Statistics for High Dimensional Data

Samuele Borsini (0001083685)
Francesco Sbrana (0001083103)
Pablo Suárez-Sucunza (0001075377)
Stella Vita Gatti (0001085816)

April 3, 2024

1 Introduction

This article aims to determine the main drivers of CEO's performance, with also a link to firm's performance, using a dataset composed by information from 177 CEO's of different organizations and 15 variables related to their compensation, work experience and firm's performance.

Chief Executive Officer's (CEO) compensation packages have been seen as a way to mitigate the conflict of interest between managers and shareholders in corporations, with a role in motivating top managers Ozkan (2011). It is thus important to determine whether there is a link between compensation and firm's performance. However, relationship between CEO compensation and firm performance is a complex and debated topic in economics and the sign of the effect of this relation in neither agreed, if this link ever exists. Some studies found a positive correlation between CEO pay and firm performance, for instance Jensen and Murphy (1990) found a positive and significant connection between CEO cash compensation and shareholder wealth, while some other found a negative connection Tariq (2010).

Although those findings are all interesting, our research question is slightly different: we tried to identify whether there exist some latent variables or main component capable of predicting CEO's performance, and thus also the economic performance of the firms.

We run our analysis using two different methods, Factor Analysis and Principal Component Analysis, which allows us to reduce the dimensionality of the dataset and to find a small number of factors or components summarizing the main drivers of CEO's performance.

2 Data

We have a dataset containing information about about salary, experience and education of the CEO, as well as firm's performance, market valuation and other related measures, for CEOs of 177 firms. The following table presents the summary statistics for all the variables in the dataset. Because of the different scale of some of the variables, we do use standardized measures during the two types of analyses we carry out.

Table 1: Summary statistics

Variable	Mean	SD	Min	Med	Max
salary	865.864	587.589	100.0	707.0	5299.0
age	56.429	8.422	33.0	57.0	86.0
college	0.972	0.166	0.0	1.0	1.0
grad	0.531	0.500	0.0	1.0	1.0
comten	22.503	12.295	2.0	23.0	58.0
ceoten	7.955	7.151	0.0	6.0	37.0
sales	3529.463	6088.654	29.0	1400.0	51300.0
profits	207.831	404.454	-463.0	63.0	2700.0
mktval	3600.316	6442.276	387.0	1200.0	45400.0
lsalary	6.583	0.606	4.6	6.6	8.6
lsales	7.231	1.432	3.4	7.2	10.8
lmktval	7.399	1.133	6.0	7.1	10.7
comtensq	656.684	577.123	4.0	529.0	3364.0
ceotensq	114.124	212.566	0.0	36.0	1369.0
profmarg	6.420	17.861	-203.1	6.8	47.5

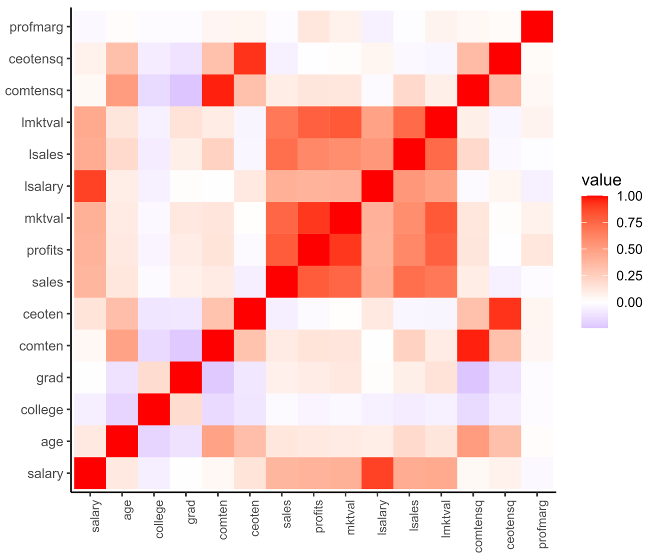

Figure 1: Correlation heatmap

Subset of variables We exclude the dummies *grad* and *college* from our upcoming analysis since we cannot use them for the Factor analysis and, to be consistent, we do not employ them either in the principal component analysis. *profmarg* is also left out given that it has low correlation with the other variables. We have *sales*, *salary* and *mktval*, and their respective logarithmic transformations, but we decided to keep the variables in levels. We also drop *age* as we already have some measures of experience and we believe that taking into account the latter, the former would bring little added value to the analysis. The experience related variables are years in the company (*comtem* and *comtemsq*) and years as the CEO (*ceoten* and *ceotensq*). We drop the squares as *ceoten* and *comten* are more correlated with the other variables.

The final subset of selected variables is then: *sales*, *salary*, *mktval*, *profits*, *ceoten* and *comten*.

3 Principal Component Analysis

In the Principal Component Analysis framework, the high-dimensionality of the data is exploited through the construction of variables (called Principal Components, $\boldsymbol{W} = [W_1, W_2, \dots, W_d]'$) which are linear combinations of the original data ($\boldsymbol{X} = [X_1, X_2, \dots, X_d]'$):

$$W_k = \sum_{j=1}^{d} \eta_{kj} X_j$$

where the coefficients η_{kj} are called "loadings". The linear combinations should satisfy the following properties:

1. Uncorrelation:
$$C(W_k, W_s) = 0 \quad \text{if } k \neq s$$
2. They should be ordered in decreasing order according to their variability:
$$V(W_1) \geq V(W_2) \geq \cdots \geq V(W_d)$$
3. The total variance of the two systems have to be the same:
$$\sum_{i=1}^{d} V(W_i) = \sum_{i=1}^{d} V(X_i)$$

Assuming that the three properties are satisfied, then (by 2 and 3) it is possible to summarize the original data (d variables) with fewer linear combination (r variables), achieving what is also known as dimension reduction.

In order to have the above-written properties verified, it is required to impose constraint on the loadings η_{kj}. In PCA, this is achieved by choosing the first vector of loadings $\boldsymbol{\eta}_1$ as the direction of most variability, i.e. as the solution of the following maximization problem:

$$\max_{\boldsymbol{\phi} \in \mathbb{R}^d} \big[V(\boldsymbol{\phi}'\boldsymbol{X})\big], \quad \text{s.to } \boldsymbol{\phi}'\boldsymbol{\phi} = 1$$

The FOCs for the problem are:

$$\begin{cases} 2V(\boldsymbol{X})\boldsymbol{\phi} - 2\lambda\boldsymbol{\phi} = \boldsymbol{0} \\ \boldsymbol{\phi}'\boldsymbol{\phi} = 1 \end{cases} \implies \begin{cases} V(\boldsymbol{X})\boldsymbol{\phi} = \lambda\boldsymbol{\phi} \\ \boldsymbol{\phi}'\boldsymbol{\phi} = 1 \end{cases}$$

Notice that the first equation implies that $\boldsymbol{\eta}_1$ and λ are, respectively, an eigenvector and an eigenvalue of $V(\boldsymbol{X})$. By the second condition we can set the norm of $\boldsymbol{\eta}_1$ equal to 1. Finally:

$$V(\boldsymbol{X})\boldsymbol{\eta}_1 = \lambda\boldsymbol{\eta}_1 \implies \boldsymbol{\eta}_1'V(\boldsymbol{X})\boldsymbol{\eta}_1 = \boldsymbol{\eta}_1'\lambda\boldsymbol{\eta}_1 \implies V(\boldsymbol{\eta}_1'\boldsymbol{X}) = \lambda\boldsymbol{\eta}_1'\boldsymbol{\eta}_1$$

and the LHS is just the variance of the linear combination W_1, while the RHS is equal to λ. Thus, the first vector of loadings $\boldsymbol{\eta}_1$ is just the eigenvector corresponding to the biggest eigenvalue of $V(\boldsymbol{X})$, which is itself the variance of the first Principal Component. Similarly, we can find the second vector of loadings as:

$$\max_{\boldsymbol{\phi}\in\mathbb{R}^d}\big[V(\boldsymbol{\phi}'\boldsymbol{X})\big], \quad \text{s.to } \boldsymbol{\phi}'\boldsymbol{\phi} = 1 \text{ and } \boldsymbol{\eta}_1'\boldsymbol{\phi} = 0$$

where the new constraint will imply the first property we where looking for. Intuitively, this second problem will lead us to the eigenvector asociated to the second largest eigenvalue of $V(\boldsymbol{X})$, and so on and so forth.

Therefore, we can just consider the spectral decomposition of $V(\boldsymbol{X})$ (henceforth called Σ):

$$\Sigma = \Gamma\Lambda\Gamma'$$

where Λ is a diagonal matrix containing all the eigenvalues of Σ in descending order, and Γ has as nth column the eigenvector associated with the nth largest eigenvalue.

Notice, it is important to standardize the variables (i.e. to use the correlation of $\boldsymbol{X}$ rather than the covariance one), especially when the units of measure of the variables are not equal.

In PCA the aim is to choose the number of components in order to retain as small a set as possible and at the same time to have a sufficient number to provide a good representation of the data. The components are derived in order of variance, therefore we choose the first components such that the proportion of variance retained is large enough. One of the main criteria of choosing the number of principal components is to examine the scree plot, shown in 2.

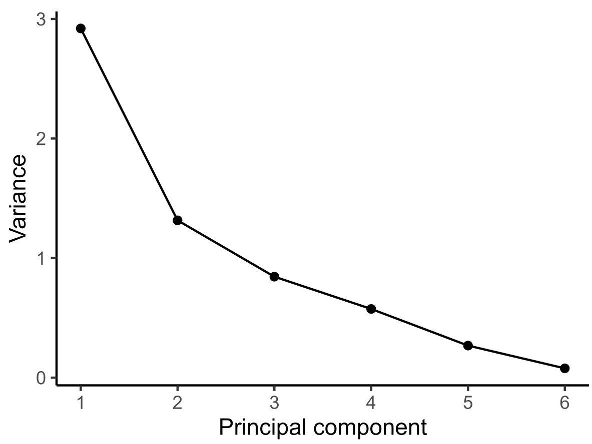

Figure 2: Scree plot

Table 2: PC standard deviations and variance explained

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	1.709	1.147	0.919	0.758	0.518	0.279
Proportion of Variance	0.487	0.219	0.141	0.096	0.045	0.013
Cumulative Proportion	0.487	0.706	0.847	0.942	0.987	1.000

The scree plot shows that the point after which the eigenvalues decrease more slowly corresponds to 2 components. Also the PCA variation summary suggests to use 2 components, since the proportion of variance explained is about 70.6%. Adding more components after this point would explain relatively little more of variance. As last evidence for our choice, the classic rule of thumb that states to choose the components with an eigenvalue bigger than 1 would pick exactly 2 components. The Jolliffe's version of this criteria (that lowers the cut-off point at 0.7) would suggest to select the first 4 PCs, yet, since all the other criteria are satisfied with 2 of them, we will focus our analysis only on the first two.

To be complete, we could also retain three rather than two components, since the variation explained by the third component is still quite significant and that in the scree plot there is no any evident edge. However, we argue that 2 components is the best choice, also because the third component would be not easily interpretable, as shown in the following part.

Table 3: PCA coefficients, rescaled coefficients and score coefficients

	Coefficients			Rescaled coefficients			Score coefficients		
Variable	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
salary	0.335	0.11	0.753	0.573	0.126	0.692	0.196	0.096	0.819
comten	0.118	0.653	-0.534	0.202	0.749	-0.49	0.069	0.57	-0.581
ceoten	0.022	0.734	0.322	0.037	0.842	0.296	0.013	0.64	0.35
sales	0.517	-0.119	-0.12	0.884	-0.136	-0.11	0.302	-0.103	-0.131
profits	0.554	-0.072	-0.14	0.946	-0.082	-0.129	0.324	-0.062	-0.152
mktval	0.547	-0.054	-0.104	0.935	-0.061	-0.095	0.32	-0.047	-0.113

The PCA coefficients provided in the frist column of table 3 illustrate the contribution of each variable to the first three principal components. As said, our model will include only two components, so the third one is shown as a reference. We observe that variables with higher positive coefficients make a more substantial positive contribution to the respective principal component, while those with higher negative coefficients contribute more negatively. For instance, in PC1, *salary, sales, profits*, and *mktval* exhibit relatively high positive coefficients, indicating significant positive contributions, whereas *comten* and *ceoten* have positive coefficients, but comparatively lower. In PC2, *comten* and *ceoten* display notably positive coefficients, indicating strong positive contributions, while *sales, profits,* and *mktval* show negative coefficients, albeit less significant. Additionally, *salary* has a small positive coefficient.

The coefficients in PCA are typically rescaled, as in the second column of table 3, to ensure that variables contributing more significantly to the principal components, which explain the most variation in the data, have larger coefficients compared to those contributing less. This rescaling helps emphasize the relative importance of variables in each principal component, making it easier to identify the key drivers of variation in the data. These rescaled coefficients are calculated as:

$$\eta_{kj}^{*} = \sqrt{\lambda_j}\eta_{kj} \quad k = 1, ..., d; \quad j = 1, ..., d$$

We observed that the first principal component shows a strong positive correlation with profits and market value, while its correlation with variables such as *ceoten* and *comten* is comparatively lower. This suggests that PC1 primarily represents the overall performance and valuation of the company, with variations in *profits, sales,* and *mktval* exerting a significant influence on its composition. Conversely, the second principal component exhibits a positive correlation with variables *comten* and *ceoten*, indicating a strong association with executive experience and tenure, possibly reflecting seniority or stability within the executive team. Additionally, PC2 shows a negative correlation with *sales, profits*, and *mktval*, suggesting that these variables contribute less to the composition of PC2 compared to executive tenure variables.

In addition, it is usual to standardize the component scores to have unit variance, hence:

$$\tilde{\eta}_{ik} = \frac{\eta_{ik}}{\sqrt{\lambda_k}} = \frac{\eta_{ik}^{*}}{\lambda_k}$$

These standardized components are referred to as score coefficients, displayed in the third column of table 3. Indeed, they represent the coefficients of each variable to compute the score on every component.

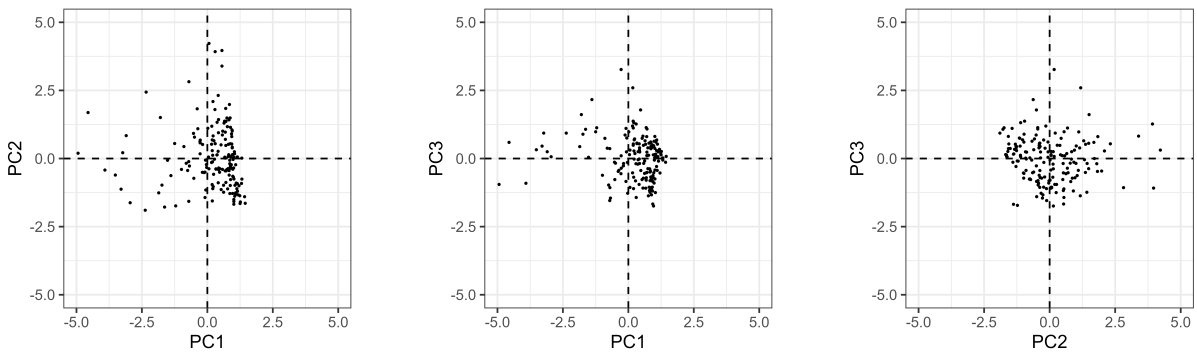

Figure 3: PC scatter plots

Finally, scatter plots serve as valuable tools for summarizing data and unveiling hidden patterns that might not be immediately apparent in the original dataset. In our analysis, we observed that PC1 displays greater variability compared to other principal components. Additionally, distinct clusters emerge on the right side of PC1, indicating groups of observations with similar characteristics. On the contrary, outliers are noticeable on the left side of PC1, representing observations with unique characteristics compared to the rest of the data. Clusters are mostly evident for PC1 and PC2 when the y-axis represents the third component. Therefore, we can establish that PC3 exhibits very low variance, reinforcing the decision to focus on the first two principal components, which capture the majority of variability in the dataset.

4 Factor model analysis

Similarly to PCA, in Factor analysis we want to retrieve from the data some latent variables (called factors in this framework) that can summarize the data. The model specification is the following:

$$\boldsymbol{X} = \boldsymbol{\mu} + \Lambda \boldsymbol{f} + \boldsymbol{u} \tag{1}$$

where $\boldsymbol{X}$ is the vector of d observed variables, $\boldsymbol{\mu}$ is their mean vector, $\boldsymbol{f}$ is the vector of q latent factors, $\boldsymbol{u}$ is the vector of d specific factors and Λ is a $d \times q$ matrix of factor loadings. We make the following assumptions:

1. $E(\boldsymbol{f}) = \boldsymbol{0}$ and $E(\boldsymbol{f}\boldsymbol{f}') = I_q$
2. $E(\boldsymbol{u}) = \boldsymbol{0}$, $E(\boldsymbol{u}\boldsymbol{u}') = \Psi$ (which is diagonal), and $E(\boldsymbol{f}\boldsymbol{u}') = 0_{q \times d}$

From the assumptions and (1) we can decompose the variance of the data Σ as:

$$\Sigma = \Lambda\Lambda' + \Psi$$

and we can notice that the LHS has $d(d+1)/2$ distinct elements, while the RHS has $d(q+1)$ parameters to estimate. Thus, a factor model could suffer from identification problems if q is too large. To reduce the number of parameters to estimate, the matrix $\Lambda'\Psi^{-1}\Lambda$ it is imposed to be diagonal. These constraints reduce the number of parameters by $q(q-1)/2$.

There are different methods that allow us to estimate the model in (1). We will use MLE because it makes possible to run tests on the number of factors. Under normality assumption of the observed variables, the log-likelihood function can be written as:

$$l(\boldsymbol{X}, \Lambda, \Psi) = \frac{n}{2} \ln |(\Lambda\Lambda' + \Psi)^{-1}| - \frac{n}{2} \mathrm{tr}[(\Lambda\Lambda' + \Psi)^{-1} S]$$

where S is the observed covariance matrix. Since analytical solutions do not exist for this maximization problem (w.r.t. Λ and Ψ), we need to rely on numerical methods.

Another important thing to notice is that if we apply an orthogonal transformation G to the factors, we would have:

$$\Sigma = \Lambda G' G \Lambda' + \Psi$$

yet since $G'G = I_q$ by orthogonality, then the new model would be indistinguishable from the original one. Therefore, we will use orthogonal rotations (and oblique ones, that allows for correlation among the factors) to interpret the factors.

Lastly, to get an estimate of the factors we can use two methods:

- Bartlett's method (which is based on MLE and requires the normality assumption and the one we will use):

$$\hat{\boldsymbol{f}}_i = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}\boldsymbol{x}_i$$

- Thomson's method (which is based on the Bayesian framework):

$$\hat{\boldsymbol{f}}_i = (I_q + \hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}\boldsymbol{x}_i$$

Firstly, 4 presents the percentage of variance explained by and the p-value of the test of sufficiency of three different factor models, using respectively 1, 2 and 3 factors. The statistics shown for the 1-factor model suggest that this specification is not a good representation of the data, given the rejection of the test and the low share of variance explain (44.8&). Whereas, with an increase of around 20% in variance explained and a large p-value (0.36) we think that the 2-factors model provides a meaningful and interpretable representation of the underlying structure of the data, explaining a substantial amount of variance in the original variables.

Table 4: Variance explained by and test of sufficiency of different factor models

	1-factor	2-factor	3-factor
p-value	0.0	0.366	- [1]
% variance explained	44.8	63.600	76.7

Given these results, we will use as main specification the 2-factors model, yet we will show the also for the 3-factors model as a reference

In order to assess the adequacy of our model's fit and its ability to capture the correlations among observed variables, we calculate a matrix for comparing the observed correlation matrix with the estimated correlation matrix (table 5).

Table 5: Observed-Estimated correlation matrix difference

	salary	comten	ceoten	sales	profits	mktval
salary	0.807	-0.070	0.000	0.053	-0.010	0.016
comten	-0.070	0.878	0.000	0.004	0.003	-0.007
ceoten	0.000	0.000	0.005	0.000	0.000	0.000
sales	0.053	0.004	0.000	0.335	0.001	-0.006
profits	-0.010	0.003	0.000	0.001	0.040	0.000
mktval	0.016	-0.007	0.000	-0.006	0.000	0.122

Larger differences suggest that the model may not accurately capture the correlations between certain variables, while smaller differences indicate a closer fit between the observed and estimated correlation matrices. In our analysis, the values outside the main diagonal are close to zero, having a third decimal number magnitude. This observation affirms the effectiveness of the two-factor model employed in capturing the relationships among the variables under consideration.

[1] Notice that the p-value is not displayed for the 3-factors model, since (by setting $d = 6$ and $q = 3$) we have exact identification (i.e. the number of parameters estimated, $d(q+1) - q(q-1)/21 =$, is exactly equal to the number of distinct elements of the covariance matrix, $d(d+1)/2 = 21$) and therefore the degrees of freedom are equal to 0.

We proceed the analysis using two types of rotation: Varimax, an orthogonal rotation method where factors are constrained to be uncorrelated, and Oblimin, an oblique rotation method where factors are allowed to be correlated. In Varimax rotation, the factor loadings represent correlations between factors and observed variables. However, Varimax relaxes the assumption of uncorrelated factors, which may be unrealistic in some cases, aiming to improve the fit of the model.

The table 6 displays the factor loadings resulting from Varimax rotation, which aims at factors with a few large loadings and many near-zero loadings. Indeed, loadings for Factors 1 and 2 are relatively weak. *comten* shows a moderate loading on Factor 2, with weaker loading on Factors 1. *ceoten* exhibits a strong loading on Factor 2, accompanied by weaker loading on Factors 1. *sales* demonstrates a strong loading on Factor 1, while *profits* and *mktval* also show strong loadings on Factor 1. Weak loadings are observed for Factors 2 across these variables.

Table 6: Varimax factor loadings

	2-factor model		3-factor model		
Variable	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
salary	0.401	0.178	0.268	0.062	0.949
comten	0.123	0.328	0.118	0.339	-0.016
ceoten	-0.085	0.994	-0.129	0.981	0.123
sales	0.815	0.002	0.795	0.014	0.175
profits	0.978	0.062	0.973	0.089	0.134
mktval	0.933	0.087	0.911	0.106	0.164

The table 7 shows the factor loadings resulting from Oblimin rotation, which aims at simple structure with low correlation between factors. Similarly to the previous case, loadings for Factors 1 and 2 are minimal. *comten* displays moderate loadings on Factors 1 and 2. *ceoten* demonstrates a strong loading on Factor 2, accompanied by weaker loading on Factors 1. Sales exhibits a strong loading on Factor 1, whereas *profits* shows an almost perfect loading on Factor 1, along with weaker loading on Factors 2. *mktval* also displays a strong loading on Factor 1, with weaker loading on Factors 2.

Table 7: Oblimin factor loadings

	2-factor model		3-factor model		
Variable	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
salary	0.414	0.151	0.004	0.004	0.986
comten	0.148	0.319	0.187	0.333	-0.089
ceoten	-0.006	0.997	-0.007	0.996	0.007
sales	0.813	-0.053	0.786	-0.058	0.057
profits	0.980	-0.003	0.999	0.005	-0.030
mktval	0.937	0.024	0.927	0.026	0.011

Table 8: Communalities and uniqueness of the 2-factor model

	salary	comten	ceoten	sales	profits	mktval
Communalities	0.193	0.122	0.995	0.665	0.96	0.878
% variance explained	3.200	2.000	16.600	11.100	16.00	14.600
Uniqueness	0.807	0.878	0.005	0.335	0.04	0.122

Then, we examine the communalities and the uniqueness of each variable. The communality represents the proportion of variance in each variable that is accounted for by the extracted factors. It indicates how well each variable is explained by the factors in the model. For example, *salary* has a communality of 0.977, suggesting that 97.7% of its variance is explained by the extracted factors.

The second row provides the percentage of total variance in the dataset explained by each variable. For instance, *comten* explains 2.1% of the total variance in the dataset, while *ceoten* explains 16.6%.

Finally, the uniqueness represents the proportion of variance in each variable that is not explained by the factors. It indicates the extent to which each variable's variance is unique or specific to that variable. For example, *comten* has a uniqueness value of 0.871, indicating that 87.1% of its variance is not accounted for by the factors. Whereas, *ceoten* has a uniqueness value of 0.005, suggesting that only 0.5% of its variance is not explained by the factors.

5 Conclusions

In this paper we conducted two different types of analysis, FA and PCA, coming up with similar results. In both cases the dataset dimension is reduced up to 2 factors and 2 principal components and also the interpretations are close. The first element represents an overall performance of the company, while the second could be interpreted as a measure of the CEO's experience, thus suggesting that those latent variables or components are the main drivers of CEO's performance in firms.

Bibliography

Ozkan, N. (2011). Ceo compensation and firm performance: An empirical investigation of uk panel data. *European Financial Management*, *17*(2), 260–285.

Jensen, M. C., & Murphy, K. J. (1990). Ceo incentives: It's not how much you pay, but how.

Tariq, U. (2010). Ceo compensation: Relationship with performance and influence of board of directors.