

Random Return Variables (Draft)

Charlotte Maia

November 21, 2010

This vignette introduces the rrv package, a package for modelling portfolio returns as random variables. There's a strong emphasis on modelling portfolios as functions of weight, and using empirical distributions.

Introduction

This package is largely inspired by the early work of Markowitz (1952, 1959). Markowitz considered the characteristics of portfolio returns, where both portfolio returns and individual security returns are regarded as random variables. He emphasized selecting efficient sets of portfolios, giving a good compromise between expected return and variance, under the assumption that security returns are often dependent. He then creates geometric representations of portfolios (especially portfolios with three securities), using isomean lines and isovariance curves. Here, we follow from Markowitz, however with a few key differences:

1. **We shall regard portfolios as functions of weight.**
More precisely, we shall regard a portfolio as a function, that maps a vector of weights to a random (portfolio) return variable.
2. **We shall model returns using empirical distributions.**
3. Extending point (1), instead of modelling weights in a cartesian space, we shall use a triangular space.
4. Also extending point (1), we shall model conditional parameters of portfolio return (e.g. expected portfolio return) as functions of weight.
5. Extending point (2), whilst we shall consider return variance (as a measure of symmetric risk), we shall also consider quantile return (as a measure of asymmetric risk).

This package is still at an early stage. The author is hoping to later add support for:

- Constrained optimisation.
- Modelling isoquantile curves (i.e. the quantile equivalent to isovariance curves).
- Triangular contour plots.

We shall make use of the dataset from Markowitz (1959), it gives discounted returns for nine securities over an eighteen year period.

```
> x = markowitz1959data ()  
> x
```

	Year	Am.T.	A.T. & T.	U.S.S.	G.M. A.T. & Sfe	C.C.	Bdn.	Frstn.	S.S.	
1	1937	-0.305	-0.173	-0.318	-0.477	-0.457	-0.065	-0.319	-0.400	-0.435
2	1938	0.513	0.098	0.285	0.714	0.107	0.238	0.076	0.336	0.238
3	1939	0.055	0.200	-0.047	0.165	-0.424	-0.078	0.318	-0.093	-0.295
4	1940	-0.126	0.030	0.104	-0.043	-0.189	-0.077	-0.051	-0.090	-0.036
5	1941	-0.280	-0.183	-0.171	-0.277	0.637	-0.187	0.087	-0.194	-0.240
6	1942	-0.003	0.067	-0.039	0.476	0.865	0.156	0.262	1.113	0.126
7	1943	0.428	0.300	0.149	0.255	0.313	0.351	0.341	0.580	0.639
8	1944	0.192	0.103	0.260	0.290	0.637	0.233	0.227	0.473	0.282
9	1945	0.446	0.216	0.419	0.216	0.373	0.349	0.352	0.229	0.578
10	1946	-0.088	-0.046	-0.078	-0.272	-0.037	-0.209	0.153	-0.126	0.289
11	1947	-0.127	-0.071	0.169	0.144	0.026	0.355	-0.099	0.009	0.184
12	1948	-0.015	0.056	-0.035	0.107	0.153	-0.231	0.038	0.000	0.114
13	1949	0.305	0.038	0.133	0.321	0.067	0.246	0.273	0.223	-0.222
14	1950	-0.096	0.089	0.732	0.305	0.579	-0.248	0.091	0.650	0.327
15	1951	0.016	0.090	0.021	0.195	0.040	-0.064	0.054	-0.131	0.333
16	1952	0.128	0.083	0.131	0.390	0.434	0.079	0.109	0.175	0.062
17	1953	-0.010	0.035	0.006	-0.072	-0.027	0.067	0.210	-0.084	-0.048
18	1954	0.154	0.176	0.908	0.715	0.469	0.077	0.112	0.756	0.185

Expected values and covariances for the dataset, are given in appendices A and B respectively. These allow us to verify some of the plots here.

Portfolios as Functions of Weight

Portfolio return Y , which is itself a random variable, can be regarded as the dot product of a vector of weights \mathbf{w} and a vector random variable \mathbf{X} , with the weights summing to one. Equivalently, it can also be regarded as the weighted average of multiple (often dependent) random variables X_1, X_2, \dots, X_k .

$$\begin{aligned}
 Y &= \mathbf{w}\mathbf{X} \\
 &= w_1X_1 + w_2X_2 + \dots + w_kX_k
 \end{aligned}$$

Traditionally, for k random variables, we would model our weights in cartesian space, with $k - 1$ dimensions. e.g. For three weights w_1, w_2, w_3 , we could use a two dimensional cartesian space, with points in that spaces represented by (w_1, w_2) , where we partly ignore w_3 , However, we shall mainly use a triangular space.

Before we continue, we need to clarify two important notions, function valued functions and random variable valued functions. Perhaps the most common example of a function valued function is differentiation. When we differentiate a function, differentiation itself can be regarded as a function, say diff , that maps a function to a function. So $f' = \text{diff}(f)$. Just as functions can return functions, functions can also return random variables. Perhaps the most common example is the mean (of random variables). The mean of the elements of a vector variable random, can be regarded as function, that maps a vector random variable to a scalar random variable. So $\bar{\mathbf{X}} = \text{mean}(\mathbf{X})$.

In our case, we will construct a portfolio g , from a set of historical returns. We have a portfolio constructor C_g , which is a function, that maps a matrix of historical returns \mathbf{x} (the realised values of \mathbf{X}) to a portfolio. We shall regard a portfolio as a function that maps a vector of weights to portfolio return. So:

$$\begin{aligned}
 g &= C_g(\mathbf{x}) \\
 Y &= g(\mathbf{w})
 \end{aligned}$$

The portfolio is based on the weighted sum given earlier, and treats the random variables as constants (constant in the sense that their distributions are constant). Exactly what g , \mathbf{X} and Y are, is discussed later. For now, let's take things a step further, and derive conditional parameters of Y .

Following the notion that a portfolio is a function of weight, we can also compute the expected value of portfolio return as a function of weight. This function is constructed from a portfolio, so:

$$\begin{aligned} f_{\mathbb{E}} &= C_{\mathbb{E}}(g) \\ \mathbb{E}(Y|\mathbf{w}) &= f_{\mathbb{E}}(\mathbf{w}) \end{aligned}$$

We are going to be unorthodox, and denote variance using \mathbb{V} and quantiles as \mathbb{Q} . We compute can compute them (and almost any conditional parameter) in the same manner, so:

$$\begin{aligned} f_{\mathbb{V}} &= C_{\mathbb{V}}(g) \\ f_{\mathbb{Q}} &= C_{\mathbb{Q}}(g, p) \\ \mathbb{V}(Y|\mathbf{w}) &= f_{\mathbb{V}}(\mathbf{w}) \\ \mathbb{Q}(Y|\mathbf{w}) &= f_{\mathbb{Q}}(\mathbf{w}) \end{aligned}$$

Note that the p in the constructor for quantile return is probability, and is a number between zero and one.

Using `rrv` we can construct a portfolio for the two investment (or two security) case, using say, the first two securities in the dataset. We can then go on to create functions for expected return, return variance and quantile return.

```
> g = portfolio (x [,2:3])
> portfolio.names (g)
[1] "Am.T."      "A.T. & T."

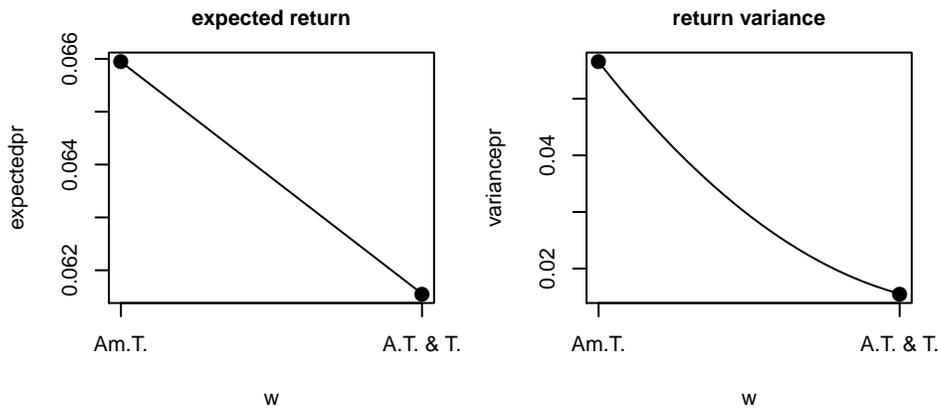
> fe = expectedpr (g)
> fv = variancepr (g)
> fq.25 = quantilepr (g, 0.25)
> fq.50 = medianpr (g)
> fq.75 = quantilepr (g, 0.75)
```

If we want, we can compute the expected return of a portfolio, which contains equal weighting over both investments.

```
> fe (c (0.5, 0.5) )
[1] 0.06375
```

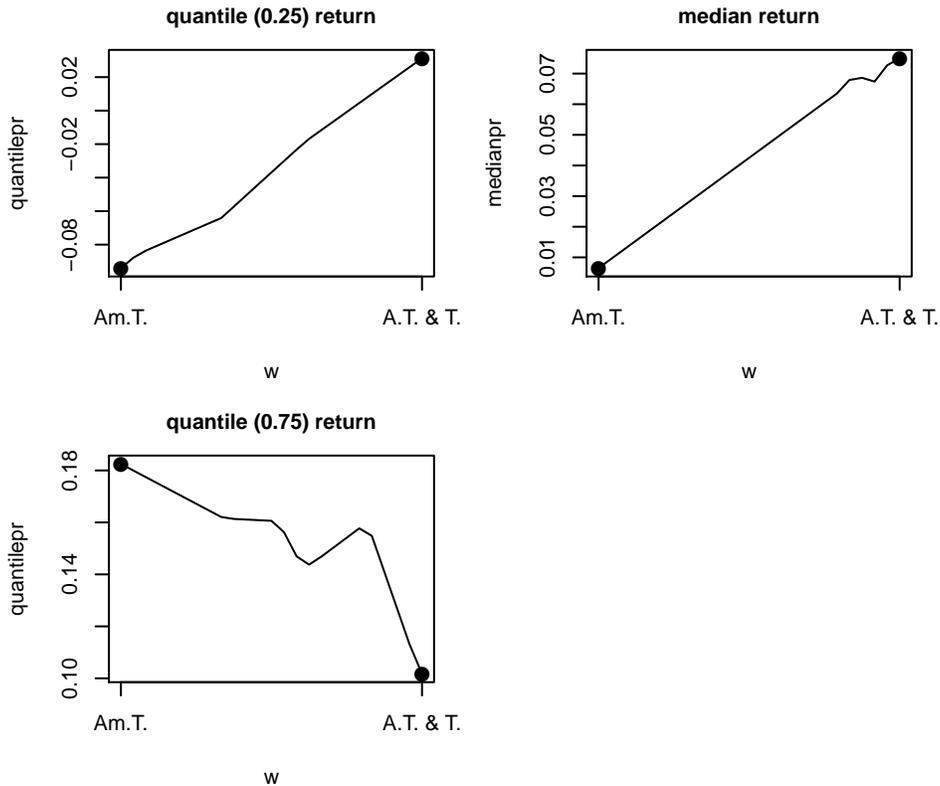
We can also plot the expected return and return variance.

```
> plot (fe, main="expected return")
> plot (fv, main="return variance")
```



As well as, the 0.25, 0.5 and 0.75 quantiles.

```
> plot (fq.25, main="quantile (0.25) return")
> plot (fq.50, main="median return")
> plot (fq.75, main="quantile (0.75) return")
```



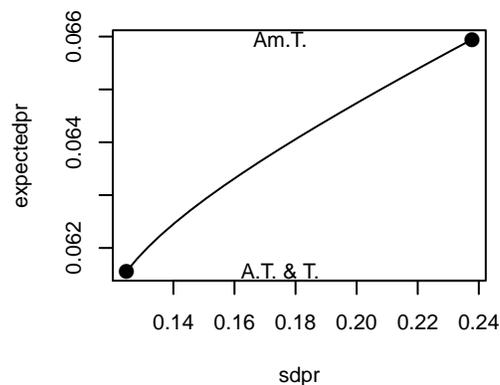
Note, that whilst both variance and quantiles can be used as measures of risk. An important difference is that, in general, low variance values are good, where as in general, high quantile values are good. Also note, that many financial distributions are asymmetric, hence portfolio return that is very attractive with respect to quantile values with $p < 0.5$, may be very unattractive with respect to quantile values with $p > 0.5$.

Many textbooks consider standard deviation versus expected return. Currently, the rrv package does not offer a standard function for this, plus the implementation is not fully vectorised, so:

```

> s = seq (0, 1, length=20)
> re = rsd = numeric (20)
> fsd = sdpr (g)
> for (i in 1:20)
  {
    w = c (1 - s [i], s [i])
    re [i] = fe (w)
    rsd [i] = fsd (w)
  }
> plot (rsd, re, type="l", xlab="sdpr", ylab="expectedpr")
> points (rsd [c (1, 20)], re [c (1, 20)], pch=16, cex=1.5)
> meanrsd = mean (rsd)
> labs = portfolio.names (g)
> text (meanrsd, re [1], labs [1])
> text (meanrsd, re [20], labs [2])

```



Visualising Portfolios over Triangular Spaces

We have suggested that weight exists in a triangular space (or equivalently, that portfolios are triangular functions). This principle is most intuitive in the case of three investments.

The examples used in the previous section, works using three investments as well. Hopefully contour plots will be implemented soon. Currently, this uses heat maps, with bright colours representing high values and dark colours representing low values.

Note that each plot, uses a separate scale (in order to achieve maximum colour variation). Hence, two points with the same colour, each from a separate plot, may represent quite different values.

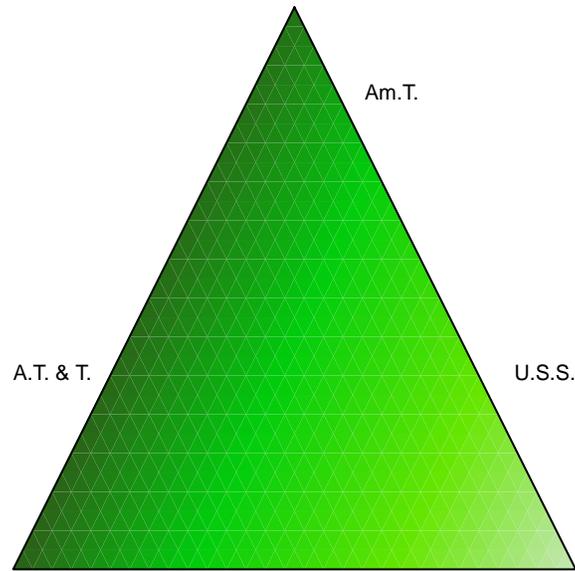
```

> g = portfolio (x [,2:4])
> portfolio.names (g)
[1] "Am.T."      "A.T. & T." "U.S.S."

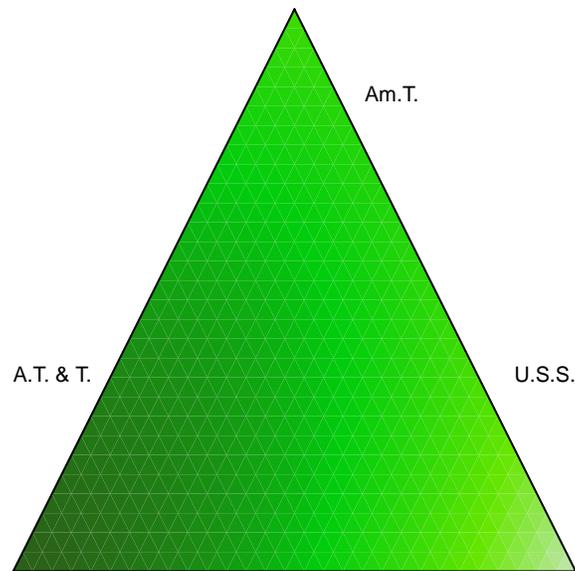
> fe = expectedpr (g)
> fv = variancepr (g)
> fq.25 = quantilepr (g, 0.25)
> fq.50 = medianpr (g)
> fq.75 = quantilepr (g, 0.75)

> #expected portfolio return
> plot (fe)

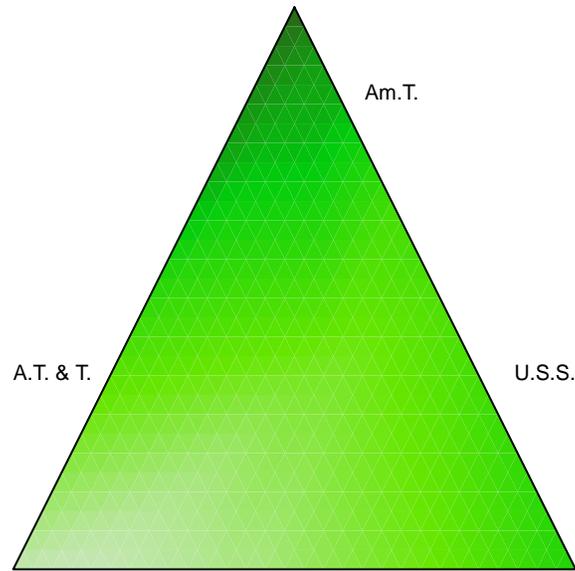
```



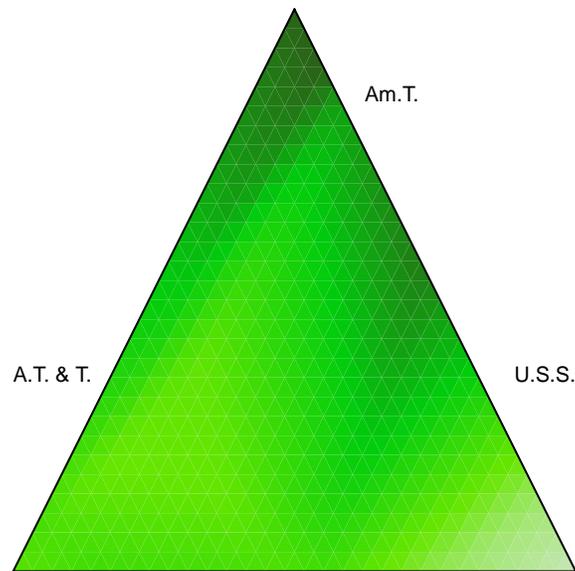
```
> #portfolio return variance  
> plot (fv)
```



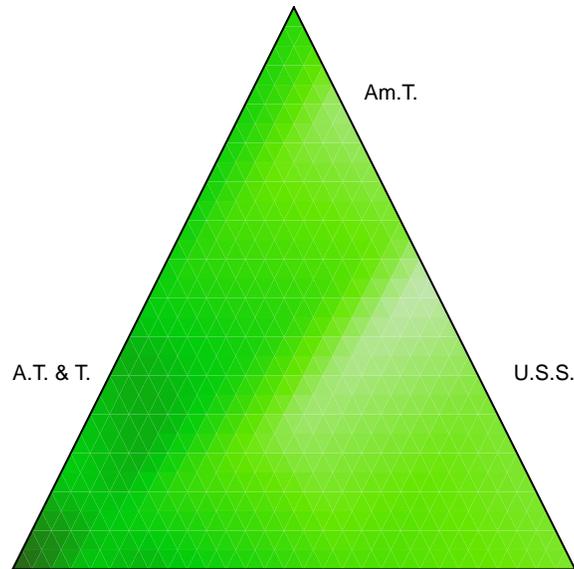
```
> #quantile (0.25) portfolio return  
> plot (fq.25)
```



```
> #median portfolio return  
> plot (fq.50)
```



```
> #quantile (0.75) portfolio return  
> plot (fq.75)
```

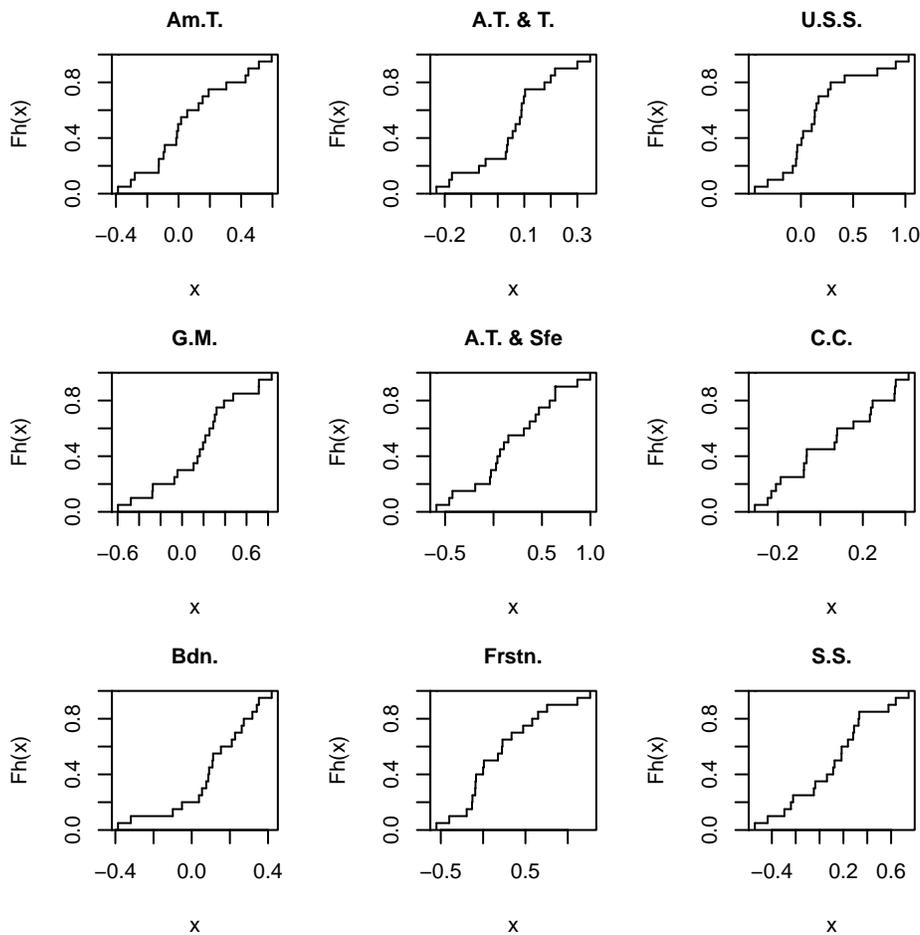


Modelling Historical Returns with Empirical Distributions

Up to this point, we haven't really discussed what \mathbf{X} and Y really are. This package assumes that random variables have an almost arbitrary distribution that can be modelled using an empirical cumulative distribution function (ECDF). In the case of vector random variables, we assume a multivariate ECDF, however at present this package does not make use of multivariate ECDFs (even though we will use the `mecdf` function).

ECDFs are modelled using realised values of a random variable, and we can plot their distributions. Using all our data:

```
> for (j in 2:10)
  plot (mecdf (x [,j], continuous=FALSE), main=names (x) [j])
```



Note that at the time of writing, there's a minor error in the `mecdf` package when computing continuous ECDFs (which is the default in the univariate case).

Modelling Portfolio Returns with Empirical Distributions

Roughly speaking, we can compute a vector of realised values for Y , which we shall denote \mathbf{y} . This uses a trivial extension to the expression used at the beginning of section two (on modelling portfolios as function). Remember that \mathbf{x} is a matrix (not a regular vector).

$$\mathbf{y} = \mathbf{x}\mathbf{w}$$

For clarity, let \mathbf{r}_i be the i th row vector of \mathbf{x} , then

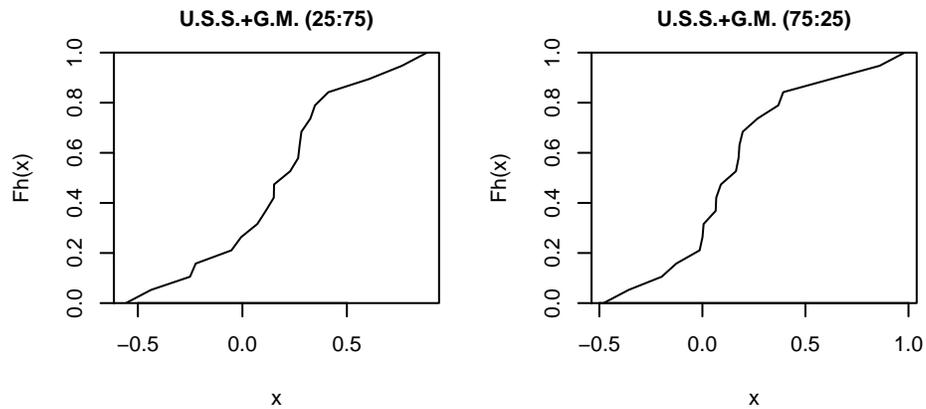
$$\begin{aligned} y_1 &= \mathbf{w}\mathbf{r}_1 \\ y_2 &= \mathbf{w}\mathbf{r}_2 \\ &\vdots \\ y_n &= \mathbf{w}\mathbf{r}_n \end{aligned}$$

In this package, we can create `rrv` (random return variable) objects from matrices (or any object that can be converted to a matrix). We can also create `rprv` objects (random portfolio return objects), using an `rrv` object and a vector of weights. `rprv` objects are what is returned by portfolio functions and are based on the expression above.

We can create a portfolio, then compute (and plot) different rprv objects.

```
> g = portfolio (x [,4:5])  
> portfolio.names (g)  
[1] "U.S.S." "G.M."
```

```
> prefix = paste (portfolio.names (g), collapse="+")  
> plot (g (c (0.25, 0.75) ), main=paste (prefix, "(25:75)" )  
> plot (g (c (0.75, 0.25) ), main=paste (prefix, "(75:25)" )
```



References

- Markowitz, H.M. (1952). Journal of Finance. Portfolio Selection.
Markowitz, H.M. (1959). Cowles Foundation. Portfolio Selection Efficient Diversification of Investments.

APPENDIX A: Expected Security Returns

Am.T.	A.T. & T.	U.S.S.	G.M.	A.T. & Sfe	C.C.	Bdn.
0.066	0.062	0.146	0.175	0.198	0.055	0.124
Frstn.	S.S.					
0.190	0.116					

APPENDIX B: Security Returns Covariance

	Am.T.	A.T. & T.	U.S.S.	G.M.	A.T. & Sfe	C.C.	Bdn.	Frstn.	S.S.
Am.T.	0.057	0.023	0.030	0.052	0.017	0.034	0.026	0.042	0.038
A.T. & T.	0.023	0.016	0.020	0.026	0.009	0.011	0.015	0.027	0.022
U.S.S.	0.030	0.020	0.091	0.066	0.047	0.014	0.012	0.073	0.044
G.M.	0.052	0.026	0.066	0.101	0.055	0.031	0.022	0.096	0.040
A.T. & Sfe	0.017	0.009	0.047	0.055	0.135	0.014	0.024	0.108	0.048
C.C.	0.034	0.011	0.014	0.031	0.014	0.044	0.012	0.031	0.023
Bdn.	0.026	0.015	0.012	0.022	0.024	0.012	0.029	0.032	0.020
Frstn.	0.042	0.027	0.073	0.096	0.108	0.031	0.032	0.155	0.056
S.S.	0.038	0.022	0.044	0.040	0.048	0.023	0.020	0.056	0.084