What components determine stock market returns in the 1990's?

M. Pojarliev, W. Polasek

Institute of Statistics and Econometrics,
University of Basel, CH-4051 Basel, Switzerland

Abstract: Stock market returns reflect influences from many sources, such as local and global financial variables, exchange rates, the business outlook and macroeconomic policy variables. How much forecasting ability is contained in these variables?
We examine the stock index returns of the US, UK, German and Japanese markets and apply a principal components (PC) analysis to find out if various sources of information can be combined into a few factors. We also include proxy variables for volatility in the PC model. The data suggest that about 5-6 factors can be used to approximate the predictive power of the original regression model. We find some similar factors in the models of the different countries. We show that PC models for stock returns can explain about 50% of the variation measured by R^2.

1 Introduction

Despite the debate on the efficiency of stock markets, there exists a demand for portfolio managers to forecast future returns for portfolio decisions. Previous research has shown that multi-factor models have supplanted the traditional CAPM in explaining an asset's returns. The driving force behind the multi-factor models has been the development of the arbitrage pricing theory (APT), (Ross, 1976). The idea of the APT model was to find a few factors that could explain the returns of a collection of stocks in a portfolio. Thus, we will try to explain in this paper the monthly returns of the Dow Jones, FTSE, DAX and Nikkei indices with principal components regression models.
We will concentrate on the following research questions. First, how much of the predictability (R^2 measure) of stock index returns can be explained by a principal components model? We will use the squared returns of certain variables as proxies for the volatility component, since it was shown in Polasek and Ren (1999) that multiple VARCH-M models can improve the predictability considerably. Second, which sources of economic risk are most important for explaining the predictabilit

of extracting factors from the regression. In Section 3 we apply the PC model to the stock returns of four countries. In a final section we summarize our findings.

2 Principal components regression

Principal components analysis can be used to reduce the number of regressors in the regressor matrix. This is particularly useful in regressions with many variables. Let us consider a normal linear regression model with centered variables, i.e. $\bar{y} = 0$, $\bar{x} = 0$, for $i = 1, .., p$, where we have an $n \times 1$ vector $\mathbf{y}$ of dependent variables and an $n \times p$ regressor matrix $\mathbf{X}$:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \epsilon \sim \mathbf{N}[\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I}_n]. \tag{1}$$

Consider now a principal components transformation of the regressors in $\mathbf{X}$:

$$\mathbf{W} = \mathbf{XG} \quad or \quad \mathbf{WG}' = \mathbf{X}, \tag{2}$$

where the covariance matrix of $\mathbf{X}$, with $\mathbf{G}$ the principal components loading matrix, is given as

$$\mathbf{\Sigma} = Var(\mathbf{X}) = \mathbf{X}'\mathbf{X}/n \tag{3}$$

and has an eigenvalue decomposition of the form

$$\mathbf{\Sigma} = \mathbf{G\Lambda G}' \tag{4}$$

with $\mathbf{\Lambda} = diag(l_1, ...l_p)$ a diagonal matrix with the eigenvalues of $\mathbf{\Lambda}$, and $\mathbf{G}$ the matrix of eigenvectors having the property $\mathbf{G}'\mathbf{G} = \mathbf{I}_p$. Inserting the principal components regression into the regression model we obtain

$$\mathbf{y} = \mathbf{WG}'\beta + \epsilon \quad or \quad \mathbf{y} = \mathbf{W}\alpha + \epsilon \tag{5}$$

with the new coefficient vector $\alpha = \mathbf{G}'\beta$. Since the columns of $\mathbf{W}$ are orthogonal, i.e. $\mathbf{W}'\mathbf{W} = \mathbf{G}'\mathbf{X}'\mathbf{XG} = n\mathbf{\Lambda}$, the least squares estimator of α does not change if only a subset of the variables in $\mathbf{W}$ is selected. The OLS estimate of α is given by

$$\hat{\alpha} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{Wy} \tag{6}$$

and since $\mathbf{W}' = (\mathbf{w}_1, ..., \mathbf{w}_n)$ is a decomposition of $\mathbf{W}$ into n row vectors $\mathbf{w}_i$, we can calculate the OLS estimates in (5) for each component simply as

$$\hat{\alpha}_i = \frac{\mathbf{w}_i\mathbf{y}}{nl_i} \tag{7}$$

where l_i is the i^{th} eigenvalue of $\mathbf{X}'\mathbf{X}/n$. Because the $\mathbf{W}$ matrix is orthogonal we have a simple ANOVA decomposition of $\mathbf{y}'\mathbf{y}$, the total sum of squares:

$$\mathbf{y}'\mathbf{y} = \sum_{\mathbf{i=1}}^{\mathbf{p}} \hat{\alpha}\mathbf{w_i}'\mathbf{w_i} + \hat{\varepsilon}'\hat{\varepsilon} \tag{8}$$

with the OLS residual vector being $\hat{\varepsilon} = \mathbf{y} - \mathbf{W}\hat{\alpha}$. This decomposition implies also a nice decomposition of the R^2 into p sub R^2 contributions $(R^2_{\alpha i})$ measures of the principal components regression:

$$R^2 = \frac{\mathbf{y}'\mathbf{y} - \hat{\varepsilon}'\hat{\varepsilon}}{\mathbf{y}'\mathbf{y}} = \sum_{i=1}^{p} nl_i\hat{\alpha}_i^2/\mathbf{y}'\mathbf{y} = R^2_{\alpha 1} + ... + R^2_{\alpha p}. \tag{9}$$

Thus, $l\alpha_i^2$ is the proportion of the variance of $\mathbf{y}$ (i.e. $\mathrm{Var}(\mathbf{y}) = \mathbf{y}'\mathbf{y}$ since $\bar{y} = 0$) which can be explained by the i^{th} principal components regressor. Note that OLS estimates $\hat{\alpha}_i$ can be tested for significance with the ordinary t-test:

$$t_{\alpha_i} = \frac{\hat{\alpha}_i}{s.e.(\alpha_i)} \sim t_{n-p-1}$$

with the standard error of the estimator

$$s.e.(\alpha_i) = \frac{nl_i(n-p-1)}{\hat{\varepsilon}'\hat{\varepsilon}}.$$

There are as many principal components as there are original variables. Several criteria show which components should be retained (see Mardia et al., 1992).
The *principal components loadings* $\mathbf{G}$ are shown in the next chapter for the US model. The loadings are the coefficients of the principal components transformation. They provide us with a convenient summary of the influence of the original variables on the new components and form a basis for interpreting factors.

3 Country models

A linear regression model using the variables from table 1 and the standardized monthly returns of the Dow Jones index as the dependent variable yields a multiple R^2 of 0.6511.

To reduce the number of variables in the regressor matrix, we apply *principal components analysis.* We are using the first five principal components in the PC regression according to Kaiser's criteria (see Mardia et al., 1992): only components whose eigenvalues are over the average

Symbol	Description
I10y	10 year Treasury bond
E	100/(price-earnings ratio)
X	returns of trade weighted currency index, 1990=100
I3m	middle rate of the three month Treasury bill
SM	RSI for the returns of the Dow Jones index, basis 1993
BM	RSI for the yields of the 10 year Treasury bond, bas.1993
EAR	aggregate earnings revisions, 6 m. moving average, in %
CPI	returns of the consumer price index less food
DAX	returns of the DAX index
USDDM	exchange rate of the US $ to the DM
MSCIE	returns of the MSCI Europe 15
vI3m	$I3m^2$: volatility of I3m
vEAR	$EAR.^2$ volatility of EAR
vDAX	DAX^2: volatility of DAX
vCPI	CPI^2: volatility of CPI
vMSCIW	$MSCIW^2$: volatility of MSCI World

Table 1: Standardized, monthly time series from November 1994 until April 1999 (source: Data Stream)

(0.98) should be retained as "factors". These factors explain 80% of the variance of the original exogenous variables. Using the R^2 decomposition we can calculate the R^2 of the PC regression by summing the first five proportions.
As opposed to the factors in the APT, our components are statistical constructs. The advantage of this methodology is that the principal components loadings help us to interpret the factors and see which sources of information are most important for forecasting returns. In the APT type of models, the factors have to be chosen a priori and the importance of information is not empirically quantified.
Table 2 lists the loadings of the five components for the Dow Jones returns. We have chosen factor names so that they are useful for interpretation. In the first factor, which is most important (as it explains 41% of the variance of the original variables), only the term structure of interest rates has influence. The rest of the variables have weights less than 0.05 and that is why we called this factor "*Term Structure*" (TS). The second factor is called "*Volatility*" because the largest weights are from the squared returns of the MSCI world index, the DAX index and the consumer price index. This result confirms previous research (Polasek and Ren, 1999) that volatilities play an important role in determining the stock returns. The name "*World information*" refers mainly to European returns and the volatility of the MSCI world index. The

name	**TS**	**Volatil.**	**World inf.**	**Inflation.**	**P/E**
*lag*I10y	**0.45**	0.13	-0.03	-0.09	0.08
*lag*E	-0.04	-0.07	0.08	0.08	**-0.79**
*lag*X	-0.01	0.12	-0.28	-0.31	0.03
*lag*I3m	**0.55**	0.13	0.10	-0.15	0.00
*lag*SM	-0.05	-0.08	0.21	0.05	-0.01
*lag*BM	**0.43**	-0.15	0.07	0.09	-0.19
*lag*EAR	0.02	-0.03	0.16	-0.13	0.17
*lag*CPI	0.00	0.36	-0.05	**0.56**	-0.12
*lag*DAX	0.00	0.19	**0.44**	-0.24	-0.05
*lag*US$DM	0.00	-0.24	0.14	0.24	**0.51**
*lag*MSCIE	0.01	-0.31	**0.69**	0.18	0.00
vI3m	**-0.55**	0.12	0.10	-0.14	-0.02
vEAR	0.05	0.04	-0.05	0.17	-0.04
vDAX	0.00	**0.44**	0.15	-0.09	0.09
vCPI	0.01	**0.37**	-0.01	**0.53**	0.13
vMSCIW	-0.01	**0.50**	**0.33**	-0.18	0.00
$R^2_{\alpha i}$ in %	16.86	11.32	10.23	8.04	7.03

Table 2: The factor loadings of the principal components analysis for the returns of the Dow Jones index (bold values emphasizes important variables)

largest weights in the factor *"CPI"* or *"Inflation"* are the returns of the consumer price index and its squared returns. The CPI is often used to measure inflation. The largest weight in the last factor *"P/E"* stems from the inverse of the price-earnings ratio (see Table 1). Table 3 gives the summary of the principal components regression for the returns of the Dow Jones index. The PC models for the other countries yield 6 factors and an R^2 of 0.74 for the UK model, 7 factors and an R^2 of 0.52 for the German model and 5 factors and an R^2 of 0.44 for the Japanese model. In contrast to the US model a factor "Exchange rate" is present in all other countries models which indicates the importance of the US$.

Now we write the orthogonal decomposition in (5) for the five components of the PC analysis of the Dow Jones index as

$$\hat{y}_t = \hat{\alpha}_1 w_{1t} + ... + \hat{\alpha}_5 w_{5t}, \quad t = 1, ..., 54 \tag{10}$$

Figure 1 plots the 5 contributions $\hat{A}_{it} = \hat{\alpha}_i w_{it}$ for $i = 1, ..., 5.$ for each time point. It is interesting to note that the largest deviation (those of observations 47 and 49) can be explained by only 3 factors.

	coef.	std.err	t.stat	p.value
IS	-13.5763	3.0965	-4.3844	0.0001
Volatility	-0.3400	0.0939	-3.6212	0.0007
World inf.	0.3281	0.0809	4.0586	0.0002
CPI	0.2409	0.0666	3.6180	0.0007
P/E	-1.5961	0.5473	-2.9162	0.0054
Residual St. Err. = 0.6916, Multiple R^2 =0.5578				
N = 54, F-stat. = 12.3618 on 5 and 49 df, p-value = 0				

Table 3: Principal Components (PC) regression for Dow Jones returns

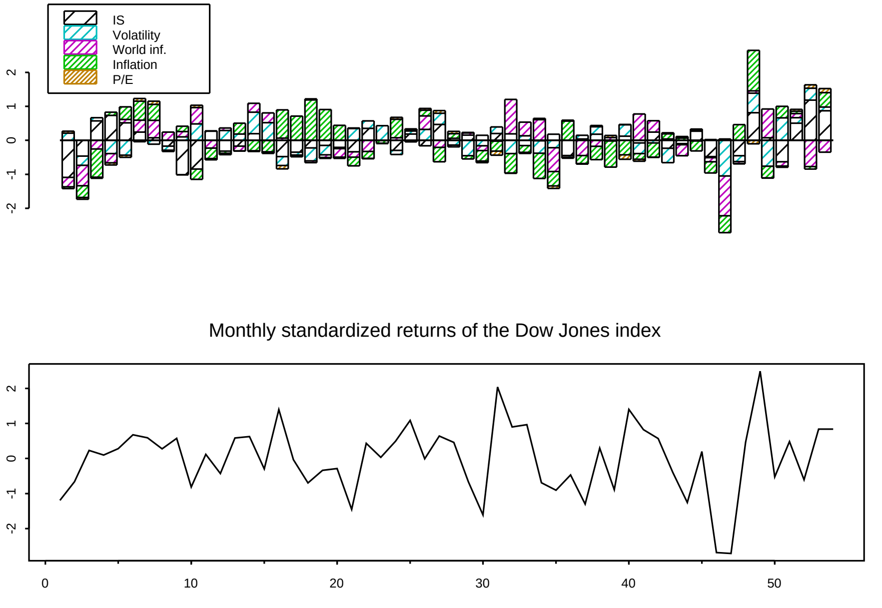

Figure 1: Estimated factor contributions of the US model

4 Conclusions

Applying principal components regression to stock returns, we have reduced the number of variables in the regressor matrix from 14-16 to 5-7 components. To interpret these factors, we have used names of the variables with the largest weights in the factor loadings. The analysis shows that the Dow Jones index influences all other markets and the factor *"World information"* determines strongly the stock market fluctuations of the UK, Germany and Japan. Recall that the MSCI North America index has more than 50% weight in the MSCI World index.

There are some differences in the country models: the UK stock market is more sensitive to local information than the German and the Japanese stock markets. The term structure of interest rates in the USA has a big influence on the returns of the Dow Jones index. The exchange rate of local currency to the US $ seems to influence the local stock market. It will be interesting to see whether the US $ has such an influence on the other stock markets after the establishment of the Euro. The factor *"Volatility"* is present in all four models. In general we can say that in the second half of the 1990's the stock market returns were determined by the term structure of the interest rates, the exchange rate of local currency to the US $, other stock markets, the volatility of fundamental variables and inflation.

Our research confirms previous research of Ferson and Harvey (1991), which found that significant predictability for stock returns exists. A related paper will report on the use of principal components models for active portfolio management.

References

FERSON, W. and HARVEY, C. (1991): Sources of Predictability in Portfolio Returns. *Financial Analysts Journal, May-June, 49–56.*

MARDIA, K. V. and KENT, J. T. and BIBBY, J. M. (1992): Multivariate Analysis. Academic Press, London

POJARLIEV, M. and POLASEK, L. (2000): Portfolio construction using multivariate time series forecasts. *University of Basel, ISO, Working paper.*

POLASEK, W. and REN, L. (1999): Volatility analysis during the Asia crisis: A multivariate GARCH-M model for stock returns in the US, Germany and Japan. *WWZ-Discussion Paper 9907, December 1999.*

ROSS, S. (1976): The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic theory, 13, 341–360.*

, ,