

$$R_t^{active} = \sum_{i=1}^{k} R_{i,t} + u_t,$$

where $R_{i,t}$ represents the contribution of a variable i of the time period t and u_t is the residual in that period.

Across T periods, the active return can be expressed by a product of the optimized linking coefficient b_t^{opt} and the individual contribution of each of the k attributes. The adjusted contribution of each of the k variables i, $\hat{R}_{i,t}$, is expressed by

$$\hat{R}_{i,t} = b_t^{opt} R_{i,t}.$$

Thus, the overall active return R_{active} can be decomposed into

$$R_{active} = \sum_{t=1}^{T} \sum_{i=1}^{k} \hat{R}_{i,t} + U,$$

where U is the residual across T periods.

4.4. Multi-Period Regression Tools

We can also conduct a multi-period regression analysis via the `pa` package.

```
> rb.multi <- regress(quarter, date.var = "date",
+                  ret.var = "return",
+                  reg.var = c("sector", "growth",
+                    "size"),
+                  benchmark.weight = "benchmark",
+                  portfolio.weight = "portfolio")
> rb.multi

Period starts:                    2010-01-01
Period ends:                      2010-03-01
Methodology:                      Regression
Securities in the portfolio:      200
Securities in the benchmark:      1000
```

Regression-based analysis can be applied to a multi-period data frame by calling the same method `regress`. By typing the name of the class object `rb.multi` directly, a short summary of the analysis is provided, showing the starting and ending period of the analysis, the methodology, and the average number of securities in both the portfolio and the benchmark.

```
> summary(rb.multi)

Period starts:                    2010-01-01
Period ends:                      2010-03-01
Methodology:                      Regression
Avg securities in the portfolio:  200
Avg securities in the benchmark:  1000

Returns
$Raw
                 2010-01-01 2010-02-01 2010-03-01
Sector               0.0032     0.0031     0.0002
Growth               0.0005     0.0009    -0.0001
Size                 0.0029     0.0295     0.0105
Residual             0.0081    -0.0172    -0.0302
Portfolio Return    -0.0291     0.0192     0.0298
Benchmark Return    -0.0438     0.0029     0.0494
```

```
Active Return        0.0147     0.0163    -0.0196

$Aggregate
                 2010-01-01, 2010-03-01
Sector                           0.0065
Growth                           0.0013
Size                             0.0433
Residual                        -0.0392
Portfolio Return                 0.0190
Benchmark Return                 0.0064
Active Return                    0.0127
```

The regression-based summary shows that the contribution of each input variable in addition to the basic information about the portfolio. The summary suggests that the active return of the portfolio in 2010 is 1.27%. The `Residual` number indicates the contribution of the interaction among various variables including `sector`, `growth`, and `growth`.

Visual representation of relative performance of a portfolio against its benchmark is best viewed across a longer time span. We use the data frame `year` for illustrative purposes.

```
> rb.multi2 <- regress(year, date.var = "date",
+                ret.var = "return",
+                reg.var = c("sector", "growth",
+                  "size"),
+                benchmark.weight = "benchmark",
+                portfolio.weight = "portfolio")
> returns(rb.multi2, type = "linking")

$Raw
              2010-01-01 2010-02-01 2010-03-01
Sector            0.0035     0.0034     0.0002
Growth            0.0005     0.0010    -0.0001
Size              0.0031     0.0320     0.0109
Residual          0.0088    -0.0187    -0.0312
Active Return     0.0159     0.0177    -0.0203
              2010-04-01 2010-05-01 2010-06-01
Sector            0.0017     0.0044     0.0077
Growth            0.0001     0.0002     0.0004
Size              0.0145     0.0041     0.0020
```

```
Residual          -0.0043     0.0346     0.0201
Active Return      0.0122     0.0433     0.0304
              2010-07-01 2010-08-01 2010-09-01
Sector             0.0016     0.0051    -0.0023
Growth            -0.0005     0.0005    -0.0006
Size               0.0066     0.0000     0.0100
Residual          -0.0333     0.0189    -0.0229
Active Return     -0.0256     0.0246    -0.0158
              2010-10-01 2010-11-01 2010-12-01
Sector             0.0016    -0.0048    -0.0084
Growth            -0.0011    -0.0004     0.0010
Size               0.0024     0.0143     0.0057
Residual           0.0149     0.0192    -0.0253
Active Return      0.0179     0.0282    -0.0270

$Aggregate
              2010-01-01, 2010-12-01
Sector                        0.0137
Growth                        0.0011
Size                          0.1056
Residual                     -0.0193
Active Return                 0.1014
```

We obtained an object *rb.multi2* of class *regressMulti* based on the data set from January 2010 through December 2010. The portfolio outperformed the benchmark by 10.1% over this period. Based on the regression model, `size` contributed to the majority of the active return.

Figure 3 displays both the cumulative portfolio and benchmark returns from January 2010 through December 2010. The portfolio, which consisted of high `value` securities in January, consistently outperformed the benchmark in 2010. Outperformance in May and June helped the overall positive active return in 2010 to a large extent.

5. Brinson as Regression

Another way to think about the analysis as Brinson et al. [2] have done is to consider it in the context of a regression model. Conducting a Brinson attribution is similar to running a linear regression without the intercept term. Estimated coefficients will then be the mean return of each level of

```
> plot(rb.multi2, var = "sector", type = "return")
```

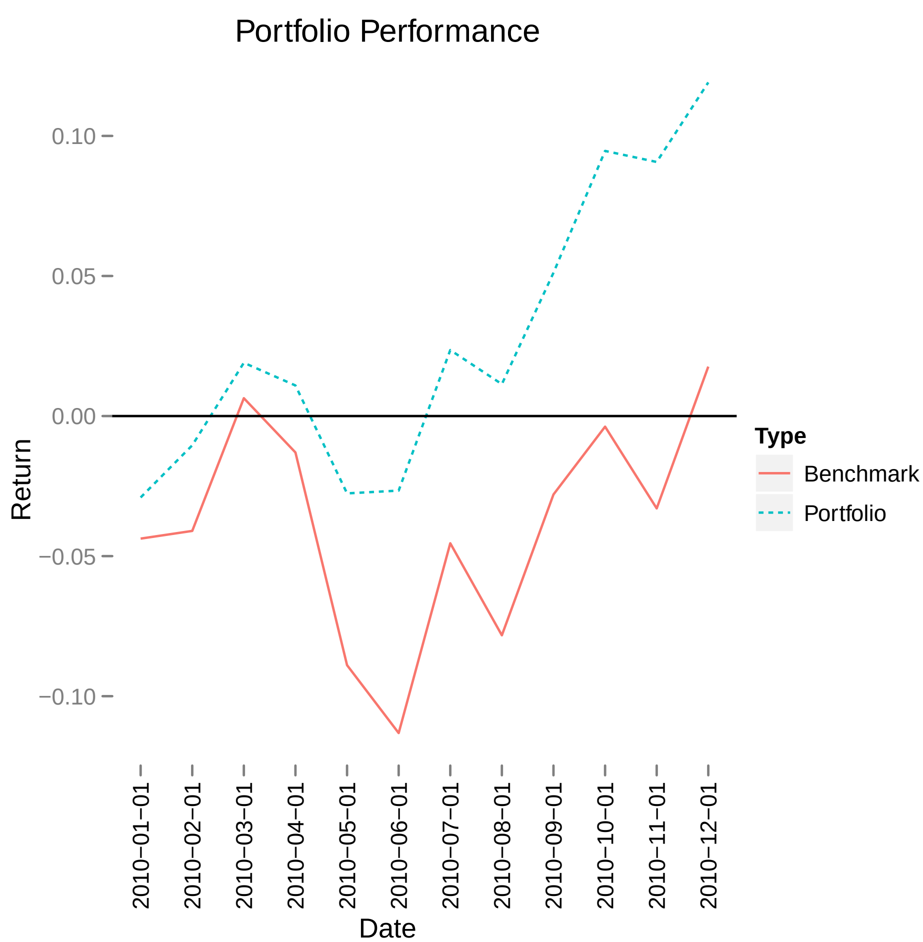

Figure 3: The plot displays both the cumulative portfolio and benchmark returns from January 2010 through December 2010. The portfolio, which consisted of high `value` securities in January, consistently outperformed the benchmark in 2010. Outperformance in May and June helped the overall positive active return in 2010 to a large extent.

the category specified in the universe, a.k.a. the factor return of each level. The mean return of each level also appears in the Brinson analysis. The equivalent to the allocation effect for the universe in the Brinson model is the sum of the product of the estimated coefficient and the active weight of each level of the category.

Applying a regression framework to the allocation effect from the Brinson model,

$$\begin{aligned} R_{allocation} &= \sum_{j=1}^{N} W_j^P R_j^B - \sum_{j=1}^{N} W_j^B R_j^B \\ &= (\mathbf{W}^P - \mathbf{W}^B)\prime\mathbf{f}, \end{aligned}$$

where $\mathbf{W}^P$ is a column vector indicating the portfolio weight of each level within the category specified by the manager; $\mathbf{W}^B$, a column vector indicating the benchmark weight of each level, and $\mathbf{f}$ is the column vector which has benchmark return of all the levels of the category. Assuming that in this case, the benchmark is the universe and the portfolio holdings are all from the benchmark, $\mathbf{R^B}$ can be estimated by regressing returns on the levels of the category specified by the portfolio manager:

$$\mathbf{r}_n = \mathbf{X}_{n,p}\mathbf{f} + \mathbf{U},$$

where

- $\mathbf{r}_n$ is a column vector of length n. Each element in $\mathbf{r}_n$ represents the return of a security in the portfolio.

- $\mathbf{X}_{n,p}$ is an n by p matrix where n refers to the number of securities in the portfolio and p refers to the number of levels within the category specified.

- $\mathbf{f}$ is the estimated coefficients on the regression without the intercept term. The estimated coefficients are the mean return for each of the levels in the category.

- $\mathbf{U}$ is the column vector with all the residual terms.

Since $\mathbf{R}^B$ is the same as $\mathbf{f}$, the allocation effect in the Brinson model is a special case of the regression approach.

In order to estimate the selection effect in the Brinson model, one can calculate the mean return of each level within the category specified in both the portfolio and the benchmark under a regression framework and use the benchmark weights to calculate the selection effect.

$$
\begin{aligned}
R_{selection} &= \sum_{j=1}^{N} W_j^B R_j^P - \sum_{j=1}^{N} W_j^B R_j^B \\
&= \mathbf{W^B}\prime(\mathbf{f^P} - \mathbf{f^B}),
\end{aligned}
$$

where $\mathbf{W^B}$ is the column vector with the benchmark weight of each level within the category specified; $\mathbf{f^P}$ and $\mathbf{f^B}$ are the column vectors indicating the mean return of the portfolio and that of the benchmark, respectively. As mentioned above, $\mathbf{f^P}$ and $\mathbf{f^B}$ can be estimated by running a linear regression without the intercept term with respect to stocks in the portfolio and benchmark separately. Hence, the selection effect in the Brinson model can be calculated by using linear regression.

Interaction effect is the difference between a portfolio's actual return and the sum of the allocation and selection effects.

An numerical example of showing that the Brinson model is a special case of the regression approach is as follows.

Suppose that an equity portfolio manager has a portfolio named `test` with the universe as the benchmark.

```
> data(test)
> test.br <- brinson(x = test, date.var = "date",
+                    cat.var = "sector",
+                    bench.weight = "benchmark",
+                    portfolio.weight = "portfolio",
+                    ret.var = "return")
> returns(test.br)

                   2010-01-01
Allocation Effect    -0.00034
Selection Effect     -0.00425
Interaction Effect    0.00101
Active Return        -0.00359
```

When we apply the standard single-period Brinson analysis, we obtain an active return of -35.9 bps which can be further decomposed into allocation (-3.4 bps), selection (-42.5 bps), and interaction (10.1 bps).

We can also show the allocation effect by running a regression model based on sector only.

```
> test.reg <- regress(x =test,
+                     date.var = "date",
+                     ret.var = "return",
+                     reg.var = "sector",
+                     benchmark.weight = "benchmark",
+                     portfolio.weight = "portfolio")
> returns(test.reg)

                 2010-01-01
Sector             -0.00034
Residual           -0.00325
Portfolio Return   -0.01621
Benchmark Return   -0.01263
Active Return      -0.00359
```

The contribution from sector based on the regression approach (-3.4 bps) matches the allocation effect from the Brinson model as shown above.

However, in order to calculate the selection effect from the regression approach, we need to apply another regression model to a universe limited to the securities held in the portfolio. Using the factor returns from the regress class object, `test.reg`, and those from the linear regression, `lm.test`, we can obtain the selection effect (-42.5 bps) via the regression approach.

```
> lm.test <- lm(return ~ sector - 1,
+               data = test[test$portfolio != 0, ])
> lm.test$coefficients

     sectorEnergy   sectorMaterials
         -0.03561          -0.05146
sectorIndustrials   sectorConDiscre
          0.00194          -0.00533
 sectorConStaples  sectorHealthCare
         -0.02514           0.04327
 sectorFinancials    sectorInfoTech
```

```
         -0.02376          -0.02376
   sectorTeleSvcs   sectorUtilities
          0.00916          -0.03878

> exposure(br.single, var = "sector")[ ,2] %*%
+   (lm.test$coefficients - test.reg@coefficients)

        [,1]
[1,] 0.00653
```

6. Conclusion

In this paper, we describe two widely-used methods for performance attribution – the Brinson model and the regression-based approach, and provide a simple collection of tools to implement these two methods in `R` with the `pa` package. We also show that the Brinson model is a special case of the regression method. A comprehensive package, `portfolio` [5], provides facilities to calculate exposures and returns for equity portfolios. It is possible to use the `pa` package based on the output from the `portfolio` package. Further, the flexibility of `R` itself allows users to extend and modify these packages to suit their own needs and/or execute their preferred attribution methodology. Before reaching that level of complexity, however, `pa` provides a good starting point for basic performance attribution.

[1] C. Bacon. *Practical Portfolio Performance Measurement and Attribution*. John Wiley & Sons, Ltd., 2 edition, 2008.

[2] G. Brinson, R. Hood, and G. Beebower. Determinants of Portfolio Performance. *Financial Analysts Journal*, 42(4):39–44, Jul.–Aug. 1986. URL `http://www.jstor.org/stable/4478947`.

[3] G. Brinson, B. Singer, and G. Beebower. Determinants of Portfolio Performance II: An Update. *Financial Analysts Journal*, 47(3):40–48, May–Jun. 1991. URL `http://www.jstor.org/stable/4479432`.

[4] O. Davies and D. Laker. Multi-Period Performance Attribution Using the Brinson Model. *Journal of Performance Measurement*, Fall, 2001.

[5] J. Enos and D. Kane. Analysing Equity Portfolios in R. *R News*, 6(2): 13–19, MAY 2006. URL `http://CRAN.R-project.org/doc/Rnews`.

[6] A. Frongello. Linking Single Period Attribution Results. *Journal of Performance Measurement*, Spring, 2002.

[7] R. Grinold. Attribution, Modeling asset characteristics as portfolios. *The Journal of Portfolio Management*, page 19, Winter 2006.

[8] T. Lord. The Attribution of Portfolio and Index Returns in Fixed Income. *Journal of Performance Measurement*, Fall:45–57, 1997.

[9] J. Menchero. An Optimized Approach to Linking Attribution Effects over Time. *Journal of Performance Measurement*, 5(1):36–42, 2000.

[10] J. Menchero. Multiperiod Arithmetic Attribution. *Financial Analysts Journal*, 60(4):76–91, 2004.

[11] J. Menchero, A. Morozov, and P. Shepard. The Barra Global Equity Model (GEM2). *MSCI Barra Research Notes*, Sept. 2008.

[12] B. Singer and D. Karnosky. The General Framework for Global Investment Management and Performance Attribution. *The Journal of Portfolio Management*, Winter:84–92, 1995.