# Regression with Principal Components

## Financial Econometrics (M524) Lecture Notes

Ke-Li Xu

Indiana University

October 17, 2022

# Multiple predictive regression

- Predictive regression with *multiple predictors:*

$$\begin{aligned} y_t &= \beta_0 + \underset{1\times(K+1)}{X'_{t-1}} \cdot \underset{(K+1)\times 1}{\beta} + u_t, \\ &= \beta_0 + x'_{t-1} \underset{K\times 1}{\beta_1} + \underset{1\times 1}{\beta_2}\, g_{t-1} + u_t, \end{aligned} \tag{1}$$

  where $X_{t-1}$ contains all predictors ($(K+1)$ by 1).
- We can be interested in $\beta$ ($\beta = 0$ means joint predictability using all predictors).
- In many cases, we are interested in a particular predictor (called $g_{t-1}$), with other predictors controled. (Is the predictive ability of $g_{t-1}$ already contained in other predictors? If yes, then controling would kill the predictability.)
- Thus the interest is in the single parameter $\beta_2$.

- The most natural method to estimate $\beta_2$ is to run multiple regression (1).
- However, this approach (kitchen-sink regression) can be problematic.

- 1. $x_{t-1}$ can be collinear, or close to be collinear (which is highly likely if you throw many variables into $x_{t-1}$, one of the issues when you have big data)
- 2. Even if collinearity is not a problem, the regression results of (1) are untrustable when $K$ is large (another issue of big data)

- We illustrate these issues using the manager sentiment example.

**Table 5**

Comparison with economic variables.

Panel A reports the in-sample estimation results for the univariate predictive regressions of the monthly excess market return on one of the lagged economic variables, $Z_t^k$,

$$R_{t+1}^m = \alpha + \psi Z_t^k + \varepsilon_{t+1}, \qquad k = 1, \ldots, 15,$$

where $R_{t+1}^m$ is the monthly excess aggregate stock market return (in percentage), and $Z_t^k$ is one of the 14 individual economic variables given in the first 14 rows of the first column or the ECON factor which is the first principal component factor extracted from the individual economic variables. See Section 2.2 for detailed definitions for economic variables. Panel B reports the in-sample estimation results for the bivariate predictive regressions on both the lagged manager sentiment index $S_t^{MS}$ and $Z_t^k$,

$$R_{t+1}^m = \alpha + \beta S_t^{MS} + \psi Z_t^k + \varepsilon_{t+1}, \qquad k = 1, \ldots, 15.$$

We report the regression coefficients, Newey–West $t$-statistics, and $R^2$s. The sample period is 2003:01–2014:12.

| | Panel A: Univariate regressions $R_{t+1}^m = \alpha + \psi Z_t^k + \varepsilon_{t+1}$ | | | Panel B: Bivariate regressions $R_{t+1}^m = \alpha + \beta S_t^{MS} + \psi Z_t^k + \varepsilon_{t+1}$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\psi$ (%) | $t$-stat | $R^2$ (%) | $\beta$ (%) | $t$-stat | $\psi$ (%) | $t$-stat | $R^2$ (%) |
| DP | 0.11 | 0.20 | 0.08 | −1.26 | −3.58 | 0.11 | 0.23 | 9.83 |
| DY | 0.31 | 0.63 | 0.61 | −1.24 | −3.54 | 0.25 | 0.56 | 10.1 |
| EP | −0.22 | −0.48 | 0.30 | −1.42 | −3.39 | 0.38 | 0.77 | 10.5 |
| DE | 0.21 | 0.42 | 0.26 | −1.34 | −3.37 | −0.25 | −0.49 | 10.1 |
| SVAR | −0.96 | −2.05 | 5.72 | −1.18 | −3.45 | −0.85 | −1.89 | 14.2 |
| BM | 0.20 | 0.49 | 0.25 | −1.33 | −3.52 | 0.43 | 1.04 | 10.9 |
| NTIS | 0.84 | 1.76 | 4.33 | −1.10 | −3.16 | 0.45 | 0.97 | 10.9 |
| TBL | −0.41 | −1.63 | 1.04 | −1.22 | −3.40 | −0.15 | −0.59 | 9.88 |
| LTY | −0.54 | −1.99 | 1.79 | −1.37 | −3.85 | −0.75 | −2.75 | 13.1 |
| LTR | 0.31 | 0.69 | 0.58 | −1.29 | −3.60 | 0.42 | 0.96 | 10.8 |
| TMS | 0.16 | 0.63 | 0.16 | −1.39 | −3.52 | −0.36 | −1.27 | 10.4 |
| DFY | −0.26 | −0.46 | 0.43 | −1.31 | −3.68 | −0.44 | −0.86 | 10.9 |
| DFR | 0.57 | 0.91 | 2.02 | −1.19 | −3.46 | 0.36 | 0.62 | 10.5 |
| INFL | 0.45 | 1.08 | 1.27 | −1.26 | −3.66 | 0.45 | 1.19 | 11.0 |
| ECON | 0.13 | 0.29 | 0.12 | −1.30 | −3.64 | 0.30 | 0.69 | 10.4 |

- There is actually *perfect collinearity* among 14 macro variables (DP, DY, ..., INFL)
- Correlation analysis looks fine

$>>$corr(macro)

| | DP | DY | EP | DE | SVAR | BM | NTIS | TBL | LTY | LTR | TMS | DFY | DFR | INFL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DP** | 1.00 | 0.97 | -0.62 | 0.78 | 0.60 | 0.29 | -0.54 | -0.40 | -0.40 | 0.06 | 0.24 | 0.85 | 0.02 | -0.28 |
| **DY** | 0.97 | 1.00 | -0.66 | 0.80 | 0.49 | 0.25 | -0.50 | -0.42 | -0.43 | 0.01 | 0.25 | 0.83 | 0.15 | -0.28 |
| **EP** | -0.62 | -0.66 | 1.00 | -0.98 | -0.30 | 0.31 | 0.31 | 0.17 | -0.17 | 0.06 | -0.34 | -0.68 | -0.23 | 0.14 |
| **DE** | 0.78 | 0.80 | -0.98 | 1.00 | 0.41 | -0.17 | -0.40 | -0.25 | 0.02 | -0.03 | 0.34 | 0.78 | 0.19 | -0.19 |
| **SVAR** | 0.60 | 0.49 | -0.30 | 0.41 | 1.00 | 0.14 | -0.44 | -0.16 | -0.03 | 0.18 | 0.18 | 0.66 | -0.23 | -0.28 |
| **BM** | 0.29 | 0.25 | 0.31 | -0.17 | 0.14 | 1.00 | 0.19 | -0.36 | -0.60 | 0.11 | 0.05 | 0.12 | -0.07 | -0.20 |
| **NTIS** | -0.54 | -0.50 | 0.31 | -0.40 | -0.44 | 0.19 | 1.00 | -0.17 | 0.05 | 0.01 | 0.26 | -0.59 | 0.11 | 0.04 |
| **TBL** | -0.40 | -0.42 | 0.17 | -0.25 | -0.16 | -0.36 | -0.17 | 1.00 | 0.64 | -0.03 | -0.85 | -0.24 | -0.09 | 0.19 |
| **LTY** | -0.40 | -0.43 | -0.17 | 0.02 | -0.03 | -0.60 | 0.05 | 0.64 | 1.00 | -0.17 | -0.13 | -0.13 | 0.01 | 0.24 |
| **LTR** | 0.06 | 0.01 | 0.06 | -0.03 | 0.18 | 0.11 | 0.01 | -0.03 | -0.17 | 1.00 | -0.08 | 0.03 | -0.42 | -0.13 |
| **TMS** | 0.24 | 0.25 | -0.34 | 0.34 | 0.18 | 0.05 | 0.26 | -0.85 | -0.13 | -0.08 | 1.00 | 0.22 | 0.12 | -0.08 |
| **DFY** | 0.85 | 0.83 | -0.68 | 0.78 | 0.66 | 0.12 | -0.59 | -0.24 | -0.13 | 0.03 | 0.22 | 1.00 | 0.13 | -0.36 |
| **DFR** | 0.02 | 0.15 | -0.23 | 0.19 | -0.23 | -0.07 | 0.11 | -0.09 | 0.01 | -0.42 | 0.12 | 0.13 | 1.00 | -0.11 |
| **INFL** | -0.28 | -0.28 | 0.14 | -0.19 | -0.28 | -0.20 | 0.04 | 0.19 | 0.24 | -0.13 | -0.08 | -0.36 | -0.11 | 1.00 |

- VIF (variance inflation factor) analysis shows the problem.

```
>>res=ols(macro(:,1),[ones(144,1) macro(:,2:4)]);
>> res.rsqr
ans =
1
>> res.beta
ans =
0.0000
-0.0000
1.0000
1.0000
```

- Examining the columns 1, 3 and 4, we actually have DP=EP+DE.

(Actually there is another collinear group: LTY-TBL=TMS)

## Notes on how to detect collinearity

Three methods to detect collinearity

1. Pairwise correlation matrix
2. VIF analysis. Regress each predictor $x_j$ on all other predictors, and get the R-square $R_j^2$. $R_j^2$ close to one indicates a problem. (Rule of thumb: $\text{VIF}_j = \frac{1}{1-R_j^2} > 10$.)
3. PC (principle components) analysis. Examine the eigenvalues of $(T-1)^{-1}\sum_{t=2}^{T} x_{t-1}x_{t-1}'$, where eigenvalues $\lambda_1, \lambda_2, ..., \lambda_K$ are sorted in decreasing order. The condition number

$$\kappa = \sqrt{\lambda_1/\lambda_K}.$$

A large $\kappa$ indicates a problem (Rule of thumb: $\kappa > 30$)

# Principal Components (PCs)

How do they work?

1. Get the eigenvalues and eigenvectors of

$$(T-1)^{-1}\sum_{t=2}^{T}(x_{t-1}-\overline{x})(x_{t-1}-\overline{x})'$$

(sample variance matrix). [MATLAB: `[E,v] = eig(S)`]

- Eigenvalues $\lambda_1, \lambda_2, ..., \lambda_K$ are sorted in decreasing order. Eigenvectors $w_1, w_2, ..., w_K$ are sorted in the same order. Eigenvectors should be such that

$$W'W = I_K,$$

where $W = (w_1, ..., w_K)$. In other words, eigenvectors are orthnormal: $w_i' w_j = 1$ if $i = j$, $w_i' w_j = 0$ if $i \neq j$.

2. All $K$ PCs:

$$PC_1 = \begin{pmatrix} x_1' w_1 \\ x_2' w_1 \\ \vdots \\ x_{T-1}' w_1 \end{pmatrix}, \; PC_2 = \begin{pmatrix} x_1' w_2 \\ x_2' w_2 \\ \vdots \\ x_{T-1}' w_2 \end{pmatrix}, \; ..., \; PC_K = \begin{pmatrix} x_1' w_K \\ x_2' w_K \\ \vdots \\ x_{T-1}' w_K \end{pmatrix}.$$

- Use all 14 variables in the manager sentiment data:

```
>>[E,v] = eig(cov(macro));
>>[v,ind] = sort(diag(v),'descend');
>>v
v =
 12.4086
 5.2660
 3.7567
 3.1870
 1.0268
 0.3405
 0.2420
 0.0867
 0.0373
 0.0058
 0.0008
 0.0005
 0.0000
 -0.0000
```

- Now we drop two variables (DE and LTY) and keep all other 12 variables

```
>>macro_good=macro(:,[1:3,5:8,10:14]);
>>[E,v] = eig(cov(macro_good));
>>[v,ind] = sort(diag(v),'descend');
>> v
v =
 12.3864
 5.1539
 3.5195
 3.1641
 0.4326
 0.2710
 0.1426
 0.0815
 0.0333
 0.0051
 0.0008
 0.0004
```

## Predictive regression with PCs

- Instead of running the naive multiple regression,

$$y_t = \beta_0 + x_{t-1}' \underset{K\times 1}{\beta_1} + \underset{1\times 1}{\beta_2} g_{t-1} + u_t,$$

  we do the following.

- Regression with $J$ PCs:

$$y_t = \beta_0 + f_{t-1}' \underset{J\times 1}{\gamma} + \underset{1\times 1}{\beta_2} g_{t-1} + \text{error}_t, \tag{2}$$

  where $f_{t-1}$ ($J$ by 1) contains the first $J$ PCs of $x_{t-1}$ :

$$f_{t-1} = \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ \vdots \\ f_{J,t-1} \end{pmatrix} = \begin{pmatrix} \text{PC}_{1,t-1} \\ \text{PC}_{2,t-1} \\ \vdots \\ \text{PC}_{J.t-1} \end{pmatrix}$$

- $f_{t-1}$ contains most information of $x_{t-1}$, but with a (much) smaller dimension. ($J$ is much smaller than $K$)
- It's legal to use all $K$ PCs (but nobody does it), which essentially reduces to the kitchen-sink regression.

- Regression with first PC:

$$y_t = \beta_0 + \underline{\gamma_1 PC_{1,t-1}} + \beta_2 g_{t-1} + \text{error}_t.$$

- Regression with first two PCs:

$$y_t = \beta_0 + \underline{\gamma_1 PC_{1,t-1} + \gamma_2 PC_{2,t-1}} + \beta_2 g_{t-1} + \text{error}_t.$$

- Regression with first three PCs:

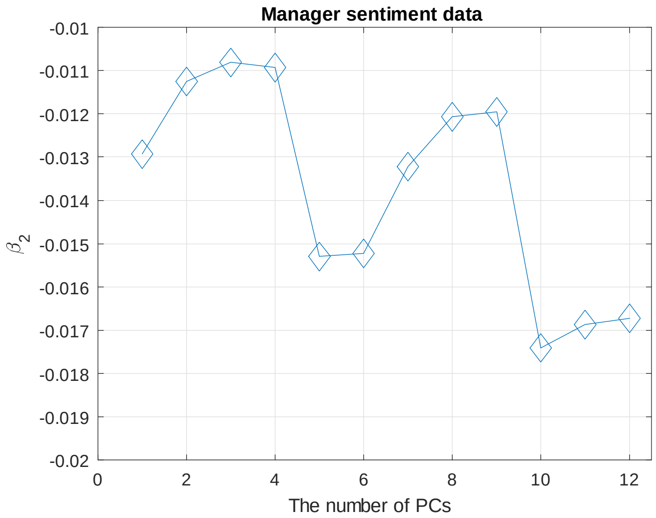
$$y_t = \beta_0 + \underline{\gamma_1 PC_{1,t-1} + \gamma_2 PC_{2,t-1} + \gamma_3 PC_{3,t-1}} + \beta_2 g_{t-1} + \text{error}_t,$$

  etc.

- The interest is still in $\beta_2$.

- Now we apply to the 12 controled variables.
- We want to see how the estimate of $\beta_2$ changes with the number of PCs used.

```
[E,v] = eig(cov(macro_good));
[v,ind] = sort(diag(v),'descend');
E = E(:,ind(1:12));
for j=1:12;
   pcaf = macro_good*E(:,1:j);
   res_pc=ols(er(2:end),[ones(143,1) MS(1:end-1)
pcaf(1:end-1,:)]);
   b2(j)=res_pc.beta(2);
end;
```




Kitchen-sink regression OLS (12 predictors): $\beta_2 = -0.0167$.

## How to estimate $\beta_1$ using PCs?

- If we only use 1 PC, revisiting (2):

$$\begin{aligned} y_t &= \beta_0 + f_{t-1} \underset{1\times 1}{\gamma} + \underset{1\times 1}{\beta_2} g_{t-1} + \text{error}_t \\ &= \beta_0 + x'_{t-1} \underbrace{w_1 \gamma}_{=\beta_1^{PC(1)}} + \beta_2 g_{t-1} + \text{error}_t . \end{aligned}$$

- If we use $J$ PCs,

$$\begin{aligned} y_t &= \beta_0 + f'_{t-1} \underset{J\times 1}{\gamma} + \underset{1\times 1}{\beta_2} g_{t-1} + \text{error}_t \\ &= \beta_0 + (W' x_{t-1})' \underset{J\times 1}{\gamma} + \beta_2 g_{t-1} + \text{error}_t \\ &= \beta_0 + x'_{t-1} \underbrace{W \gamma}_{=\beta_1^{PC(J)}} + \beta_2 g_{t-1} + \text{error}_t , \end{aligned} \quad (3)$$

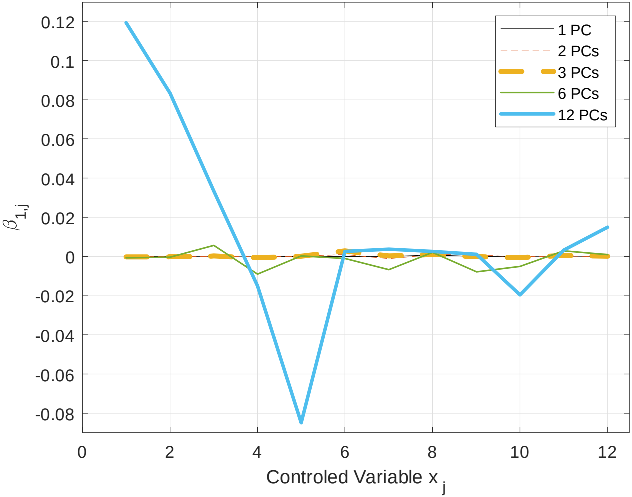
where $W'$ is $J \times K$ (the $j$-th row is $w'_j$).

- If we use all $K$ PCs, then this approach recovers the kitchen sink regression (the naive OLS regression).
- To show this, assume there is no intercept and $g_{t-1}$ (which makes notations easier),

$$\begin{aligned}
\widehat{\beta}_1^{PC(K)} &= W\widehat{\gamma} \\
&= W(\sum_{t=2}^{T} f_{t-1}f_{t-1}')^{-1}\sum_{t=2}^{T} f_{t-1}y_t \\
&= W(\sum_{t=2}^{T} W'x_{t-1}x_{t-1}'W)^{-1}\sum_{t=2}^{T} W'x_{t-1}y_t \\
&= (\sum_{t=2}^{T} x_{t-1}x_{t-1}')^{-1}\sum_{t=2}^{T} x_{t-1}y_t = \widehat{\beta}_1^{OLS}.
\end{aligned}$$

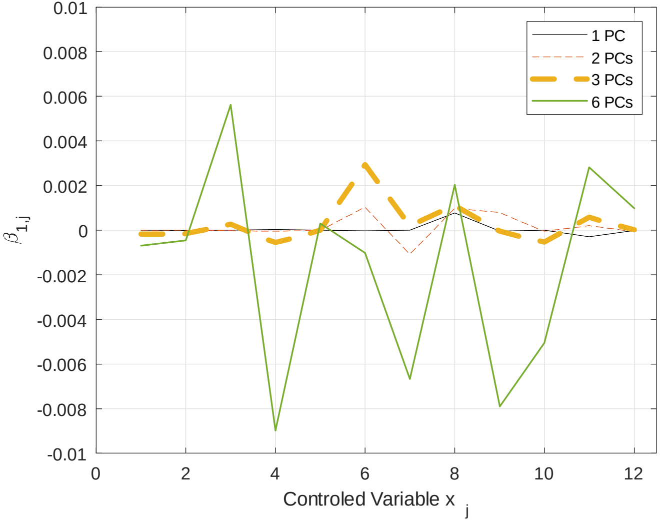
- The key is that $W$ is $K \times K$ matrix, so that we can take the inverse.

- Now come back the manager sentiment example. We use model (3) to estimate $\beta_1$, with $J$ PCs.
- The following graph shows for $J = 1, 2, 3, 6, 12$. (Recall $J = 12$ equals the kitchen-sink regression.)




>> res=ols(er(2:end),[ones(143,1) MS(1:end-1) macro_good(1:end-1,:)]);
>> res.beta'
0.9500, -0.0167, **0.1196, 0.0835, 0.0334, -0.0150, -0.0849, 0.0025, 0.0037, 0.0025, 0.0010, -0.0195, 0.0032, 0.0151**

- Zoom in the last graph to see $\beta_1$ estimates more clearly for $J = 1, 2, 3, 6$. ($J = 12$ is dropped in the following graph)



# Theoretical justification

- $K-$dim random vector $x_t$ (Recall that the actually used is $x_{t-1}$, and thus $f_{t-1}$)
- The covariance matrix
$$Cov(x_t) = \underset{K\times K}{\Sigma}.$$
- Principal Component Analysis (PCA): using *a few (J)* linear combinations of $x_t$ to explain the structure of $\Sigma$.
- Linear combinations:
$$f_{jt} = \underset{1\times K}{w_j'} \underset{K\times 1}{x_t}$$
where $j = 1, \cdots, J$ (a small number).

How to get the first 3 PCs?

1. Looking for $w_1$ :

$$\max_{w_1 \in \mathbb{R}^K} \underbrace{w_1' \Sigma w_1}_{= Var(f_{1t})}$$

$$\text{s.t. } w_1' w_1 = 1.$$

   You can stop here if you only want one PC.

2. Looking for $w_2$ :

$$\max_{w_2 \in \mathbb{R}^K} \underbrace{w_2' \Sigma w_2}_{= Var(f_{2t})}$$

$$\text{s.t. } w_2' w_2 = 1 \text{ and } Cov(f_{2t}, f_{1t}) = 0.$$

3. Looking for $w_3$ :

$$\max_{w_3 \in \mathbb{R}^K} \underbrace{w_3' \Sigma w_3}_{= Var(f_{3t})}$$

$$\text{s.t. } w_3' w_3 = 1 \text{ and } Cov(f_{3t}, f_{1t}) = Cov(f_{3t}, f_{2t}) = 0.$$

- In above, we use the standardization: $w_j' w_j = 1$.
- Otherwise we can always multiply $w_1$ by a large number, to make $w_1' \Sigma w_1$ larger.
- Note that $w_j$ are not usual "weights" (which require that each weight is non-negative, and all weights sum to one).

- There is a beatiful theory in linear algebra to make finding weights $w_j$ easy.
- Let $e_1, ..., e_K$ be eigenvectors of $\Sigma$ corresponding to eigenvalues $\lambda_1, ..., \lambda_K$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$.

*Spectral decomposition* of $\Sigma$:

$$\Sigma = PDP'$$

where $P = (e_1, ..., e_K)$ and $D = diag(\lambda_1, ..., \lambda_K)$.

Since $\Sigma$ is symmetric, we can make $P$ to be an orthogonal matrix (i.e. $PP' = P'P = I_K$).

**Theorem**:

$$w_j = e_j,$$

for $1 \leq j \leq K$.

**Proof of the Theorem.**

- Note that
$$\max_{w'w=1} w'\Sigma w = \lambda_1 \text{ (attained when } w_1 = e_1\text{)}.$$
- It follows from
$$w'\Sigma w = w'PDP'w \overset{z=P'w}{=} z'Dz = \sum_{k=1}^{K} \lambda_k z_k^2 \leq \lambda_1 \sum_{k=1}^{K} z_k^2 = \lambda_1,$$
since $z'z = w'PP'w = w'w = 1$. This means no $w$ can make $w'\Sigma w$ larger than $\lambda_1$.
- This upper bound $\lambda_1$ is actually *attained* when $w_1 = e_1$ :
$$e_1'\Sigma e_1 = e_1'PDP'e_1 = \iota_1'D\iota_1 = \lambda_1.$$

Here $\iota_k$ is the $k$-th column of $I_K$, for $k = 1, 2, ..., K$.

$$\begin{aligned} \iota_1' &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \\ \iota_2' &= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \end{aligned}$$

**Proof of the Theorem (continued).**
Now search for $w_2$.

- Since $Cov(f_{2t}, f_{1t}) = Cov(w_2' x_t, e_1' x_t) = w_2' \underbrace{\Sigma e_1} = w_2' \underbrace{\lambda_1 e_1}$, so requiring $Cov(f_{2t}, f_{1t}) = 0$ means requiring $w_2' e_1 = 0$.
- So

$$w_2' \Sigma w_2 = w_2' PDP' w_2 \overset{z=P'w_2}{=} z' Dz = \sum_{k=2}^{K} \lambda_k z_k^2 \leq \lambda_2 \sum_{k=2}^{K} z_k^2 = \lambda_2.$$

Above we use properties of $z = P' w_2$ :

1. $z = (0, z_2, ..., z_K)$.
2. $|z| = 1$.

- This upper bound $\lambda_2$ is actually *attained* when $w_2 = e_2$ :

$$e_2' \Sigma e_2 = e_2' PDP' e_2 = \iota_2' D \iota_2 = \lambda_2.$$

The proof is done by proceeding similarly. $\square$

## Properties of PCs

- All PCs are uncorrelated:

$$Cov(f_{jt}, f_{kt}) = Cov(w_j' x_t, w_k' x_t) = w_j' \underbrace{\Sigma w_k} = w_j' \underbrace{\lambda_k w_k} = 0,$$

  for $j, k = 1, ..., K$.
- The variances of PCs are $\lambda_j$:

$$Var(f_{jt}) = w_j' \underbrace{\Sigma w_j} = w_j' \underbrace{\lambda_j w_j} = \lambda_j.$$

**A trivial case** (when PCA are not quite useful)

- If $\Sigma$ is diagonal, $\Sigma = diag(\sigma_1^2, ..., \sigma_K^2)$, then $(e_1, ..., e_K) = (\iota_1, ..., \iota_K)$.
- In this case, the $i$-th PC is just the element of $x_t$ with $i$-th largest variance.
- Thus, the PCs are not useful in dimension reduction (the first $J$ PCs doesn't contain information more than $J$ variables)

- Determining the number of PCs: *scree plot*
- Look for an elbow (or a bend; a point with the sharpest/smallest angle) in the scree plot.
- Selecting $r = 2$ in the manager sentiment data