Generalized Linear Models For The Covariance Matrix of Longitudinal Data

How To Lift the "Curses" of Dimensionality and Positive-Definiteness?

Mohsen Pourahmadi
Division of Statistics
Northern Illinois University

Department of Statistics
UW, Madison

April 5, 2006

Outline

I. Prevalence of Covariance Modeling / GLM

II. Correlated Data; Example, Sample Cov. Matrix

III. Linear and Log-Linear Covariance Models

IV. Generalized Linear Models (GLM)

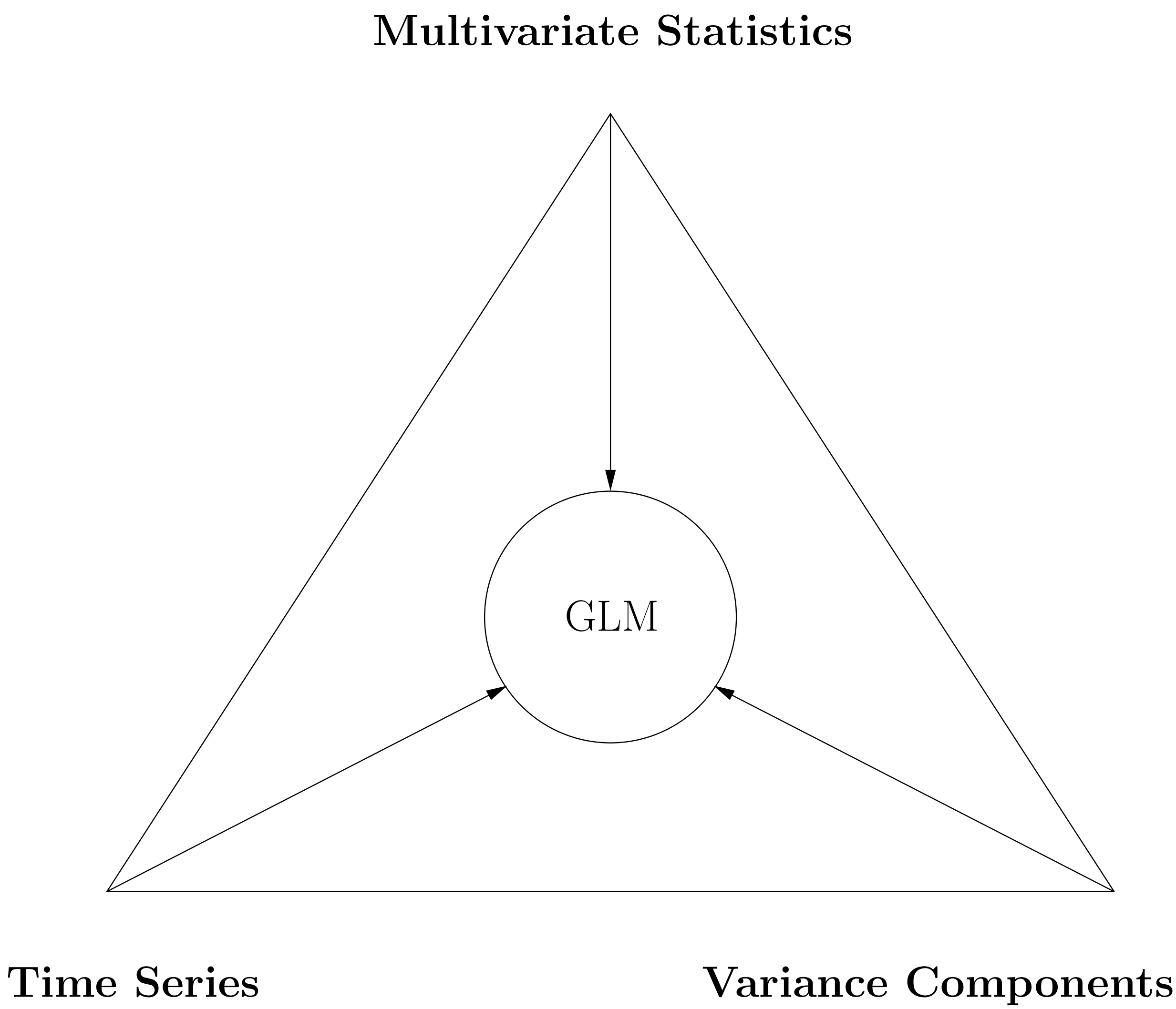
- Motivation (Link Function)
- Model Formulation (Regressogram)
- Estimation and Diagnostics
- Data Analysis

V. Bayesian, Nonparametric, LASSO, . . .

VI. Conclusion

I. Prevalence of Cov. Modeling / GLM

- Covariance matrices have been studied for over a century.
- Parsimonious cov. is needed for efficient est. and inference in regression and time series analysis, for prediction, portfolio selection, assessing risk in finance (ARCH-GARCH), $\cdots$.



- Nelder and Wedderburn's (1972) GLM unifies
  - normal linear regressions (Legendre, 1805; Gauss, 1809),
  - logistic (probit, ...) binary regressions, Poisson regressions, log-linear models for contingency tables,
  - variance component estimation using ANOVA sum of squares,
  - joint modelling of mean and dispersion (Nelder & Pregibon, 1987)
  - survival function (McCullagh & Nelder, 1989),
  - spectral density estimation in time series using periodogram ordinates (Cameron & Tanner, 1987),
  - generalized additive models (Hastie & Tibshirani, 1990); non-parametric methods,
  - hierarchical GLMs (Lee & Nelder, 1996),
  - Bayesian GLMs (Dey et al. 2000).

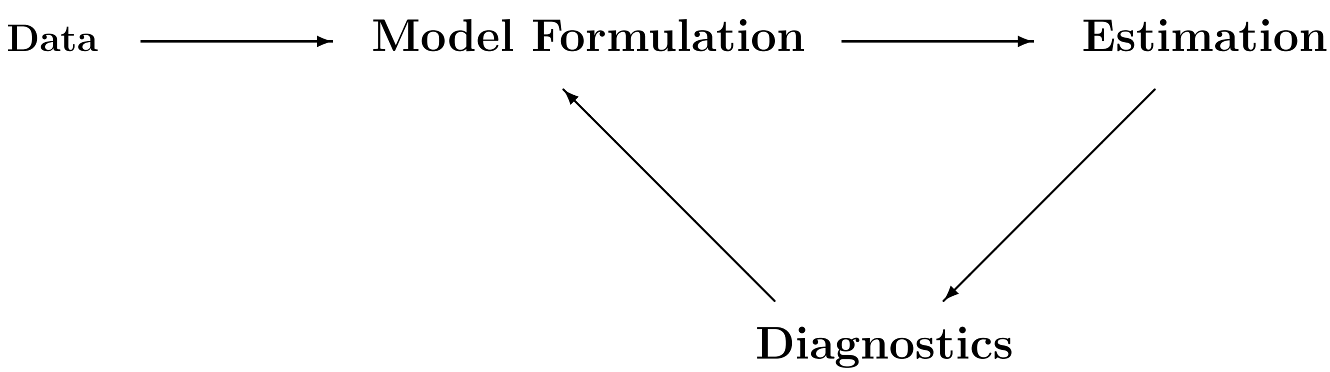
•• The Success of GLM Is Mainly Due to Using

I. unconstrained (canonical) parameters,

II. models that are additive in the covariates,

III. MLE / IRWLS or their variants.

Goal: Model a covariance matrix using covariates similar to modeling the mean vector in regression analysis.



- **Generalized Linear Models** for the mean vector $\mu = E(Y)$:

$$g(\mu) = X\beta,$$

  where g acts *componentwise* on the vector μ.

– GLM for the covariance matrix

$$\Sigma = E(Y-\mu)(Y-\mu)',$$

  requires finding $g(\cdot)$ so that entries of $g(\Sigma)$ are **unconstrained**, then one may set

$$g(\Sigma) = Z\alpha.$$

- $g(\cdot)$ acting *componentwise* cannot remove the **positive-definiteness constraint**.

$$c' \; \Sigma \; c = \sum_i \sum_j c_i c_j \sigma_{ij} > 0, c_i \; real.$$

- $g(\cdot)$ is not necessarily unique, the one with the most **interpretable** parameters is preferred.

II. Correlated Data

- **Ideal Shape of Correlated Data:** Many Short Time Series.

Occasions

Units		1	2	$\cdots$	t	$\cdots$	n
	1	y_{11}	y_{12}	$\cdots$	y_{1t}	$\cdots$	y_{1n}
	2	y_{21}	y_{22}	$\cdots$	y_{2t}	$\cdots$	y_{2n}
Units	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	i	$(y_{i1}$	y_{i2}	$\cdots$	y_{it}	$\cdots$	$y_{in}) = Y_i$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	m	y_{m1}	y_{m2}	$\cdots$	y_{mt}	$\cdots$	y_{mn}

Special Cases in Increasing Order of Difficulty:

I. **Time Series Data:** $m = 1, n$ large.

II. **Multivariate Data:** $m > 1$, n small to moderate; rows are indep.
Longitudinal Data, Cluster Data.

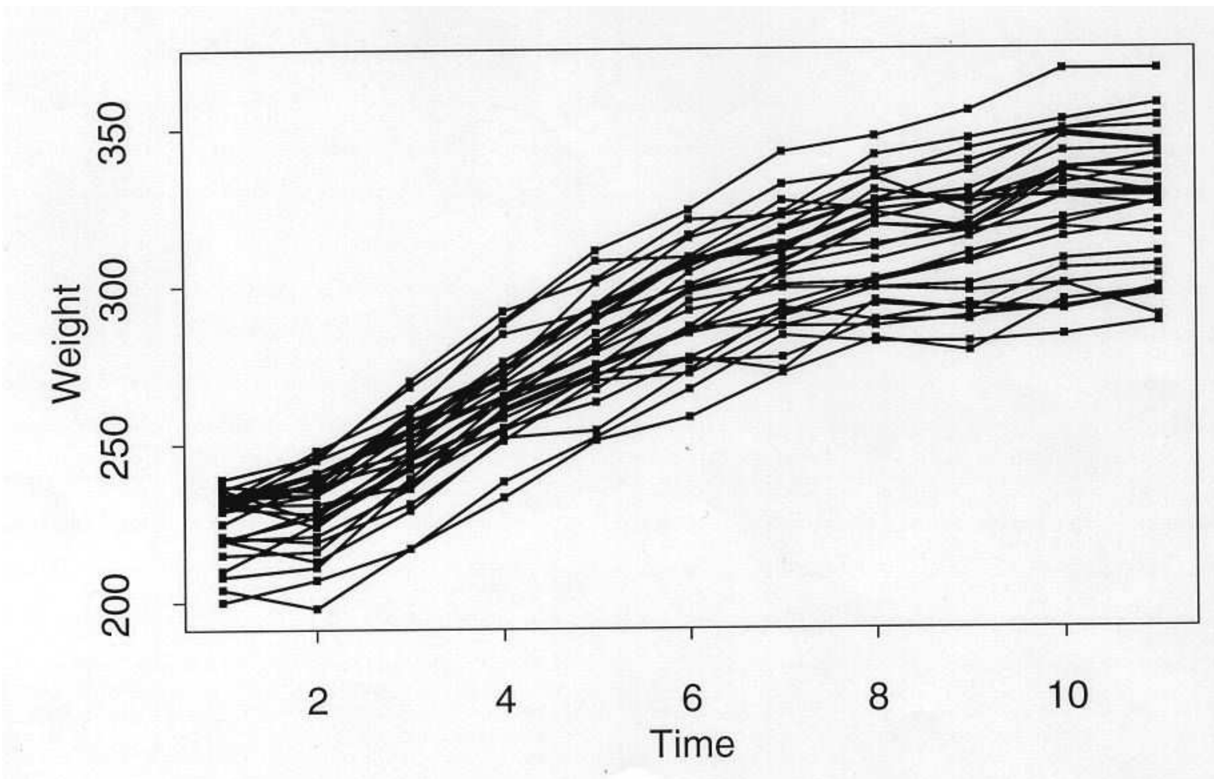
III. **Multiple Time Series:** $m > 1$, n large, rows are dependent.
Panel Data

IV. **Spatial Data:** m & n are hopefully large, rows are dependent.

- "Time" or "order" is required for the GLM / Cholesky decomposition of the covariance matrix of the data.

Example: Kenward's (1987) Cattle Data:

An experiment to study effect of treatments on intestinal parasites. $m = 30$ animals received treatment A, they were weighed $n = 11$ times, the first 10 measurements were made at two-week intervals and the final measurement was made after a one week interval. The times are rescaled to $t_j = 1, 2, \cdots, 10, 10.5$.



- Clearly, variances **increase** over time,
- Are equidistant measurements equicorrelated?
- Is the correlation matrix stationary (Toeplitz)?

TABLE 1. Sample variances are along the main diagonal and correlations are off the main diagonal.

106										
.82	155									
.76	.91	165								
.66	.84	.93	185							
.64	.80	.88	.94	243						
.59	.74	.85	.91	.94	284					
.52	.63	.75	.83	.87	.93	306				
.53	.67	.77	.84	.89	.94	.93	341			
.52	.60	.71	.77	.84	.90	.93	.97	389		
.48	.58	.70	.73	.80	.87	.88	.94	.96	470	
.48	.55	.68	.71	.77	.83	.86	.92	.96	.98	445

- The correlations **increase** along the subdiagonals (the learning effect) and **decrease** along the columns.
- **Stationary** (Toeplitz) covariance is not advisable for such data.
- **SAS PROC MIXED** and **lme** provide a long menu of covariance structures, such as CS, AR, ..., to choose from. Very popular in longitudinal data anlysis.
- How to view **larger covariance matrices**, like the 102×102 cov. matrix of the Call Center Data?

- **The Sample Covariance Matrix**

Balanced Data: $Y_1, \ldots, Y_m$ are i.i.d. $N(\mu, \Sigma)$.

Sample Cov. Matrix: $S = \frac{1}{m} \sum_{i=1}^{m} (Y_i - \bar{Y})(Y_i - \bar{Y})'$.

The **Spectral Decomposition** $PSP' = \Lambda$, plays a central role in **Reducing** the **Dimension** or the **No. of parameters** in Σ: PCA, Factor Analysis, ... (Pearson, 1901; Hotelling, 1933).

R. Boik (2002). Spectral models for covariance matrices. *Biometrika*, 89, 159-182.

Eigenvalues:

- **Improving** S
  - **Stein's Estimator** (1961+): Shrinks the eigenvalues of S to reduce the **risk**.

  In finance and microarray data, usually $n >> m$, and S is **singular**.

  (Ledoit et al., 2000+): $\hat{\Sigma} = \alpha S + (1 - \alpha) I, \quad 0 \leq \alpha \leq 1$.

  Ledoit & Wolf (2004). **Honey, I shrunk the sample covariance matrix.** *J. Portfolio Management., 4, 110-119.*

III. Linear & Log-Linear Models

History: Linear Covariance Model (LCM)

	$\Sigma = (\sigma_{ij})$	$\Sigma^{-1} = (\sigma^{ij})$
Edgeworth (1892)		Parameterized $N(0, \Sigma)$ in terms of entries of the **concentration** matrix.
Slutsky (1927)	**Banded:** Stationary MA(q)	
Yule (1927)		Banded: **Stationary** AR(p), $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$.
Gabriel (1962)		Banded: **Nonstationary** AR(p) or **ante-dependence** (AD) structure. $y_t = \phi_{t1} y_{t-1} + \phi_{t2} y_{t-2} + \varepsilon_t$,
Dempster (1972)		**Sparse:** Certain $\sigma^{ij} = 0$. Σ^{-1}, the **natural param.** of MVN. Graphical Models. Matrix completion problem in LA.
Anderson (66, 69, 73)	**Linear**	**Linear Models**

Anderson, T.W. (1973). Asym. eff. est. of cov. matrices with linear structure. *Ann. of Stat.*, 135-141.

- Anderson's **Linear Covariance Model** (LCM):

$$\Sigma^{\pm 1} = \alpha_1 U_1 + \cdots + \alpha_q U_q,$$

where U_i's are symmetric matrices (covariates) and α_i's are **constrained** parameters so that Σ is positive- definite.

  – **Every Σ has a representation as LCM:**

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \sigma_{11} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \sigma_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \sigma_{12} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

  it includes virtually all time series models, mixed models, factor models, multivariate GARCH models, … .

  – A major drawback of LCM is the *constraint* on $\alpha = (\alpha_1, \ldots, \alpha_q)$, which amounts to the **root constraint** in time series, and **nonnegative variance/coefficients** in variance components, factor analysis, etc.

- LCM and many other techniques pursue a **term-by-term** modeling of the covariance matrix, Prentice & Zhao (1991); Diggle & Verbyla (1998); Yao, Müller and Wang (2005), … .

- When the LCM est. $\hat{\Sigma}$ is **not positive-definite**, the advice is to replace its negative eigenvalues by zero. How good is this modified estimator?

- **Log-Linear Models** (LLM):
  Motivation: Σ is pd $\Leftrightarrow$ $\log \Sigma$ is real and symmetric.
  Set
  $$\log \Sigma = \alpha_1 U_1 + \cdots + \alpha_q U_q,$$
  where U_i's are as in LCM and α_i's are unconstrained.

  Q. How does one define $\log \sum$?

  Ans. $\log \Sigma = A \Leftrightarrow \Sigma = e^A = I + \frac{A}{1!} + \frac{A^2}{2!} + \cdots,$

  OR

  If $\Sigma = P'\Lambda P$, then $\log \Sigma = P' \log \Lambda P$.

– **Variance heterogeneity** (Cook and Weisberg, 1983):

  When Σ is **diagonal**, LLM reduces to regression modeling of variance heterogeneity.

– A major drawback of LLM, in general, is the lack of *statistical interpretability* of entries of $\log \Sigma$.

Ex. If $\log \Sigma = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$, then

$$\sigma_{11} = \frac{1}{2\sqrt{\Delta}} \exp\left(\frac{\alpha+\gamma}{2}\right) \left\{\sqrt{\Delta}\, u^{+} - (\alpha-\gamma)u^{-}\right\},$$

where

$$\Delta = (\alpha-\gamma)^2 + 4\beta^2,$$

$$u^{\pm} = \exp\left(\frac{\sqrt{\Delta}}{2}\right) \pm \exp\left(-\frac{\sqrt{\Delta}}{2}\right).$$

1. Leonard & Hsu (1992). Bayesian inference for a covariance matrix. *Ann. of Stat.*, 20, 1669-1696.

2. Chiu, Leonard & Tsui (1996). The matrix-logarithm covariance model. *JASA*, 91, 198-210.

3. Pinheiro & Bates (1996). Unconstrained parameterizations for variance-covariance matrices.
Stat. Comp., 289-296.

IV. GLM for Cov. Matrices

- **Motivation: Time Series & Cholesky Dec.**
  The AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t,$$

for $t = 1, 2 \ldots, n$ can be written as a **linear model**:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -\phi_1 & 1 & 0 & \cdots & \cdots & 0 \\ -\phi_2 & -\phi_1 & 1 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\phi_2 & -\phi_1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix} + \begin{bmatrix} \phi_2 & \phi_1 \\ 0 & \phi_2 \\ \cdots & \cdots \\ 0 \cdots & 0 \\ \vdots & \vdots \\ 0 \cdots & 0 \end{bmatrix} \begin{bmatrix} y_{-1} \\ y_0 \\ \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

Or

$$TY = \varepsilon + Ce.$$

Then, it follows that

$$\begin{aligned} T\text{cov}(Y)T' &= \sigma^2 I_n + \begin{pmatrix} C_1 \text{cov}(e) C_1' & 0 \\ 0 & 0 \end{pmatrix} \\ &= \text{A nearly } \textbf{diagonal} \text{ matrix.} \end{aligned}$$

- In general, ARMA models can be seen as means to "nearly" **diagonalize** a covariance matrix via a **structured unit lower triangular matrix** T. The cov. of the "initial values" is the only obstacle.

- **Reg./G.-Schmidt/Chol./Szegö/Bartlett/DL/KF**
  Regress y_t on its predecessors:

$$y_t = \phi_{t,t-1} y_{t-1} + \cdots + \phi_{t1} y_1 + \varepsilon_t,$$

y_1	y_2	y_3	$\cdots$	y_{n-1}	y_n
σ_1^2					
ϕ_{21}	σ_2^2				
ϕ_{31}	ϕ_{32}	σ_3^2			
$\vdots$	$\vdots$		$\ddots$		
ϕ_{n1}	ϕ_{n2}	$\cdots$	$\cdots$	$\phi_{n,n-1}$	σ_n^2

in matrix form

$$\begin{bmatrix} 1 & & & & \\ -\phi_{21} & 1 & & & \\ -\phi_{31} & -\phi_{32} & 1 & & \\ \vdots & & & \ddots & \\ -\phi_{n1} & -\phi_{n2} & \cdots & -\phi_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

- ϕ_{tj} and $\log \sigma_t^2$ are the unconstrained **generalized autoregressive parameters** (GARP) and **innovation variances** (IV) of Y or Σ .

- This can reduce the unintuitive task of **covariance modeling** to that of a **sequence of regressions** (with varying-order and varying-coefficients).

- **Generalized Linear Models :**

For $\sum$ pd, there are unique T and D with positive diagonal entries such that

$$T \sum T' = D.$$

Note. $\sum \longleftrightarrow (T, D)$.

Link functions: $g(\sum) = 2I - T - T' + logD$,

a symmetric matrix with unconstrained and statistically meaningful entries.

Strategy: Model T "linearly" as in Anderson (1966)

log D " " " Leonard et al. (92,96).

or replace "linearly" by parametrically/nonparam. / Bayesian $\cdots$.

Bonus: The estimate $\hat{\Sigma} = \hat{T}^{-1}\hat{D}\hat{T}'^{-1}$ is always pd, here $\hat{T}$ and $\hat{D}$ are estimates of **parsimoniously** modeled T and D.

Q. How to identify parsimonious models for (T, D) ?

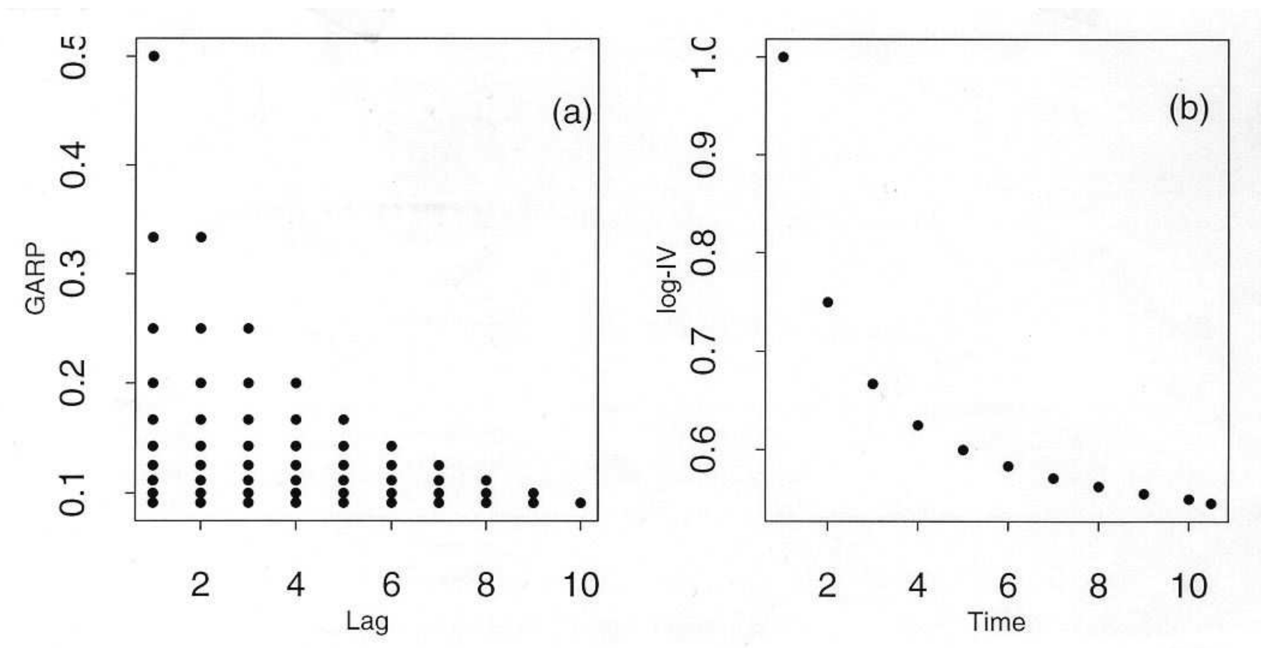
Ans. (i) Use covariates,

(ii) Shrink to zero the smaller entries of T using penalized likelihood, various priors (Smith & Kohn, 02; Huang, Liu, Pourahmadi, Liu, 06).

- **Model Formulation: Regressogram*** :

  Plays roles similar to the correlogram in time series. For a $t \geq 2$, simply plot the GARP $\phi_{t,j}$ vs the lags $j = 1, 2, \cdots, t-1$, and plot $\log \sigma_t^2$ vs $t = 1, 2, \cdots, n$.

  Ex. Compound Symmetry Covariance ($\rho = .5, \sigma^2 = 1$):



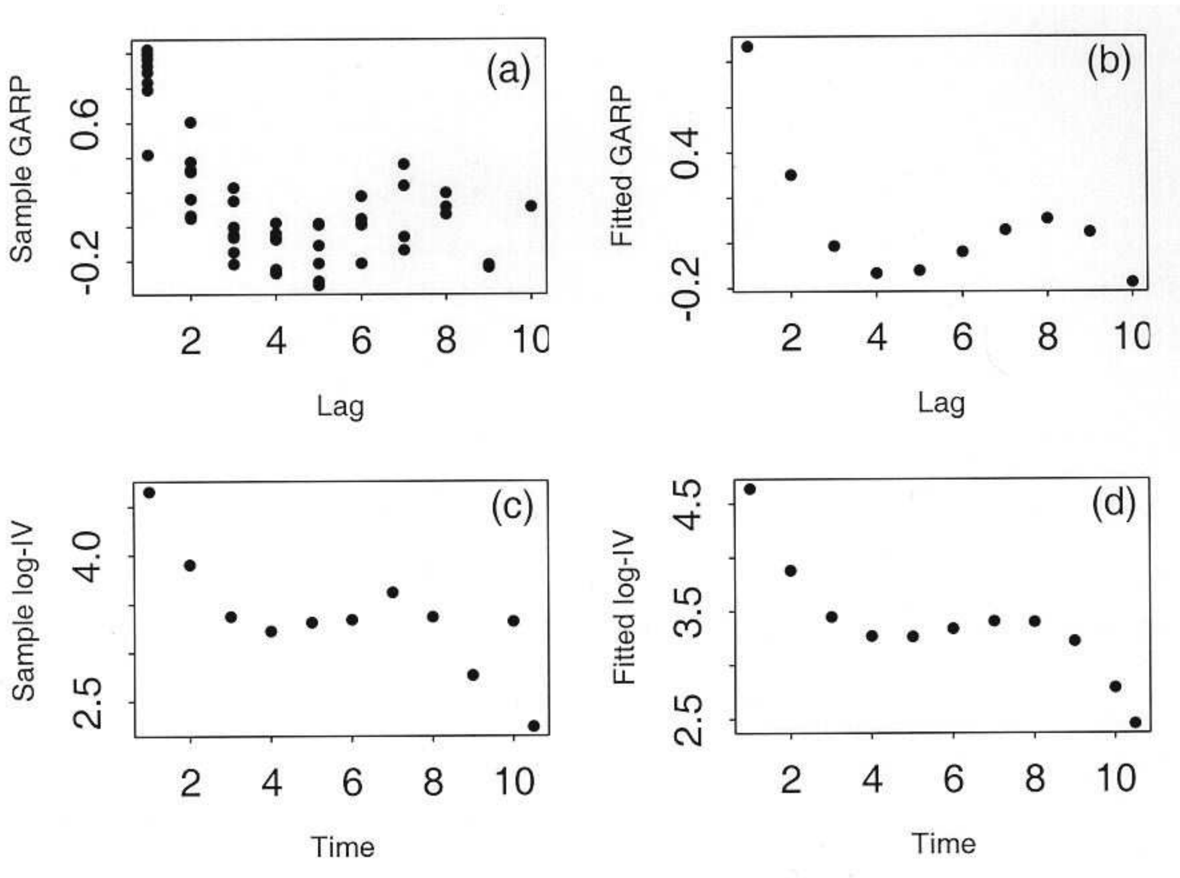
Ex. $AR(p)$, $AD(p)$.

Other Graphical Tools: Scatterplot Matrices; Variogram (Diggle, 1988); Partial Scatterplot Matrices (Zimmerman, 2000)

Lorelogram (Heagerty & Zeger, 1998).

⋮

*Tukey (1961). Curves as parameters, and touch estimation. 4th Berkeley Symp., 681-694.



Sample and Fitted Regressograms for the Cattle Data. (a) Sample GARP, (b) Fitted GARP, (c) Sample log-IV and (d) Fitted log-IV.

Example. Cattle Data

Table 2: Values of L_{max}, NO. of parameters and BIC for several models. The last four rows are from Zimmerman & Núñez-Antón (97).

Model	L_{max}	NO. of Parameters	BIC
Unstructured Σ	-1019.69	66	75.35
Poly (3,3)	-1049.01=L_1	8	70.84
Poly (3,2)	-1080.08=L_0	7	72.80
Poly (3,1)	-1131.61	6	76.09
Poly (3,0)	-121235	5	81.59
Poly (3)	-1377.43	4	92.28
Unstructured AD(2)	-1035.98	30	72.47
Structured AD(2)	-1054.13	8	71.18
Stationary AR(2)	-1062.89	3	71.20
Structured AD(2) with $\lambda_1 = \lambda_2 = 1$	-1054.20	6	70.96

Likelihood Ratio Test:

$$2(L_1 - L_0) = 62.14 \sim \chi_1^2,$$

so $(t-j)^3$ is kept in the model.

Regressogram suggests cubic models for the GARP and log IV for the cattle data with 8 param. For $t = 1, 2, \cdots, 11$, and $j = 1, 2, \cdots, t-1$.

$$\begin{cases} \log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \epsilon_{t,v}, \\ \\ \phi_{t,j} = \gamma_1 + \gamma_2 (t-j) + \gamma_3 (t-j)^2 + \gamma_4 (t-j)^3 + \epsilon_{t,d}. \end{cases}$$

In general, these and μ_t can be modeled as

$$\mu_t = x_t' \beta, \log \sigma_t^2 = z_t' \lambda, \phi_{t,j} = z_{t,j}' \gamma,$$

where $x_t, z_t, z_{t,j}$ are $p \times 1, q \times 1$ and $d \times 1$ vectors of covariates, $\beta = (\beta_1, \cdots, \beta_p)'$, $\lambda = (\lambda_1, \cdots, \lambda_q)'$ and $\gamma = (\gamma_1, \cdots, \gamma_d)'$ are parameters corresponding to the means, innovation variances and correlations.

Pourahmadi (1999). Joint mean-covariance models with applications to longitudinal data; Unconstrained parameterization. Biometrika, 86, 677-690.

- **Estimation: MLE of $\theta = (\beta', \lambda', \gamma')$:**

  The normal likelihood function has three representations corresponding to the three components of θ:

$$
\begin{aligned}
-2L(\beta, \lambda, \gamma) &= m \log|\Sigma| + \sum_{i=1}^{m} (Y_i - X_i\beta)'\Sigma^{-1}(Y_i - X_i\beta) \\
&= m \sum_{t=1}^{n} \log \sigma_t^2 + \sum_{t=1}^{n} \frac{RSS_t}{\sigma_t^2} \\
&= m \sum_{t=1}^{n} \log \sigma_t^2 + \sum_{i=1}^{m} \{r_i - Z(i)\gamma\}' D^{-1} \{r_i - Z(i)\gamma\},
\end{aligned}
$$

  where $r_i = Y_i - X_i\ \beta = (r_{it})_{t=1}^n, RSS_t$ and $Z(i)$ depend on r_i and other covariates and parameter values.

- For the estimation algorithm and asymptotic distribution of the MLE of θ, see Theorem 1 in

  Pourahmadi (2000). MLE of GLMs for MVN covariance matrix. Biometrika, 87, 425-435.

- **MLE of irregular and sparse longitudinal data**;

  Ye and Pan (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika*, to appear.

  &

  Holan and Spinka (2006).

V. **Other Developments (Bayesian, Nonparametric, LASSO, . . .)**

- **Covariate-selection** (Pan & MacKenzie, 2003). Relied on AIC & BIC, not the regressogram.

- **Random effects selection** (Chen & Dunson, 2003). Used
$$\Sigma = DLL'D.$$

- **Bayesian** (Daniels & Pourahmadi, 02; Kohn and Smith 02):
$$g(\Sigma) \sim N(\quad , \quad).$$

- **Nonparametric** (Wu & Pourahmadi, 2003). Smooth (T, D) using
$$\log \sigma_t^2 = \sigma^2(t/n),$$
$$\phi_{t,t-j} = f_j(t/n),$$
where $\sigma^2(\cdot)$ and $f_j(\cdot)$ are smooth functions on $[0, 1]$.

– Amounts to approximating T by the **varying-coefficients** AR:
$$y_t = \sum_{j=1}^{p} f_j(t/n) y_{t-j} + \sigma(t/n)\varepsilon_t.$$

– This formulation is fairly standard in the nonparametric regression literature where one pretends to observe $\sigma^2(\cdot)$ and $f_j(\cdot)$ on **finer grids** as n gets larger.

•• **Penalized likelihood** (Huang, Liu, MP & Liu, 06).

- Log-likelihood function

$$-2L(\gamma, \lambda) = \ m \log |\Sigma| + \sum_{i=1}^{m} Y_i' \Sigma^{-1} Y_i$$

- Penalized likelihood with L_p penalty,

$$-2L(\gamma, \lambda) + \alpha \sum_{t=2}^{n} \sum_{j=1}^{t-1} |\phi_{tj}|^p,$$

  where $\alpha > 0$ is a tuning parameter.

- $p = 2$, corresponds to **Ridge Regression**,

- $p = 1$,“ ” Tibshirani’s (1996) **LASSO** (**L**east **a**bsolute **s**hrinkage and **s**election **o**perator).

– Use of L_1 norm, allows LASSO to do **variable selection**–it can produce coefficients that are **exactly** zero.

– LASSO is most effective when there are a small to moderate number of moderate-sized coefficients.

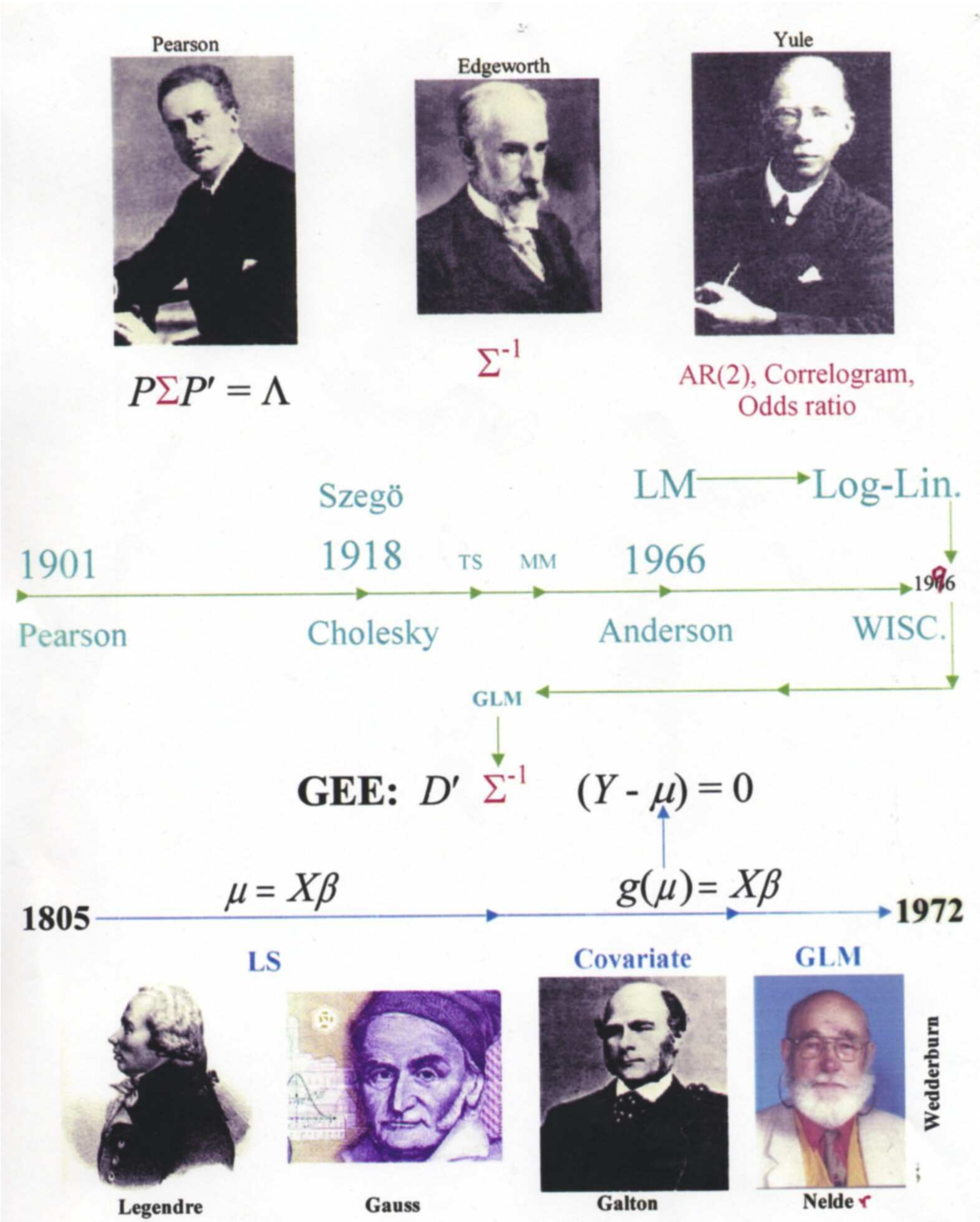
- **Bridge Regression** ($p > 0$), Frank & Friedman (1993), Fu (1998); Fan & Li (2001).

- For the Call Center Data with $n = 102$ and 5151 parameters in T, about 4144 are essentially zero.

  L. Brown et al. (2005). Statistical Analysis of a Telephone Call Center: A Queueing Science Perspective. JASA, 36-50.

•• Simultaneous Modeling of Several Covariance Matrices
(Pourahmadi, Daniels, Park, JMA, 2006).
Applications to Model-Based Clustering

Classification, Finance, $\cdots$.


Pearson
Edgeworth
Yule
PΣP′ = Λ
Σ-1
AR(2), Correlogram,
Odds ratio
Szegö
LM
Log-Lin.
1901
1918
TS
MM
1966
1976
Pearson
Cholesky
Anderson
WISC.
GLM
GEE: D′ Σ-1 (Y - μ) = 0
1805
μ = Xβ
g(μ) = Xβ
1972
LS
Covariate
GLM
Wedderburn
Legendre
Gauss
Galton
Nelder

REFERENCES

Anderson, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1**, 135-141.

Chen, Z. and Dunson, D. (2003). Random effects selection in linear mixed models. *Biometrics*, 59, 762-769.

Dempster, A.M. (1972). Covariance selection, *Biometrics*, **28**, 157-175.

Diggle, P.J., Verbyla, A.P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, **54**, 401-415.

Gabriel, K.R. (1962). Ante-dependence analysis of an ordered set of variables. *Ann. Math. Statist.*, **33**, 201-212.

Kenward, M.G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics*, **36**, 296-308.

Pan, J.X. and Mackenzie, G. (2003). Model selection for joint mean-covariance structures in longitudinal studies. *Biometrika*, 90, 239-249.

Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*, John Wiley, New York.

Pourahmadi, M. and Daniels, M. (2002). Dyanamic conditionally linear mixed models for longitudinal data. *Biometrics*, **58**, 225-231.

Roverato, A. (2000). Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika*, **87**, 99-112.

Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional data analysis for sparse longitudinal data. *JASA*, 100, 577-590.

Zimmerman, D.L. and V. Núñez-Antón (1997). Structured antedependence models for longitudinal data. In *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*, 63-76 (T.G. Gregoine, et al., eds.) Springer-Verlag, New York.