Empirical Methods in Finance MFE230E
Week 8: Principal component analysis and Kalman filter

Martin Lettau

Spring 2019

Haas School of Business

Outline

1. **PCA**
   - **PCA: Theory**
   - **PCA: Examples**
2. State space models and the Kalman filter
   - Kalman filter
   - The general state space model
3. Markov Chain Monte Carlo (MCMC) estimation of SV models

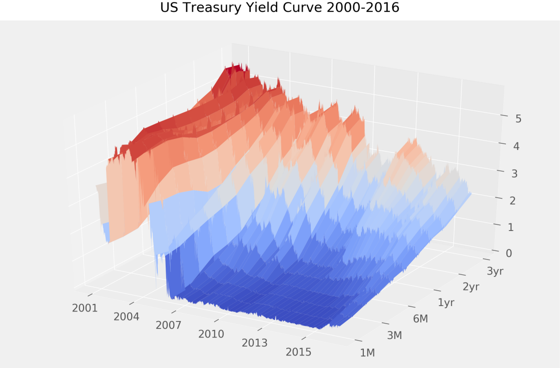
- Sometimes the amount of available data is overwhelming
- Suppose we have a "large data" set with N time series
- Example: Forecast GDP with N variables, if $N > T$ OLS does not apply
- The St. Louis Fed maintains a database with over 212,000 U.S. and international time series
- The Treasury issues bonds with 11 different maturities ranging from 1 month to 30 years

Question: How can we reduce the dimensionality and summarize the information in an efficient way?

Answer: **Principal Component Analysis (PCA)**

Reading: Ruppert ch. 17

Example: The panel of Treasury yields across time



- Can we summarize the dynamics of the panel with a few "factors"?
- If so, how many factors are needed?
- How much information are we losing?

EXAMPLE: FORECASTING WITH MANY PREDICTORS

Forecast y_{t+1} with N predictor variables $\mathbf{X}_t = (x_{1t}, \ldots, x_{Nt})'$:

$$y_{t+1} = \boldsymbol{\beta}\mathbf{X}_t + \epsilon_t$$

If N is "large" relative to T, then

- If $N > T$, the equation cannot be estimated (# of regressors $>$ # of observations)
- Out-of-sample forecasts will be poor.

$\longrightarrow$ Can we summarize the information in $\mathbf{X}_t$ to yield better forecasting factors?

- Suppose we have an N-dimensional random variable $\mathbf{x} = (x_1, \ldots, x_N)'$ with covariance matrix $\boldsymbol{\Sigma}$
- Objective: Can we summarize the variation in $\mathbf{x}$ with a **"few" linear combinations** of $\mathbf{x} = (x_1, \ldots, x_N)'$?
- "Principal component" i: $y_i = \mathbf{w}_i'\mathbf{x}$
- $\mathbf{w}_i$ are called weights, loadings or betas
- Data: $\mathbf{X} = (\mathbf{x}_{1,t}', \ldots, \mathbf{x}_{N,t}')$ is a $T \times N$ matrix
- Principle component: $y_{i,t} = \mathbf{w}_i'\mathbf{x}_t$
- Question: What are "optimal" $\mathbf{w}_i$?

PCA: Start with the covariance matrix $\boldsymbol{\Sigma}$ (alternative: correlation matrix)

Let $\mathbf{w}_i = (w_{i1}, \ldots, w_{iN})'$ and

$$y_i = \mathbf{w}_i'\mathbf{x} = \sum_{j=1}^{N} w_{ij}x_j$$

$$\text{with} \quad \mathbf{w}_i'\mathbf{w}_i = \sum_{j=1}^{N} w_{ij}^2 = 1$$

$$\Rightarrow \quad \text{Var}(y_i) = \mathbf{w}_i'\boldsymbol{\Sigma}\mathbf{w}_i$$

$$\text{Cov}(y_i, y_j) = \mathbf{w}_i'\boldsymbol{\Sigma}\mathbf{w}_j$$

Let $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)$.

PCA

Idea: Find linear combinations $\mathbf{w}_i, \mathbf{w}_j$ such that $y_i = \mathbf{w}_i'\mathbf{x}$ and $y_j = \mathbf{w}_j'\mathbf{x}$ are uncorrelated and the **variances of y_i, y_j are as large as possible**:

1. The first PC of $\mathbf{x}$ is the linear combination $y_1 = \mathbf{w}_1'\mathbf{x}$ that maximizes $\text{Var}(y_1)$ subject to the constraint that $\mathbf{w}_1'\mathbf{w}_1 = 1$
2. The second PC of $\mathbf{x}$ is the linear combination $y_2 = \mathbf{w}_2'\mathbf{x}$ that maximizes $\text{Var}(y_2)$ subject to the constraint that $\mathbf{w}_2'\mathbf{w}_2 = 1$ and $\text{Cov}(y_1, y_2) = 0$
3. The jth PC of $\mathbf{x}$ is the linear combination $y_j = \mathbf{w}_j'\mathbf{x}$ that maximizes $\text{Var}(y_j)$ subject to the constraint that $\mathbf{w}_j'\mathbf{w}_j = 1$ and $\text{Cov}(y_i, y_j) = 0$ for $i = 1, \ldots, j-1$

Some linear algebra:

Recall: Any non-singular square matrix can be decomposed using its eigenvectors and eigenvalues

The vector $\mathbf{v}$ is an **eigenvector** of the $N \times N$ matrix $\mathbf{A}$ if

$$\mathbf{Av} = \lambda \mathbf{v}$$

and λ is its associated **eigenvalue**.

An $N \times N$ non-singluar matrix has N eigenvectors and N eigenvalues.

Let $\mathbf{V}$ be the $N \times N$ matrix of eigenvectors and $\mathbf{\Lambda}$ be a diagonal matrix with the corresponding eigenvalues on the diagonal. The we can write any symmetric non-singular $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'.$$

Note that since the eigenvectors are orthogonal the matrix $\mathbf{V}$ is an orthogonal matrix (i.e. $\mathbf{V}\mathbf{V}' = \mathbf{I}$ and $\mathbf{V}' = \mathbf{V}^{-1}$).

This property implies that

$$\mathbf{\Lambda} = \mathbf{V}'\mathbf{A}\mathbf{V}.$$

The eigenvalue equation for a matrix $\mathbf{A}$ is

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. The eigenvalues can be computed by solving $\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}$, which is a polynomial of degree n.

Example:

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix}$$

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det\left(\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

$$= (2-\lambda)\left[(3-\lambda)(9-\lambda) - 16\right]$$

$$= -\lambda^3 + 14\lambda^2 - 35\lambda + 22$$

$$= -(\lambda - 11)(\lambda - 2)(\lambda - 1) = 0$$

The eigenvalues are are $\lambda_1 = 11, \lambda_2 = 2, \lambda_3 = 1$.

The eigenvectors can be found by solving the linear equations

$$\mathbf{Av} = \lambda \mathbf{v}$$

for each eigenvalue.

Eigenvectors: $v_1 = [1, 0, 0]'$, $v_2 = [0, 2, -1]'$ and $v_3 = [0, 1, 2]'$

Why do we care about all of this??

- The variance-covariance matrix of $\mathbf{x}$, $\boldsymbol{\Sigma}$, is by construction symmetric and positive definite. Thus
$$\boldsymbol{\Lambda} = \mathbf{V}'\boldsymbol{\Sigma}\mathbf{V}$$
- Wlog, order the eigenvalues so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$. Consider the largest eigenvalue λ_1 with eigenvector $\mathbf{v}_1$:
$$\text{Var}(y_1) = \mathbf{v}_1'\boldsymbol{\Sigma}\mathbf{v}_1 = \lambda_1$$
- The eigenvalue is equal to the variance of the linear combination $\mathbf{v}_1'\mathbf{x}$.
- Thus, the "optimal" $\mathbf{w}_1$ in $y_{1,t} = \mathbf{w}_1'\mathbf{x}_t$ is the first eigenvector $\mathbf{w}_1 = \mathbf{v}_1$

- More generally:

$$\text{Var}(y_j) = \mathbf{v}_j' \boldsymbol{\Sigma} \mathbf{v}_j = \lambda_j$$

$$\frac{\text{Var}(y_i)}{\sum_{j=1}^{N} \text{Var}(y_j)} = \frac{\lambda_i}{\lambda_1 + \ldots + \lambda_N}$$

- The (scaled) eigenvalues indicate the variance share of the PCs
- These results make PCA very easy to implement. Just compute the eigenvalues and eigenvalues of the variance-covariance matrix!

PCA: Summary

- Compute eigenvalues λ_i and eigenvectors $\mathbf{v}_i$, such that $\lambda_{i+1} \geq \lambda_i$
- Then the i-th PC, $y_{i,t}$ is

$$y_{i,t} = \mathbf{v}_i' \mathbf{x}_t$$

- The i-th PC captures a share of $\lambda_i / \sum \lambda_j$ of the total time series variation of $\mathbf{X}$
- Note: Traditional PCA applies the eigenvalue decomposition to the variance-covariance matrix of $\mathbf{x}$, $\mathbf{\Sigma} = \mathrm{E}[\mathbf{x}'\,\mathbf{x}] - \mathrm{E}[\mathbf{x}']\,\mathrm{E}[\mathbf{x}]$
- Alternative: Apply PCA to the correlation matrix
- The covariance and correlation matrices remove means of the data
- Finance: Means of factors are important!

  $\longrightarrow$ Use 2nd moment matrix $E[\mathbf{x}_t'\,\mathbf{x}_t]$ or normalized 2nd moment matrix
- Lettau and Pelger (2018): Importance of means in PCA in finance

- Next: Applications

- Size and value premium: 25 Size/BM portfolios
- Recall: Fit of CAPM and 3-factor Fama-French model for 25 Size/BM portfolios
- Excess returns:

	BM1	BM2	BM3	BM4	BM5
ME1	3.32	9.43	9.42	11.91	12.97
ME2	6.11	9.28	10.51	10.89	11.61
ME3	6.21	9.55	9.02	10.47	12.04
ME4	7.64	7.43	8.38	10.08	9.93
ME5	6.12	6.43	6.64	5.87	7.87

- Apply PCA to 2nd moment matrix $E[\mathbf{x}_t' \mathbf{x}_t]$
- First 8 eigenvalues $E[\mathbf{x}_t' \mathbf{x}_t]$: [44.94, 3.43, 1.81, 0.66, 0.38, 0.28, 0.25, 0.19]
- Look at first 4 PCs
- Compare to MKT, SMB, HML factors
- SMB and HML are linear combinations of Size/BM portfolios with ad hoc weights

Application: 25 Size/BM portfolios

	PC1	PC2	PC3	PC4
% of Var	88.39	6.75	3.55	1.30
Mean	7.29	5.13	6.18	4.87
Std.dev.	15.00	15.00	15.00	15.00
SR	0.49	0.34	0.41	0.32

	Mkt-RF	SMB	HML
Mean	6.35	2.95	3.82
Std.dev.	15.23	10.46	9.69
SR	0.42	0.28	0.39

Correlation of FF and PC factors

	PC1	PC2	PC3	PC4
Mkt-RF	0.93	0.15	-0.29	-0.08
SMB	0.59	-0.59	0.48	0.17
HML	-0.15	0.58	0.75	-0.15

PCA: 25 Size/BM portfolios

PC Loadings

	BM1	BM2	BM3	BM4	BM5
ME1	0.27	0.25	0.22	0.21	0.22
ME2	0.26	0.22	0.20	0.20	0.22
ME3	0.23	0.20	0.18	0.18	0.20
ME4	0.21	0.18	0.18	0.17	0.20
ME5	0.14	0.14	0.13	0.14	0.16

	BM1	BM2	BM3	BM4	BM5
ME1	-0.42	-0.31	-0.16	-0.13	-0.07
ME2	-0.26	-0.10	0.00	0.05	0.05
ME3	-0.20	0.03	0.12	0.16	0.17
ME4	-0.09	0.14	0.20	0.21	0.26
ME5	0.07	0.20	0.26	0.32	0.30

	BM1	BM2	BM3	BM4	BM5
ME1	-0.05	0.11	0.18	0.26	0.35
ME2	-0.23	0.00	0.10	0.17	0.28
ME3	-0.31	-0.07	0.03	0.10	0.21
ME4	-0.38	-0.15	-0.05	0.03	0.11
ME5	-0.41	-0.25	-0.17	-0.04	0.01

	BM1	BM2	BM3	BM4	BM5
ME1	-0.35	-0.14	-0.05	-0.05	-0.06
ME2	0.01	0.16	0.19	0.15	0.04
ME3	0.10	0.24	0.19	0.16	0.08
ME4	0.06	0.17	0.15	0.09	-0.08
ME5	-0.09	-0.06	-0.05	-0.16	-0.72

- How well do PCs explain the cross-section of returns?
- Run TS regressions where the factors are PCs

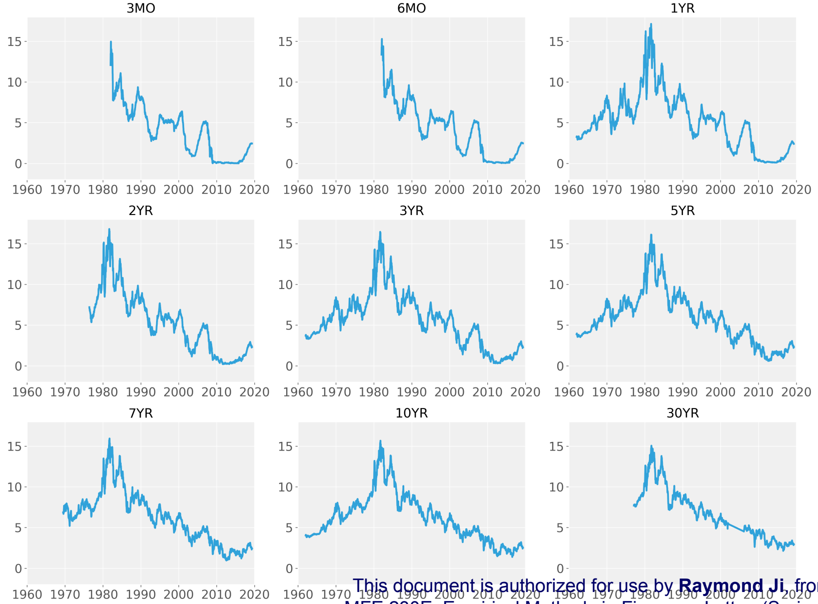
$$R^e_{i,t} = \alpha_i + \boldsymbol{\beta}_i \mathrm{PC}_t + e_t$$

	RMSE (% p.a.)
Mkt	3.30
Mkt, SMB	2.98
Mkt, HML	1.99
Mkt, SMB, HML	1.68

	RMSE (% p.a.)
PC1	2.75
PC1, PC2	2.16
PC1, PC2, PC3	1.56
PC1, PC2, PC3, PC4	1.38

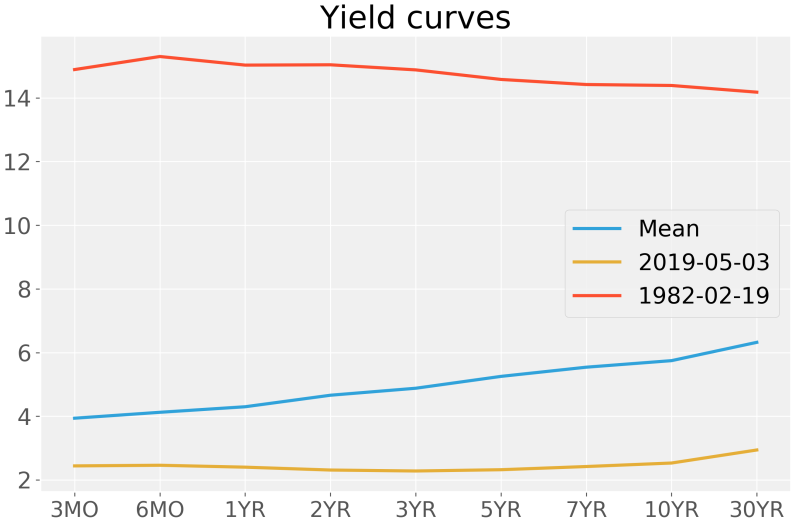
SMB and HML can be interpreted as proxies for PCs with pre-specified instead of estimated “optimal” loadings

Application: Treasury yields (weekly data 1980–2019)



Application: Treasury yields

	3MO	6MO	1YR	2YR	3YR	5YR	7YR	10YR	30YR
mean	3.94	4.12	4.30	4.66	4.88	5.25	5.54	5.75	6.32
std	3.13	3.23	3.29	3.36	3.32	3.19	3.10	2.97	2.82
min	0.00	0.03	0.09	0.20	0.30	0.59	0.95	1.38	2.13
25%	0.96	1.05	1.23	1.56	1.85	2.50	2.93	3.38	3.92
50%	4.04	4.29	4.36	4.63	4.67	4.97	5.14	5.24	5.99
75%	5.84	6.13	6.23	6.69	6.84	7.17	7.35	7.55	8.05
max	14.97	15.30	15.11	15.10	15.05	14.91	14.90	14.84	14.68


Yield curves
14
12
10
8
6
4
2
3MO
6MO
1YR
2YR
3YR
5YR
7YR
10YR
30YR
Mean
2019-05-03
1982-02-19

	3MO	6MO	1YR	2YR	3YR	5YR	7YR	10YR	30YR
3MO	9.81	10.11	10.25	10.40	10.19	9.64	9.23	8.73	8.28
6MO	10.11	10.45	10.61	10.78	10.57	10.01	9.59	9.08	8.62
1YR	10.25	10.61	10.82	11.03	10.83	10.28	9.86	9.34	8.90
2YR	10.40	10.78	11.03	11.32	11.15	10.65	10.24	9.73	9.33
3YR	10.19	10.57	10.83	11.15	11.02	10.57	10.19	9.70	9.35
5YR	9.64	10.01	10.28	10.65	10.57	10.21	9.88	9.45	9.18
7YR	9.23	9.59	9.86	10.24	10.19	9.88	9.60	9.20	8.99
10YR	8.73	9.08	9.34	9.73	9.70	9.45	9.20	8.85	8.70
30YR	8.28	8.62	8.90	9.33	9.35	9.18	8.99	8.70	7.95

Eigenvalues: 93.60, 1.67, 0.12, 0.02, 0.01, 0.0, 0.0, 0.0, 0.0

	3MO	6MO	1YR	2YR	3YR	5YR	7YR	10YR	30YR
3MO	1.00	1.00	1.00	0.99	0.98	0.96	0.95	0.94	0.91
6MO	1.00	1.00	1.00	0.99	0.99	0.97	0.96	0.94	0.92
1YR	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.96	0.93
2YR	0.99	0.99	1.00	1.00	1.00	0.99	0.98	0.97	0.95
3YR	0.98	0.99	0.99	1.00	1.00	1.00	0.99	0.98	0.97
5YR	0.96	0.97	0.98	0.99	1.00	1.00	1.00	0.99	0.98
7YR	0.95	0.96	0.97	0.98	0.99	1.00	1.00	1.00	0.99
10YR	0.94	0.94	0.96	0.97	0.98	0.99	1.00	1.00	1.00
30YR	0.91	0.92	0.93	0.95	0.97	0.98	0.99	1.00	1.00

Eigenvalues: 9.33, 0.18, 0.01, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0

Recall: Explained variance = eigenvalues

PCA on covariance matrix

	value	Difference	Proportion	Cumulative value	Cumulative Proportion
0	93.58	91.90	0.98	93.58	0.98
1	1.68	1.56	0.02	95.26	1.00
2	0.12	0.09	0.00	95.37	1.00
3	0.02	0.02	0.00	95.40	1.00

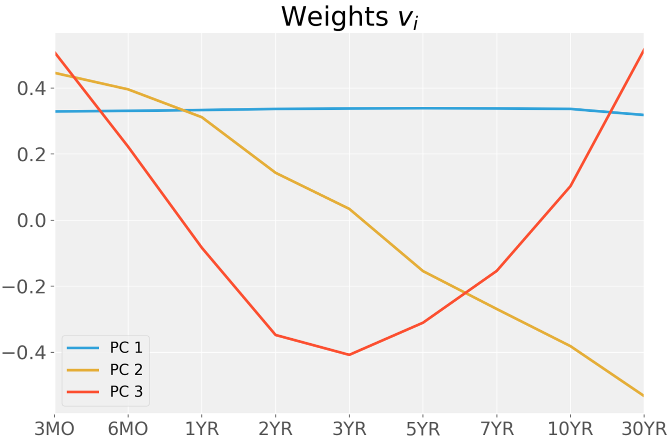
PCA on correlation matrix

	value	Difference	Proportion	Cumulative value	Cumulative Proportion
0	9.33	9.15	0.98	9.33	0.98
1	0.18	0.17	0.02	9.51	1.00
2	0.01	0.01	0.00	9.53	1.00
3	0.00	0.00	0.00	9.53	1.00

PCA: Treasury yields – PCs

Recall: i-th PC $y_i = \mathbf{v}_i'\mathbf{x}$

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9
3MO	0.33	0.45	0.51	0.51	0.31	0.24	0.10	0.12	0.02
6MO	0.33	0.40	0.22	-0.15	-0.43	-0.47	-0.25	-0.44	-0.04
1YR	0.33	0.31	-0.08	-0.53	-0.25	0.15	0.29	0.58	-0.08
2YR	0.34	0.14	-0.35	-0.24	0.39	0.22	-0.33	-0.19	0.58
3YR	0.34	0.03	-0.41	0.01	0.35	0.02	0.03	-0.23	-0.74
5YR	0.34	-0.15	-0.31	0.33	-0.06	-0.41	0.61	-0.04	0.33
7YR	0.34	-0.27	-0.15	0.37	-0.20	-0.14	-0.59	0.50	-0.05
10YR	0.34	-0.38	0.10	0.06	-0.46	0.63	0.10	-0.34	-0.01
30YR	0.32	-0.53	0.52	-0.38	0.37	-0.26	0.03	0.05	-0.02


Weights v_i
0.4
0.2
0.0
−0.2
−0.4
PC 1
PC 2
PC 3
3MO
6MO
1YR
2YR
3YR
5YR
7YR
10YR
30YR

Application: Treasury yields

- Most of the variation in yields can be summarized by 1-3 principal components
- The first PC (the "level") is responsible for 90% of the variation in yields
- The second ("slope") and third ("curvature") add some explanatory power
- Further details: Prof. Stanton's fixed income class
- PCA is useful in other asset classes as well
- Ruppert ch. 17: Equity
- Lettau, Maggiori and Weber (*Journal of Financial Economics*, 2016): Exchange rates, equity portfolios, commodities

PCA LIMITATIONS

- PCA is a useful **statistical** method to summarize information
- PCA cannot answer the important question of **what economic forces cause prices to move**
- Example: The level factor affects all yields.
  Question: What is the level factor? Current monetary policy, expectations about future interest rates, inflation expectations, market risk aversion,...?
- Active academic literature that tries to link PCAs to economic variables

Ludvigson and Ng (*Review of Financial Studies*, 2009): Use information in 132 monthly economic series to forecast bond returns

Categories:

- Output and income
- Employment and hours
- Retail, sales, and orders
- Compensation and labor costs
- Capacity utilization
- Financial: interest rates and spreads, stock market and foreign exchange indicators

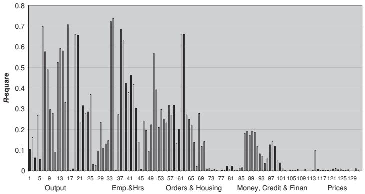


Figure 1
Marginal *R*-squares for F_1
Note: Chart shows the *R*-square from regressing the series number given on the *x*-axis onto F_1. See the Appendix for a description of the numbered series. The factors are estimated using data from 1964:1 to 2003:12.

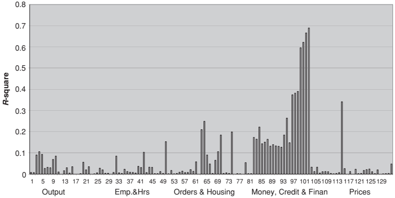


Figure 2
Marginal *R*-squares for F_2
Note: See Figure 1.

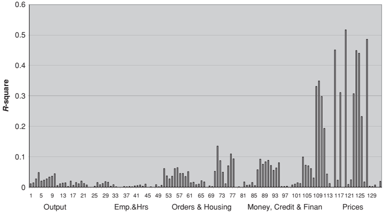


Figure 3
Marginal *R*-squares for F_3
Note: See Figure 1.

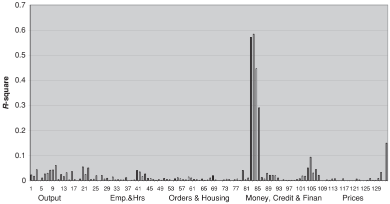


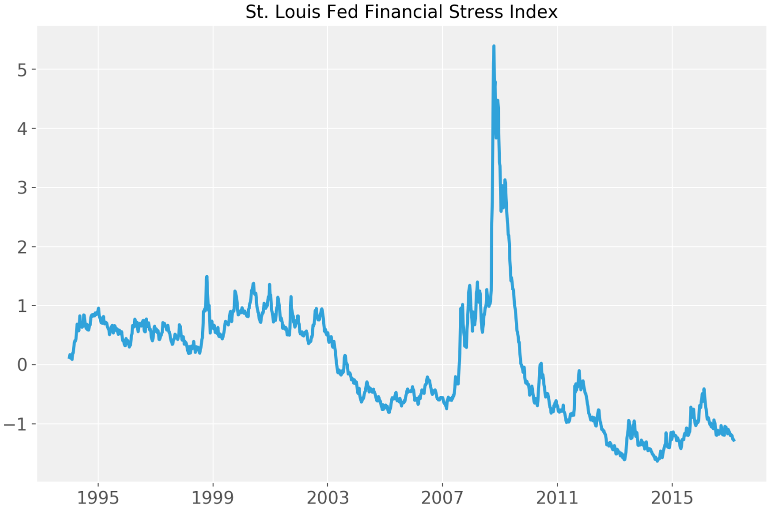
Figure 5
Marginal *R*-squares for F_8
Note: See Figure 1.

St. Louis Fed:

> *The STLFSI measures the degree of financial stress in the markets and is constructed from 18 weekly data series: seven interest rate series, six yield spreads and five other indicators. Each of these variables captures some aspect of financial stress. Accordingly, as the level of financial stress in the economy changes, the data series are likely to move together.*

https://fred.stlouisfed.org/series/STLFSI

Other applications of PCA: Financial stress index

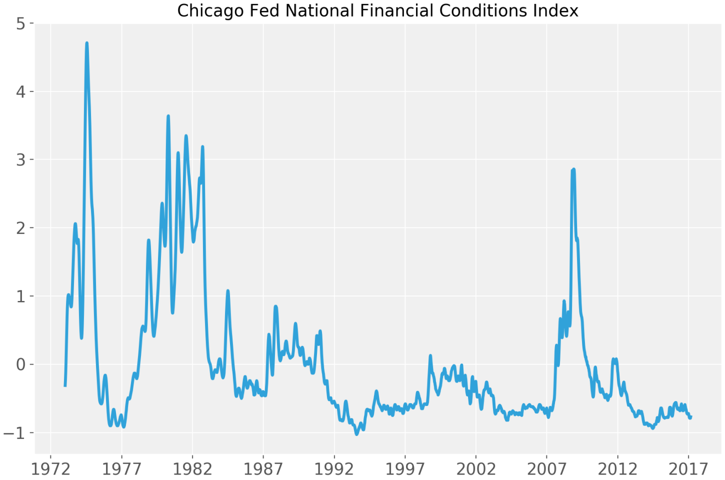


Federal Reserve Bank of Chicago:

> The NFCI is a weighted average of 105 indicators of risk, credit, and leverage in the financial system—each expressed relative to its sample average and scaled by its sample standard deviation. As such, a zero value for the NFCI can be thought of as the U.S. financial system operating at historical average levels of risk, credit, and leverage. The ANFCI removes the variation in these indicators attributable to economic activity, as measured by the three-month moving average of the Chicago Fed National Activity Index (CFNAI), and inflation, according to its three-month total based on the Personal Consumption Expenditures (PCE) Price Index. As such, a zero value for the ANFCI corresponds with a financial system operating at historical average levels of risk, credit, and leverage consistent with economic activity and inflation.

https://fred.stlouisfed.org/series/NFCI

Other applications of PCA: National Financial Conditions Index



- When N is large relative to T, traditional methods often break down
- PCA is a simple and useful tool to reduce the dimensionality of large data sets
- PC are linear combinations of the N time series that decompose the variance-covariance matrix
- PCs can be computed by calculating the eigenvalues and eigenvectors of the variance-covariance matrix

Summary: PCA

- The *i*th PC is the linear combination of the *i*th eigenvector and the data matrix
- The contribution of the *i*th PC is equal to its associated eigenvalue
- Treasure yield: First three PCs represent “level, slope and curvature”
- PCA can be applied to other asset classes
- Other applications: Forecasting and economic indices
- Some “big data” techniques are based on PCA
- Lettau and Pelger (2018): Incorporate APT-restrictions about time-series and cross-section of risk premia

Outline

- Model with time-varying coefficients:

$$y_t = \alpha_{s_t} + \phi_{s_t} y_{t-1} + \epsilon_t.$$

where s_t is an unobserved state variable: How to extract estimates of s_t from observations $(y_1, \ldots, y_T)$?

- This model is an example of a **filtering problem**
- Many financial problems can be written as filtering models

Introduction

- Kalman (1963) provided solutions to standard (i.e. linear normal) filtering problems, many extensions have been developed since
- Original applications in engineering: Guiding missiles in real time to target subject to changing weather conditions
- In finance, the Kalman filter can be used to extract information from noisy or imperfect data
- Reading:
  - Tsay ch. 11
  - More advanced: Hamilton ch. 13

Example 1: CAPM with time-varying coefficients

$$r_t = \alpha_t + \beta_t r_{M,t} + e_t$$
$$\alpha_{t+1} = \alpha_t + u_t$$
$$\beta_{t+1} = \beta_t + v_t$$

Example 2: Regression with ARMA errors

$$y_t = x_t' \beta + z_t$$
$$A(L) z_t = B(L) e_t$$

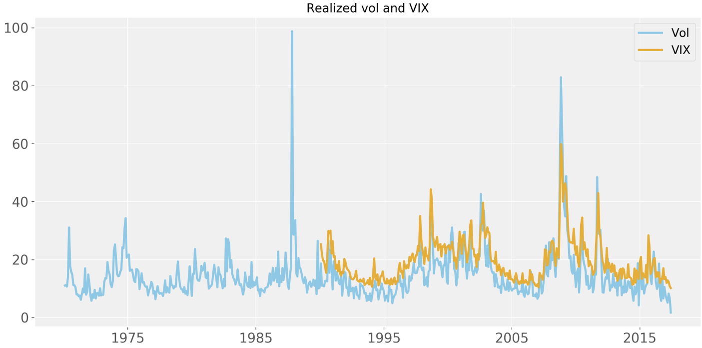
Example 3: Unobserved component models

$$y_t = \mu_t + \gamma_t + \omega_t + e_t$$

where μ_t is an unobserved trend, γ_t is an unobserved seasonal component, ω_t is an unobserved cycle components and e_t is an unobserved shock

Example 4: Missing values

Realized vol and VIX:



- Realized vol and the VIX are noisy measures of true vol
- True vol is **latent** → Need to estimate true vol

EXAMPLE: MODELING VOLATILITY

Consider the **local trend model**

$$y_t = \mu_t + e_t,\ \ e_t \sim N(0, \sigma_e^2) \tag{1}$$

$$\mu_{t+1} = \mu_t + \eta_t,\ \ \eta_t \sim N(0, \sigma_\eta^2) \tag{2}$$

- y_t is observed realized vol
- μ_t is the unobserved trend (here: random walk): true volatility
- e_t is iid noise or measurement error

The system (y_t, μ_t) is known as a **state-space model**:

- y_t is the observed data
- μ_t is the **state** at time t and is not observed
- (1) is the **measurement** equation
- (2) is the **state** equation

Assume for the moment that all model parameters are known.

Let $F_t = (y_1, \ldots, y_t)$

The Kalman filter has the following steps at each time t:

1. **Filtering**: Recover (an estimate) the state variable μ_t by removing the measurement error
2. **Prediction**: Forecast μ_{t+1} and y_{t+1} given F_t

Notation:

$$\mu_{t|s} = \mathrm{E}[\mu_t | F_s]$$

$$\boldsymbol{\Sigma}_{t|s} = \mathrm{Var}[\mu_t | F_s]$$

$$v_t = y_t - y_{t|t-1}$$

$$V_t = \mathrm{Var}(v_t | F_{t-1})$$

The Kalman filter is a recursive procedure:

1. Start with some initial conditions, e.g. $\mu_1 \sim N(\mu_{1|0}, \mathbf{\Sigma}_{1|0})$
2. Filtering: Use observation y_1 to compute $\mu_{1|1}$ and $\mathbf{\Sigma}_{1|1}$
3. Prediction: Compute the conditional mean $\mu_{2|1}$ and variance $\mathbf{\Sigma}_{2|1}$
4. Filtering: Use observation y_2 to compute $\mu_{2|2}$ and $\mathbf{\Sigma}_{2|2}$
5. Prediction: Compute the conditional mean $\mu_{3|2}$ and variance $\mathbf{\Sigma}_{3|2}$
6. Repeat Filtering and Prediction until $t = T$

At time $t-1$, we have $\mu_{t|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1}$.

Then we observe y_t. Let $v_t = y_t - y_{t|t-1}$.

$$\mu_{t|t} = \mu_{t|t-1} + \frac{\boldsymbol{\Sigma}_{t|t-1}}{V_t} v_t = \mu_{t|t-1} + K_t v_t$$

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} + \frac{\boldsymbol{\Sigma}^2_{t|t-1}}{V_t} = \boldsymbol{\Sigma}_{t|t-1}(1 - K_t)$$

$$K_t = \frac{\boldsymbol{\Sigma}_{t|t-1}}{V_t} \quad \text{is called the } \textbf{Kalman gain}$$

The Kalman gain K_t is the factor that governs the contribution of the new shock v_t to the new state variable μ_t

The Kalman filter with normal distributions: Prediction

We have the filtered state moments $\mu_{t|t}$ and $\mathbf{\Sigma}_{t|t}$.

Predict next period's moments of the state variable μ_t: $\mu_{t+1|t}$ and $\mathbf{\Sigma}_{t+1|t}$

In the local trend model

$$y_t = \mu_t + e_t$$

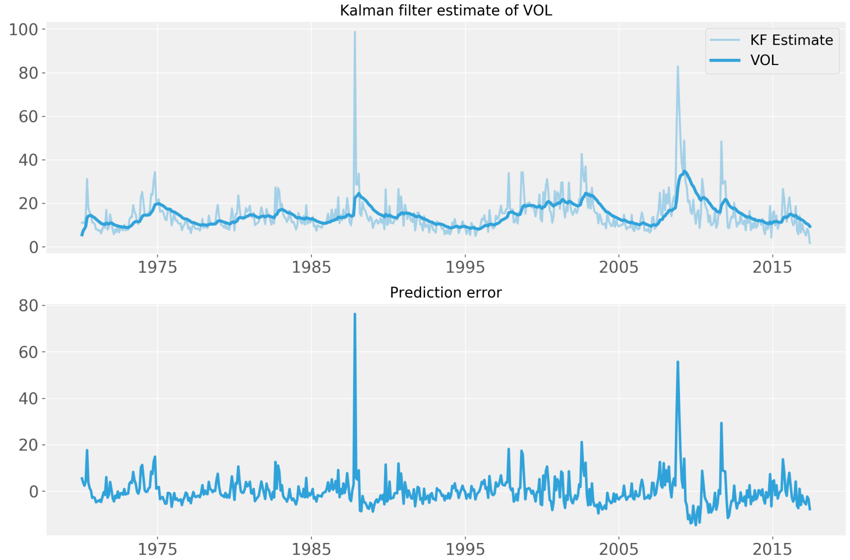
$$\mu_{t+1} = \mu_t + \eta_t$$

the forecasting equations very simple:

$$\mu_{t+1|t} = \mathrm{E}[\mu_t + \eta_t | F_t] = \mu_{t|t}$$

$$\mathbf{\Sigma}_{t+1|t} = \mathrm{Var}(\mu_{t+1}|F_t) = \mathrm{Var}(\mu_t + \eta_t | F_t) = \mathbf{\Sigma}_{t|t} + \mathbf{\Sigma}_{\eta}$$

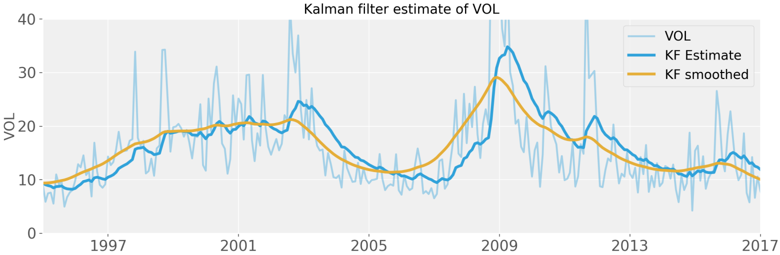
Filtered state variable: Realized vol

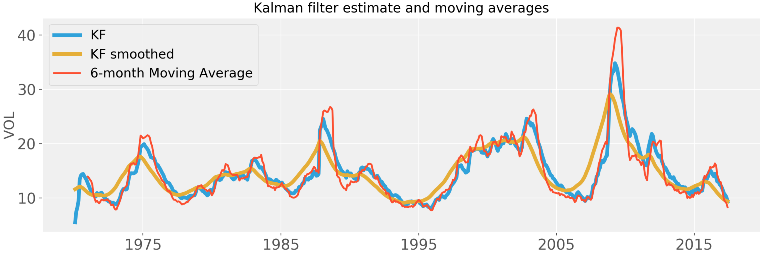


State smoothing

- So far: We estimated the state μ_t using data $y_1, \dots, y_t$
- We can also estimate μ_t using the entire sample $y_1, \dots, y_T$: $\mu_{t|T}$ are called **smoothed** state estimates (details are in Hamilton and Tsay)
- Same idea as the smoothing in Hamilton's regime switching model.

State smoothing





So far: We assumed that $\mathbf{\Sigma}_e$ and $\mathbf{\Sigma}_\eta$ were known.

Next: Estimate $\mathbf{\Sigma}_e$ and $\mathbf{\Sigma}_\eta$ by MLE

Under normality and $y_1 \sim N(\mu_{1|0}, V_1)$:

$$\begin{aligned} L(y_1, \dots, y_T | \mathbf{\Sigma}_e, \mathbf{\Sigma}_\eta) &= p(y_1 | \mathbf{\Sigma}_e, \mathbf{\Sigma}_\eta) \prod_{t=2}^{T} p(y_t | F_{t-1}, \mathbf{\Sigma}_e, \mathbf{\Sigma}_\eta) \\ &= p(y_1 | \mathbf{\Sigma}_e, \mathbf{\Sigma}_\eta) \prod_{t=2}^{T} p(v_t | F_{t-1}, \mathbf{\Sigma}_e, \mathbf{\Sigma}_\eta) \end{aligned}$$

where $v_t = y_t - \mu_{t|t-1}$.

- Thus, the log-likelihood under normality is

$$\log L(\mathbf{\Sigma}_e, \mathbf{\Sigma}_\eta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\log(V_t) + \frac{v_t^2}{V_t}\right]$$

  which can be numerically maximized to obtain estimates of $\mathbf{\Sigma}_e$ and $\mathbf{\Sigma}_\eta$.
- For the vol example we get $\hat{\mathbf{\Sigma}}_e = 7.38^2$ and $\hat{\mathbf{\Sigma}}_\eta = 0.39^2$.
- KF package for Python: `pykalman`

Let $\mathbf{y}_t$ be a $k \times 1$ data vector and $\mathbf{s}_t$ be a $m \times 1$ vector of state variables:

$$\text{state equation:} \quad \mathbf{s}_{t+1} = \mathbf{d}_t + \mathbf{T}_t \mathbf{s}_t + \mathbf{R}_t \boldsymbol{\eta}_t$$

$$\text{measurement equation:} \quad \mathbf{y}_t = \mathbf{c}_t + \mathbf{Z}_t \mathbf{s}_t + \mathbf{e}_t$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \;\; (n \times 1)$$

$$\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{H}_t) \;\; (k \times 1)$$

$$\mathbf{s}_1 \sim N(\mu_{1|0}, \boldsymbol{\Sigma}_{1|0})$$

The general Gaussian linear state-space model

- Usually $\mathbf{e}_t$ and $\boldsymbol{\eta}_t$ are assumed to be independent (can be relaxed)
- $\mathbf{T}_t, \mathbf{R}_t, \mathbf{Q}_t, \mathbf{Z}_t$ and $\mathbf{H}_t$ are called **system matrices** and depend on the application
- The system matrices depend on parameters $\boldsymbol{\theta}$ that can be estimated by MLE

The state space model in incredibly versatile:

- Any ARMA model can be written in state space form
- The state space model allows for general measurement error in linear regressions
- Regressions with time-varying coefficients can be written in state-space form

The Kalman filter steps are similar to the steps in the local trend model:

Let $\mathbf{s}_{i|j}$ and $\mathbf{\Sigma}_{i|j}$ be the condition mean and covariance matrix of $\mathbf{s}_i$ given F_j

1. Start with some initial conditions, e.g. $\mathbf{s}_1 \sim N(\mathbf{s}_{1|0}, \mathbf{\Sigma}_{1|0})$
2. Filtering: Use observation y_t to compute $\mathbf{s}_{t|t}$ and $\mathbf{\Sigma}_{t|t}$
3. Prediction: Compute the conditional mean $\mathbf{s}_{t+1|t}$ and variance $\mathbf{\Sigma}_{t+1|t}$
4. Repeat Filtering and Prediction until $t = T$
5. Compute smoothed estimates $\mathbf{s}_{t|T}$ and $\mathbf{\Sigma}_{t|T}$

The forecast error $\mathbf{v}_t$ is given by

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{c}_t - \mathbf{Z}_t \mathbf{s}_{t|t-1}$$

with a variance of

$$\mathbf{V}_t = \mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$$

The forecast error $\mathbf{v}_t$ is given by The **filtering** equations are

$$\mathbf{s}_{t|t} = \mathbf{s}_{t|t-1} + \mathbf{\Sigma}_{t|t-1}\mathbf{Z}_t'\mathbf{V}_t^{-1}$$

$$\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t-1}\mathbf{Z}_t'\mathbf{V}_t^{-1}\mathbf{Z}_t\mathbf{\Sigma}_{t|t-1}$$

The **forecasting** equations are

$$\mathbf{s}_{t+1|t} = \mathbf{d}_t + \mathbf{T}_t\,\mathbf{s}_{t|t}$$

$$\mathbf{\Sigma}_{t+1|t} = \mathbf{T}_t\mathbf{\Sigma}_{t|t}\mathbf{T}_t + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}_t'$$

Note: If all system matrices are constant, the matrix $\mathbf{\Sigma}_{t|t-1}$ converges to a constant steady state matrix $\mathbf{\Sigma}_*$. In steady state $\mathbf{V}_t$ and $\mathbf{\Sigma}_{t|t}$ are constants.

Estimation: MLE under normality

Example: CAPM with time-varying coefficients

$$r_t = \alpha_t + \beta_t r_{M,t} + w_t$$
$$\alpha_{t+1} = \alpha_t + u_t$$
$$\beta_{t+1} = \beta_t + v_t$$

$$\Rightarrow \begin{pmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

$$r_t = (1, r_{M,t}) \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + w_t$$

Write this model in state space form:

$$\mathbf{s}_{t+1} = \mathbf{d}_t + \mathbf{T}_t\,\mathbf{s}_t + \mathbf{R}_t\,\boldsymbol{\eta}_t$$
$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{Z}_t\,\mathbf{s}_t + \mathbf{e}_t$$
$$\boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$$
$$\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{H}_t)$$
$$\mathbf{s}_1 \sim N(\mu_{1|0}, \boldsymbol{\Sigma}_{1|0})$$

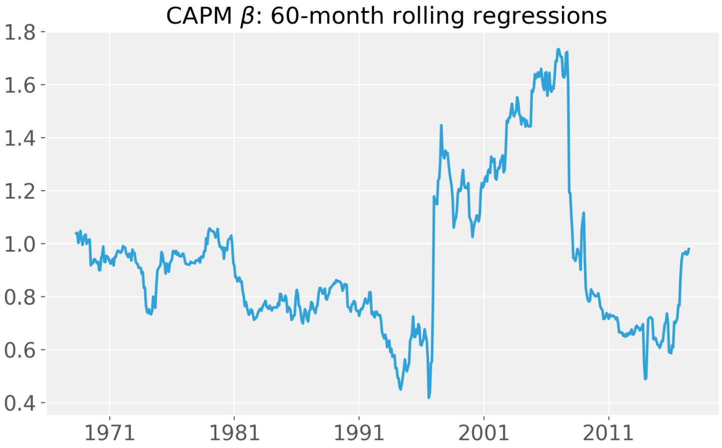
$$\mathbf{s}_t = (\alpha_t, \beta_t)'$$
$$\mathbf{y}_t = r_t$$
$$\mathbf{T}_t = \mathbf{R}_t = \mathbf{I}_2$$
$$\mathbf{d}_t = \mathbf{c}_t = 0$$
$$\mathbf{Z}_t = (1, r_{M,t})$$
$$\mathbf{H}_t = \boldsymbol{\Sigma}_w^2$$
$$\mathbf{Q}_t = \mathrm{diag}(\boldsymbol{\Sigma}_u^2, \boldsymbol{\Sigma}_v^2)$$

Monthly excess returns for S&500 index and IBM from 1963 to 2017

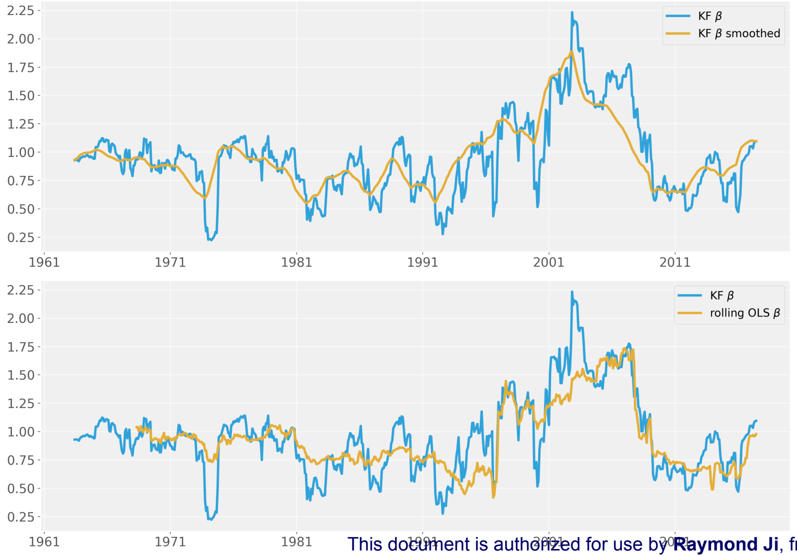
Standard CAPM OLS regression:

$$r_t = .003 + 0.94 r_{M,t} + w_t$$
$$\hat{\mathbf{\Sigma}}_w = .05^2$$
$$R^2 = 33.9\%$$

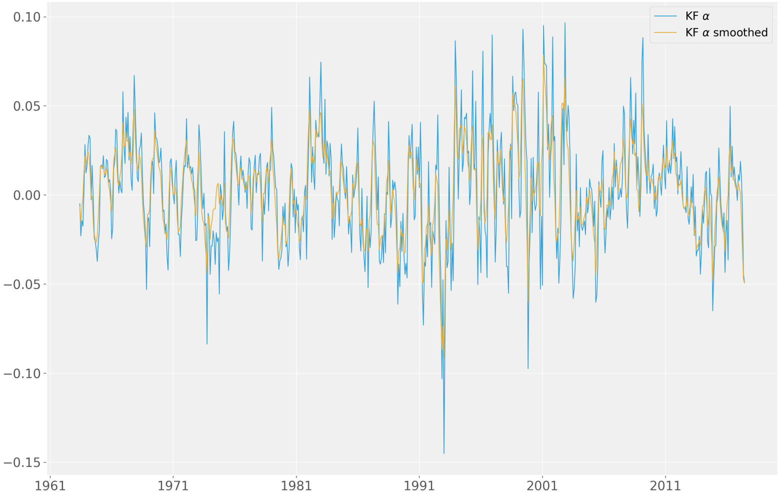
Example: CAPM with time-varying coefficients



Example: CAPM with time-varying coefficients – Beta



EXAMPLE: CAPM WITH TIME-VARYING COEFFICIENTS – ALPHA



- The KF can be extended to nonlinear dynamics and non-normal shocks
- Estimation becomes more complicated and computationally difficult
- Word of caution: The parameter vector $\boldsymbol{\theta}$ includes all elements of $\mathbf{T}_t$, $\mathbf{R}_t$, $\mathbf{Q}_t$, $\mathbf{Z}_t$ and $\mathbf{H}_t$ $\longrightarrow$ numerical maximization of likelihood function over high dimensional $\boldsymbol{\theta}$ can quickly become unmanageable, even using modern computers!
- If possible, impose restrictions on system matrices to reduce dimensionality

Risk-return trade-off: Brandt and Kang (*Journal of Financial Economics, 2004*)

Let y_t be continuously compounded excess returns with time-series dynamics

$$y_t = \mu_{t-1} + \sigma_{t-1}\varepsilon_t \quad \text{with } \varepsilon_t \sim \mathrm{N}[0,1], \tag{1}$$

where μ_t and σ_t denote the conditional mean and volatility of excess returns, respectively. Both the conditional mean and volatility are unobservable quantities, which we assume to evolve jointly as a first-order VAR process in logs:

$$\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = d + A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \sigma_{t-1} \end{bmatrix} + \eta_t \quad \text{with } \eta_t \sim \mathrm{MVN}[0, \Sigma], \tag{2}$$

where d is a 2×1 coefficient vector that relates to the long-term means of the two latent state variables, A is a 2×2 coefficient matrix, and Σ is a 2×2 covariance matrix (symmetric and nonnegative definite). Specifically, we write

$$d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \text{with}$$

$$b_{12} = b_{21} = \rho\sqrt{b_{11}b_{22}}. \tag{3}$$

Table 3
Parameter estimates
This table presents the simulated maximum likelihood estimates of the model:

$$y_t = \bar{\mu} e^{m_{t-1}} + \bar{\sigma} e^{v_{t-1}} \varepsilon_t = \bar{\mu} e^{Z_1' s_{t-1}} + \bar{\sigma} e^{Z_2' s_{t-1}} \varepsilon_t \quad \text{with } \varepsilon_t \sim \mathrm{N}[0,1]$$

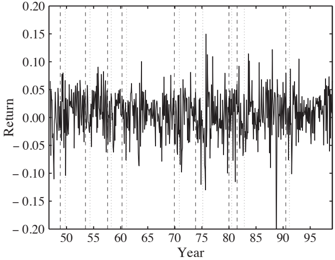
and

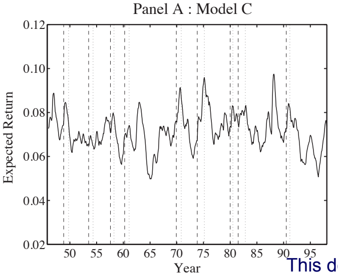
$$s_t = A s_{t-1} + \eta_t \quad \text{with } \eta_t \sim \mathrm{MVN}[0, \Sigma],$$

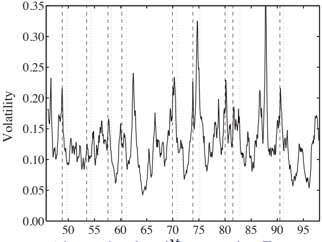
The estimates are for monthly returns on the value-weighted CRSP index in excess of the one-month Treasury bill rate from January 1946 through December 1998.

Parameters	Model A		Model B		Model C		Model D	
	Estimate	t-Stat	Estimate	t-Stat	Estimate	t-Stat	Estimate	t-Stat
a_{11}	0.8592	17.41	0.8313	9.23	0.8658	11.21	0.8677	10.92
a_{21}	−0.0531	−1.92	−0.0211	−1.43	−0.0885	−1.96	−0.1292	−1.42
a_{12}	0.1081	0.92	0.1168	1.12	0.0861	1.01	0.0947	0.71
a_{22}	0.9237	15.23	0.9110	16.32	0.8973	15.32	0.9086	18.49
b_{11}	7.60×10^{-3}	4.24	6.43×10^{-3}	2.20	5.96×10^{-3}	3.19	4.68×10^{-3}	2.76
b_{22}	0.0554	13.02	0.0561	12.22	0.0614	5.55	0.0591	5.22
ρ	−0.6345	−4.32	−0.4577	−3.21	−0.5584	−5.80	−0.5621	−5.96
ρ_μ	—	—	−0.0866	−0.49	—	—	−0.0517	−0.77
ρ_σ	—	—	—	—	−0.2541	−4.04	−0.2430	−3.67
$\bar{\mu}$	6.50×10^{-3}	3.31	6.48×10^{-3}	5.94	6.24×10^{-3}	5.11	6.24×10^{-3}	4.87
$\bar{\sigma}$	0.0377	10.32	0.0385	5.60	0.0382	8.79	0.0382	9.01
$\beta_1 = \rho\sqrt{b_{11}/b_{22}}$	−0.2350	−6.23	−0.1550	−4.11	−0.1740	−5.98	−0.1582	−3.99
$\beta_2 = \rho\sqrt{b_{22}/b_{11}}$	−1.7131	−8.34	−1.3519	−6.34	−1.7923	−9.32	−1.9975	−10.92
$a_{12} - \beta_1 a_{22}$	0.3252	3.82	0.2580	3.11	0.2422	3.47	0.2384	3.82
$a_{21} - \beta_2 a_{11}$	1.4188	5.21	1.1028	4.35	1.4623	4.11	1.6040	2.42
$a_{11} + a_{12} - a_{21} - a_{22}$	0.0967	3.33	0.0684	2.54	0.1431	2.31	0.1830	2.92
$\ln \mathscr{L}$	1,150.98		1,151.11		1,152.94		1,153.06	

EXAMPLE: NONLINEAR KALMAN FILTER







- The state space/Kalman filter model is a powerful tool for modeling financial data
- Many models can be represented in state-space form and estimated via the Kalman filter
- Even if a particular problem does not fit the standard linear specification with normal errors, the set up can usually be extended to fairly easily
- Estimation in the normal/linear framework straightforward using MLE
- Nonlinear/non-normal extensions require more complicated estimation procedures
- Potential obstacle: Too many parameters to estimate

$$r_t = \sigma_t\, u_t$$

$$\log \sigma_t - \alpha = \phi(\log \sigma_{t-1} - \alpha)) + \eta_t$$

$$u_t \sim N(0,1)$$

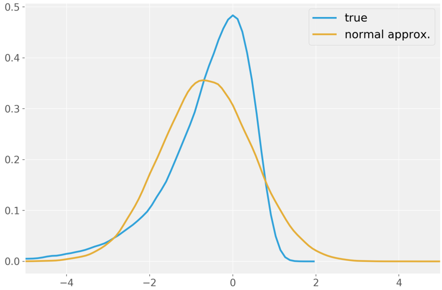
$$\eta_t \sim N(0, \sigma_\eta^2)$$

- Volatility $\log \sigma_t$ is the unobserved state
- Define: $y_t = \log(|r_t|)$, $s_t = \log(\sigma_t)$, $\xi_t = \log(|u_t|)$
- State space representation:

  Measurement equation: $y_t = s_t + \xi_t$

  Transition equation: $s_t = (1 - \phi)\alpha + \phi s_{t-1} + \eta_t$
- Complication: $\xi_t = \log(|u_t|)$ in measurement equation is not normal!
- Moments of $\xi_t = \log(|u_t|)$: $E[\xi_t] = -0.635$, $\text{Var}(\xi_t) = \pi^2/8$ and left-skewed

Quasi Maximum Likelihood (QML): Pretend that $\xi_t = \log(|u_t|)$ is normal with $E[\xi_t] = -0.635$, $Var(\xi_t) = \pi^2/8$



Note: QML is consistent and asymptotically normal

Alternative: **Simulated MLE**

KF-QML

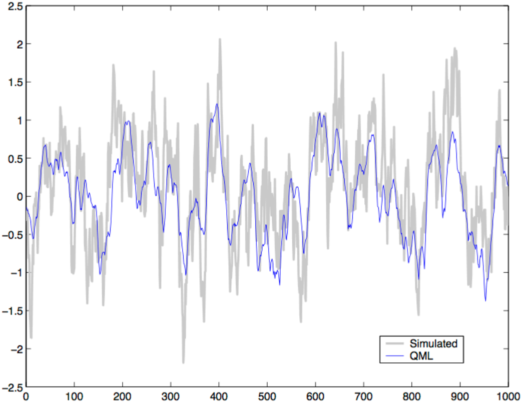


Figure 3: Simulated underlying volatility process (thick grey line) and estimated smoothed volatilities via the QML method (thin black line).

- Lecture: Intuition only (see Tsay's "Analysis of Financial Time Series", ch. 12)
- MCMC is an example of **Bayesian estimation**
- Here is the idea: Let θ_1, θ_2 and θ_3 be 3 parameters to be estimated.
- The full likelihood function $f_1(\theta_1, \theta_2, \theta_3, \mathbf{X})$ is unknown or does not exist
- But we know the 3 **conditional distributions** of each parameter given the other two parameters:
  - $f_1(\theta_1|\theta_2, \theta_3, \mathbf{X})$
  - $f_2(\theta_2|\theta_1, \theta_3, \mathbf{X})$
  - $f_3(\theta_3|\theta_1, \theta_2, \mathbf{X})$
- In practice: Draw random samples from $f_i(\cdot)$ since exact forms of $f_i(\cdot)$ are usually not known in closed form
- Python package: `pymc3` (not included in Anaconda)

Regression with AR(1) errors:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + z_t \tag{1}$$

$$z_t = \phi z_t + u_t \tag{2}$$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi, \sigma^2) \tag{3}$$

Question: What happens if $\boldsymbol{x}_t$ includes lagged dependent variables and (1) is estimated by OLS?

Note:

- If we knew ϕ, σ in(2), estimate (1): $\widehat{\boldsymbol{\beta}}$
- If we knew $\boldsymbol{\beta}$ in (1), we could back out $z_t = y_t - \mathbf{x}_t'\widehat{\boldsymbol{\beta}}$ and estimate (2): $\widehat{\phi}, \widehat{\sigma}$

Let $\theta_{2,0}$ and $\theta_{3,0}$ be some starting values. The **Gibbs sampler** is defined as

1. Draw random sample from $f_1(\theta_1|\theta_{2,0}, \theta_{3,0}, \mathbf{X})$. Denote estimator by $\theta_{1,1}$
2. Draw random sample from $f_2(\theta_2|\theta_{1,1}, \theta_{3,0}, \mathbf{X})$. Denote estimator by $\theta_{2,1}$
3. Draw random sample from $f_3(\theta_3|\theta_{1,1}, \theta_{2,1}, \mathbf{X})$. Denote estimator by $\theta_{3,1}$

Save parameters $\theta_{1,1}, \theta_{2,1}, \theta_{3,1}$ and repeat iteration with $\theta_{1,1}, \theta_{2,1}, \theta_{3,1}$ as starting values. Repeat iterations many times.

$\boldsymbol{\theta}_i = (\theta_{1,i}, \theta_{2,i}, \theta_{3,i})'$ represent a Markov chain.

The Markov chain can be used to compute an estimate of θ as well as its distribution of the estimator.

Regression with AR(1) errors:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + z_t \tag{4}$$

$$z_t = \phi z_t + u_t \tag{5}$$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi, \sigma^2) \tag{6}$$

If we knew (5), we could estimate (4). Moreover, if we knew (4), we could back out $z_t = y_t - \mathbf{x}_t' \boldsymbol{\beta}$ and estimate (5).

To implement the Gibbs sampler we need:

1. $f(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}, \phi, \sigma^2)$
2. $f(\phi | \mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2)$
3. $f(\sigma^2 | \mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \phi)$

Priors: Now suppose that we specify our initial guesses of $\theta = (\boldsymbol{\beta}', \phi, \sigma^2)$ as

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \Sigma_0)$$

$$\phi \sim N(\phi_0, \sigma_0^2)$$

$$\frac{v\lambda}{\sigma^2} \sim \chi_v^2.$$

The conditional distributions

1. $f(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \phi, \sigma^2)$
2. $f(\phi|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2)$
3. $f(\sigma^2|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \phi)$

are called **posteriors**.

Let's start with some starting values for $\theta = (\boldsymbol{\beta}', \phi, \sigma^2)$

Step 1: Compute $f(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \phi, \sigma^2)$

Given ϕ, it can be shown that $\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \phi, \sigma^2 \sim N(\boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*)$ (the formulas for $\boldsymbol{\beta}^*$ and $\boldsymbol{\Sigma}^*$ are similar to the updating equations in the Kalman filter)

Note that the form of the posterior distribution is the same as the prior. Priors for which this is the case are called **conjugate priors**.

Step 2: Compute $f(\phi|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2)$

Given $\boldsymbol{\beta}$, it can be shown that $\phi|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\phi^*, \sigma^{2*})$

Step 3: Compute $f(\sigma^2 | \mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \phi)$

It can be shown that $\frac{v\lambda + \sum_{t=2}^{n} u_t^2}{\sigma^2} \sim \chi^2_{v+(n-1)}$

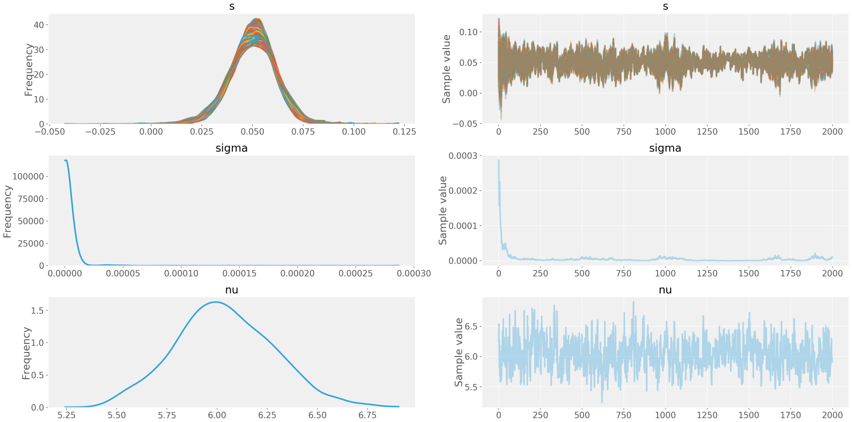
CAPM regression:

$$R_t^i = \alpha_i + \beta_i R_t^M + e_t$$

Priors:

$$\alpha \sim N(0, 1)$$
$$\boldsymbol{\beta} \sim N(0, 1)$$
$$\sigma \sim \chi^2_{10}$$

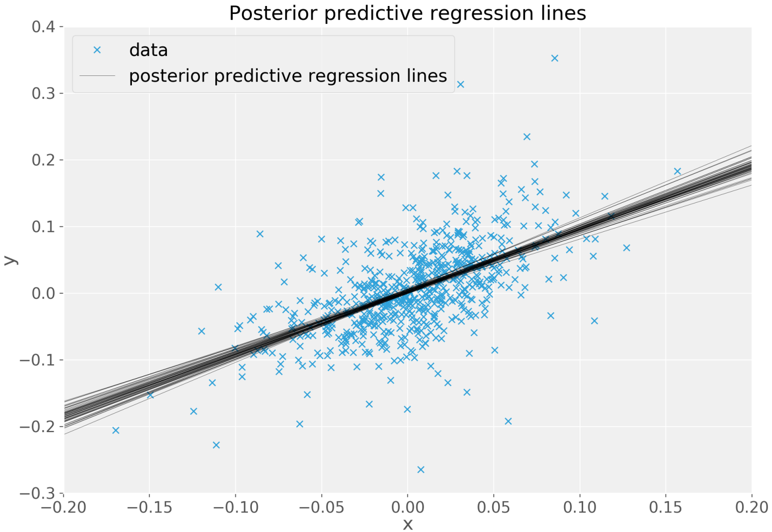
Estimation using the MCMC package `pymc3` (Implementation in lab this week)


s
Frequency
40
30
20
10
0
−0.050
−0.025
0.000
0.025
0.050
0.075
0.100
0.125
s
Sample value
0.10
0.05
0.00
−0.05
0
250
500
750
1000
1250
1500
1750
2000
sigma
Frequency
100000
75000
50000
25000
0
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
sigma
Sample value
0.0003
0.0002
0.0001
0.0000
nu
Frequency
1.5
1.0
0.5
0.0
5.25
5.50
5.75
6.00
6.25
6.50
6.75
nu
Sample value
6.5
6.0
5.5

MCMC posterior distributions

	mean	sd	mc_error	hpd_2.5	hpd_97.5
sigma	5.524e-06	1.632e-05	1.462e-06	9.308e-08	1.193e-05
nu	6.033e+00	2.458e-01	1.111e-02	5.534e+00	6.485e+00

GIBBS SAMPLING: CAPM FOR IBM



SV model for demeaned returns:

$$r_t = e_t$$

$$e_t \sim t(\nu, 0, \exp(-2s_t))$$

$$s_t \sim N(s_{t-1}, \sigma^2)$$

Priors:

$$\sigma \sim \text{Exp}(50)$$

$$\nu \sim \text{Exp}(0.1)$$

Traditional parameters: σ, ν

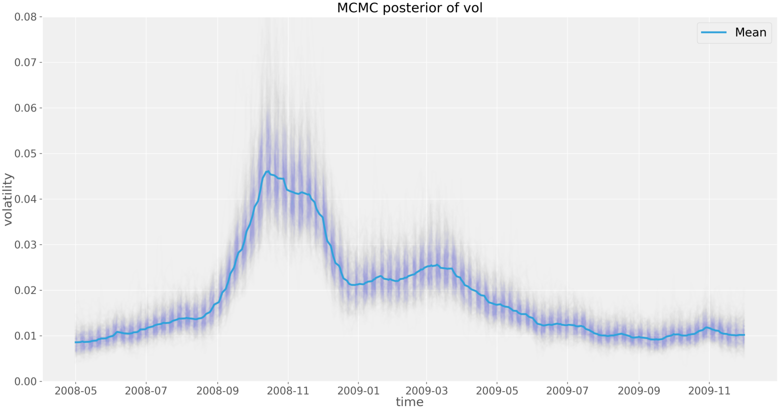
Trick: Treat $(s_1, s_2, \ldots, s_T)$ as parameters!

Number of parameters to be estimated: $T + 2$

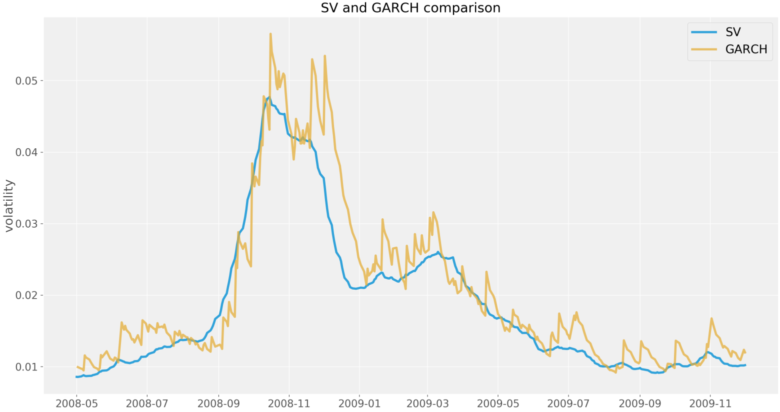
MCMC posterior distributions

	mean	sd	mc_error	hpd_2.5	hpd_97.5
nu	18.325	10.016	0.288	5.379	39.702
sigma	0.006	0.005	0.000	0.002	0.010
r_missing__0	-0.000	0.010	0.000	-0.022	0.019
s__400	-4.586	0.238	0.005	-5.045	-4.103
s__399	-4.587	0.228	0.005	-5.045	-4.161
s__398	-4.586	0.220	0.004	-5.040	-4.190
s__397	-4.594	0.216	0.005	-5.040	-4.213
s__396	-4.599	0.211	0.005	-5.040	-4.219

MCMC ESTIMATION OF SV MODELS



MCMC ESTIMATION OF SV MODELS



Summary

- Stochastic volatility is an extension of GARCH models
- Advantages: Allows volatility shocks, multivariate models
- Disadvantages: Harder to estimate
- No closed-form likelihood function, GMM has poor small sample properties
- Alternative: MCMC

