Time Series Techniques & Applications

Peter C. B. Phillips

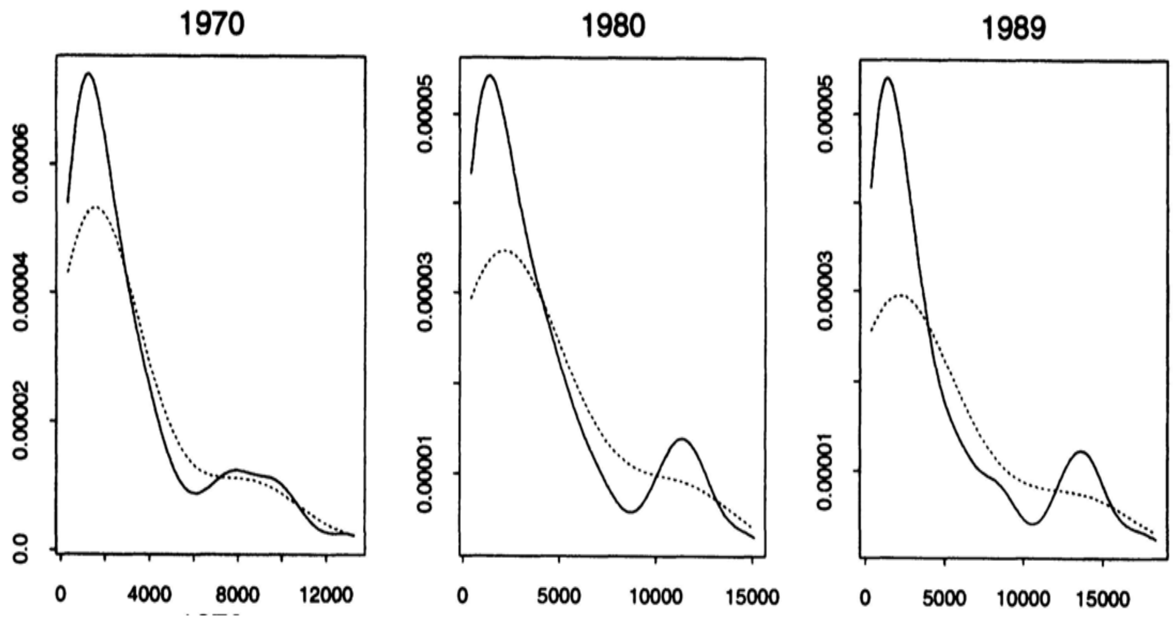
Yale Lecture Supplement Fall 2014

Modeling Trends, Trend Extraction, Automated Discovery

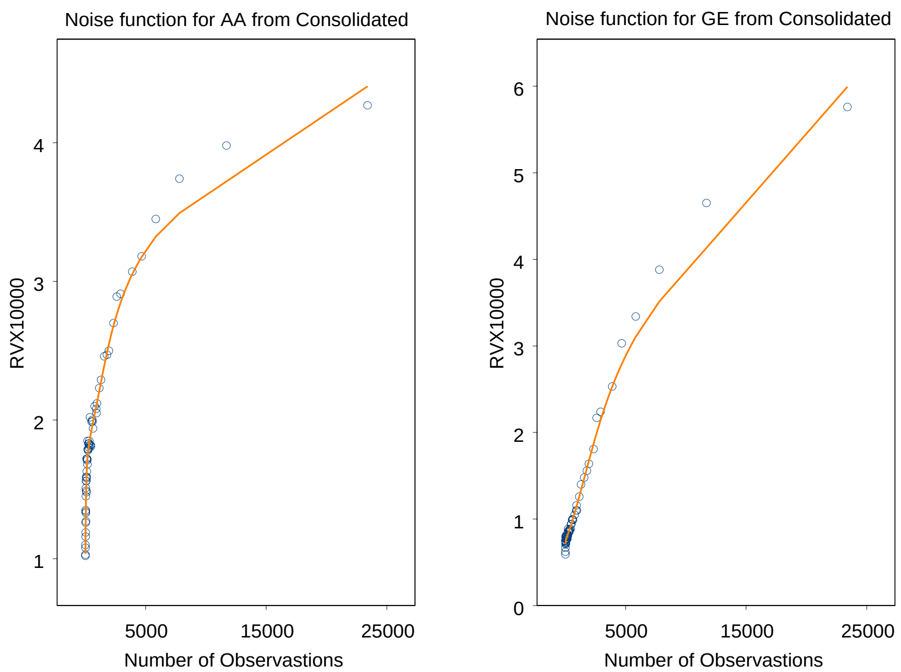
1. (2003). "Laws and Limits of Econometrics", *Economic Journal,* 113, C26-C52..
2. (2005). "Challenges of Trending Time Series Econometrics", *Mathematics and Computers in Simulation*, 68, 401-416.
3. (2005). "Automated Discovery in Econometrics". *Econometric Theory*, 21, 3-20.
4. (2009). "Econometric Theory and Practice". *Econometric Theory*, 25, 583-586.
5. (2010). "The Mystery of Trends". *Macroeconomic Review*, October Issue

Some of the Biggest Issues in Economics and Finance concern Trend

- Macroeconomics:
  - the process of economic growth
  - the study of growth convergence + divergence
  - emergent peaks
  - evolution in the distribution of world income
  - trends in world consumer culture/transportation
- Finance:
  - reconciling martingale models of efficient price determination with long run growth and long run predictability
  - modeling and predicting financial bubbles



Trends in Kernel Density Estimates of Distribution of per capita GDP in constant US dollars over 119 Countries (Bianchi, 1997, JAE)



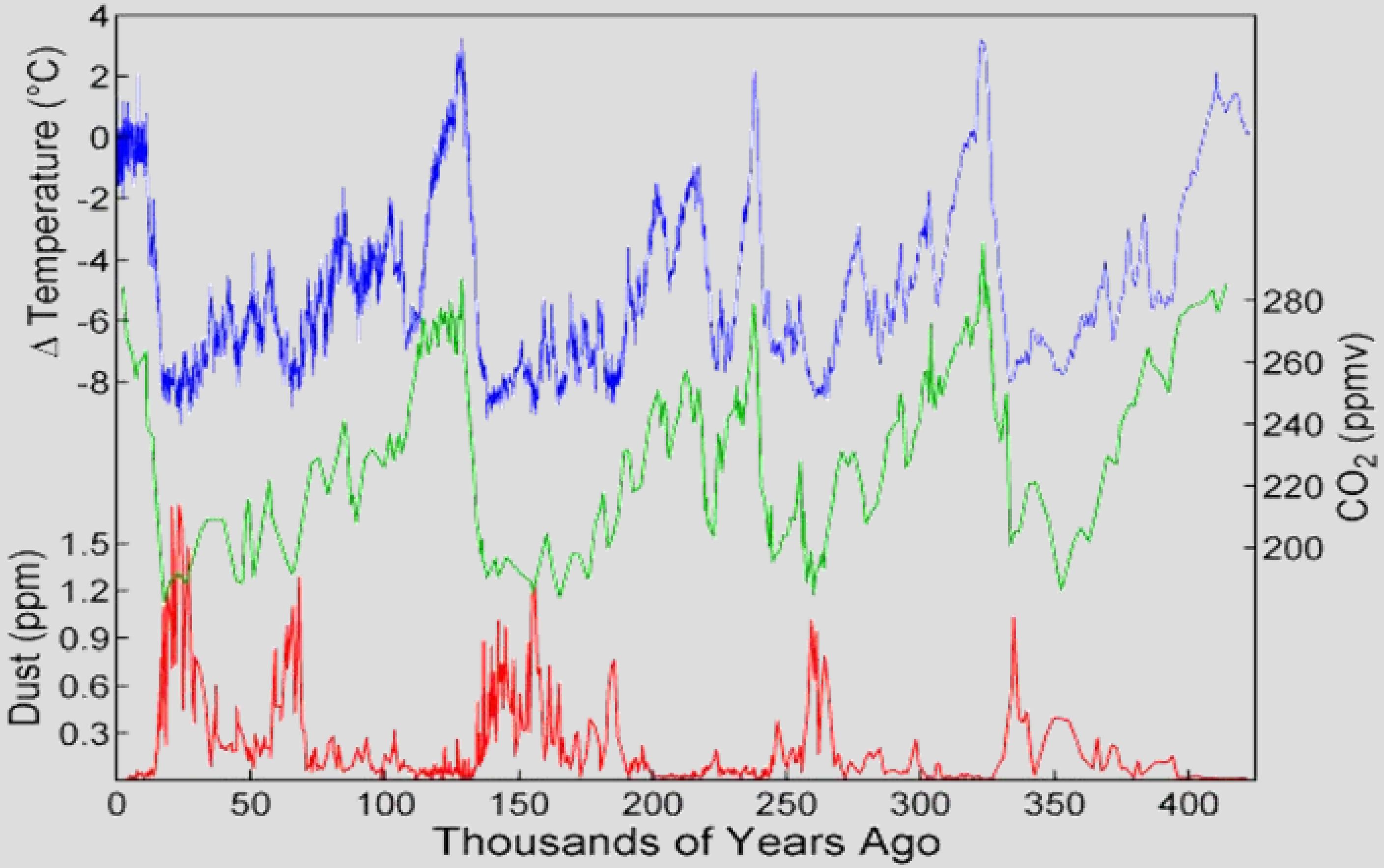
Market microstructure noise functions for AA and GE. The horizontal axis is the number of prices used to construct the realized volatility. The vertical axis is the realized volatility. Consolidated market Trade prices (November 1, 2004 to November 24, 2004)

And Other Fields

- Natural History
  - paleodiversity + history of life
  - origination and extinction of species
- Environmetrics
  - atmospheric pollution
  - climate change
  - deforestation, ozone depletion, exotic afforestation
- Human characteristics & demographics
  - athletic records
  - obesity
  - life expectancy & trends in aging

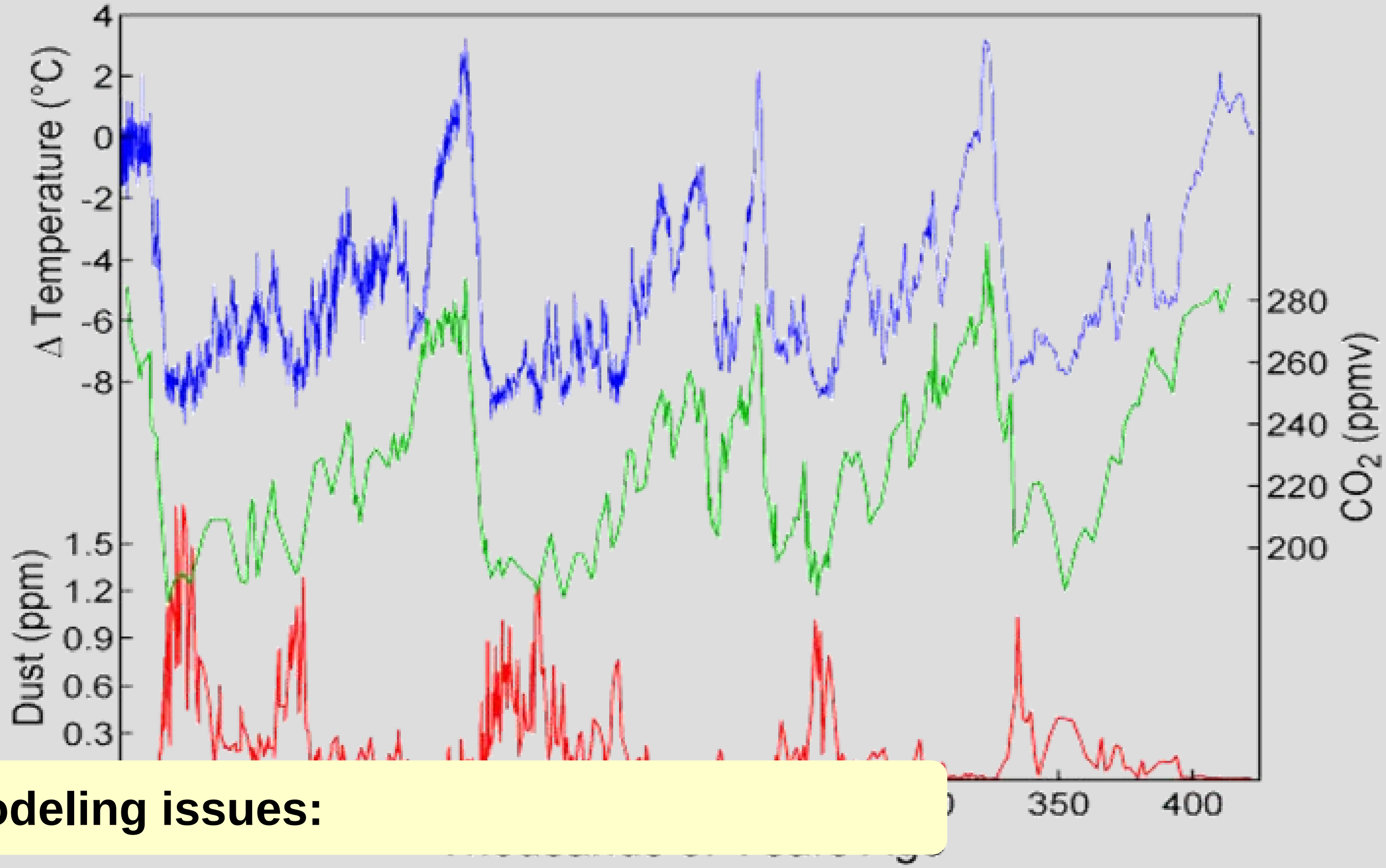
Climate Change

Climate Change: ice core data



Vostok Ice Core Data

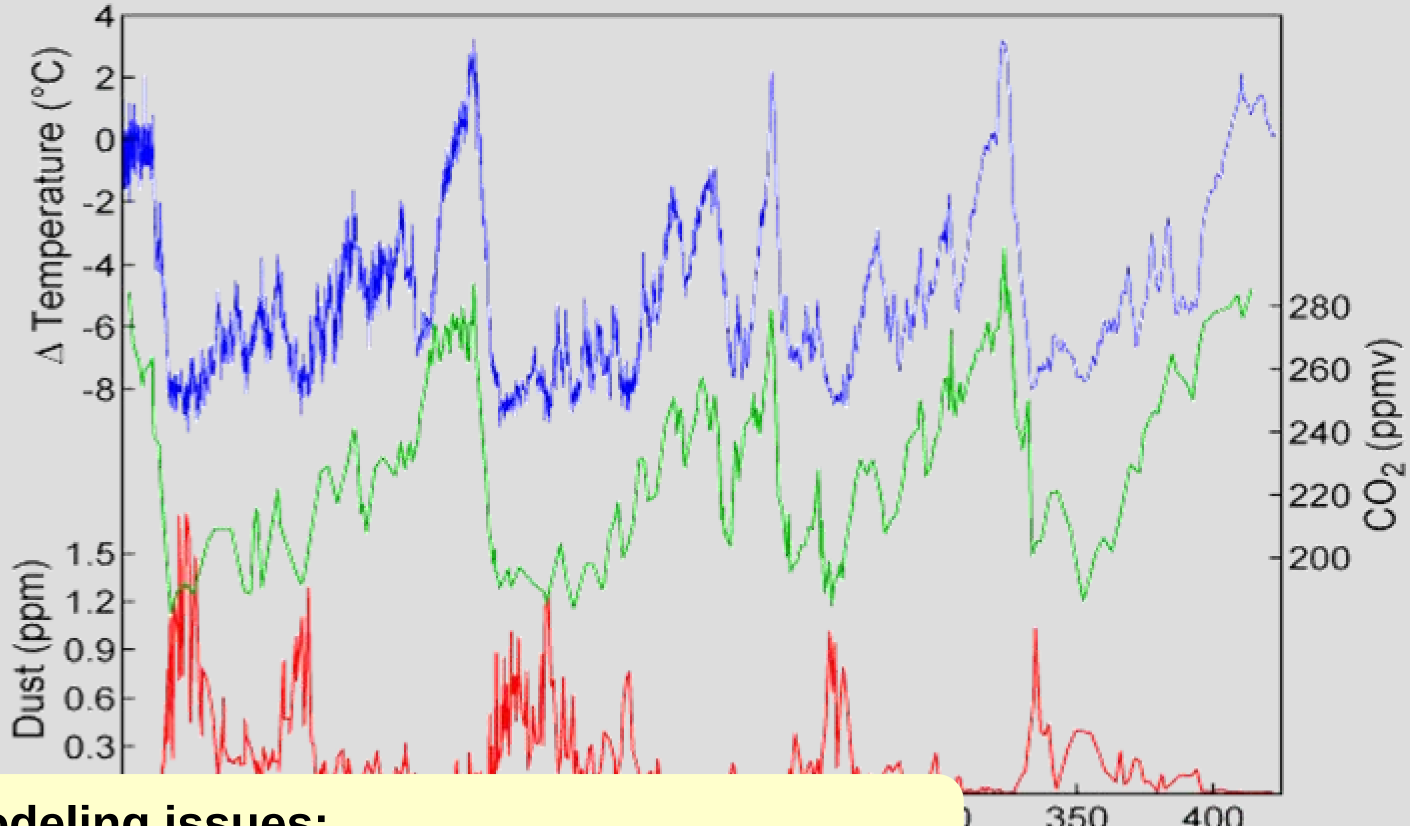
Climate Change: ice core data



Modeling issues:

Vostok Ice Core Data

Climate Change: ice core data

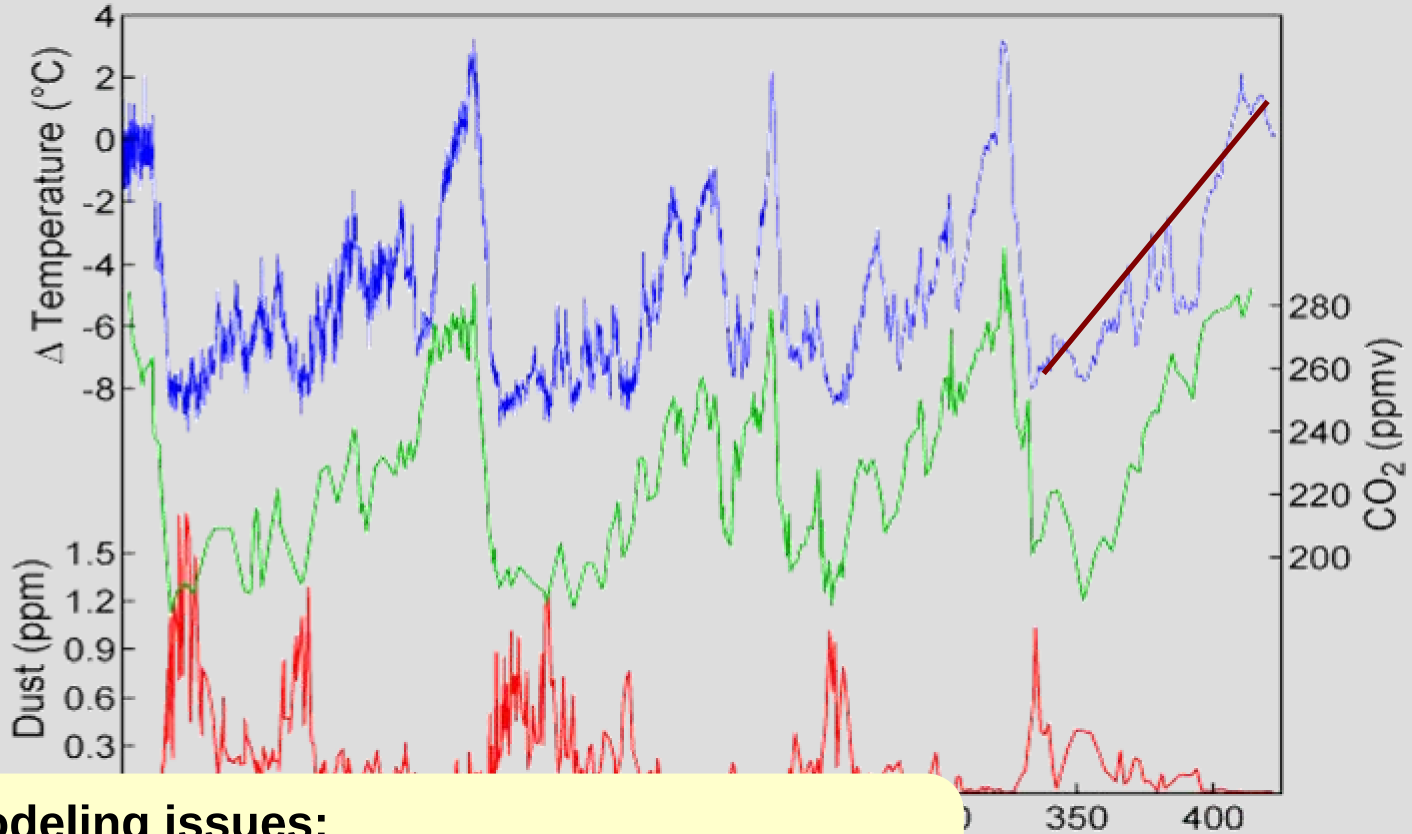


Modeling issues:

a. irregular cycle & heterogeneity

Vostok Ice Core Data

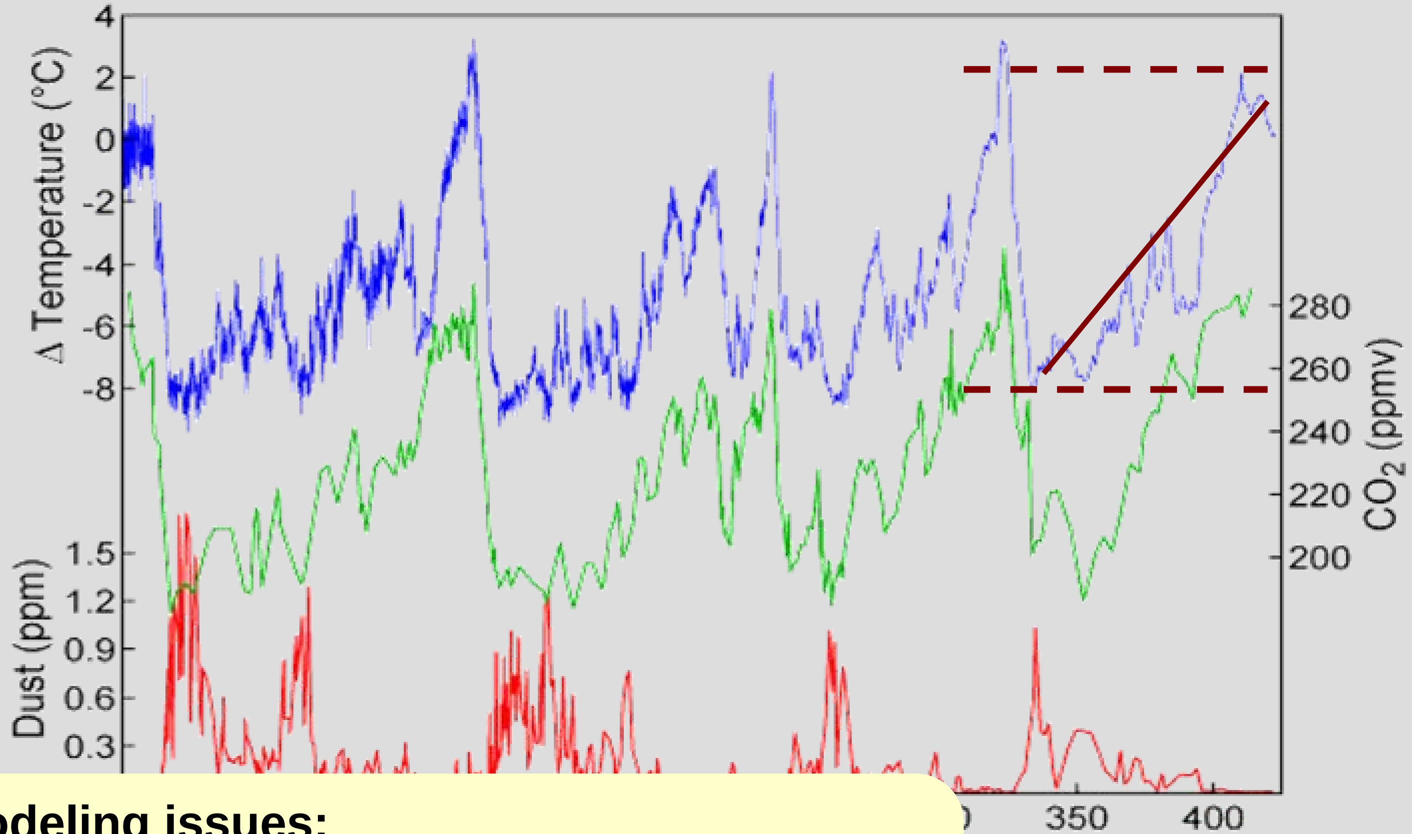
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle

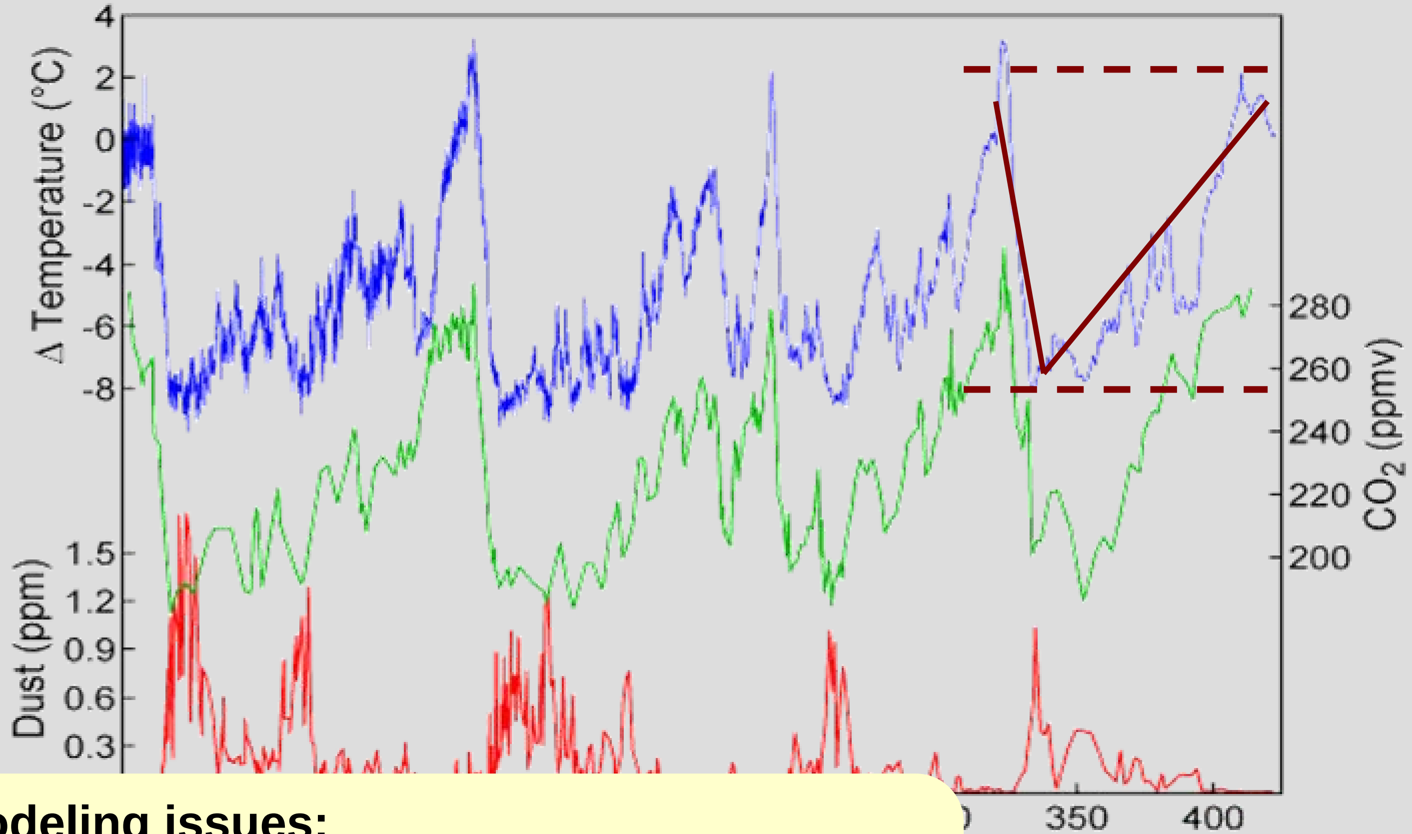
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

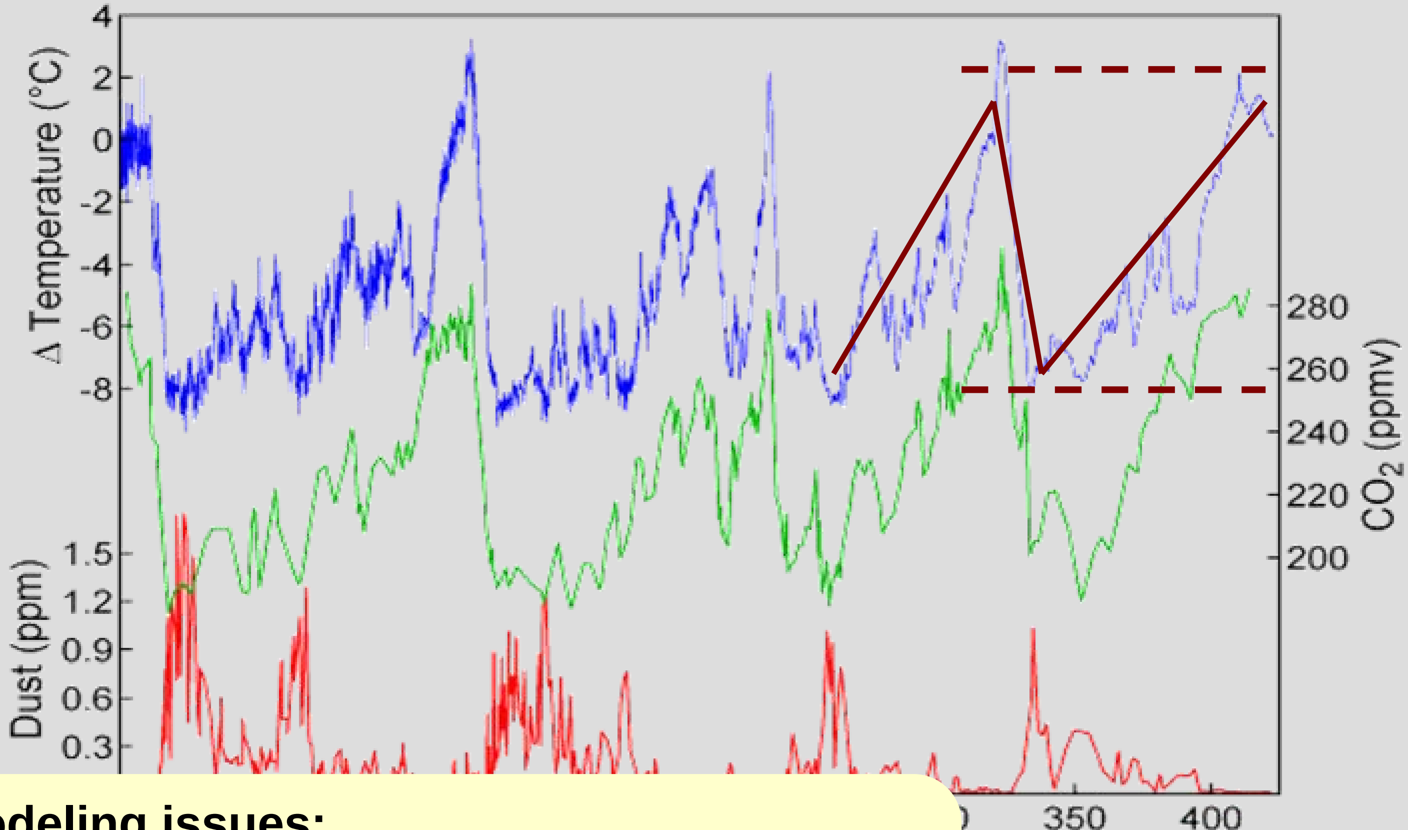
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

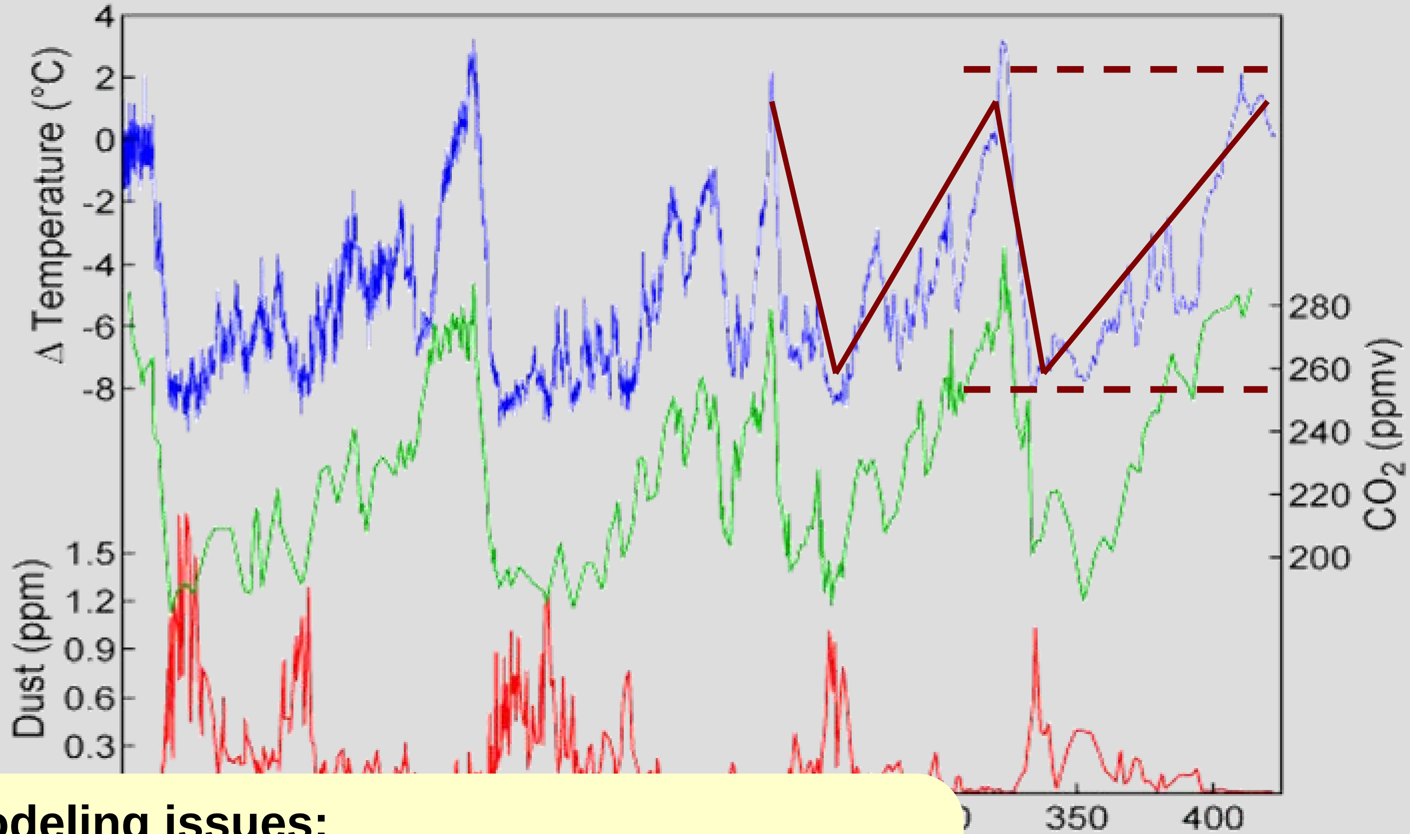
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

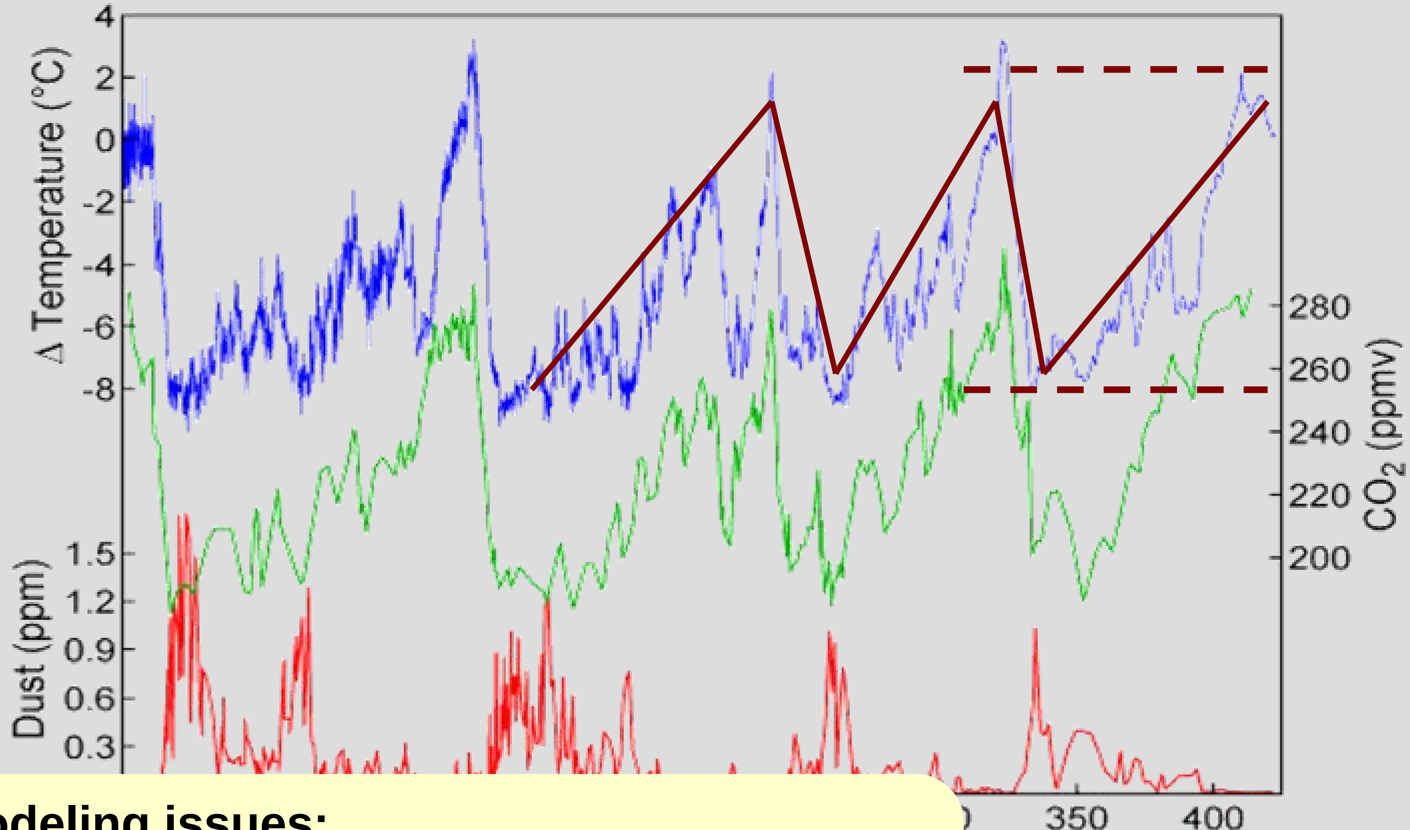
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

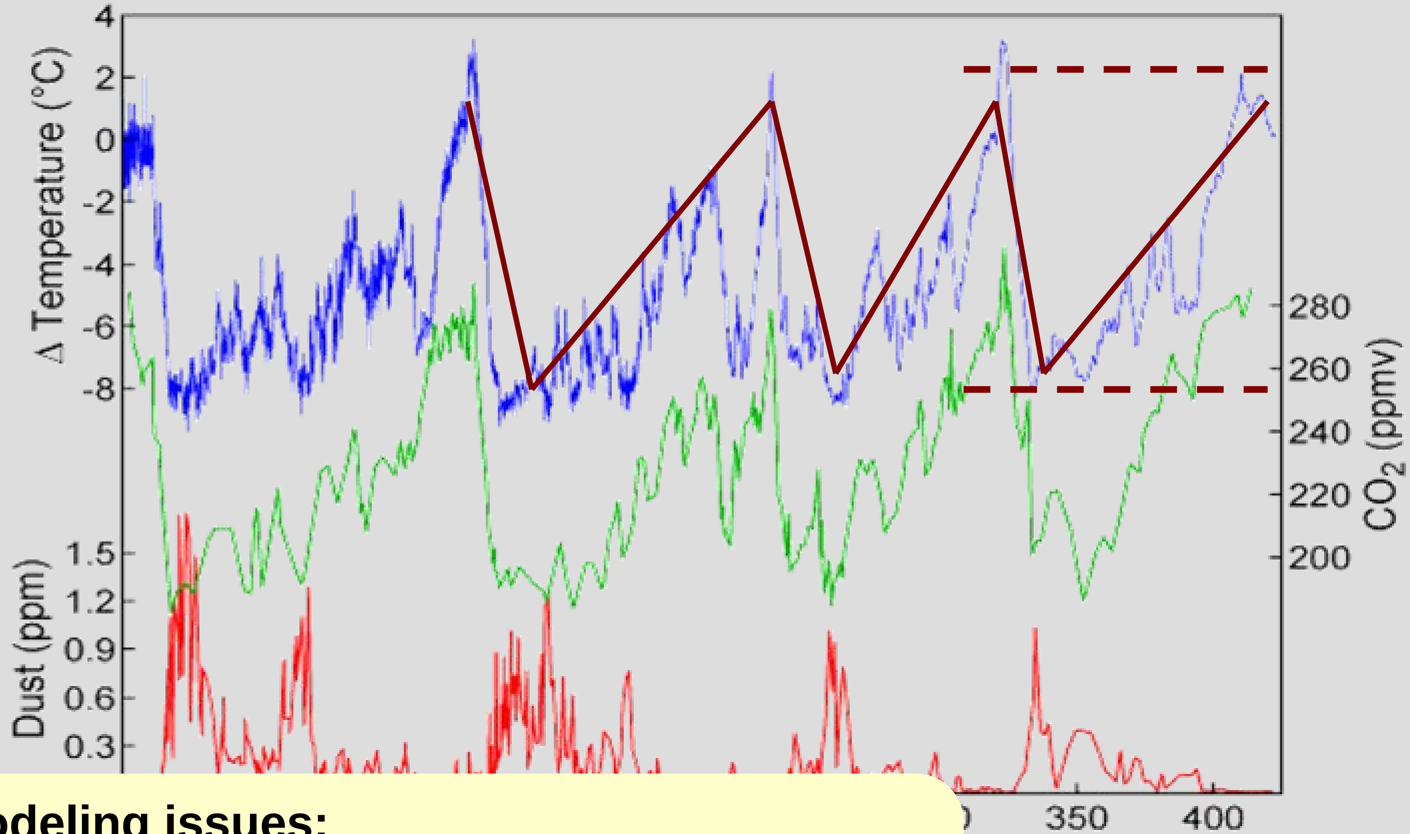
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

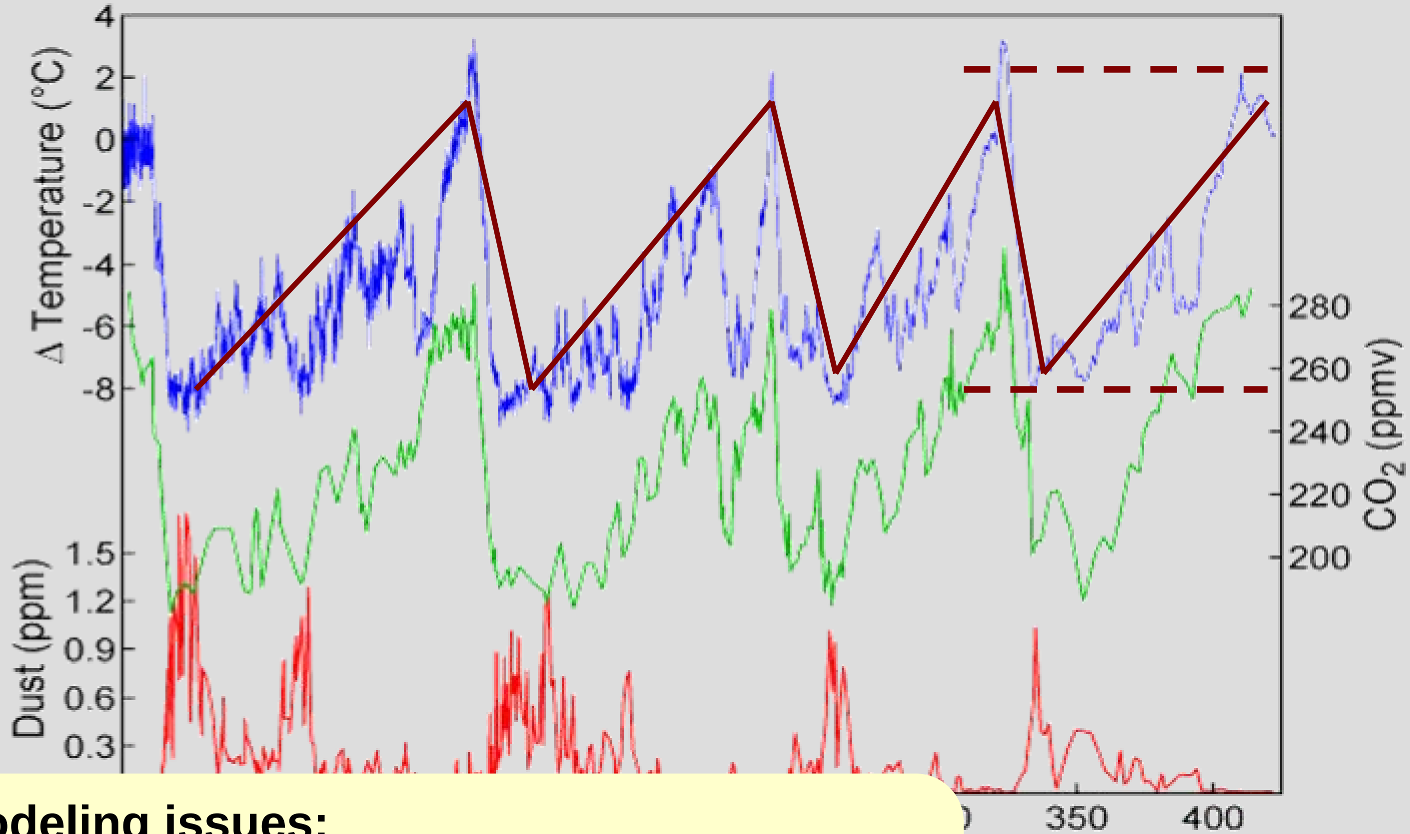
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

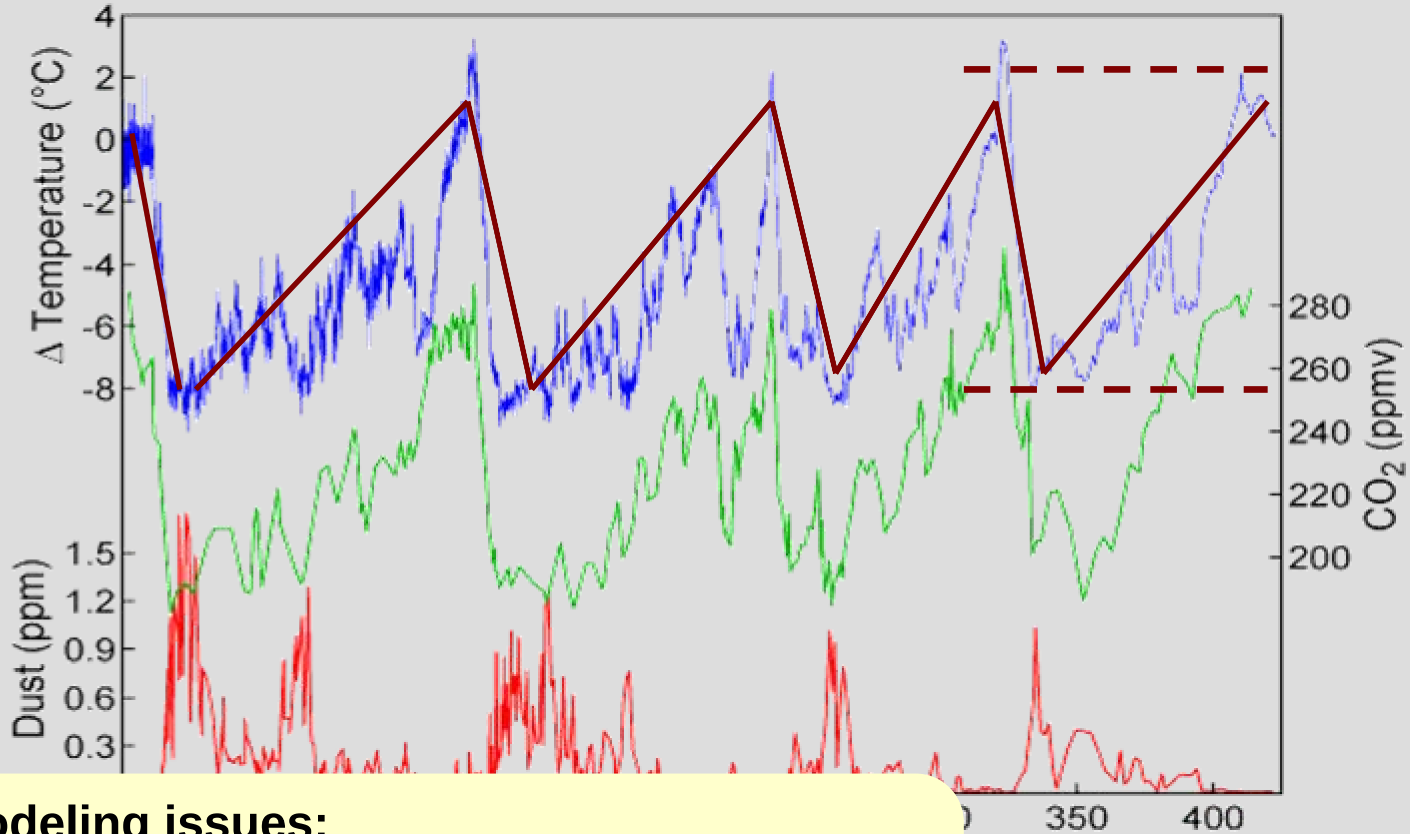
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

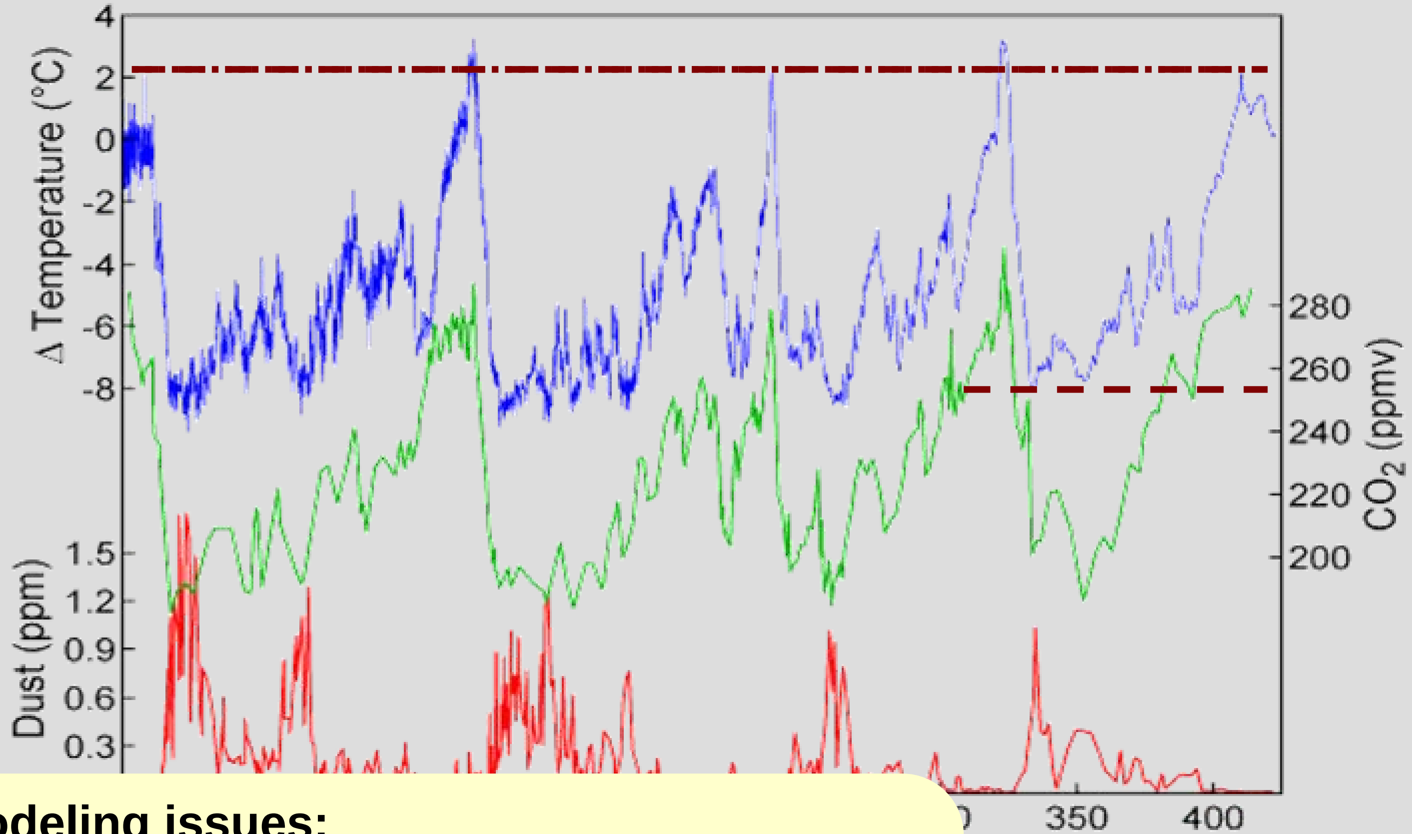
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering

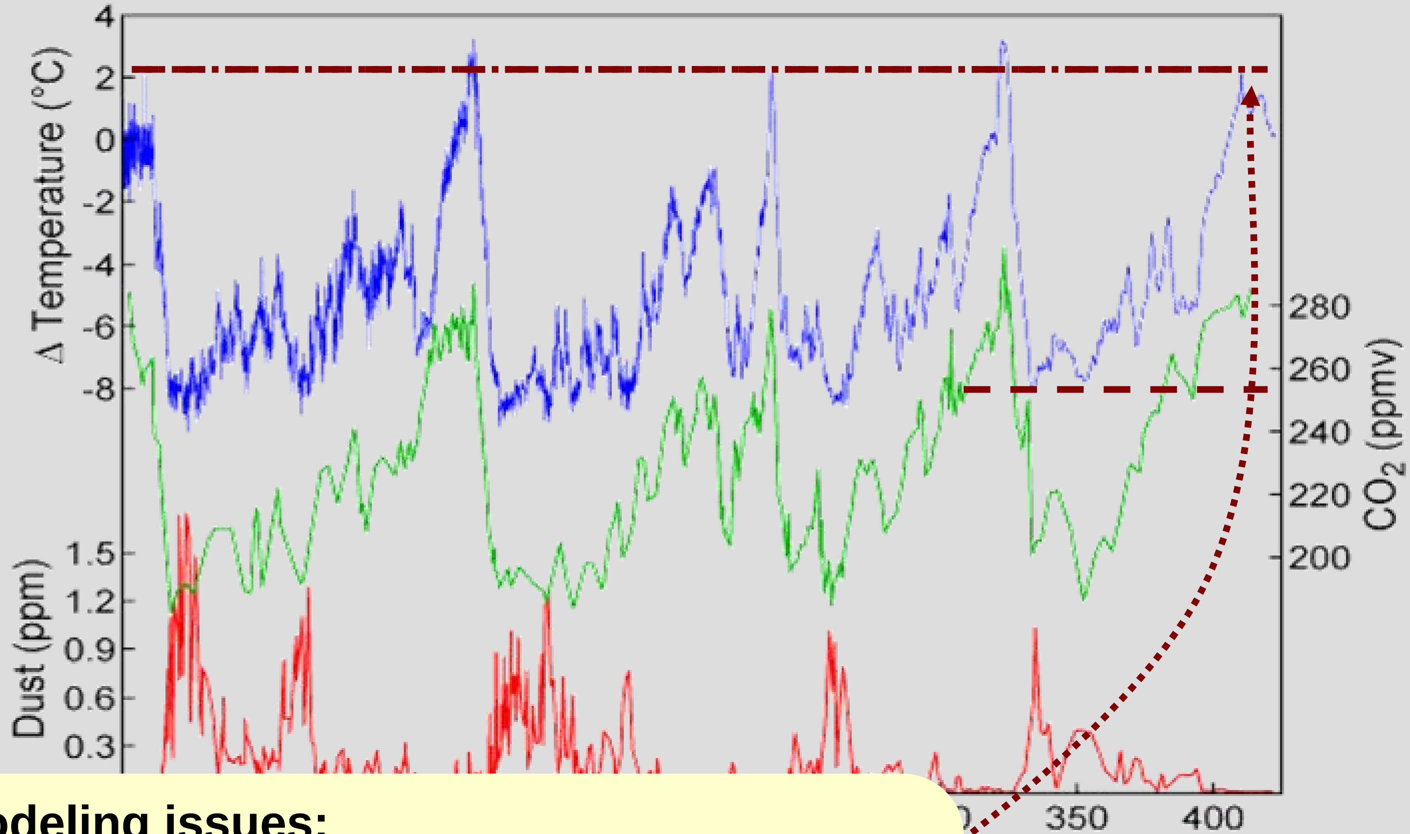
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

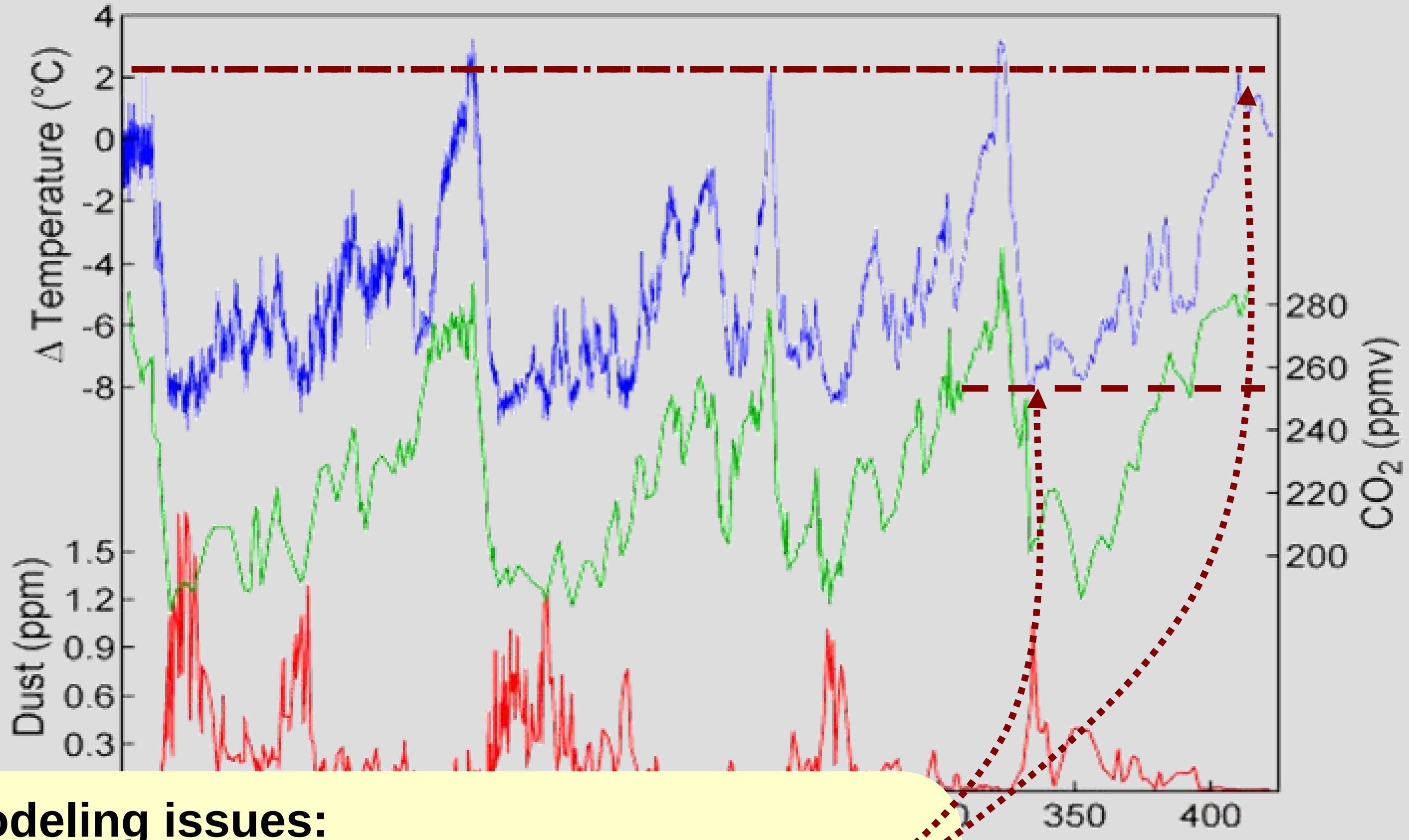
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

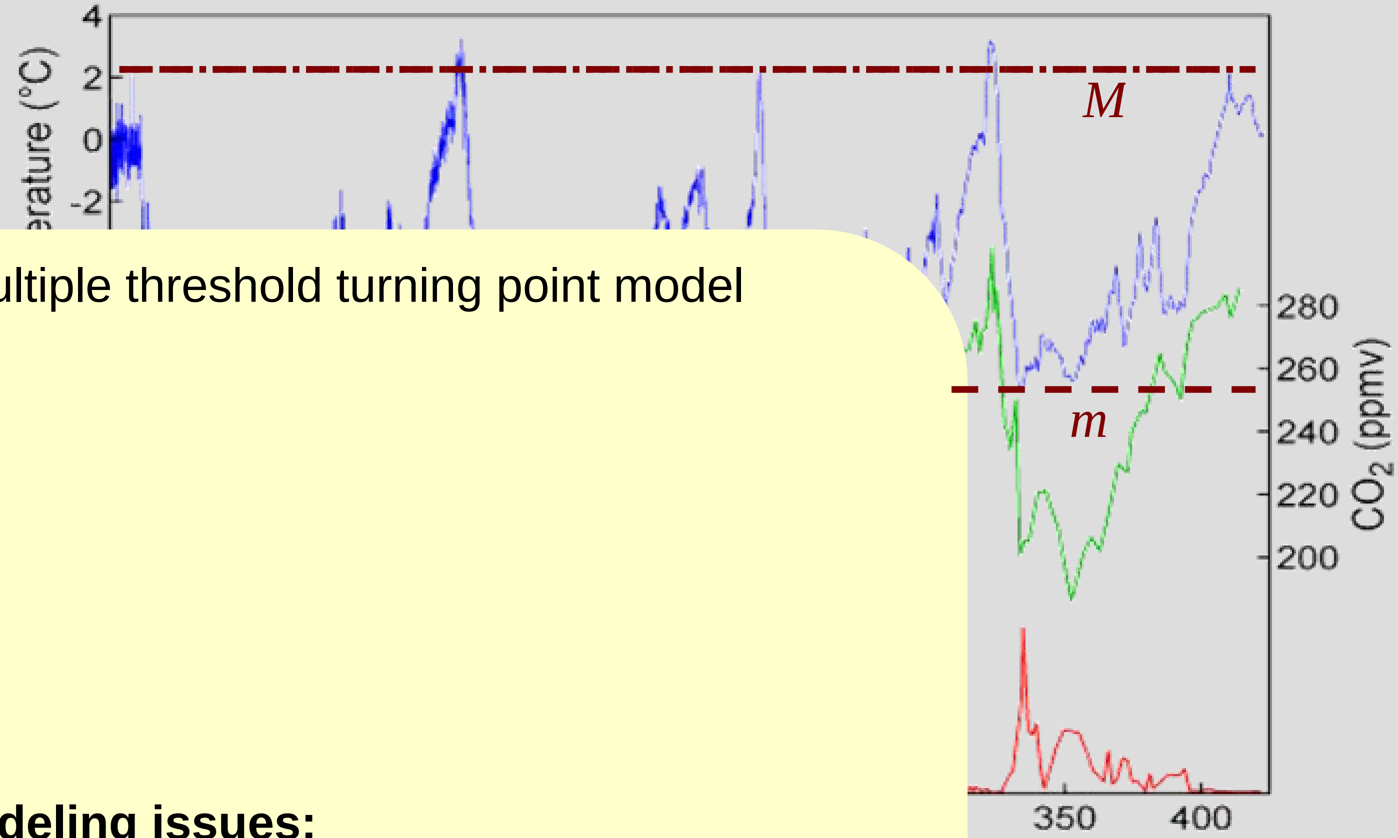
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

Climate Change: ice core data

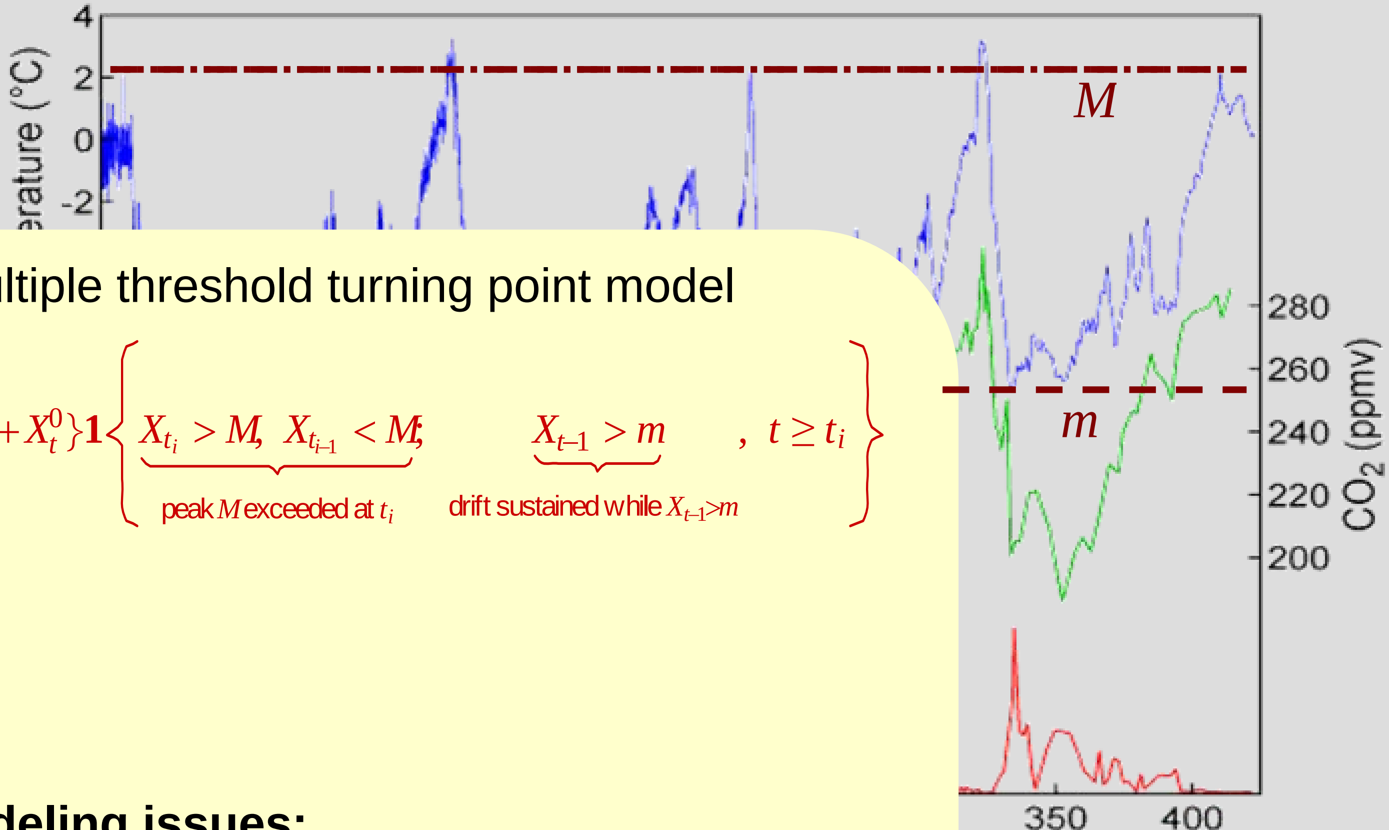


Multiple threshold turning point model

Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

Climate Change: ice core data



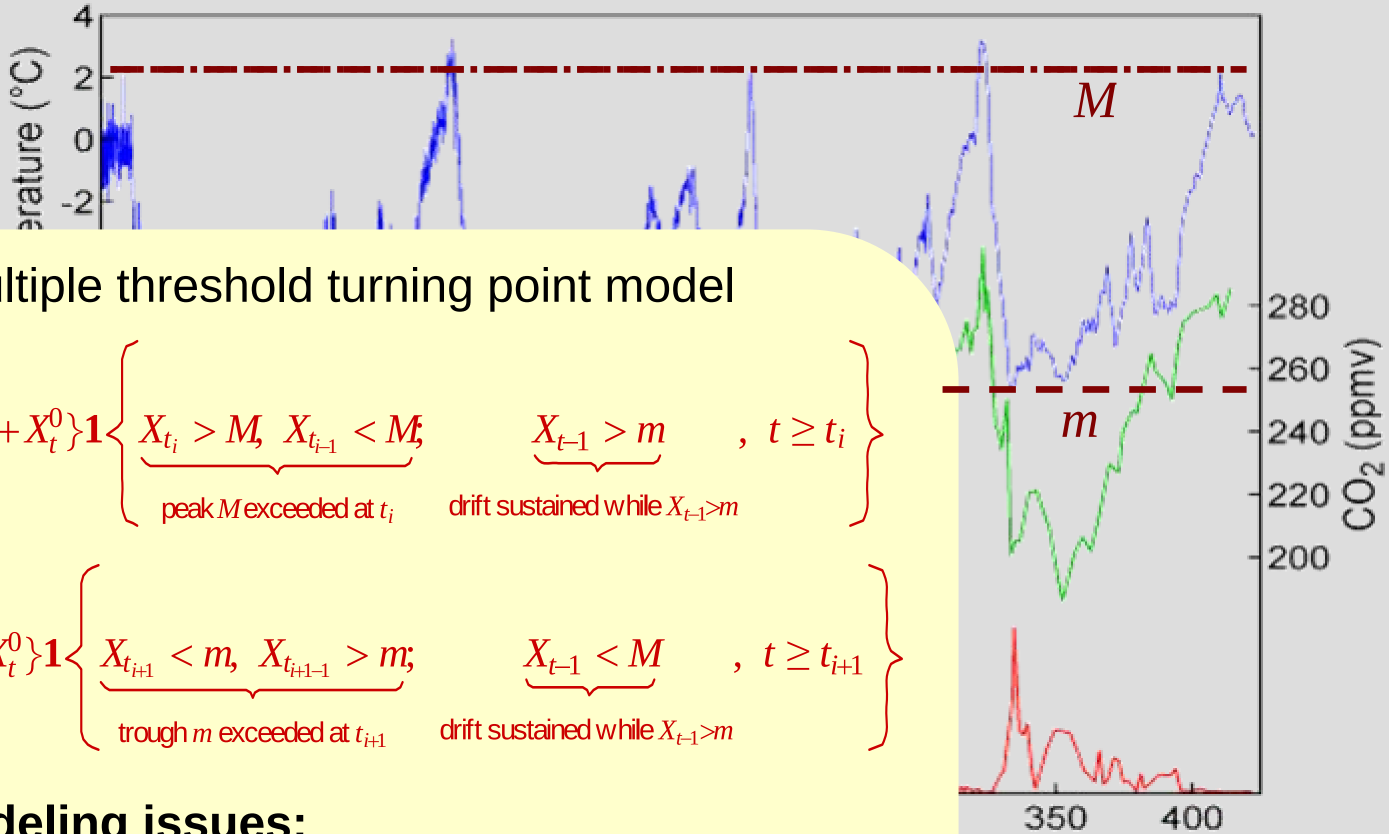
Multiple threshold turning point model

$$X_t = \{a_1 + b_1 t + X_t^0\}\mathbf{1}\left\{ \underbrace{X_{t_i} > M,\ X_{t_{i-1}} < M}_{\text{peak } M \text{ exceeded at } t_i};\ \underbrace{X_{t-1} > m}_{\text{drift sustained while } X_{t-1}>m},\ t \ge t_i \right\}$$

Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

Climate Change: ice core data



Multiple threshold turning point model

$$X_t = \{a_1 + b_1 t + X_t^0\}\mathbf{1}\left\{\underbrace{X_{t_i} > M,\ X_{t_{i-1}} < M;}_{\text{peak } M \text{ exceeded at } t_i}\quad \underbrace{X_{t-1} > m}_{\text{drift sustained while } X_{t-1}>m}\quad ,\ t \ge t_i\right\}$$

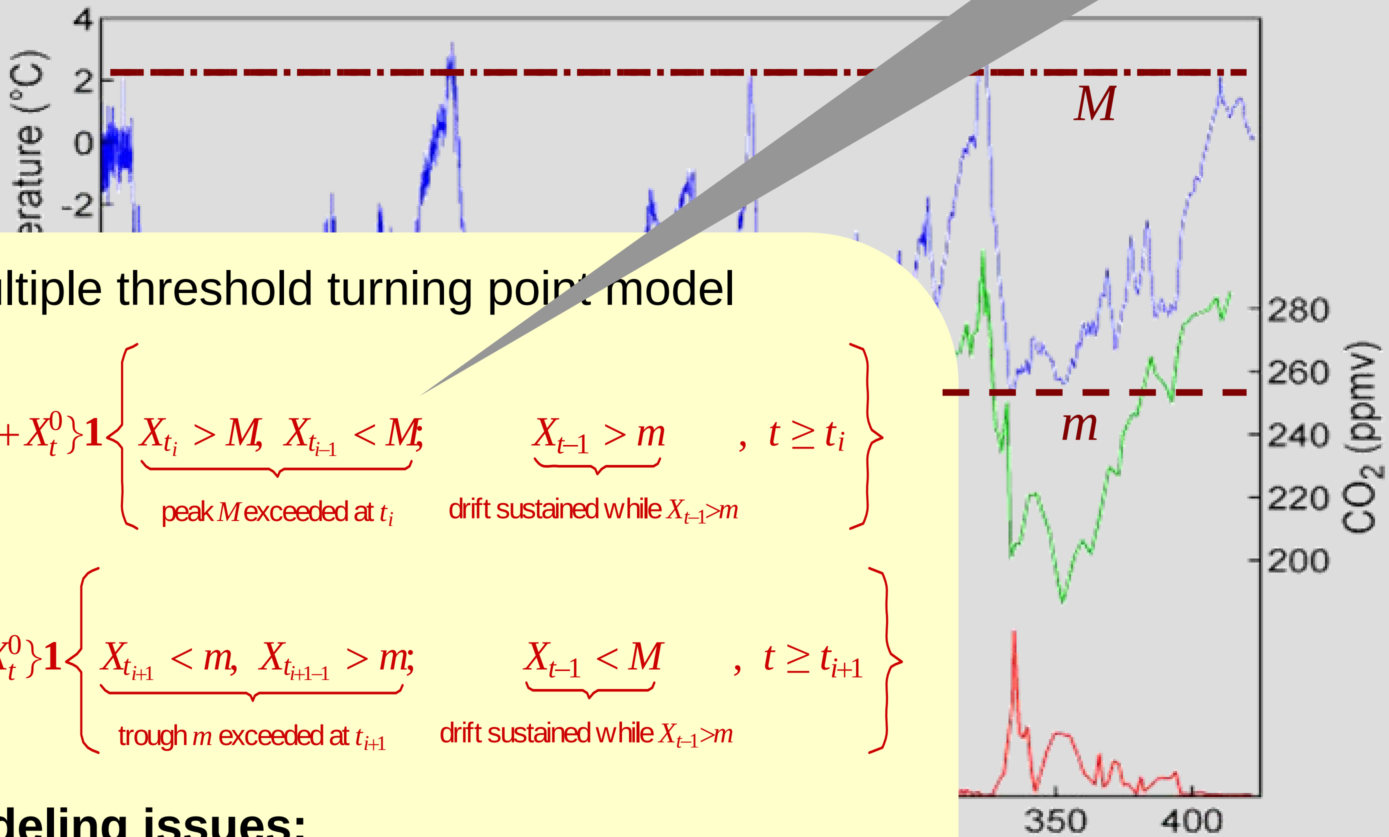
$$+\{a_2 + b_2 t + X_t^0\}\mathbf{1}\left\{\underbrace{X_{t_{i+1}} < m,\ X_{t_{i+1-1}} > m;}_{\text{trough } m \text{ exceeded at } t_{i+1}}\quad \underbrace{X_{t-1} < M}_{\text{drift sustained while } X_{t-1}>m}\quad ,\ t \ge t_{i+1}\right\}$$

Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

Climate Change: ice co

further issues:
- duration over M
- duration below m



Multiple threshold turning point model

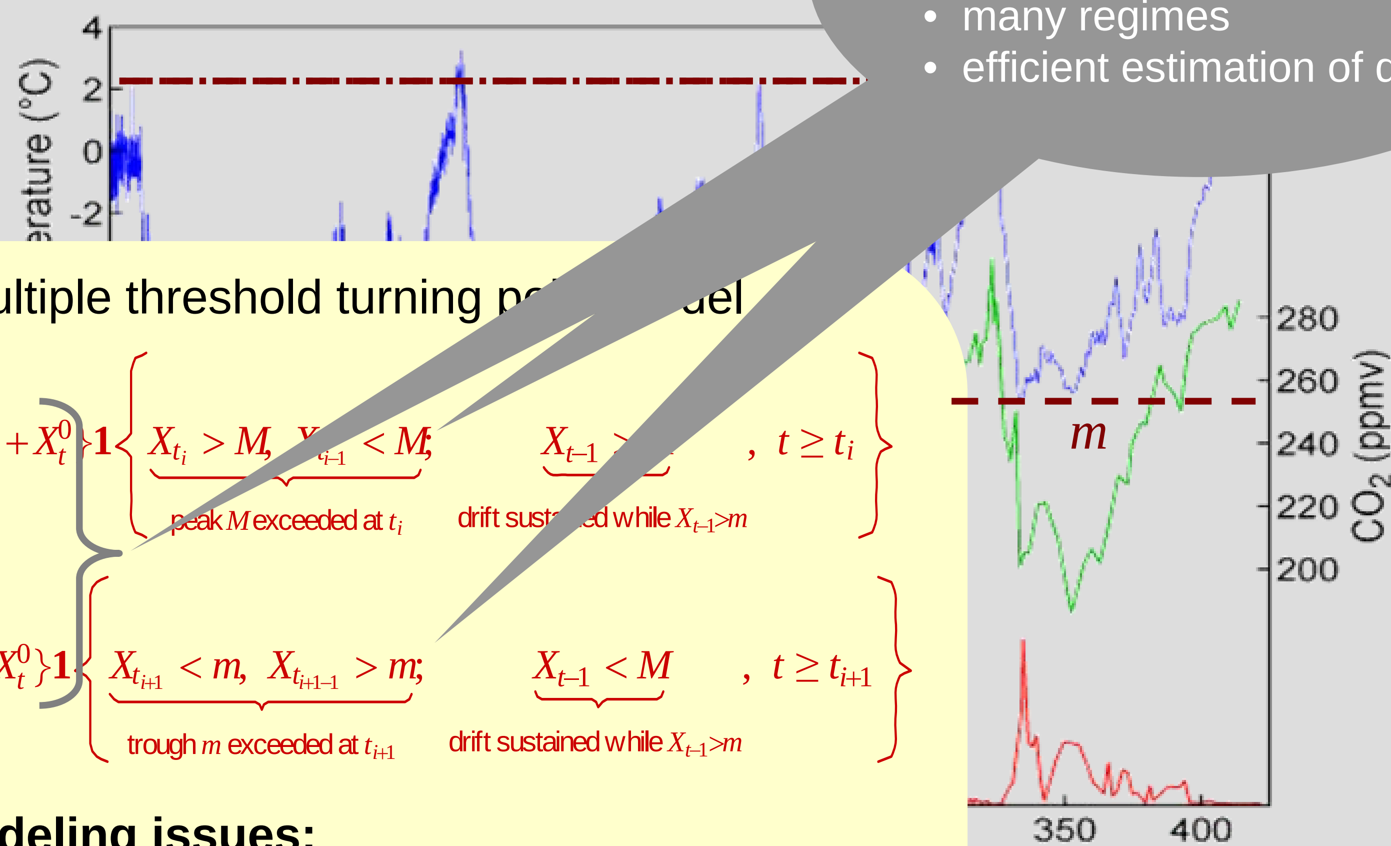
$$X_t = \{a_1 + b_1 t + X_t^0\}\mathbf{1}\left\{\underbrace{X_{t_i} > M,\ X_{t_{i-1}} < M;}_{\text{peak } M \text{ exceeded at } t_i} \quad \underbrace{X_{t-1} > m}_{\text{drift sustained while } X_{t-1} > m},\ t \geq t_i\right\}$$

$$+\{a_2 + b_2 t + X_t^0\}\mathbf{1}\left\{\underbrace{X_{t_{i+1}} < m,\ X_{t_{i+1-1}} > m;}_{\text{trough } m \text{ exceeded at } t_{i+1}} \quad \underbrace{X_{t-1} < M}_{\text{drift sustained while } X_{t-1} > m},\ t \geq t_{i+1}\right\}$$

Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

Climate Change: ice



further issues:

- duration over *M*
- duration below *m*
- many regimes
- efficient estimation of drift

Multiple threshold turning point model

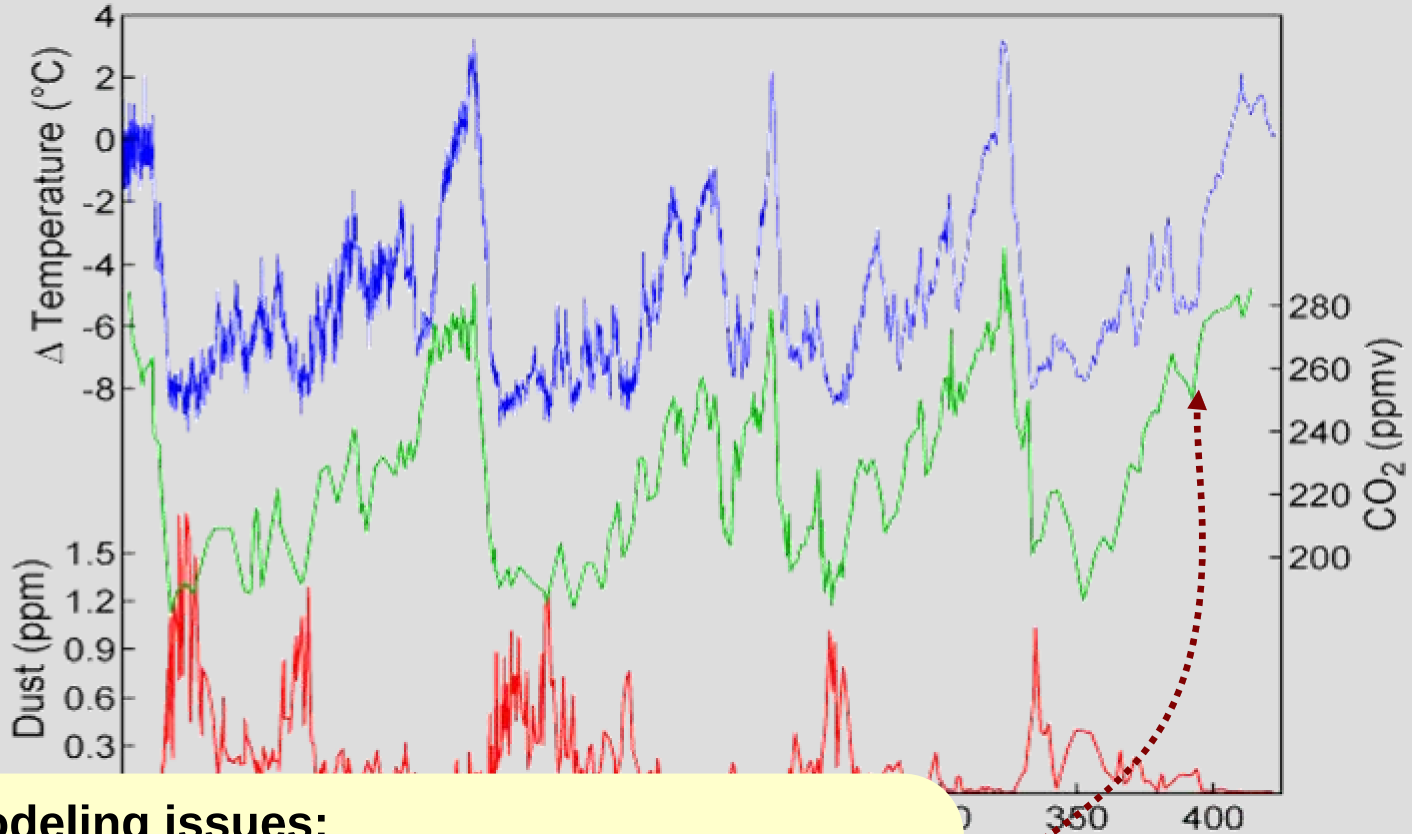
$$X_t = \{a_1 + b_1 t + X_t^0\}\mathbf{1}\Big\{\underbrace{X_{t_i} > M,\ X_{t_{i-1}} < M}_{\text{peak } M \text{ exceeded at } t_i};\ \underbrace{X_{t-1} > m}_{\text{drift sustained while } X_{t-1}>m},\ t \geq t_i\Big\}$$

$$+\{a_2 + b_2 t + X_t^0\}\mathbf{1}\Big\{\underbrace{X_{t_{i+1}} < m,\ X_{t_{i+1-1}} > m}_{\text{trough } m \text{ exceeded at } t_{i+1}};\ \underbrace{X_{t-1} < M}_{\text{drift sustained while } X_{t-1}>m},\ t \geq t_{i+1}\Big\}$$

Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points

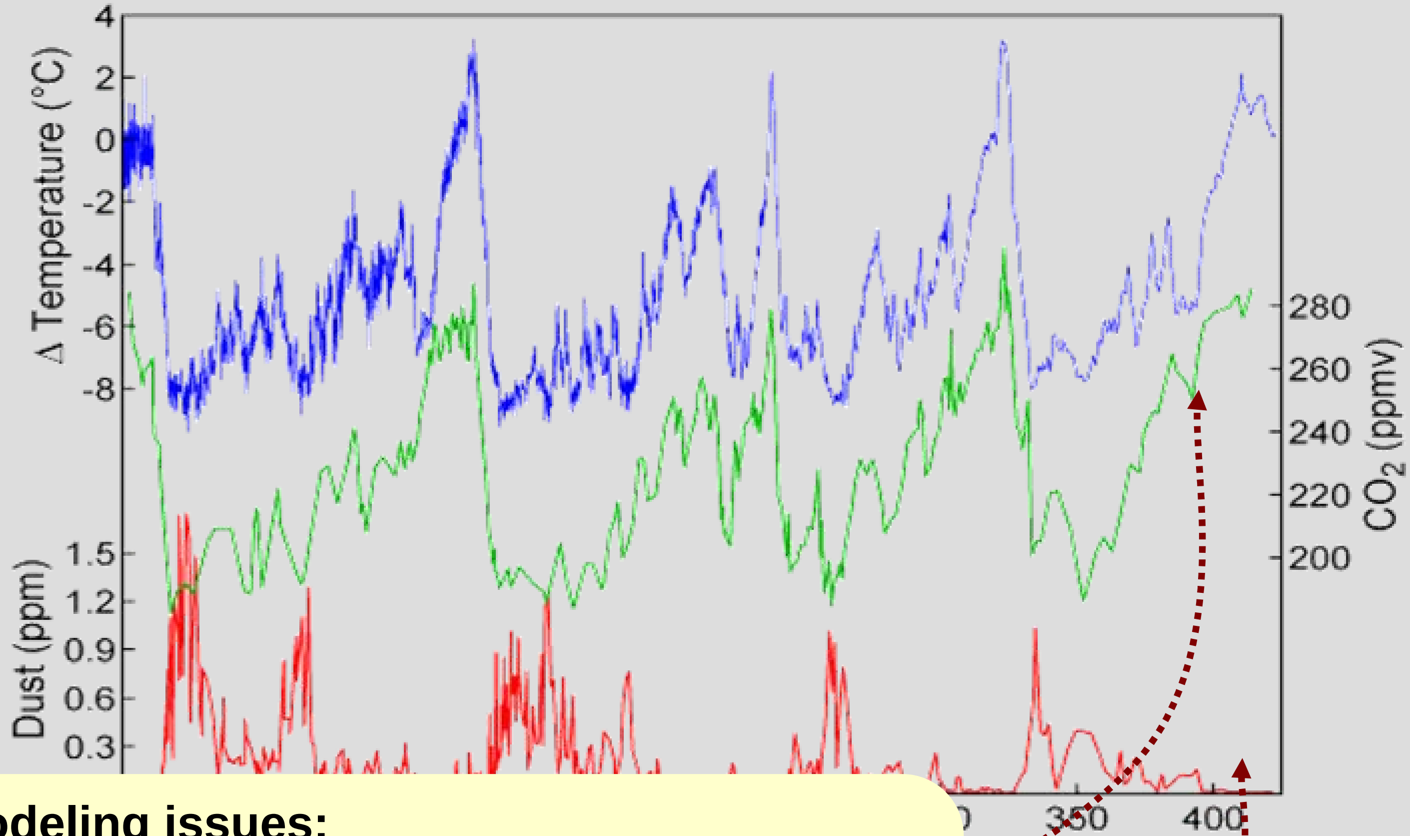
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points
e. comovement with CO_2

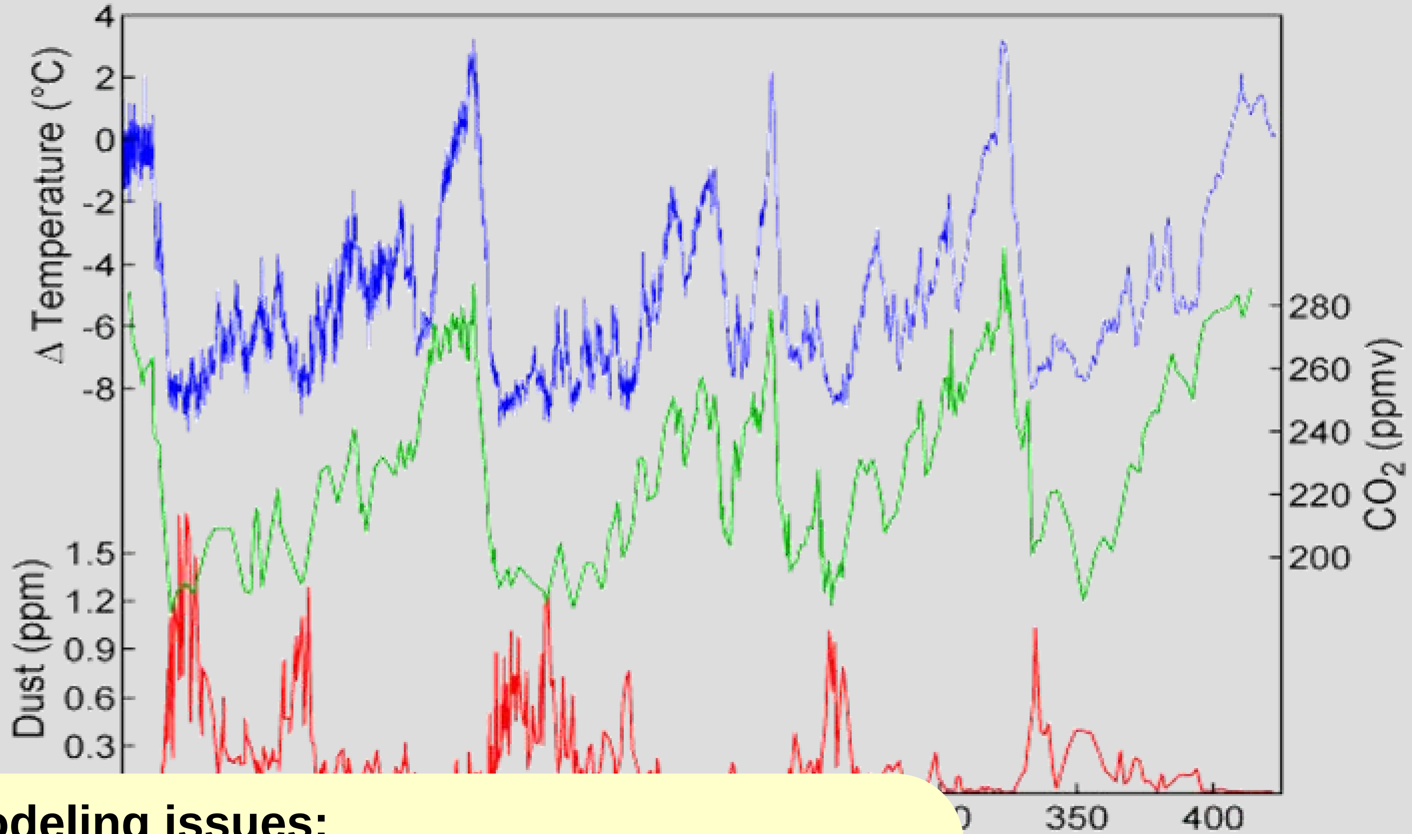
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points
e. comovement with CO_2 and Dust

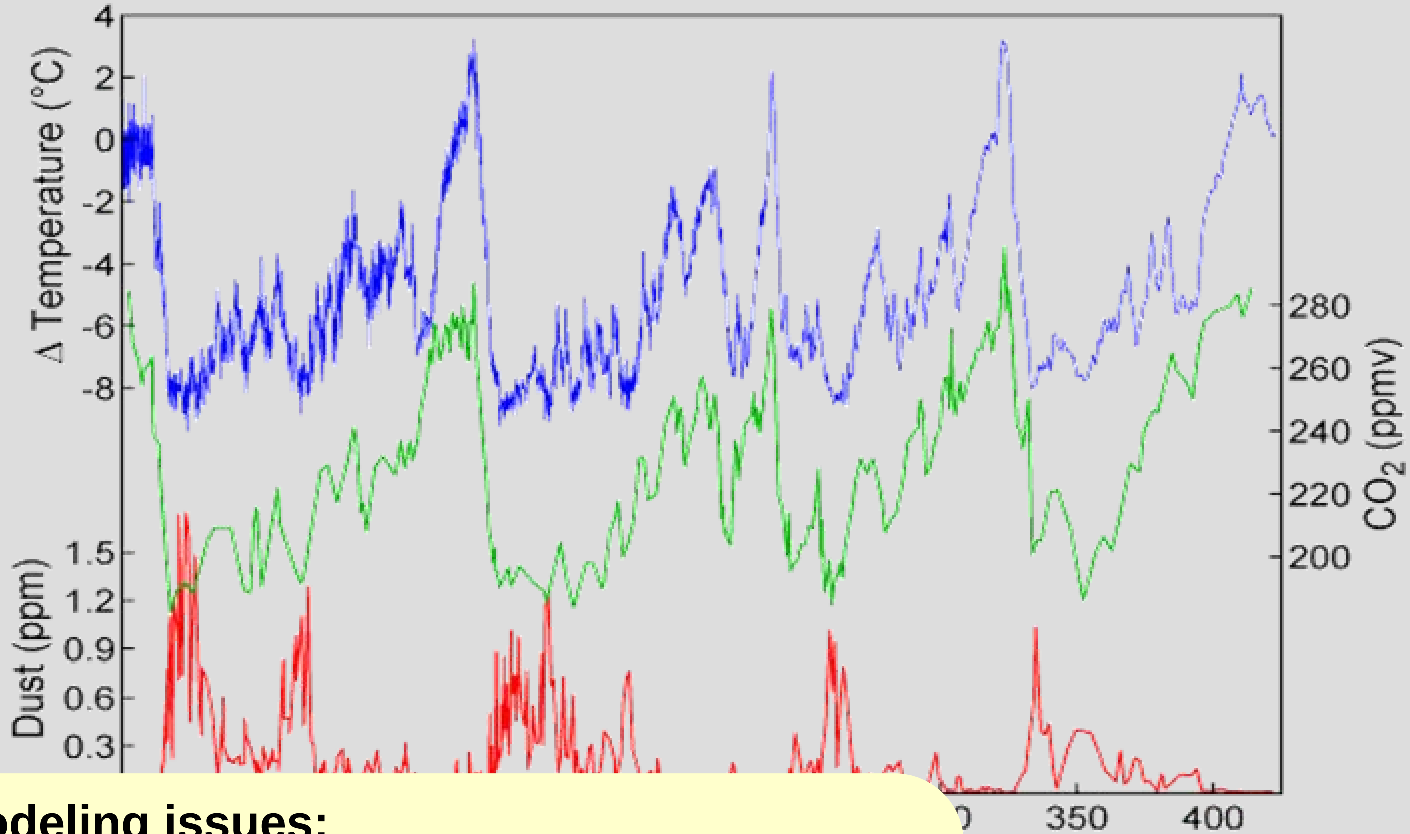
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points
e. comovement with CO_2 and Dust
f. causal anticipation

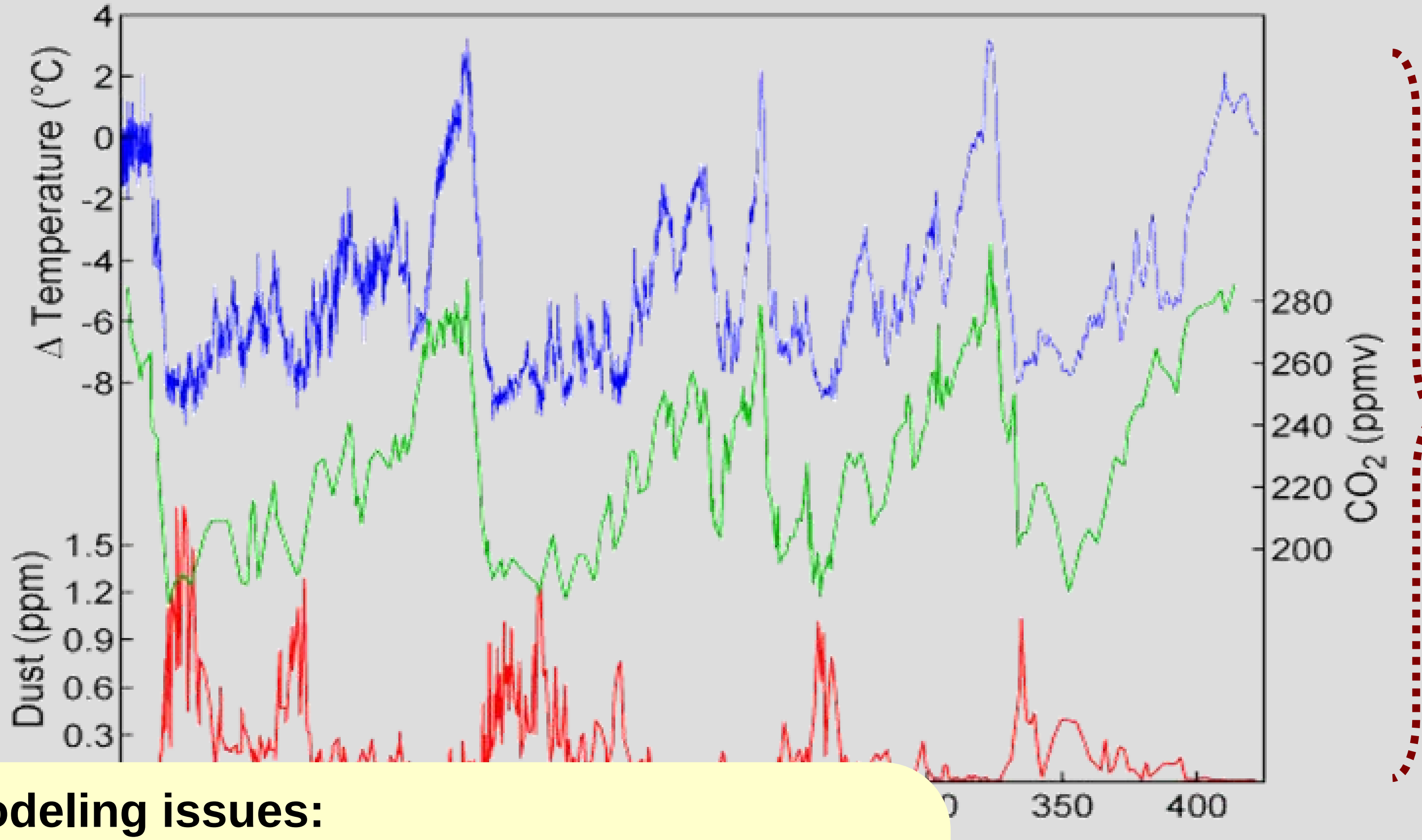
Climate Change: ice core data



Modeling issues:

a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points
e. comovement with CO_2 and Dust
f. causal anticipation

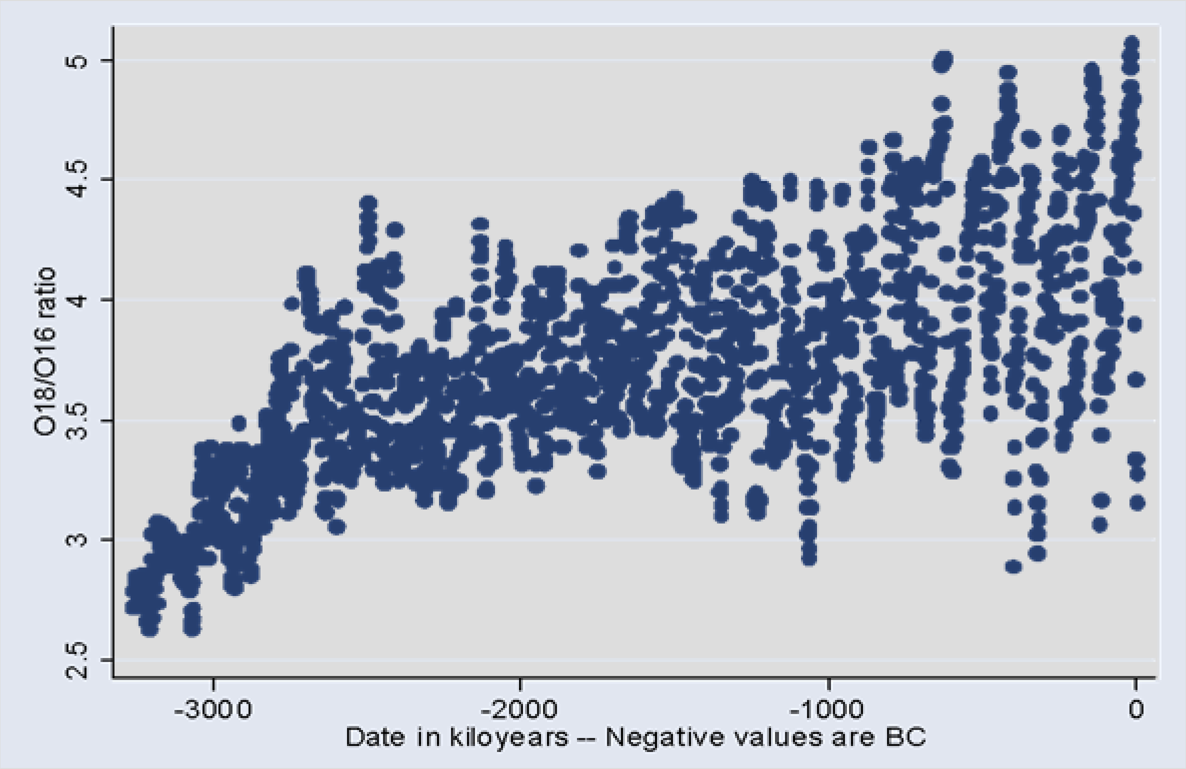
Climate Change: ice core data



Modeling issues:

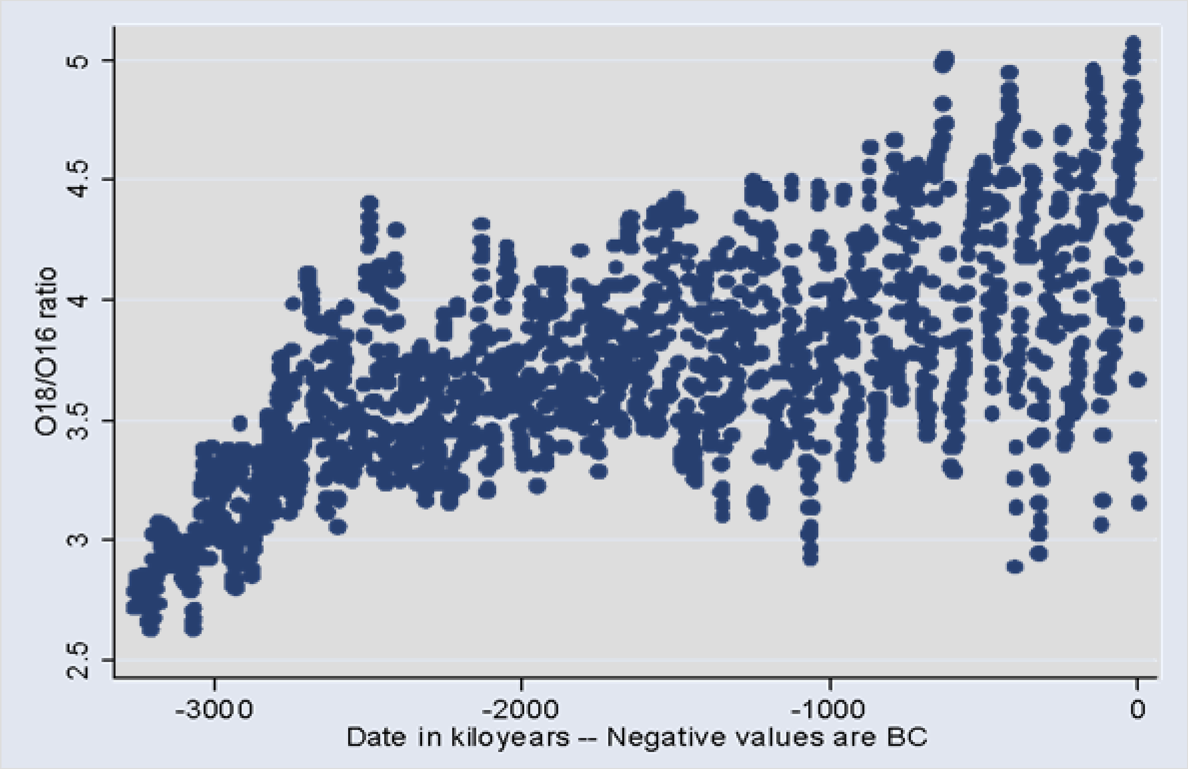
a. irregular cycle & heterogeneity
b. drift and random wandering within cycle
c. regulated drift and random wandering
d. thresholds & turning points
e. comovement with CO_2 and Dust
f. causal anticipation

Climate Change: drilling data



Deep Sea Atlantic Drilling Data

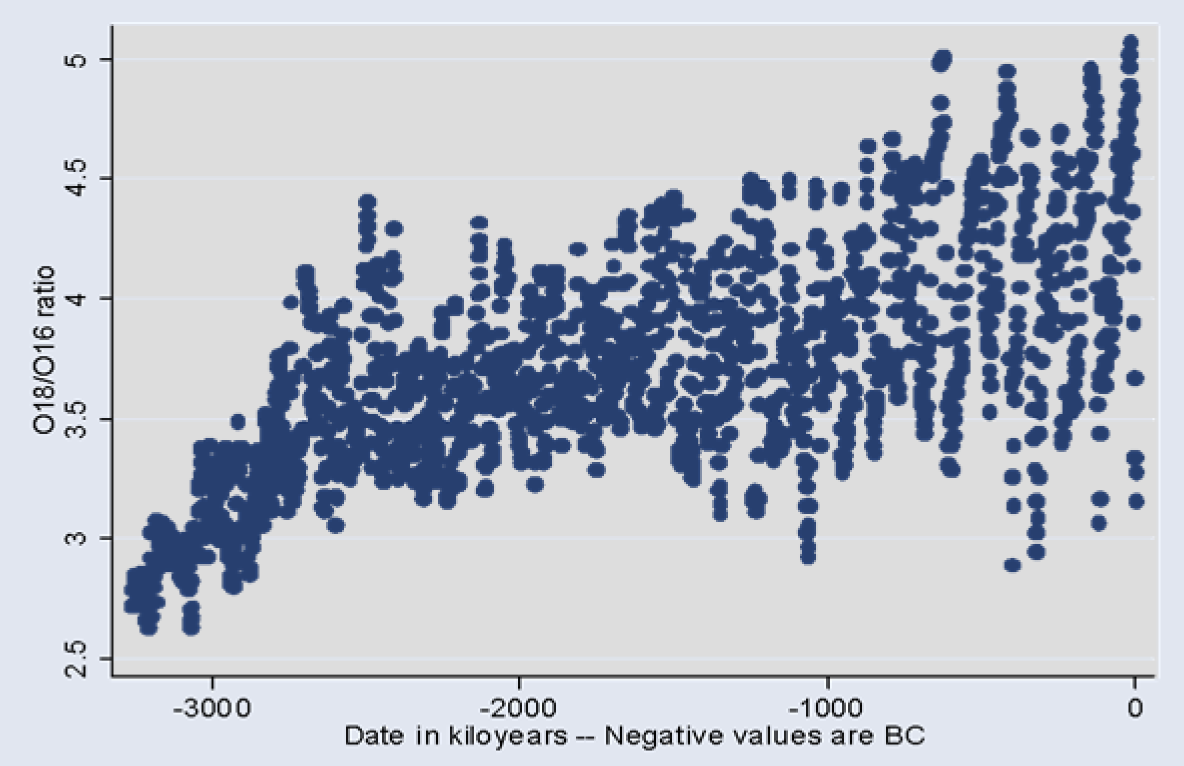
Climate Change: drilling data



Colder as O^{18} increases

Deep Sea Atlantic Drilling Data

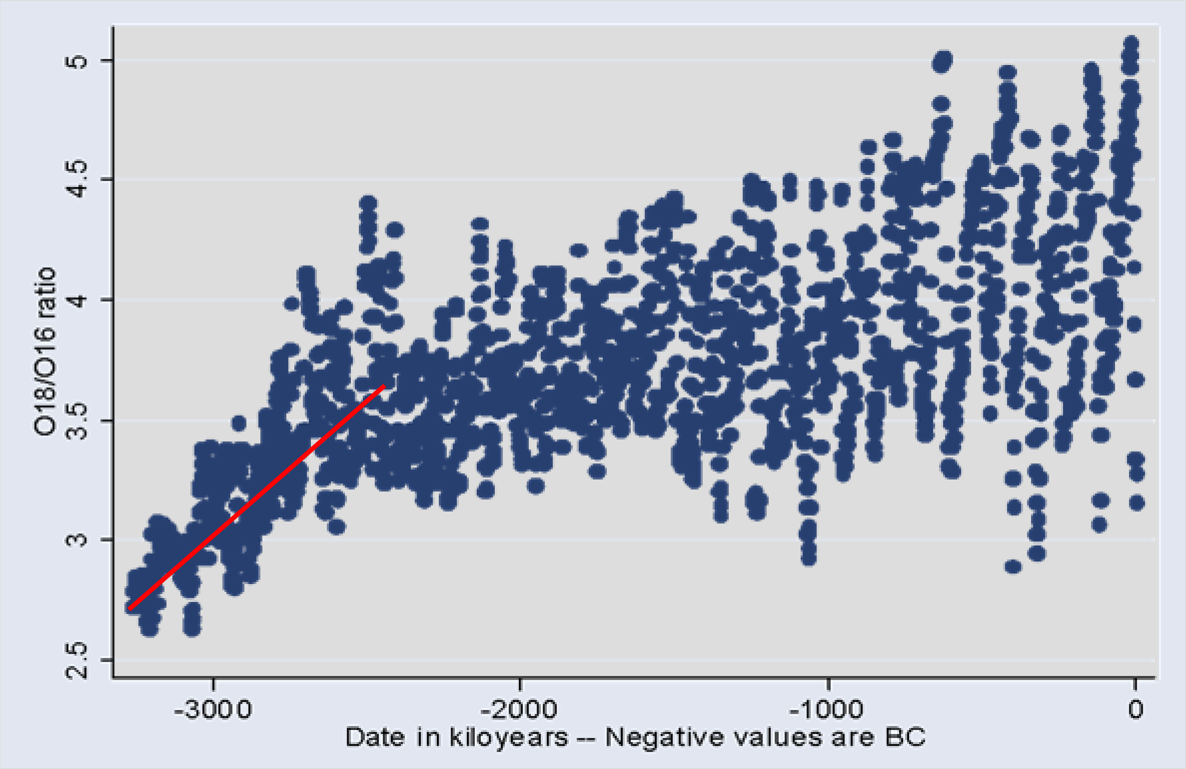
Climate Change: drilling data



Modeling issues:

a. Longer term trends & embedding ice core data

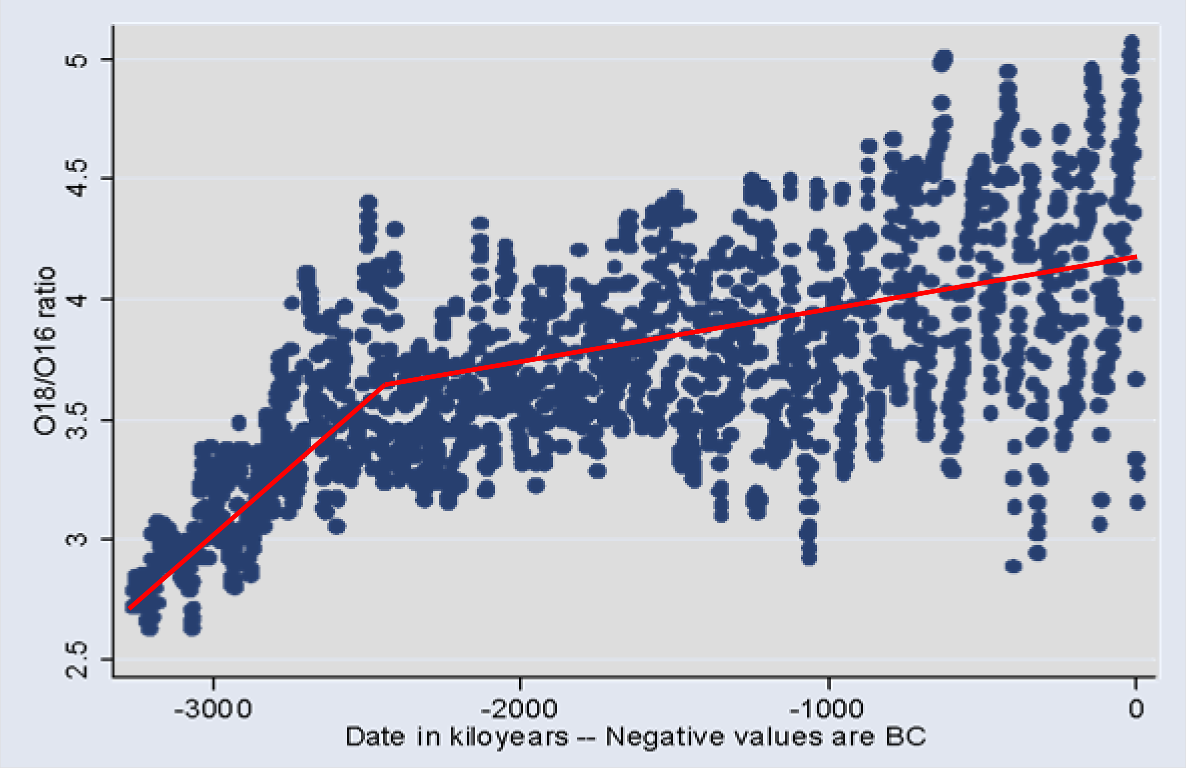
Climate Change: drilling data



Modeling issues:

a. Longer term trends & embedding ice core data

Climate Change: drilling data

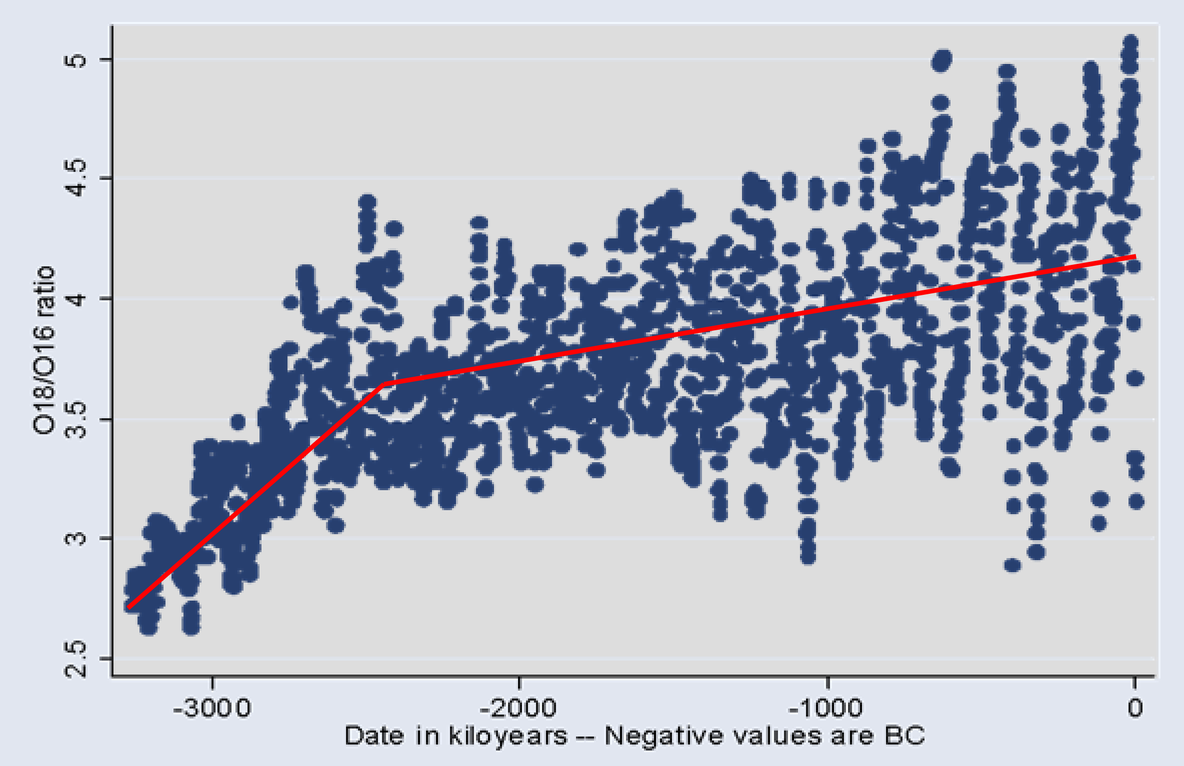


Cooling trend over 3myr

Modeling issues:

a. Longer term trends & embedding ice core data

Climate Change: drilling data



Cooling trend over 3myr

Modeling issues:

a. Longer term trends & embedding ice core data
b. Heterogeneity & measurement error

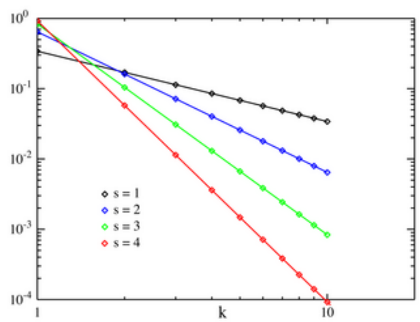
Ideas and Motivation

Basic Properties of Economic & Financial Time Series & Panels

1. Temporal dependence (first and higher moments)
2. Joint dependence - endogeneity, cross correlation
3. Nonstationarity (secular growth, random wandering behavior, long memory)
4. Individual effects + time effects - panel characteristics
5. Volatility & conditional volatility - second moment modeling
6. Heavy tails & outlier activity (Pareto Law, Zipf law; power law probability)
   (a) Income and wealth distributions in economics
   (b) Company size in finance - frequency inversely proportional to rank

Zipf Law (Harvard linguist - George Zipf)

$$f(k; s, N) = \frac{\frac{1}{k^s}}{\sum_{n=1}^{N} \frac{1}{n^s}}$$



Zipf Law probability function (log scale)

company size (few large multinationals, many small businesses)

statistical occurence of words in different languages (few special nouns, many articles)

internet trafffic & frequency of access to web pages

top income earners, earthquake size, human settlement size etc

Hill Estimator of Tail Slope Parameter

1. Pareto Tail Shape

$$\left.\begin{array}{c} P\left(X > x\right) \\ P\left(X < -x\right) \end{array}\right\} = \left\{\begin{array}{l} \frac{a}{x^{a}}\left\{1 + \frac{d}{x^{\beta}} + o\left(\frac{1}{x^{\beta}}\right)\right\} \\ \frac{b}{x^{a}}\left\{1 + \frac{d}{x^{\beta}} + o\left(\frac{1}{x^{\beta}}\right)\right\} \end{array}\right. \qquad \alpha, \beta, a, b > 0$$

2. Order Statistics

$$X_1, X_2, X_3, ..., X_j, ...X_n$$

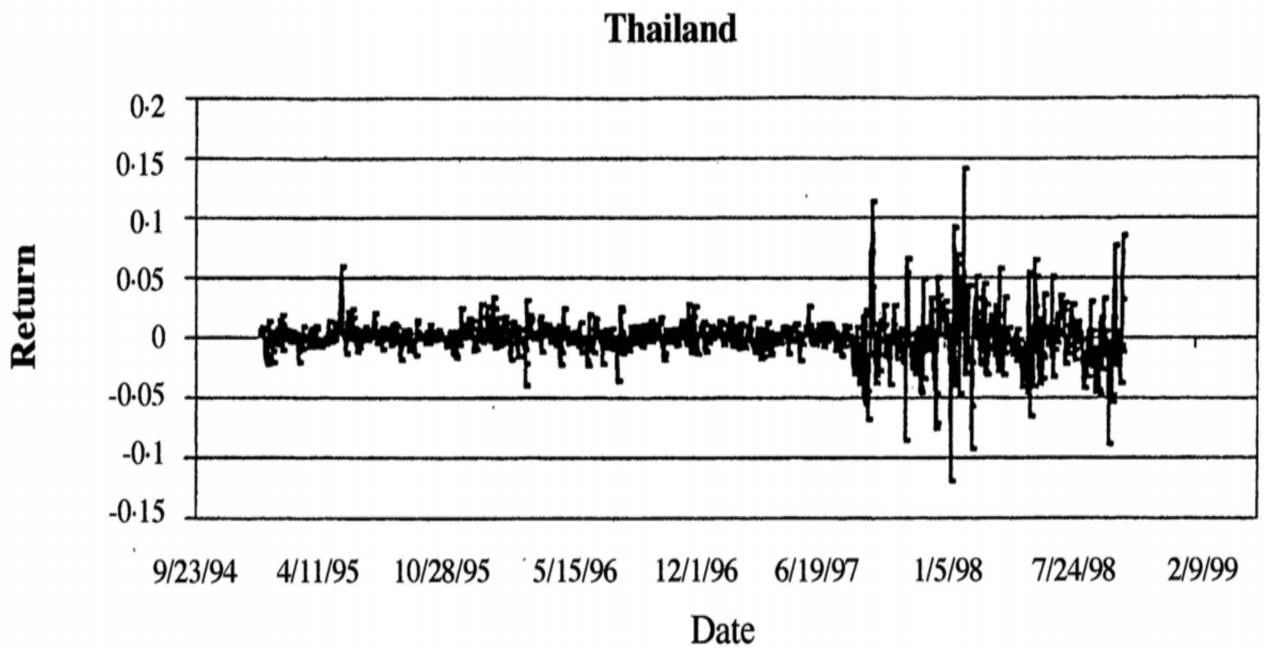
$$X_{(1)} \; < \; X_{(2)} < X_{(3)} < .... < X_{(j)} < ..... < X_{(n)}$$

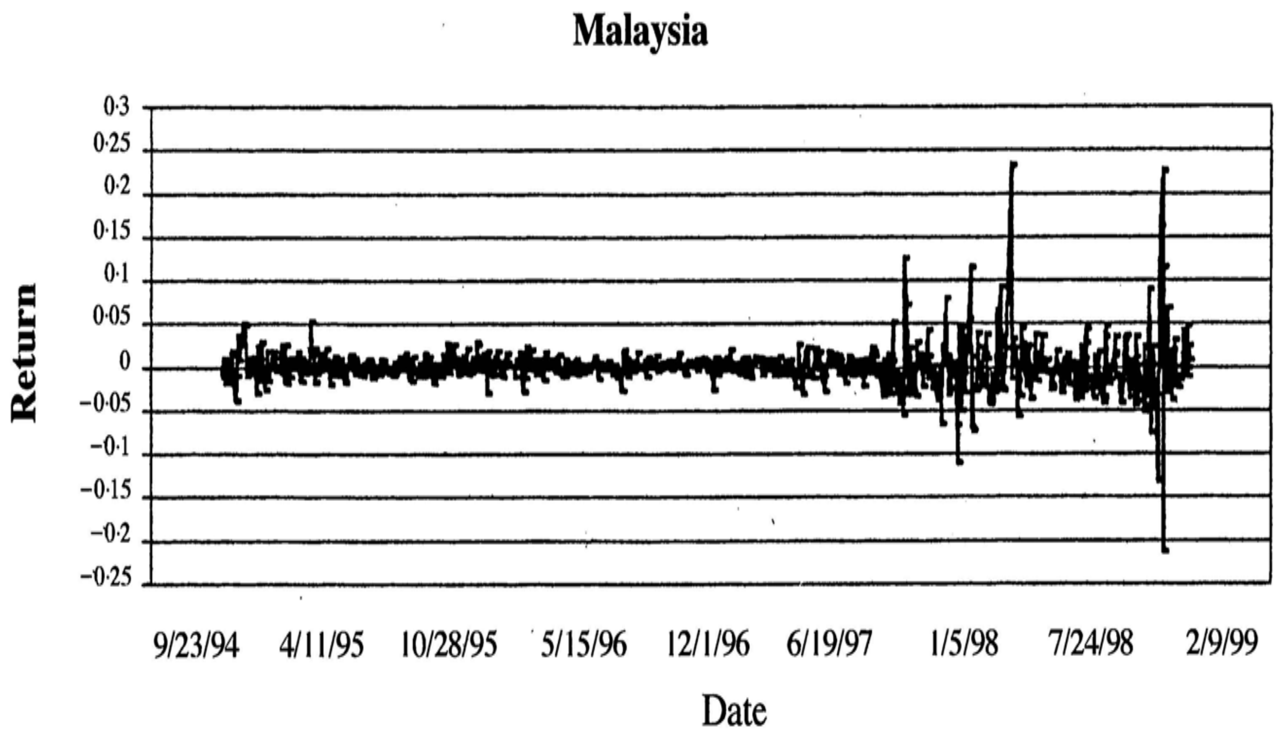
3. Hill Estimator of tail slope parameter

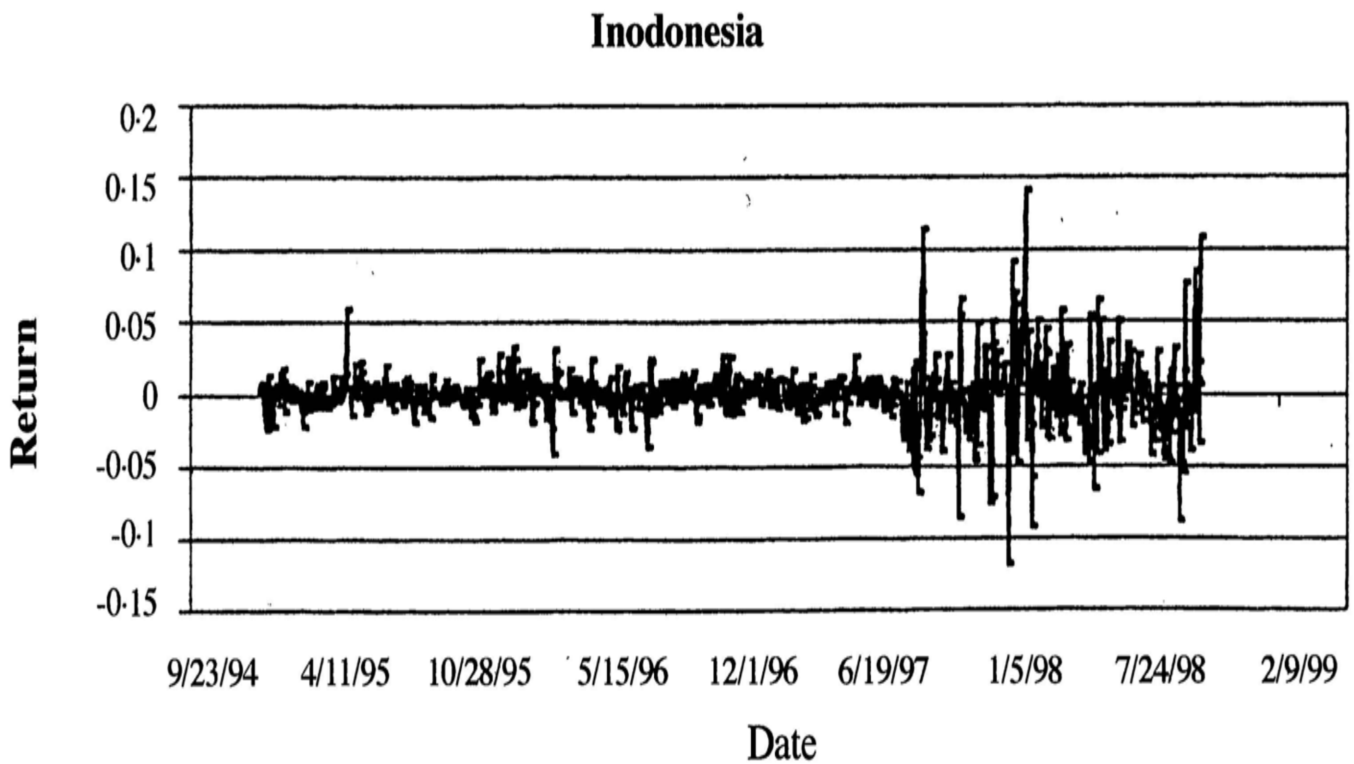
$$\hat{\alpha} = \frac{1}{\frac{1}{m+1}\sum_{k=0}^{m} \log \frac{X_{(n-k)}}{X_{(n-m)}}}, \quad m+1 \text{ largest observations}$$

4. Limit distribution

$$\sqrt{m}\left(\hat{\alpha} - \alpha\right) \rightarrow_d N\left(0, \alpha^2\right), \quad \frac{1}{m} + \frac{m^{\frac{2\beta}{2\beta+\alpha}}}{n} \rightarrow 0$$


Thailand
0·2
0·15
0·1
0·05
0
-0·05
-0·1
-0·15
Return
9/23/94
4/11/95
10/28/95
5/15/96
12/1/96
6/19/97
1/5/98
7/24/98
2/9/99
Date


Malaysia
0·3
0·25
0·2
0·15
0·1
0·05
0
−0·05
−0·1
−0·15
−0·2
−0·25
Return
9/23/94
4/11/95
10/28/95
5/15/96
12/1/96
6/19/97
1/5/98
7/24/98
2/9/99
Date


Inodonesia
Return
0·2
0·15
0·1
0·05
0
-0·05
-0·1
-0·15
9/23/94
4/11/95
10/28/95
5/15/96
12/1/96
6/19/97
1/5/98
7/24/98
2/9/99
Date

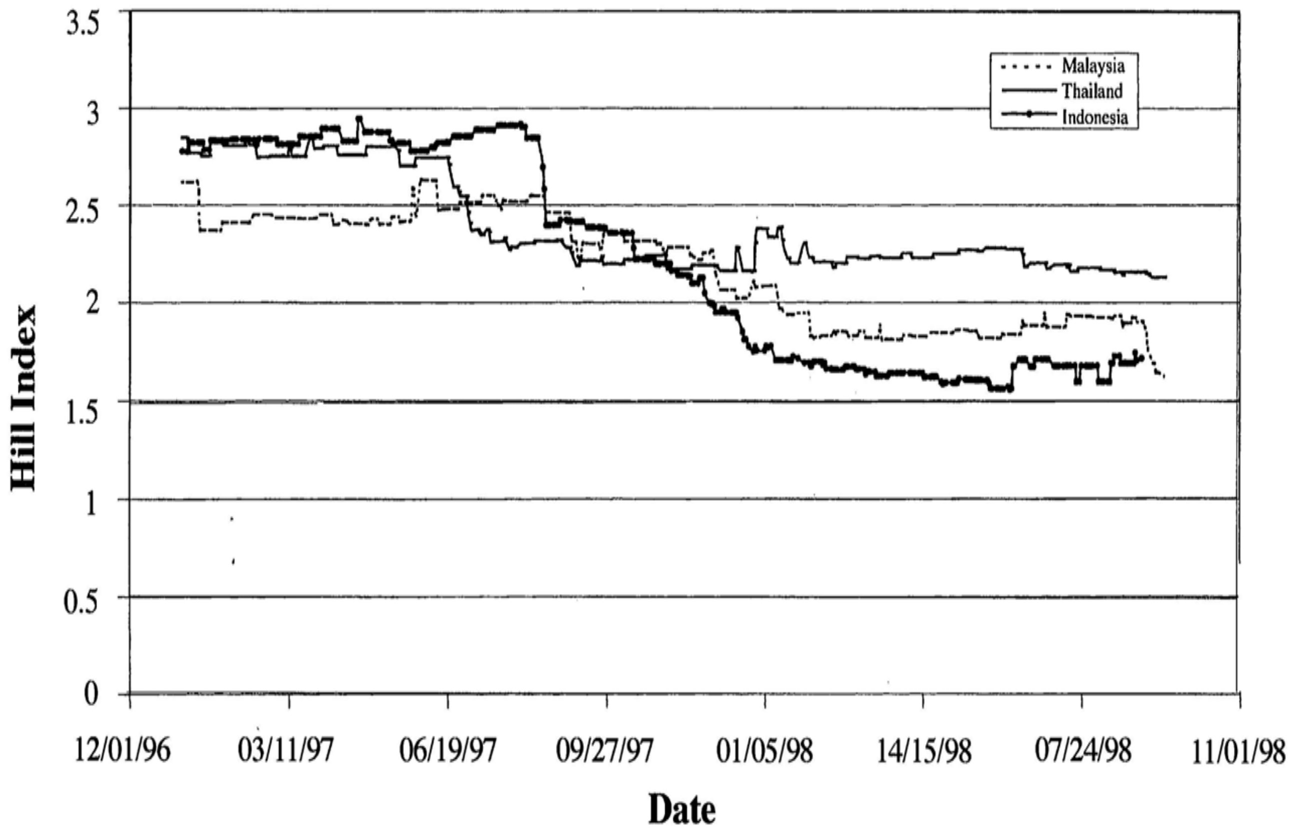
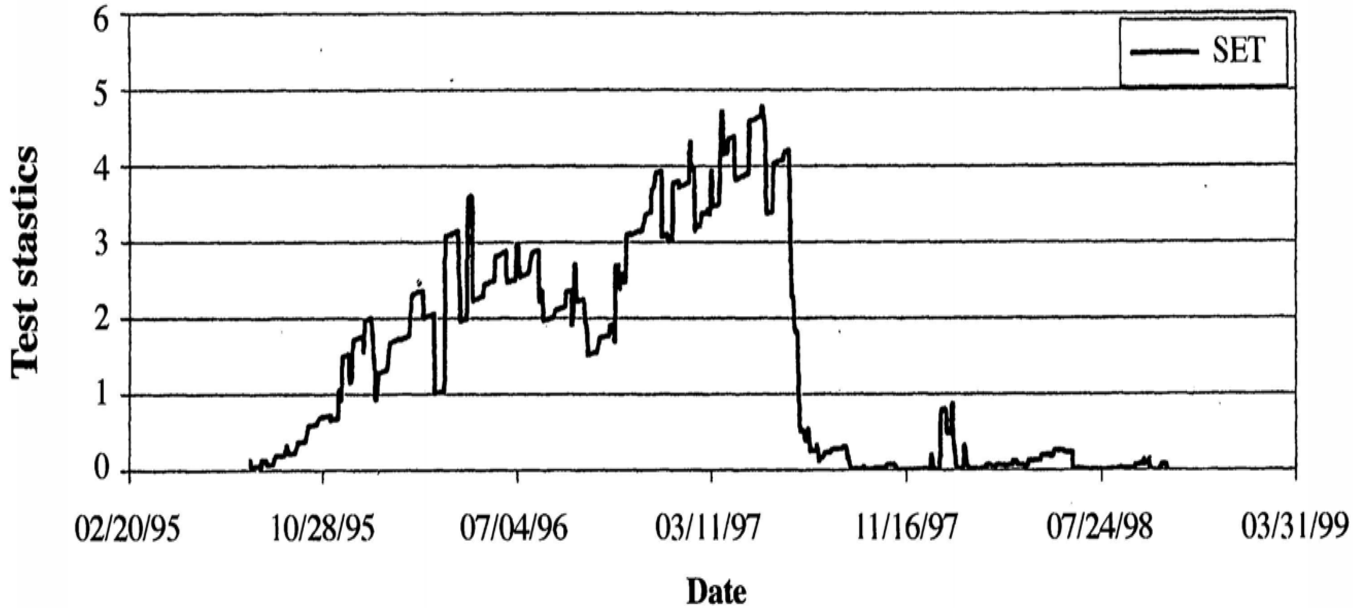


FIGURE 3
Hill's index: recursive



Test stastics
6
5
4
3
2
1
0
SET
02/20/95
10/28/95
07/04/96
03/11/97
11/16/97
07/24/98
03/31/99
Date

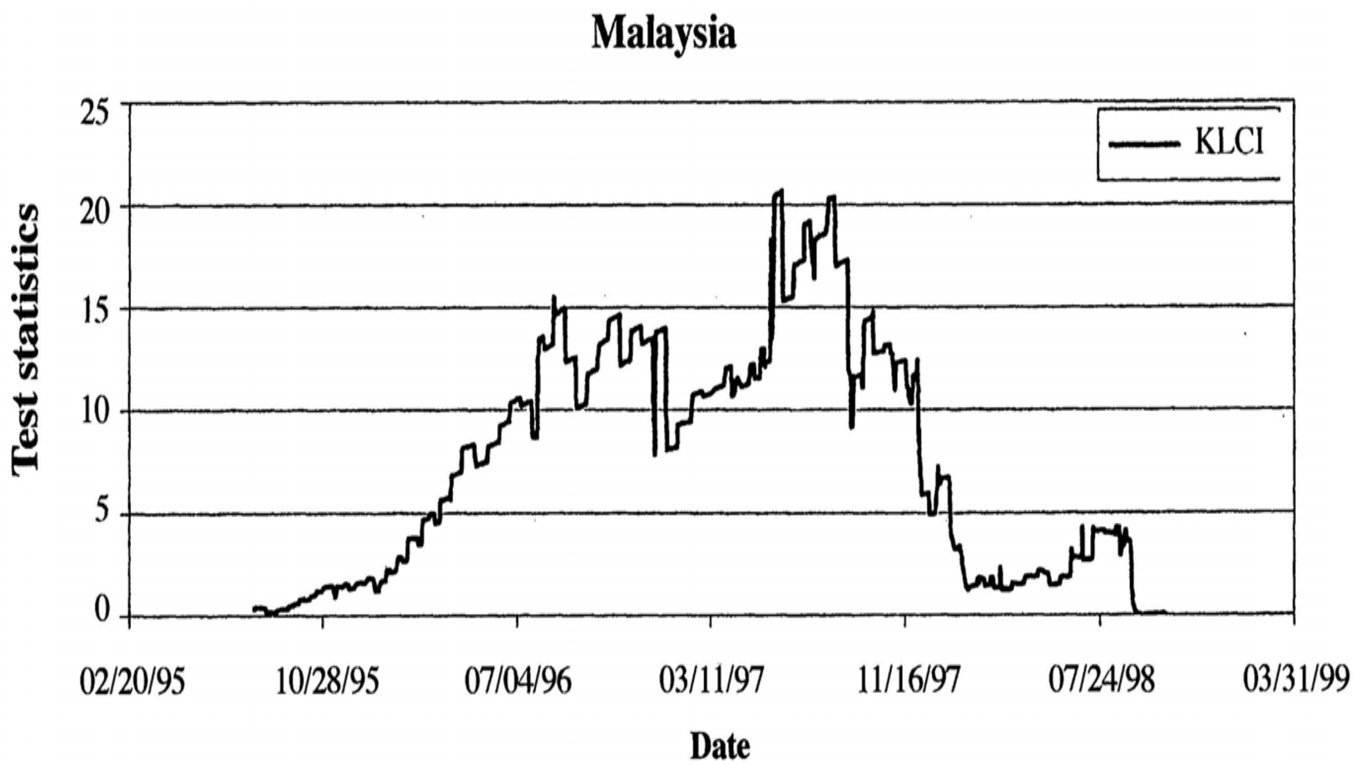

Malaysia
KLCI
Test statistics
25
20
15
10
5
0
02/20/95
10/28/95
07/04/96
03/11/97
11/16/97
07/24/98
03/31/99
Date

FIGURE 4

Recursive test (95% critical value = 1·78)

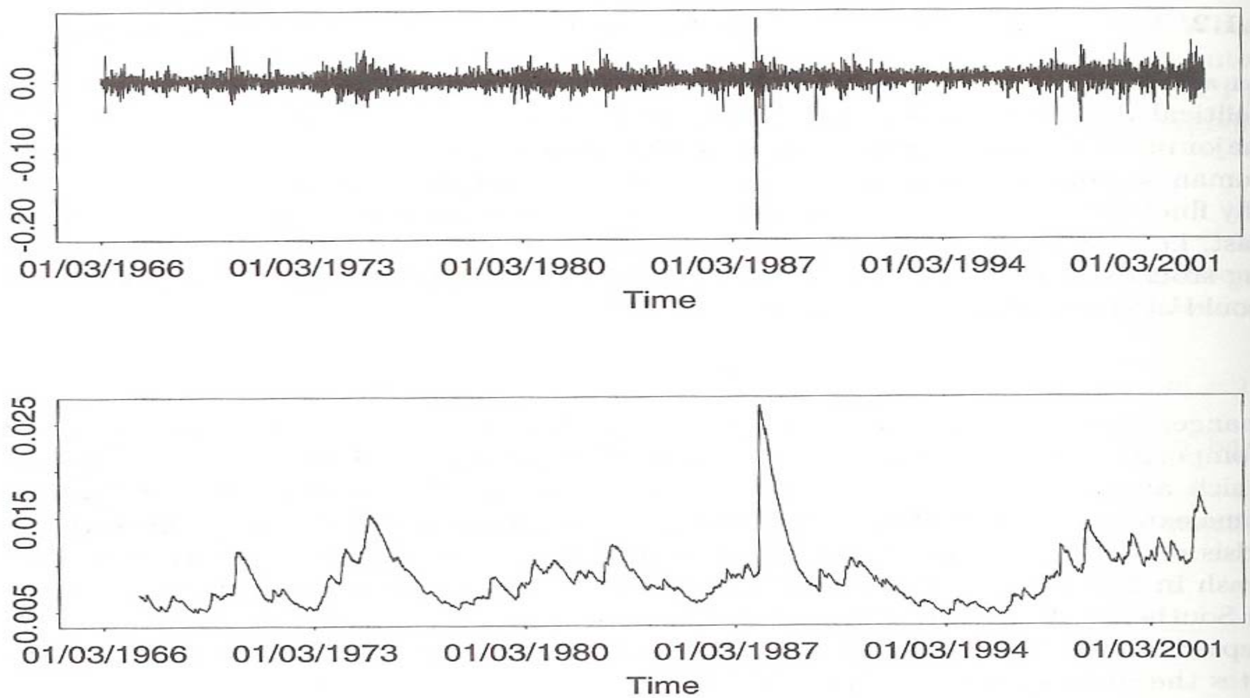


Fig. 3.4. NYSE Composite log–returns (top) and estimated volatility (bottom). For EWMA, the smoothing parameter $\lambda = 0.99$ was used. The first 300 values of $\hat{\sigma}_t$ have been discarded.

Source: Straumann, D. (2004). *Estimation in Conditionally Heteroscedastic Time Series Models.* Springer. EWMA: $\hat{\sigma}_t^2 = (1-\lambda)\, X_t^2 + \lambda \hat{\sigma}_{t-1}^2$.

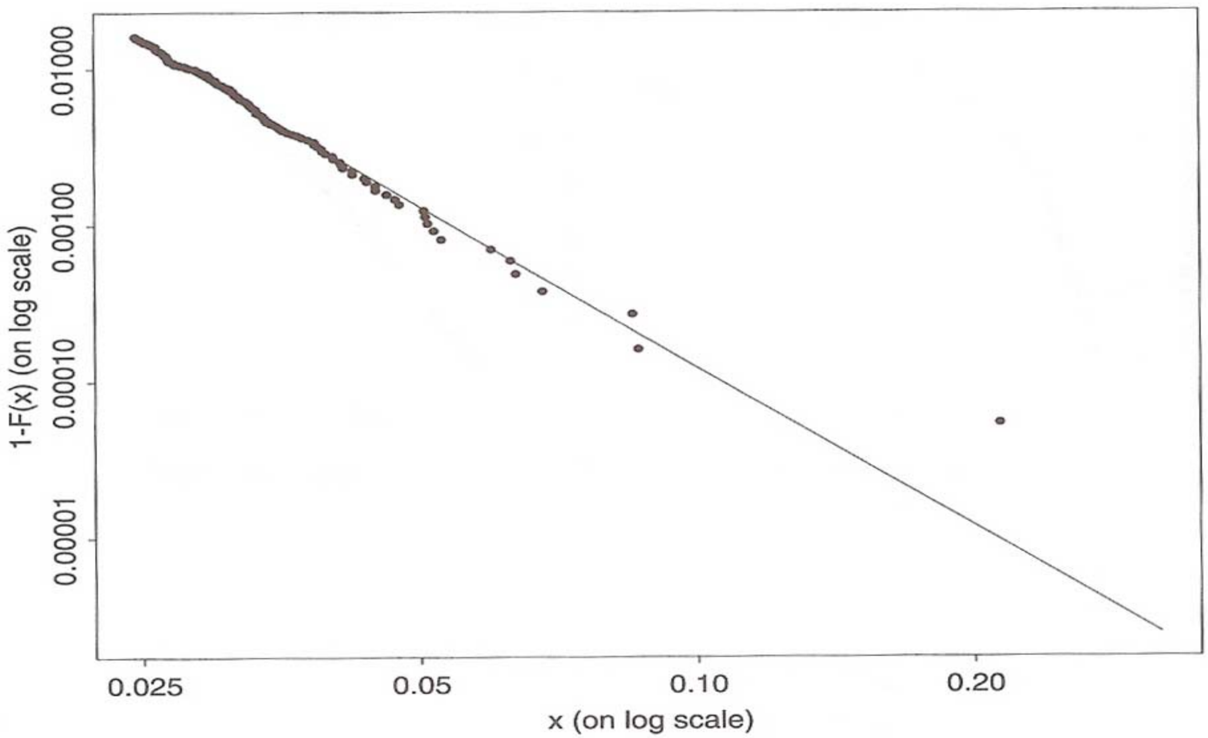


Fig. 3.6. Tail of the empirical distribution function of absolute NYSE Composite log–returns, evaluated at the 150 largest values. The solid line is a plot of the tail of the GPD (Generalized Pareto Distribution), $\bar{F}(x) = (1 + (\alpha\beta)^{-1}(x - \mu))^{-\alpha}$, with parameters $\alpha = 3.31$, $\mu = 0.00861$, $\beta = 0.00194$. The fit was obtained by application of the Splus function gpd of the EVIS Software package by McNeil [94].

Source: Straumann, D. (2004).

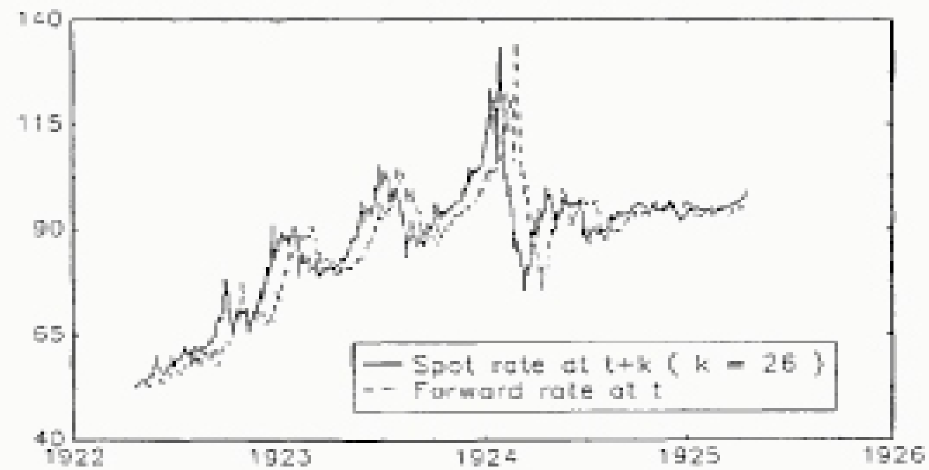


Figure 1. Belgium, FR and SR: 1922–5

Historical Daily Exchange Rate Data 1922-1925

Source: McFarland, J. W., P. C. McMahon and P. C. B. Phillips (1996). *J. Applied Econometrics*, 11, 1-23.

P. C. B. PHILLIPS, J. W. McFARLAND AND P. C. McMAHON

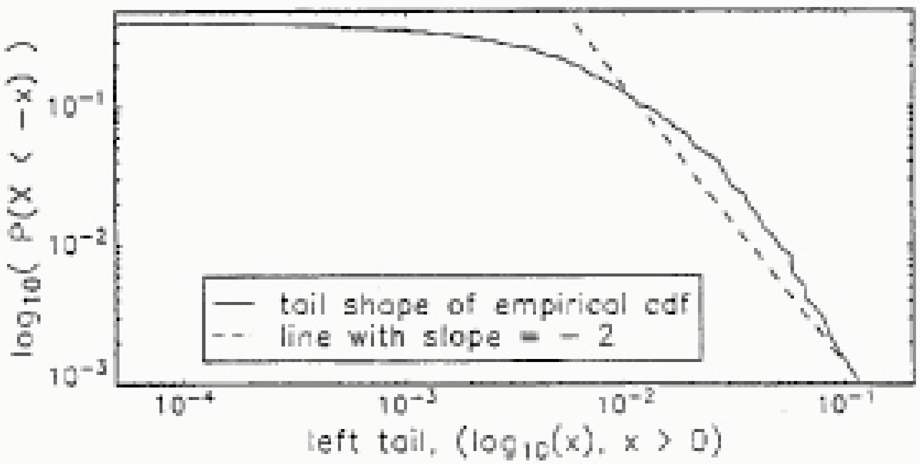


Figure 5. Empirical cdf: Belgium, forward exchange rate

Empirical cdf & Tail Slope

Table I. Point estimates of tail slope parameters

Country	s	Forward rate						Spot rate						Spread		Estimated AR order PIC/BIC
		Left tail		right tail		Two-tail		Left tail		right tail		Two-tail		$s_{t+1}-f_{t,1}$		
Belgium	15	2·803	(0·723)	3·095	(0·799)	3·106	(0·825)	2·814	(0·729)	3·121	(0·805)	3·198	(0·825)	3·383	(0·873)	
	25	2·277	(0·455)	2·171	(0·543)	3·090	(0·618)	2·294	(0·458)	2·716	(0·543)	3·012	(0·602)	3·511	(0·702)	
	50	1·857	(0·262)	2·186	(0·309)	2·570	(0·363)	1·879	(0·265)	2·179	(0·308)	2·568	(0·363)	2·927	(0·413)	4
	75	1·707	(0·197)	1·889	(0·218)	2·429	(0·281)	1·713	(0·197)	1·898	(0·219)	2·451	(0·283)	2·769	(0·319)	
	ŝ	2·092[41]	(0·326)	2·150[49]	(0·307)	2·473[71]	(0·293)	2·176[43]	(0·332)	2·316[48]	(0·334)	2·442[72]	(0·287)	3·048[66]	(0·375)	
France	15	2·420	(0·625)	4·057	(1·047)	5·050	(1·304)	2·625	(0·677)	4·309	(1·113)	3·852	(0·994)	4·721	(1·219)	
	25	2·859	(0·571)	2·926	(0·585)	3·298	(0·659)	2·877	(0·575)	3·089	(0·617)	3·460	(0·691)	4·628	(0·925)	
	50	1·985	(0·280)	2·306	(0·326)	2·935	(0·415)	2·222	(0·314)	2·362	(0·334)	2·948	(0·417)	3·633	(0·513)	4
	75	1·766	(0·203)	2·002	(0·231)	2·464	(0·284)	1·878	(0·216)	2·011	(0·232)	2·500	(0·288)	3·231	(0·373)	
	ŝ	2·445[42]	(0·377)	2·392[58]	(0·314)	2·629[71]	(0·312)	2·238[44]	(0·337)	2·269[58]	(0·297)	2·493[72]	(0·293)	3·232[66]	(0·397)	
Italy	15	3·358	(0·867)	3·576	(0·923)	3·390	(0·875)	3·374	(0·871)	3·641	(0·940)	3·265	(0·843)	3·059	(0·789)	
	25	2·822	(0·564)	3·393	(0·678)	3·308	(0·661)	2·856	(0·571)	3·271	(0·654)	3·132	(0·626)	2·969	(0·593)	
	50	2·118	(0·299)	3·364	(0·475)	3·274	(0·463)	2·077	(0·293)	3·420	(0·483)	3·208	(0·453)	3·329	(0·470)	3
	75	2·151	(0·248)	2·660	(0·307)	2·900	(0·334)	2·099	(0·242)	2·613	(0·301)	2·969	(0·342)	3·216	(0·371)	
	ŝ	2·701[39]	(0·433)	3·346[44]	(0·504)	2·890[68]	(0·350)	2·747[40]	(0·434)	3·400[44]	(0·512)	2·836[66]	(0·349)	3·546[66]	(0·436)	
USA	15	3·442	(0·887)	2·917	(0·753)	2·833	(0·736)	3·849	(0·993)	2·883	(0·744)	2·844	(0·734)	5·504	(1·421)	
	25	3·279	(0·655)	2·544	(0·508)	3·153	(0·630)	3·091	(0·618)	2·582	(0·516)	3·289	(0·657)	4·183	(0·836)	
	50	3·173	(0·448)	2·349	(0·332)	2·791	(0·394)	3·017	(0·426)	2·322	(0·328)	2·767	(0·391)	3·016	(0·300)	1
	75	2·331	(0·269)	2·101	(0·242)	2·664	(0·307)	2·370	(0·273)	2·023	(0·233)	2·550	(0·294)	2·602	(0·300)	
	ŝ	3·115[49]	(0·445)	2·207[55]	(0·297)	2·649[78]	(0·299)	2·942[46]	(0·434)	2·144[53]	(0·294)	2·634[79]	(0·296)	2·634[78]	(0·298)	

[ ] = ŝ = adaptive estimate of order statistic truncation number; ( ) = standard error of $\hat{\alpha}_r$.

Tail Slope Estimates for Exchange Rate Data

Nonstationarity + Joint Dependence in Panels

- How do we model nonstationarity and trend?
- Common convention (and convenience) of log regression on a linear trend
  - measures average growth rate
  - but no causal mechanism
  - need to penalize fit
- In panel data
  - often a multiplicity/richness of individual outcomes
  - but some sense of common factor
- Suggests some mechanism of co-dependence + common engine of growth?
  - cumulative sum - random wandering features are common
  - dynamic factor & nonlinear factor modeling

Examples

A: World income over 1950-2000 data sets:

Penn World Table data (http://pwt.econ.upenn.edu/)

OECD world Economic data

(http://www.theworldeconomy.org/publications/worldeconomy/statistics.htm)

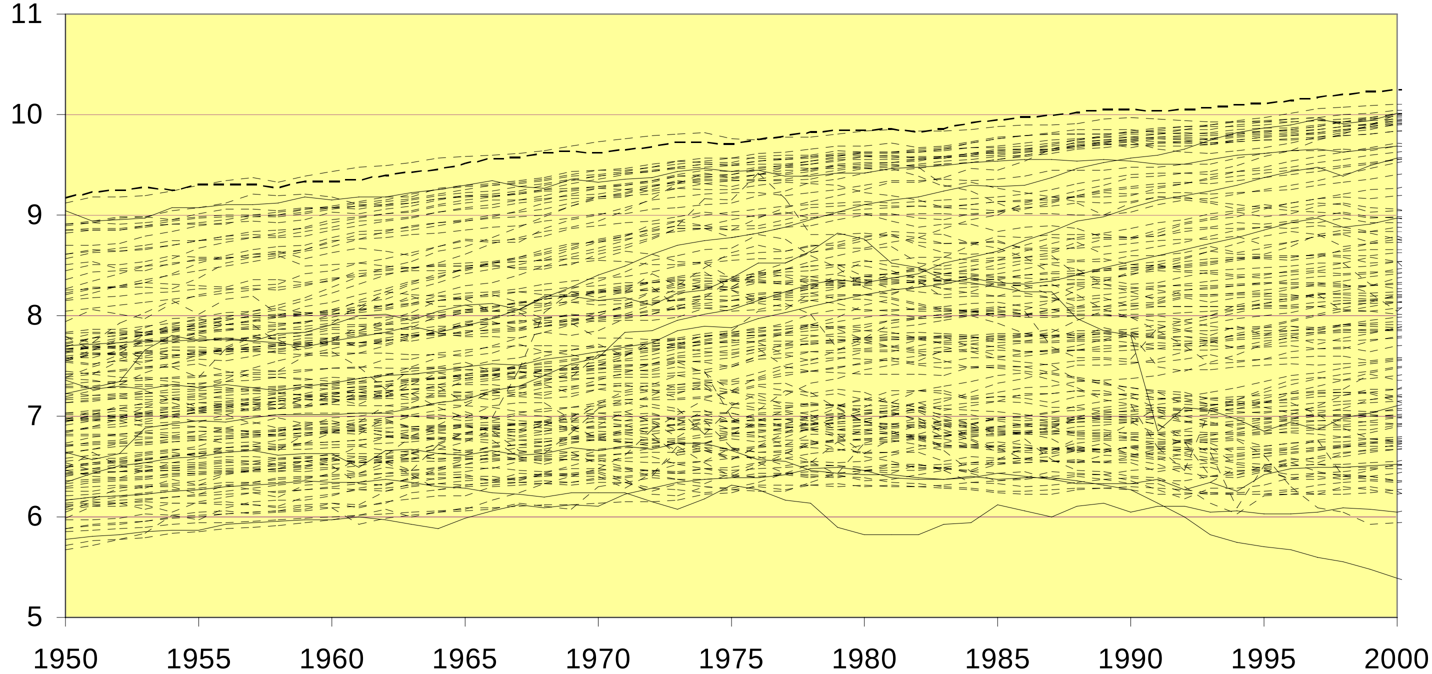
References

a. Barro, R. J. (1997), *Determinants of Economic Growth.* Cambridge Press.

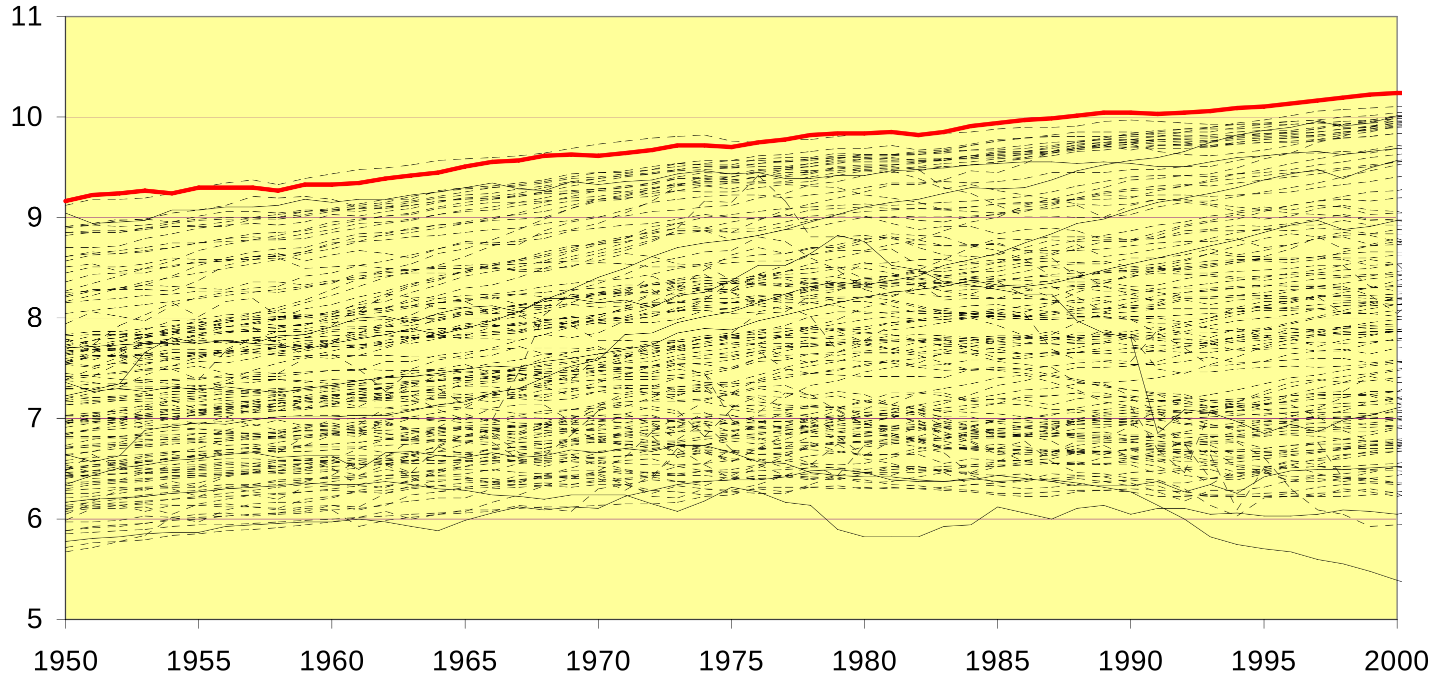
b. Barro, R. J. & X. Sala-i-Martin (1992). *J. Political Economy,* 100, 223-251.

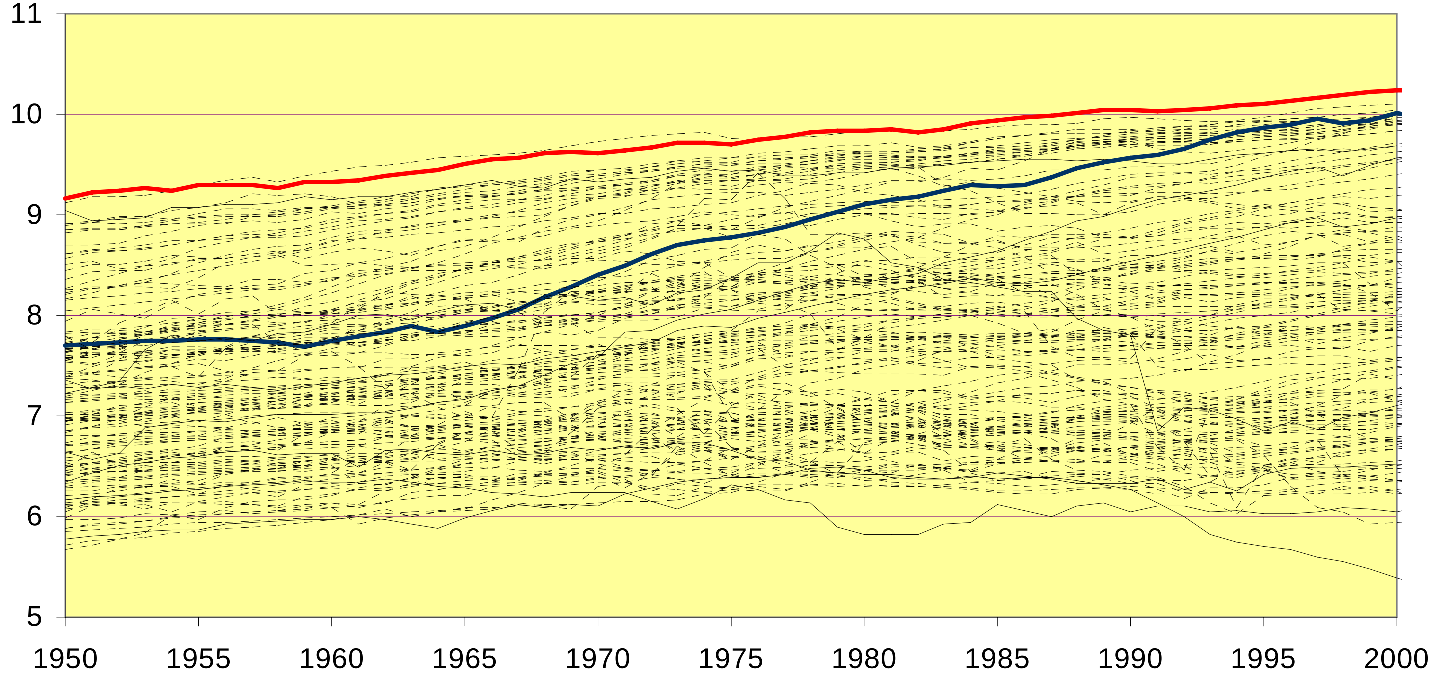
c. Barro, R. J. & X. Sala-i-Martin (1995). *Economic Growth.* McGraw Hill

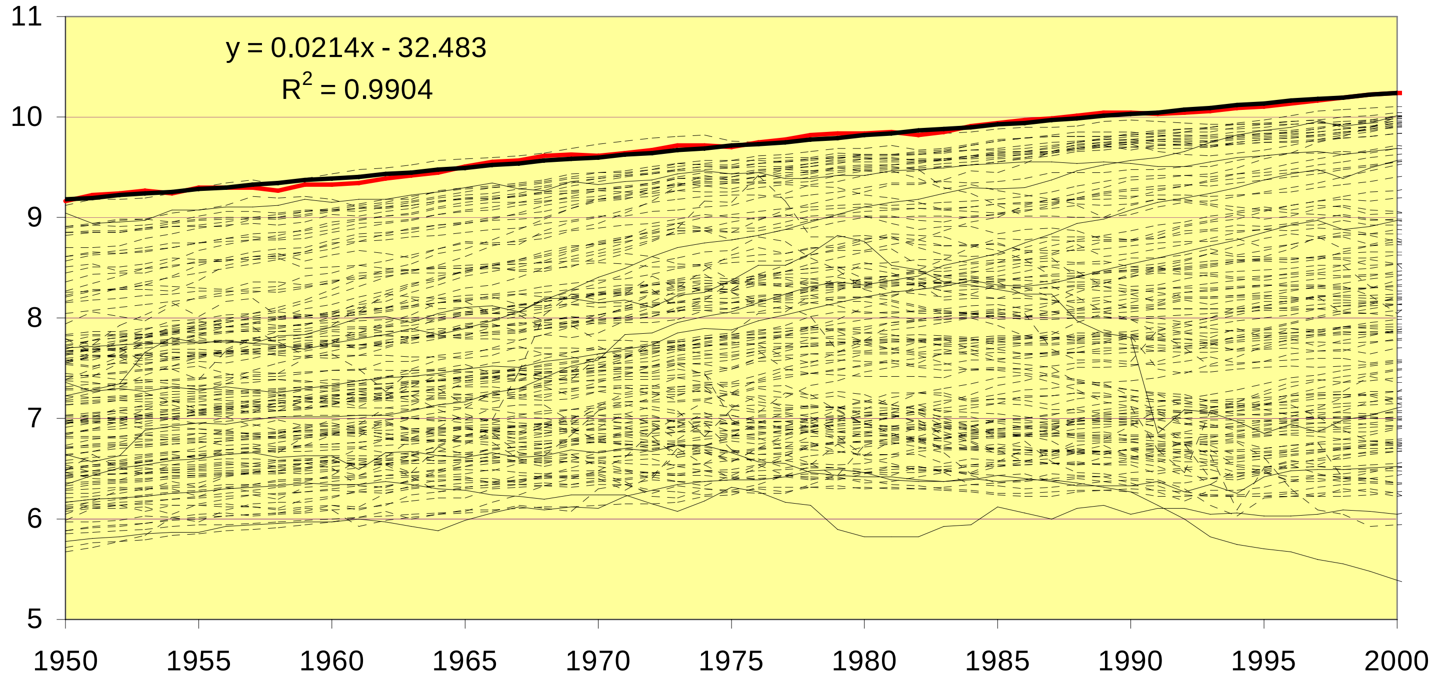
d. Phillips, P. C. B. & D. Sul (2004). Transition & Economic Growth, Cowles Discussion Paper, Yale.



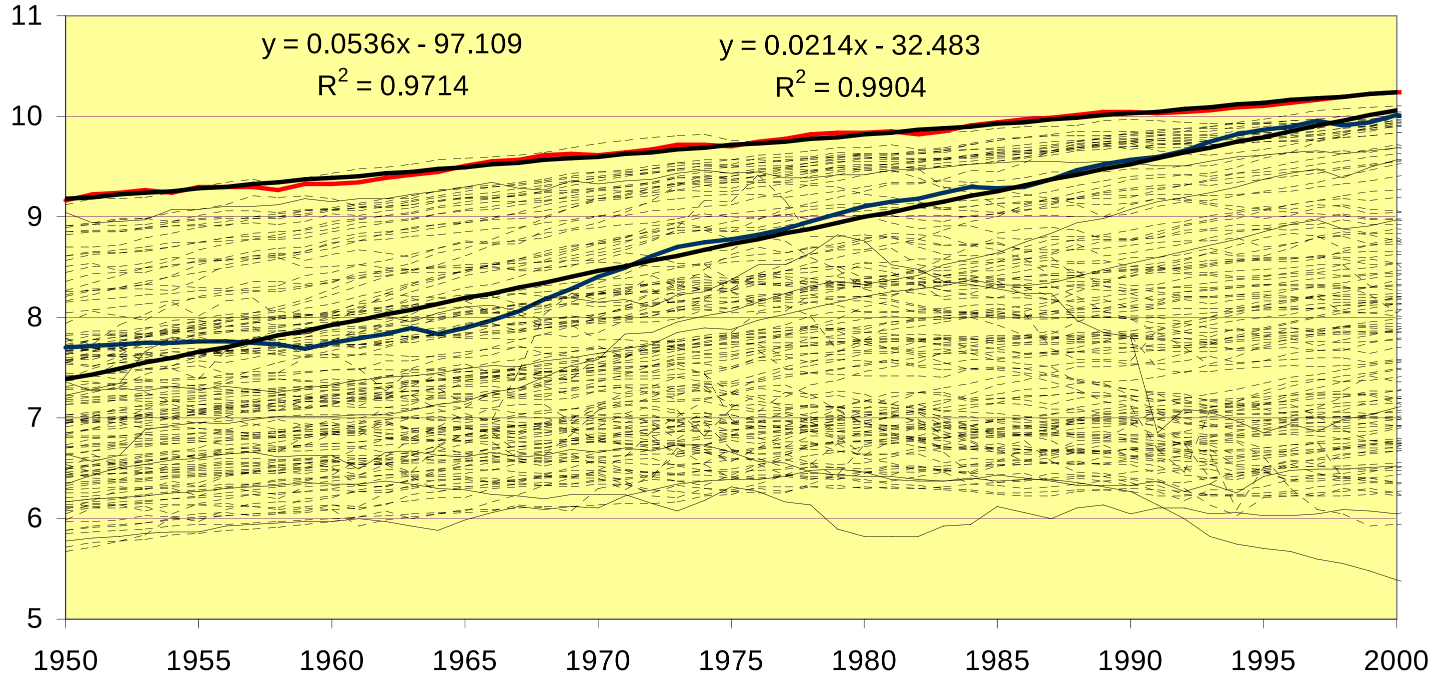
11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000


11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

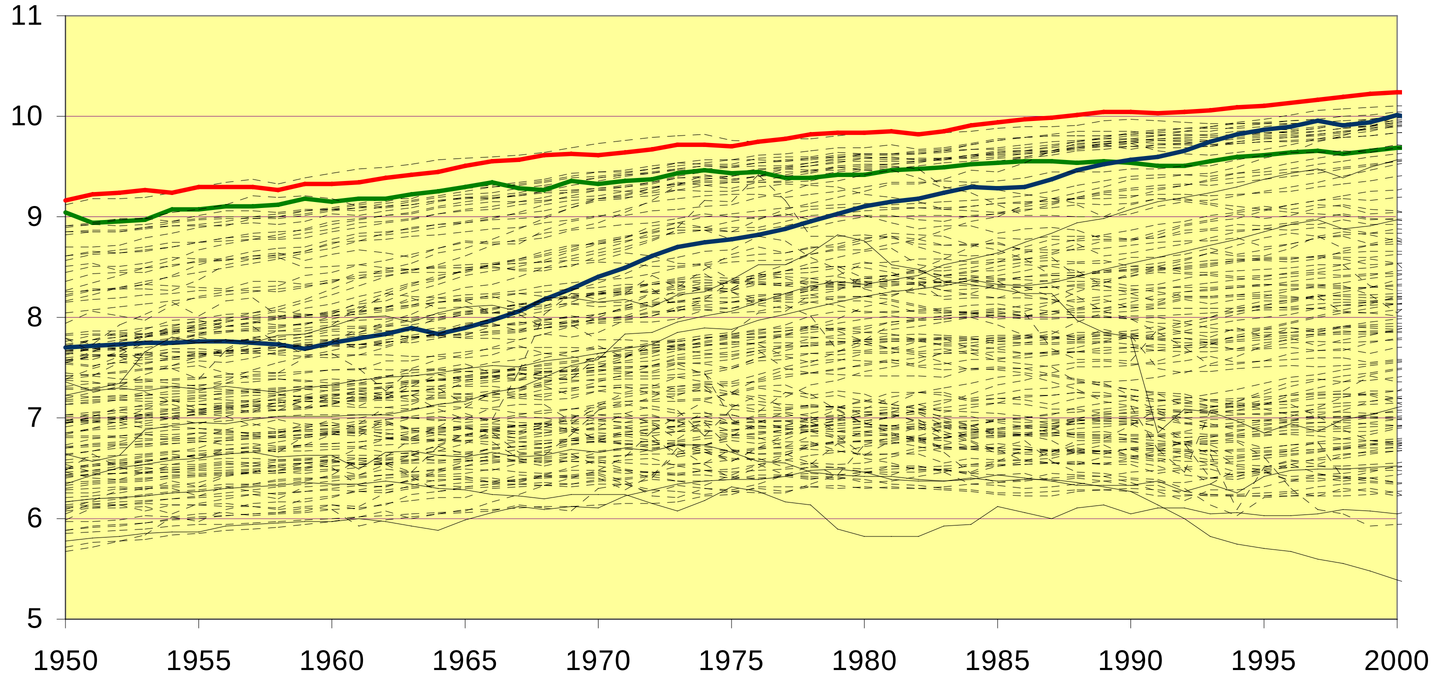

11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

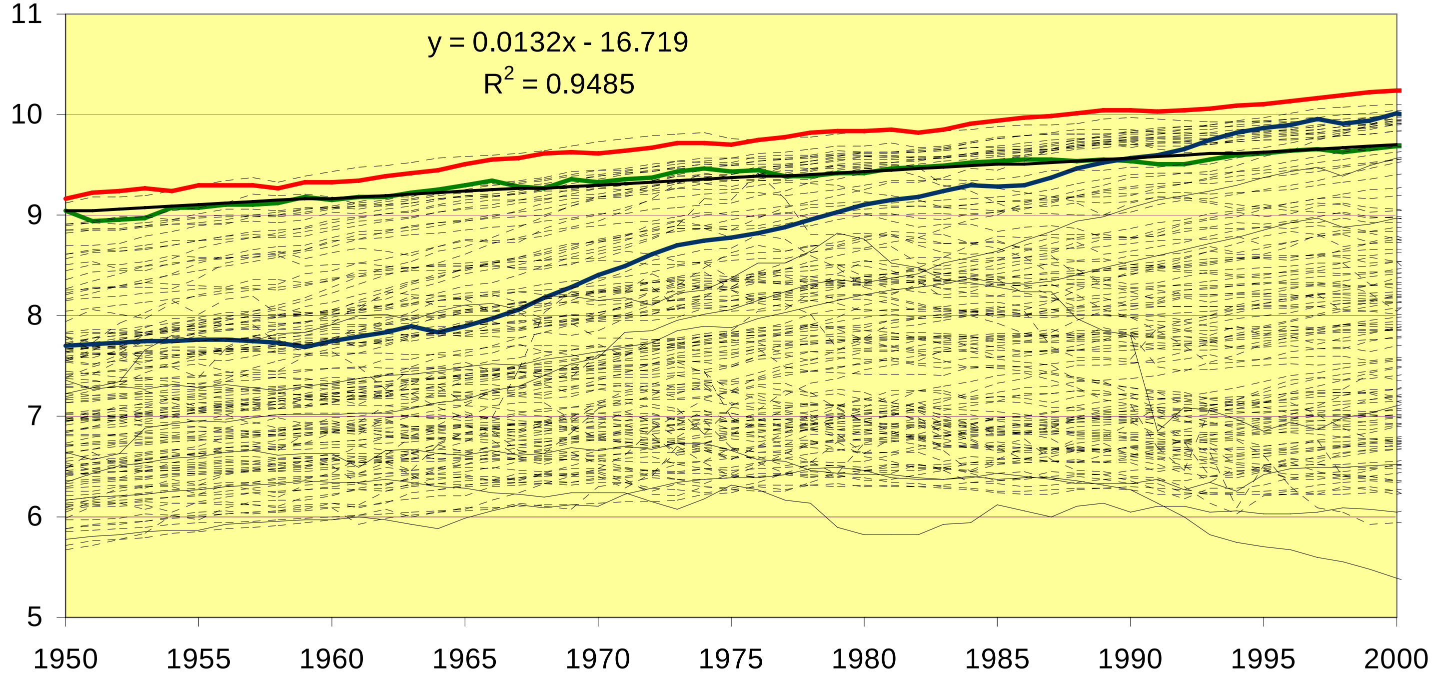


US Trend Growth

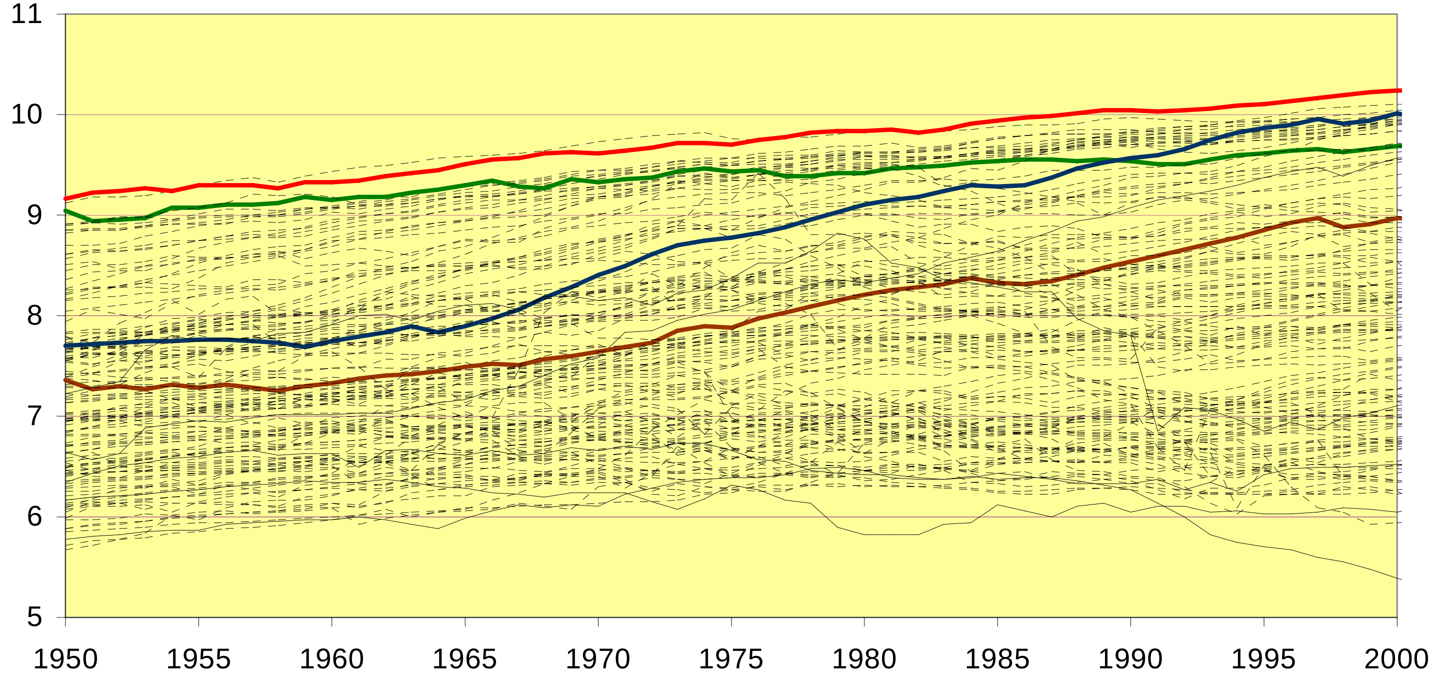


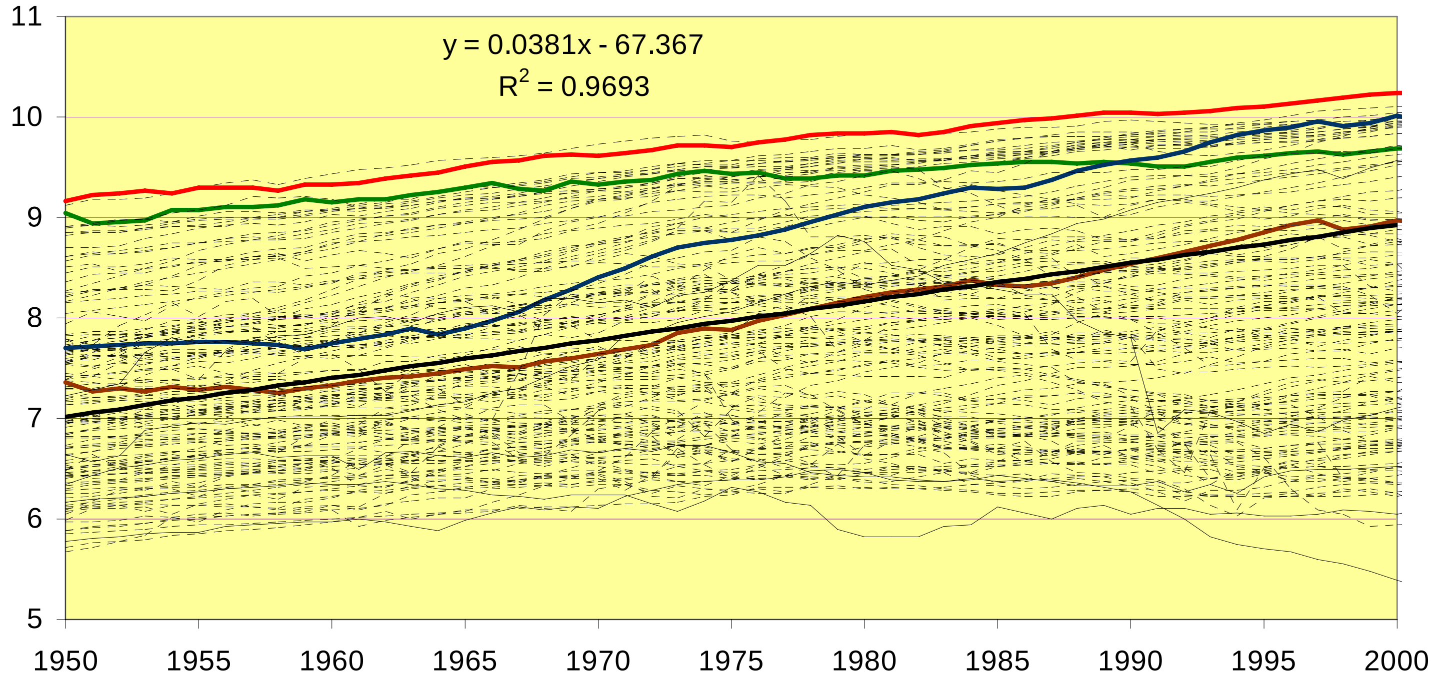
US & Singapore Trend Growth


11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

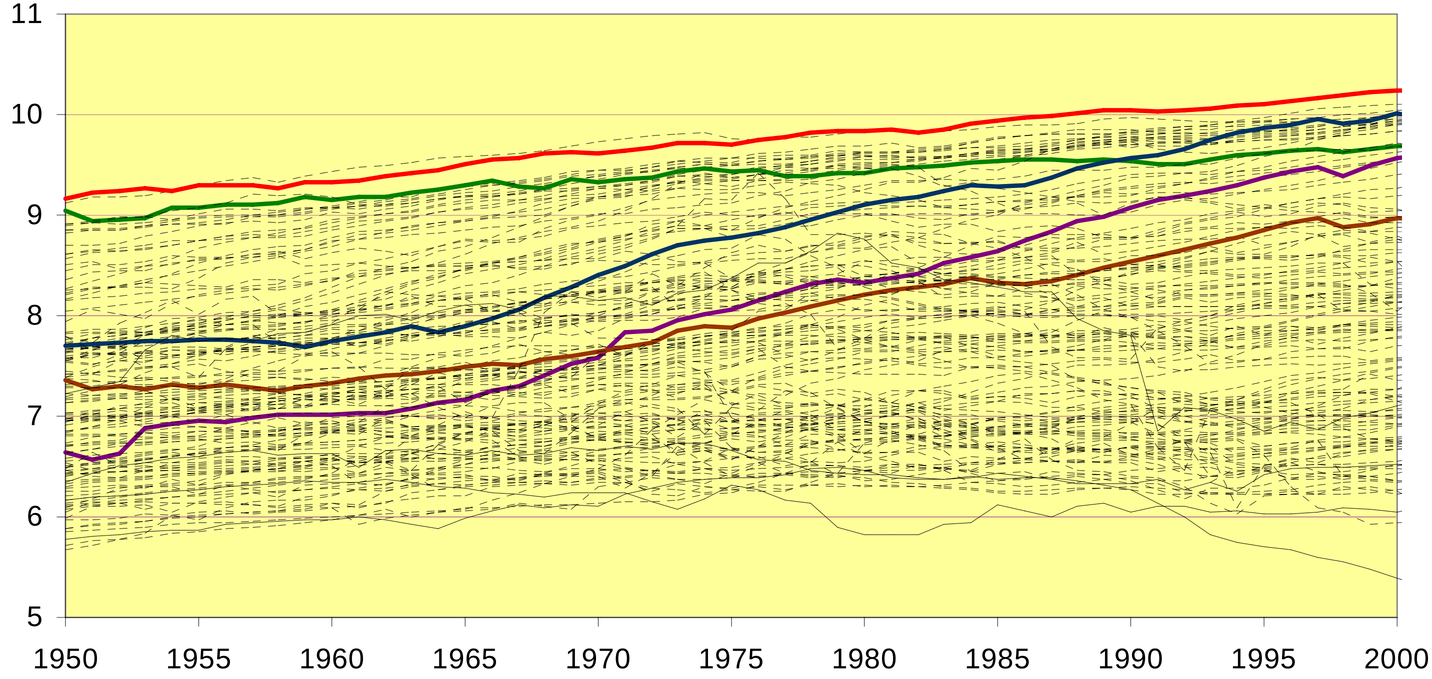


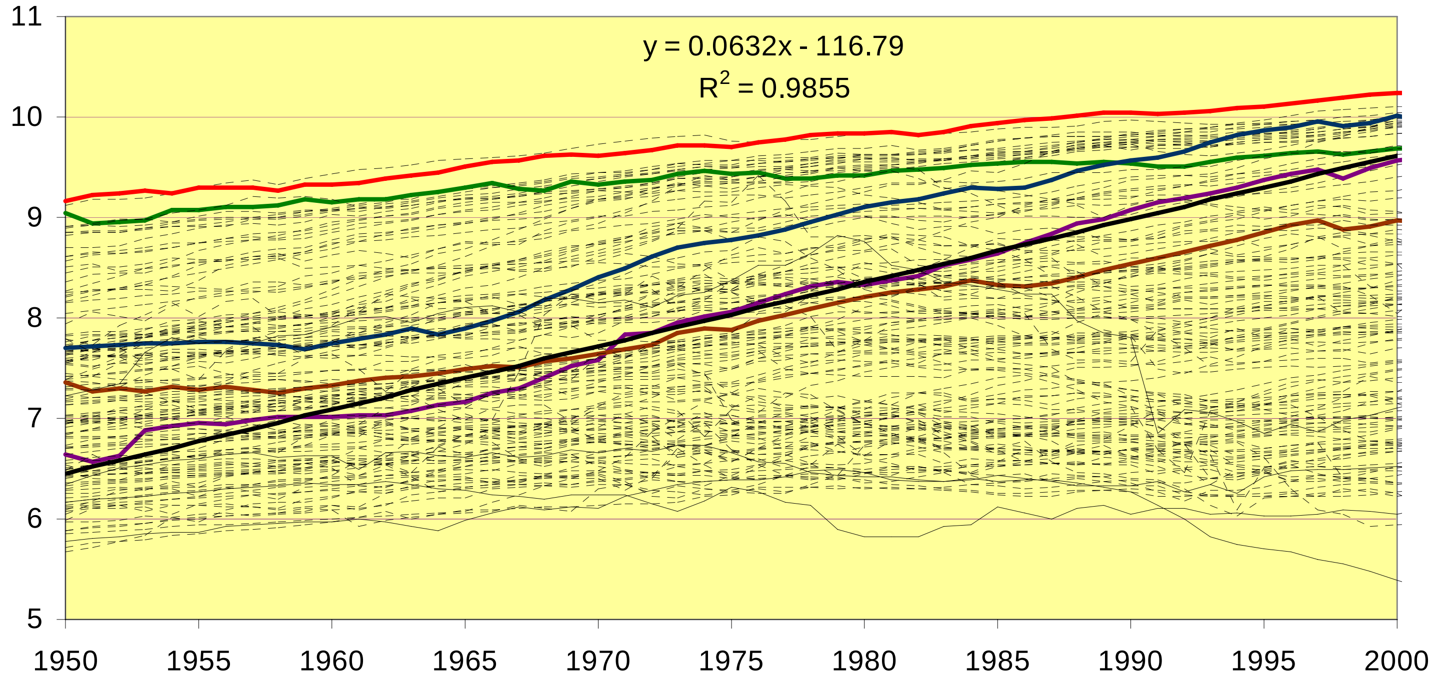
New Zealand Trend Growth


11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

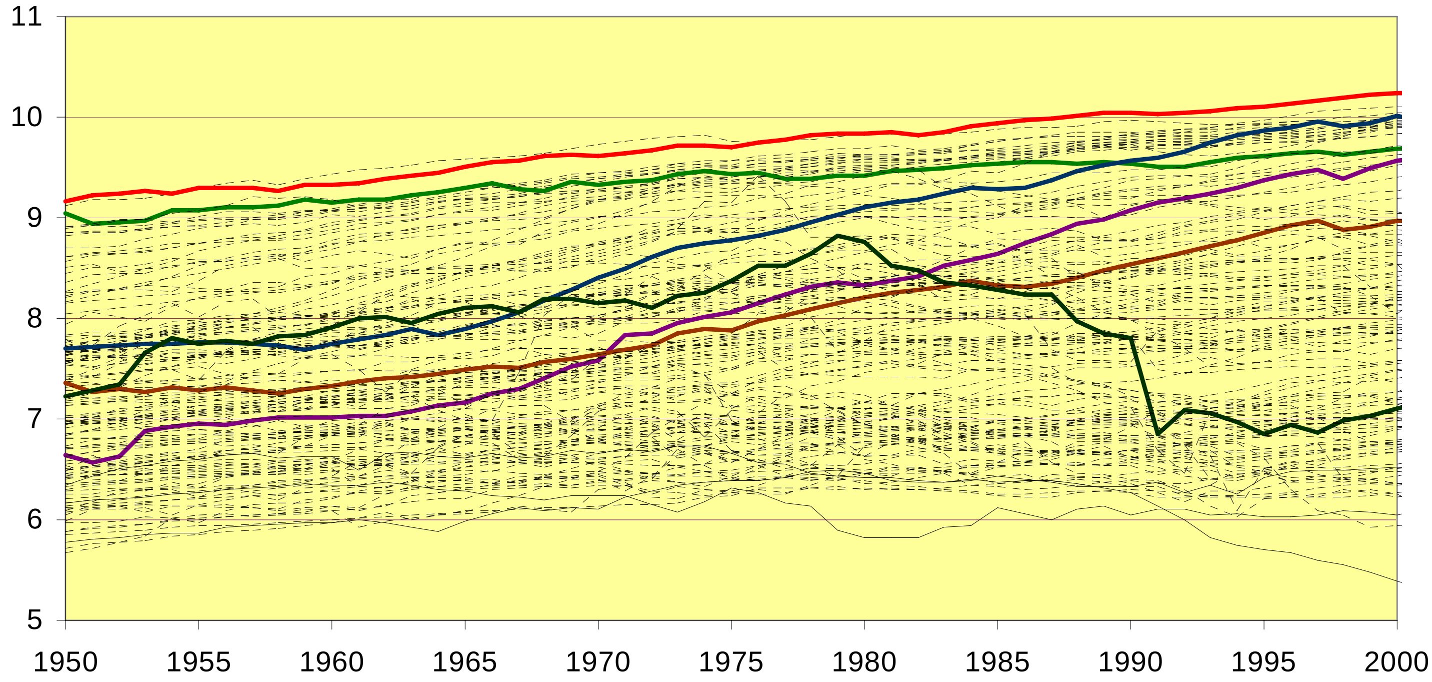


Malaysia Trend Growth

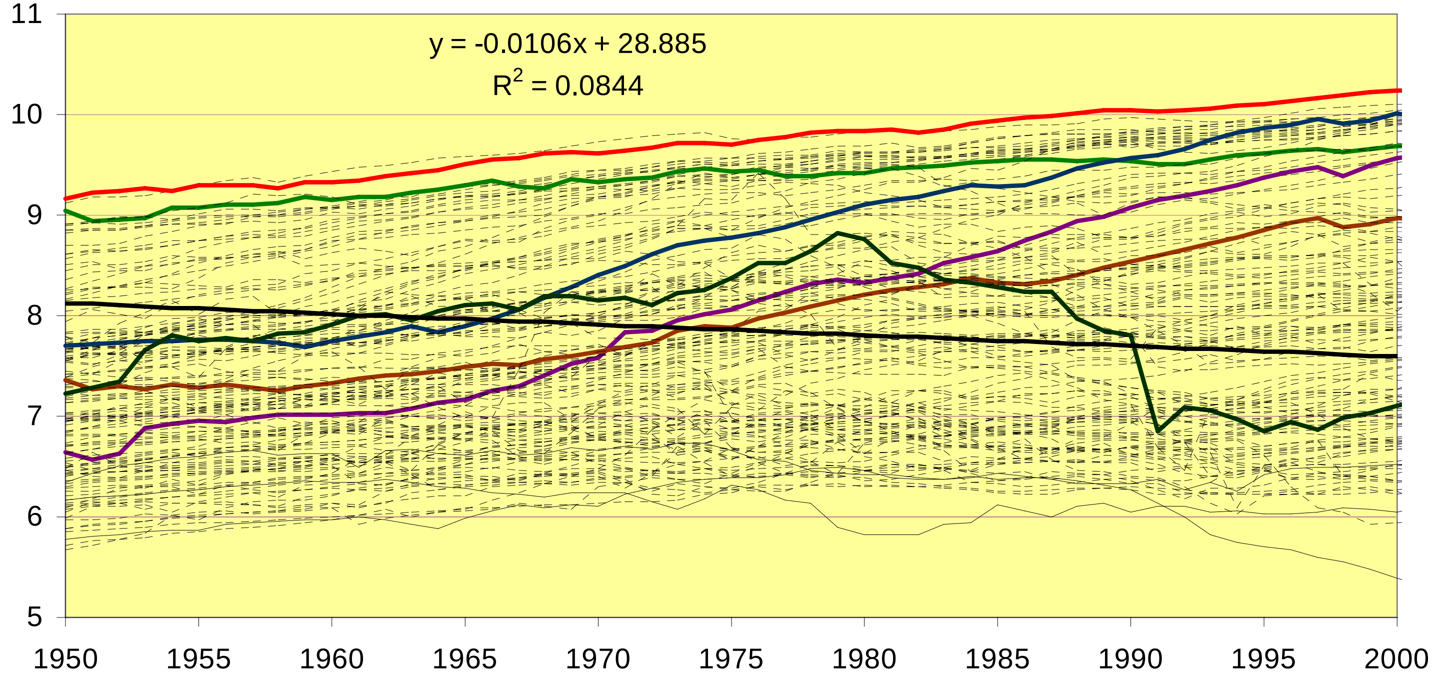

11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000



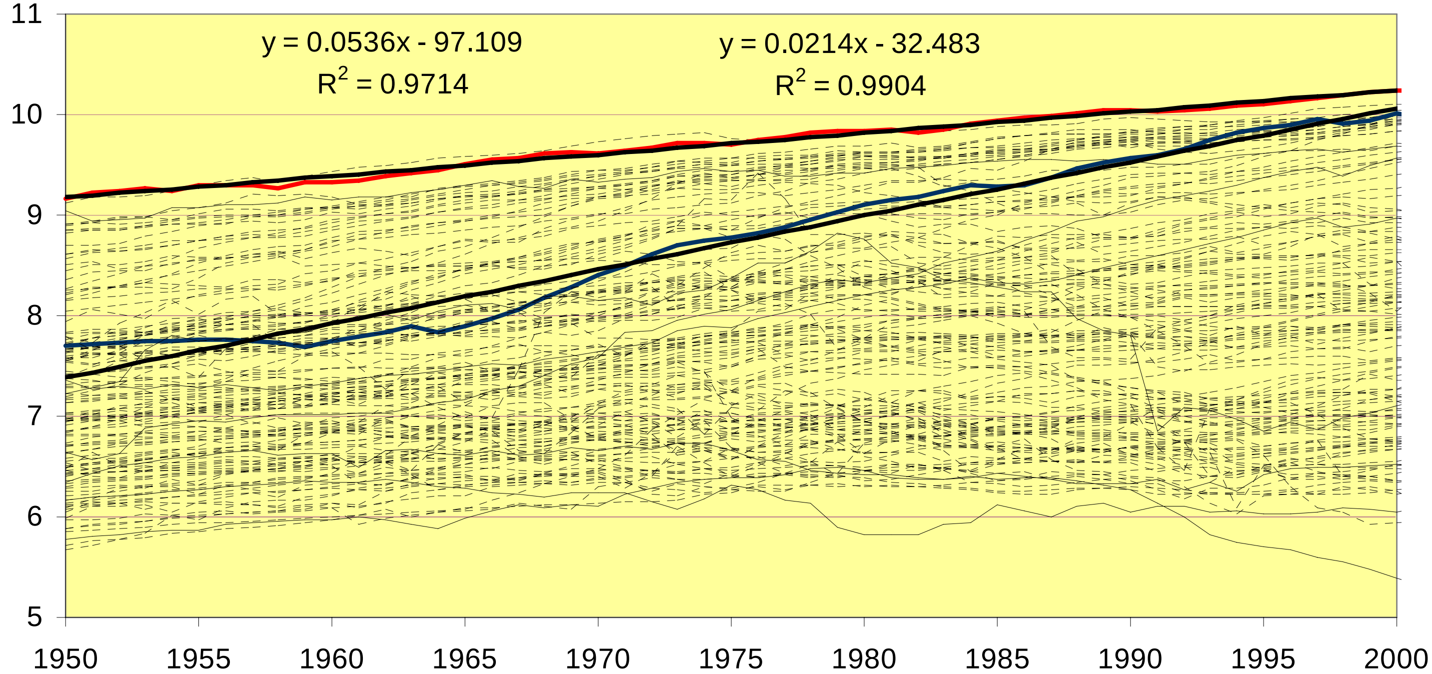
South Korea Trend Growth



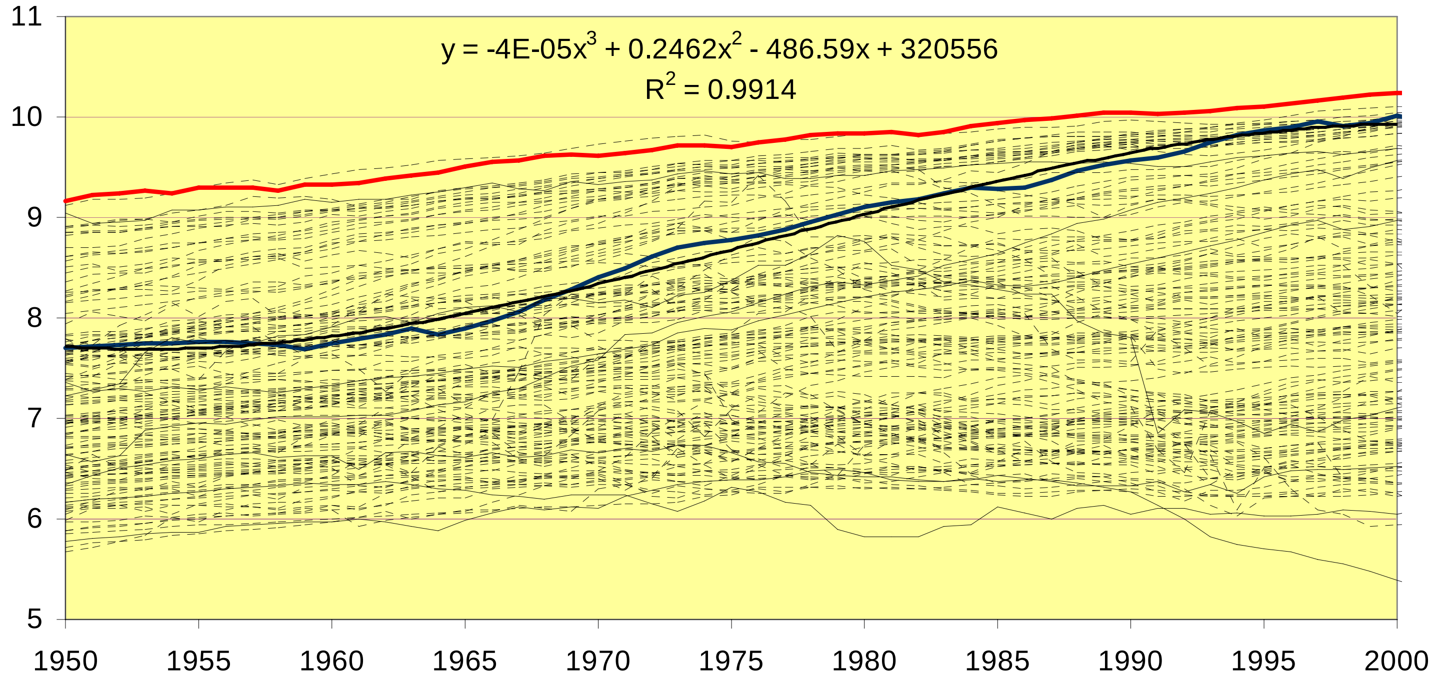
11
10
9
8
7
6
5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000



Iraq Trend Growth



How Adequate is a Linear Trend in Modeling Growth?



Polynomial Trend Growth for Singapore - high on fit, low on realism.

Need to Penalize Fit

B: Paleobiodiversity - History of Life - Example

Diversity, Origination, Extinction over 550 Million Years

- Marine fossil records - record new species (originations) & extinctions
- Total genera (G_i) appearing at some time during $[t_i, t_{i+1}]$ in relation to number of genera that first appeared (O_i) and number of genera that last appeared (E_i)

$$G_{i+1} = G_i - E_i + O_{i+1}$$

leading to

$$G_n = G_1 + \sum_{i=2}^{n} O_i - \sum_{i=1}^{n-1} E_i$$

which has cumulative sum - random wandering features.

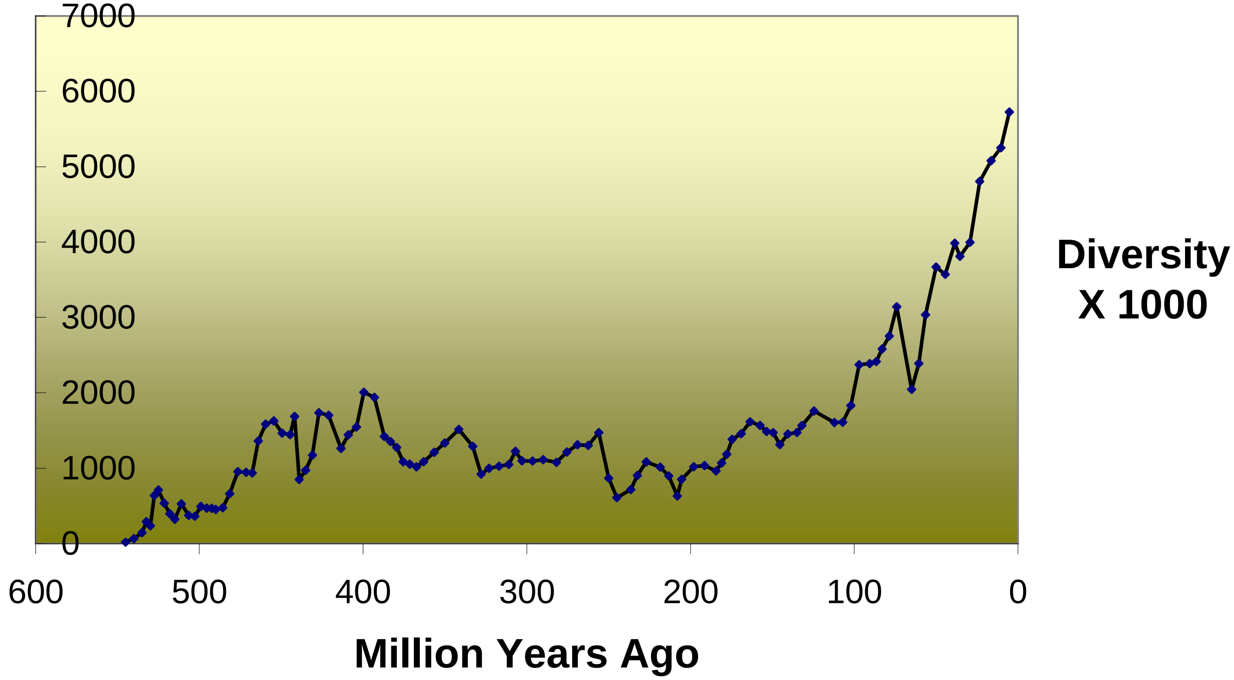
a. Sepkoski, J. J. (1997), *J. Paleontology*, 71, 533-539.

b. Cornette, J. L. and B. S. Lieberman (2004), *Proc. Nat. Acad. Sci.* 101, 187-191.

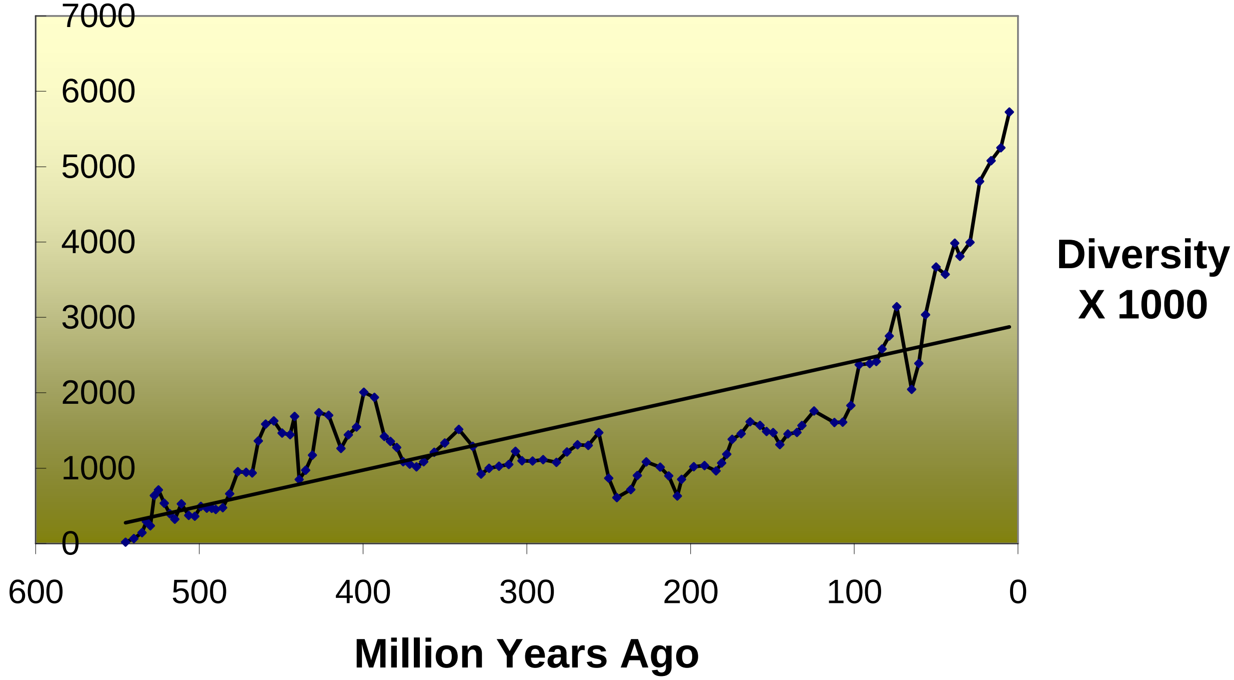
Phanerozoic Eon (543 mya to present)	Cenozoic Era (65 mya to today)	Quaternary (1.8 mya to today) Holocene (10,000 years to today) Pleistocene (1.8 mya to 10,000 yrs) Tertiary (65 to 1.8 mya) Pliocene (5.3 to 1.8 mya) Miocene (23.8 to 5.3 mya) Oligocene (33.7 to 23.8 mya) Eocene (54.8 to 33.7 mya) Paleocene (65 to 54.8 mya)
	Mesozoic Era (248 to 65 mya)	Cretaceous (144 to 65 mya) Jurassic (206 to 144 mya) Triassic (248 to 206 mya)
	Paleozoic Era (543 to 248 mya)	Permian (290 to 248 mya) Carboniferous (354 to 290 mya) Pennsylvanian (323 to 290 mya) Mississippian (354 to 323 mya) Devonian (417 to 354 mya) Silurian (443 to 417 mya) Ordovician (490 to 443 mya) Cambrian (543 to 490 mya) Tommotian (530 to 527 mya)
Precambrian Time (4,500 to 543 mya)	Proterozoic Era (2500 to 543 mya)	Neoproterozoic (900 to 543 mya) Vendian (650 to 543 mya) Mesoproterozoic (1600 to 900 mya) Paleoproterozoic (2500 to 1600 mya)
	Archaean (3800 to 2500 mya)	
	Hadean (4500 to 3800 mya)	

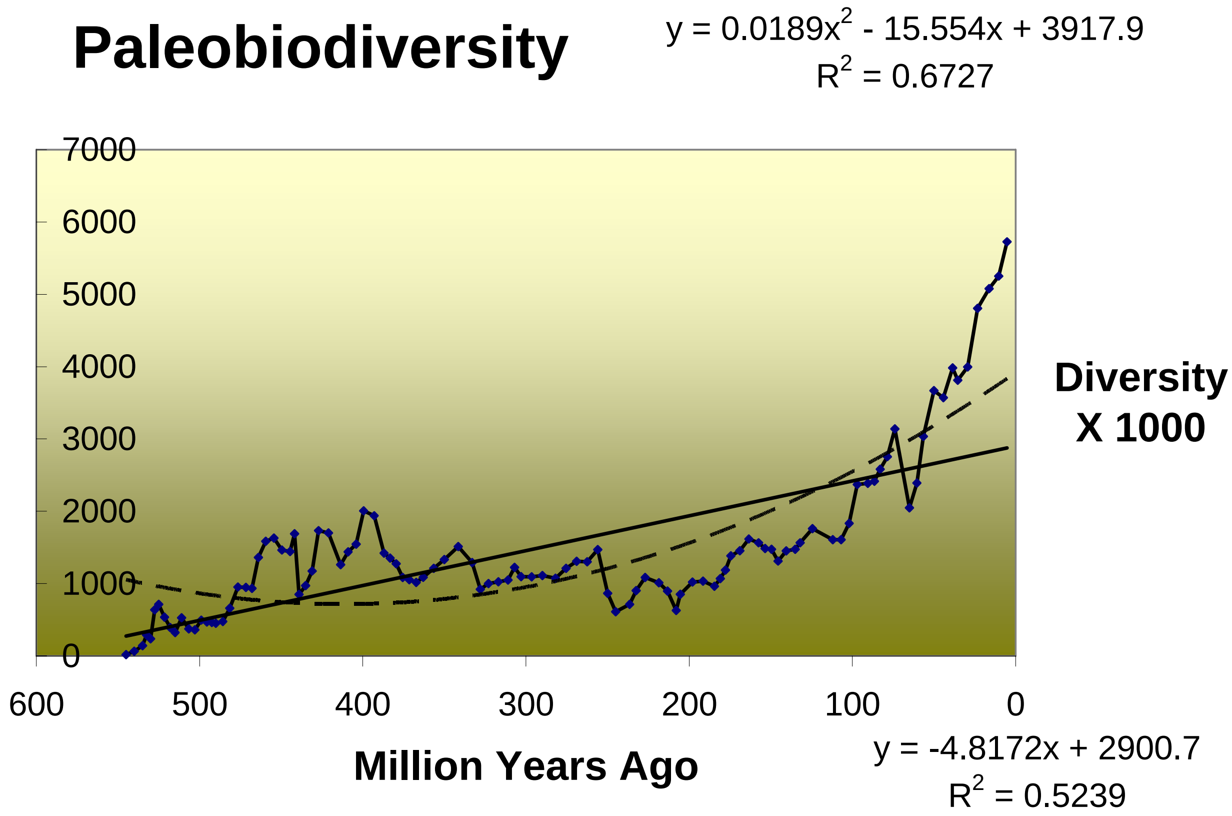
Geological Chronology

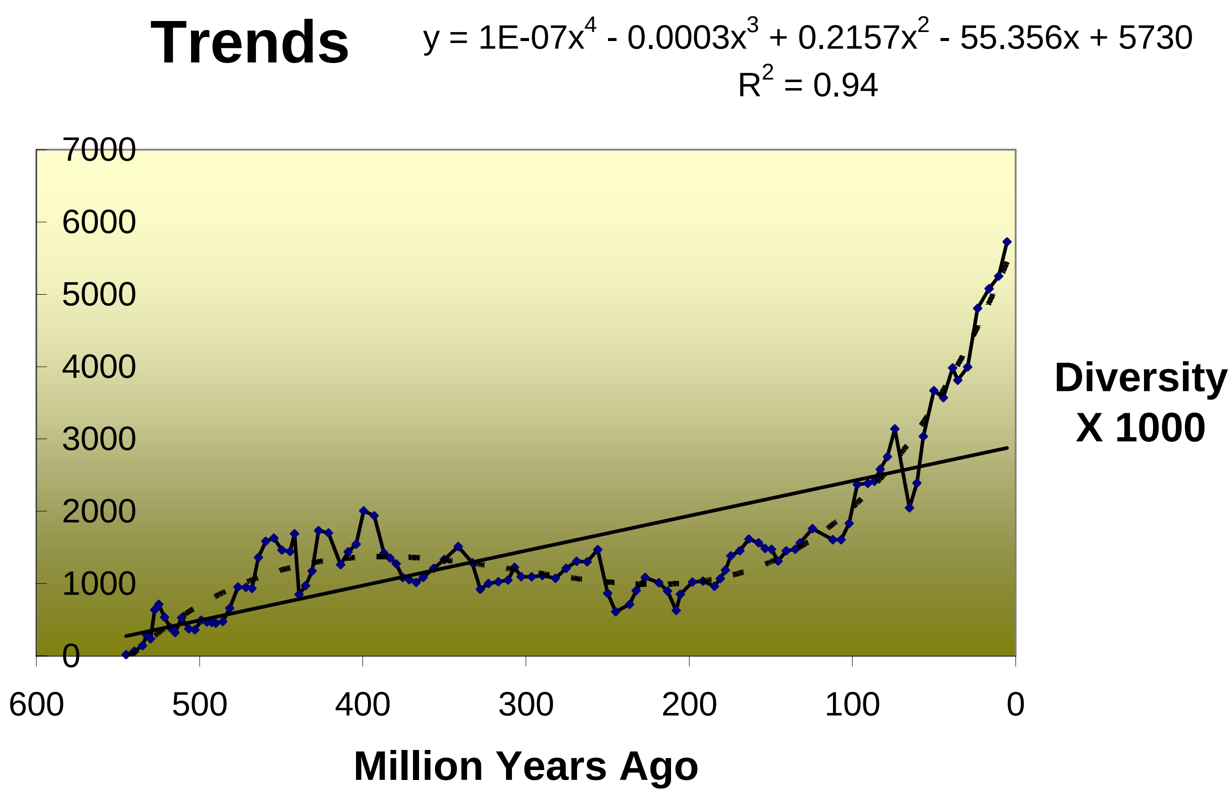
Paleobiodiversity



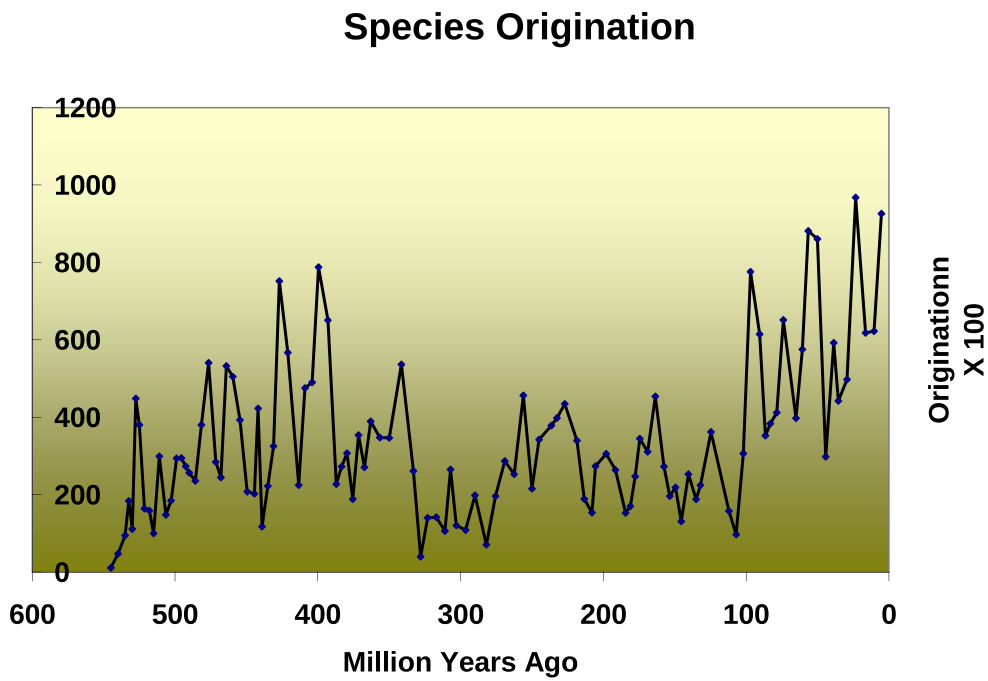
Paleobiodiversity + Linear Trend



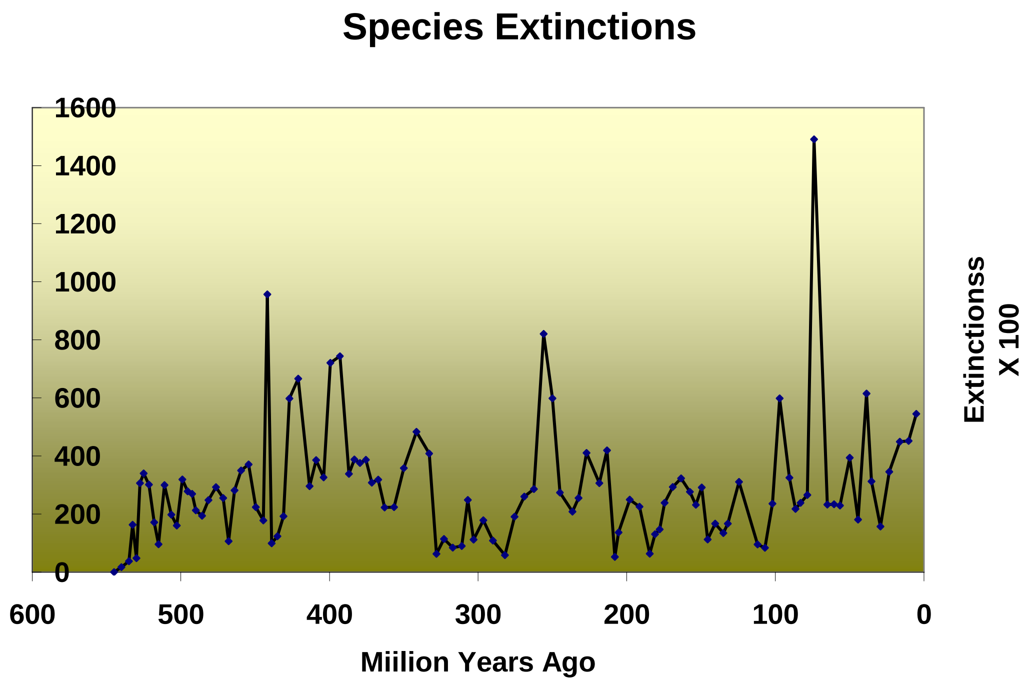

Paleobiodiversity
y = 0.0189x2 - 15.554x + 3917.9
R2 = 0.6727
7000
6000
5000
4000
3000
2000
1000
0
600
500
400
300
200
100
0
Diversity X 1000
Million Years Ago
y = -4.8172x + 2900.7
R2 = 0.5239



Trends
y = 1E-07x4 - 0.0003x3 + 0.2157x2 - 55.356x + 5730
R2 = 0.94
7000
6000
5000
4000
3000
2000
1000
0
600
500
400
300
200
100
0
Million Years Ago
Diversity X 1000



Species Origination
1200
1000
800
600
400
200
0
600
500
400
300
200
100
0
Million Years Ago
Originationn
X 100


Species Extinctions
1600
1400
1200
1000
800
600
400
200
0
600
500
400
300
200
100
0
Miilion Years Ago
Extinctionss
X 100

C: Social Trends - Divorce Rates

Effect of Societal Laws on Behavior

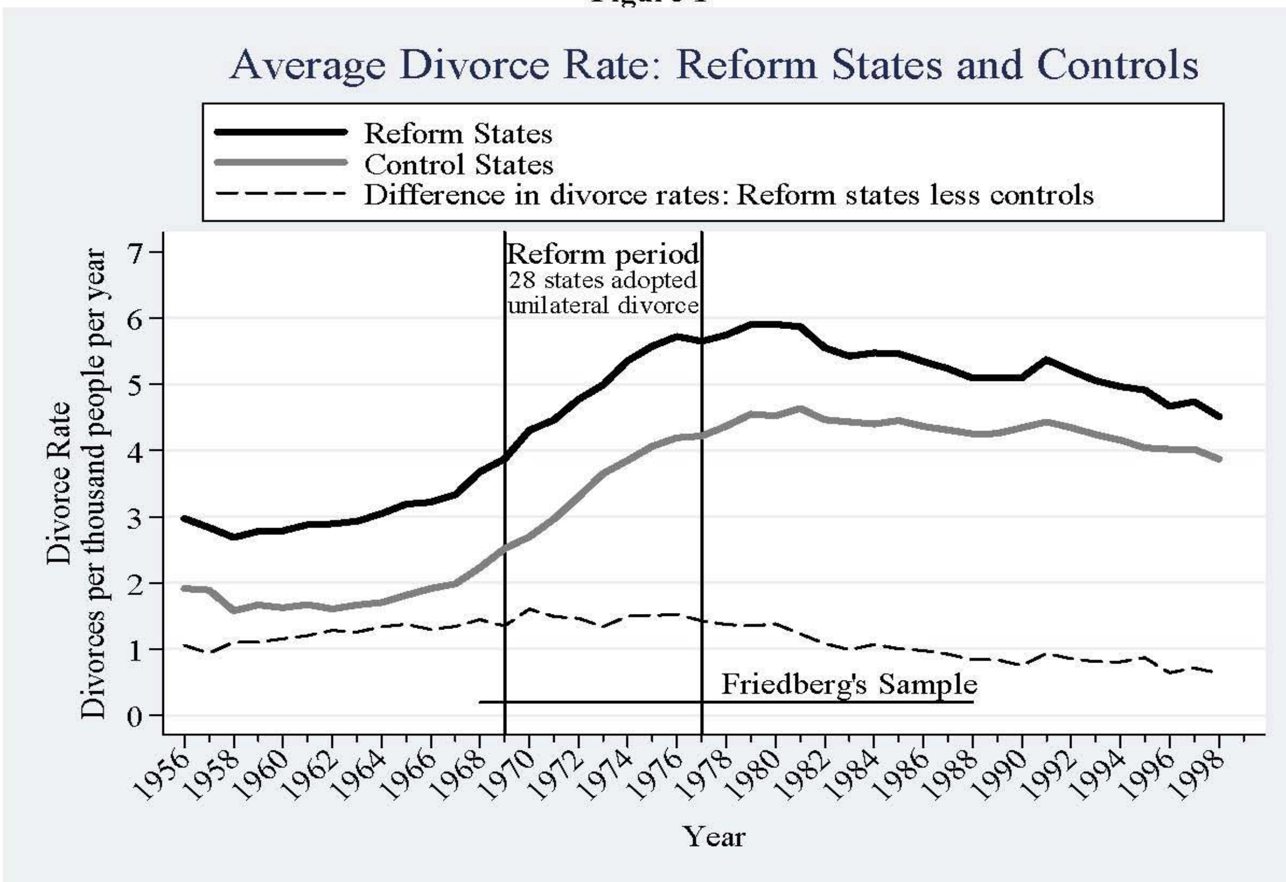
- Marital bargaining models (Becker, 1981)
- Empirical Trends in Divorce over US States (Wolfers, AER 2006)

a. effect of unilateral/no fault divorce laws

b. regime change – structural change in trend from consent divorce regime

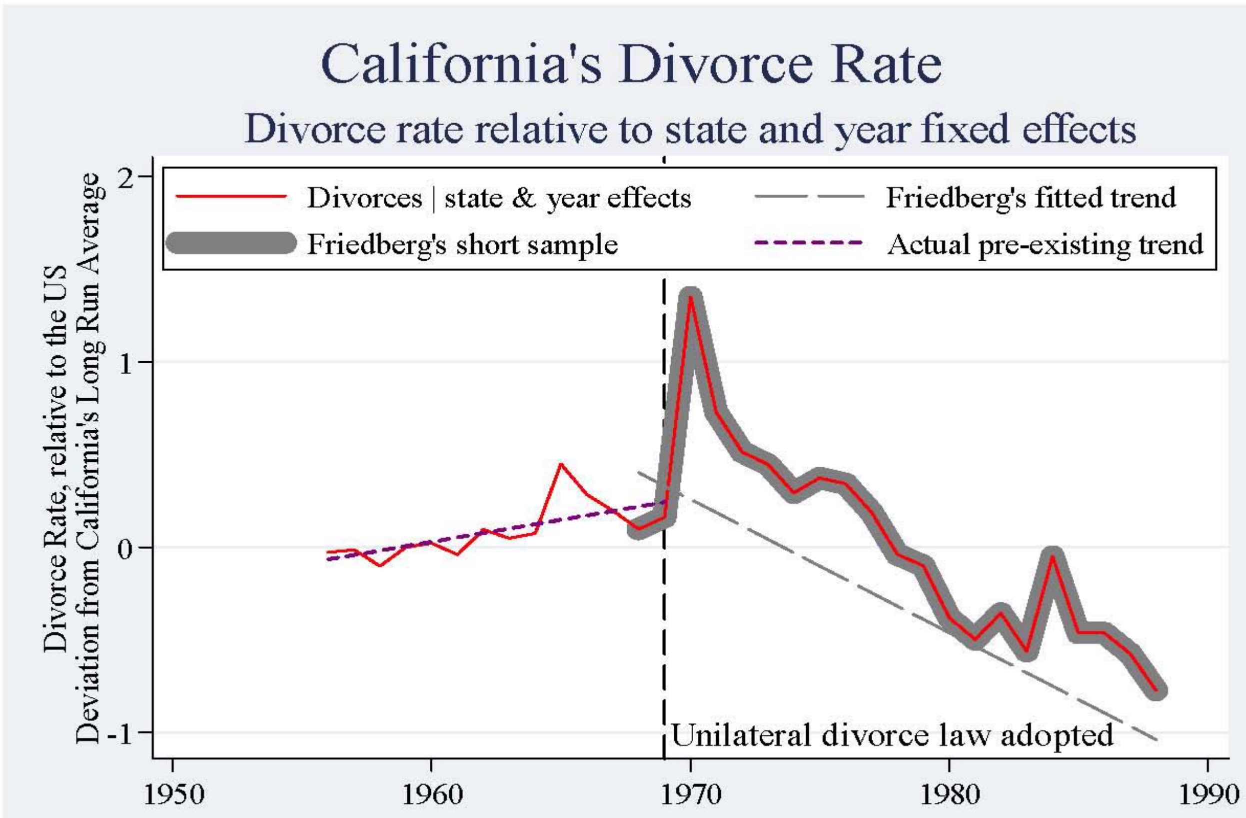
c. dynamic responses over time to regime change

Figure 1



Average Divorce Rate: Reform States and Controls
Reform States
Control States
Difference in divorce rates: Reform states less controls
Reform period
28 states adopted
unilateral divorce
Friedberg's Sample
Divorce Rate
Divorces per thousand people per year
Year
0
1
2
3
4
5
6
7
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998

Figure 5


California's Divorce Rate
Divorce rate relative to state and year fixed effects
Divorces | state & year effects
Friedberg's short sample
Friedberg's fitted trend
Actual pre-existing trend
Divorce Rate, relative to the US
Deviation from California's Long Run Average
Unilateral divorce law adopted
2
1
0
-1
1950
1960
1970
1980
1990

Modeling and Understanding Trends

- Many possible functional forms - polynomial, trigonometric polynomial, exponential, neural net
- Relatively easy to get decent fit
  - but what use is it?
  - What do the coefficients mean + how do we interpret them?
- Modeling data generating process:
  - need to evaluate models + accommodate misspecification
  - trend may well be stochastic in nature
  - if so, how does deterministic modeling cope?
  - is there a random walk or unit root in the history of life?
- When there is a trending panel - how to do we correlate the trends?

Explicit Forms of Trend Function

1. Time Polynomial or power function form with residual

$$X_t = \sum_{i=0}^{p} a_i t^i + X_t^0; \quad X_t = \sum_{i=0}^{p} a_i t^{\alpha_i} + X_t^0$$

2. General Deterministic - nonparametric forms with residual

$$X_t = f(t) + X_t^0; \quad X_t = f\left(\frac{t}{n}\right) + X_t^0$$

3. Breaking Trends + partial + multiple breaks

$$X_t = \left(\sum_{i=0}^{p_1} a_i^1 t^i\right) \mathbf{1}(t < n_1) + \left(\sum_{i=0}^{p_2} a_i^2 t^i\right) \mathbf{1}(t \geq n_1) + X_t^0$$

4. Smooth Transition functions (e.g. STAR, VECM models)

$$\begin{aligned} \Delta X_t &= A z_t(\beta) + B z_t(\beta) F(q_t, \lambda) + u_t, \quad F(q_t, \lambda) = \frac{1}{1 + e^{-\lambda_1 (q_t - \lambda_2)}} \\ z_t(\beta) &= \left(\beta' X_{t-1}, \Delta X_{t-1}, ..., \Delta X_{t-p}\right) \end{aligned}$$

5. Decay Models - evaporating trends

$$X_t = \frac{\beta}{t^\alpha} + u_t, \quad X_t = \frac{\beta}{L(t)\, t^\alpha} + u_t, \quad L(t) \text{ slowly varying at } \infty$$

6. Nonlinear factor models with trend

$$\begin{aligned} X_{it} &= \delta_{it}\mu_t, \quad \delta_{it} = \begin{cases} \delta_i + \frac{\theta_i}{L(t)t^\alpha} + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha} \to_p \delta_i & \text{idiosyncratic paths} \\ \delta + \frac{\theta_i}{L(t)t^\alpha} + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha} \to_p \delta & \text{common paths} \end{cases} \\ \mu_t &= \text{common trend/growth component} \end{aligned}$$

7. Explosive bubbles

$$X_t = \theta X_{t-1} + u_t, \quad \begin{cases} \theta > 1 & \text{pure explosive process} \\ \theta = 1 + \frac{c}{k_n} > 1, \quad k_n \to \infty & \text{mildly explosive process} \end{cases}$$

Common Stochastic Trends

1. Unit root (accumulated sum) model - $I(1)$ process

$$\Delta X_t = u_t; \quad X_t = \sum_{s=1}^{t} u_s + X_0$$

2. Multiple unit root model - $I(2)$ process

$$\begin{aligned} \Delta^2 X_t &= u_t; \text{ or } \Delta X_t = v_t, \quad \Delta v_t = u_t \text{ so that} \\ X_t &= \sum_{s=1}^{t} \left(\sum_{j=1}^{s} u_s + \Delta X_0 \right) + X_0 \\ &= \sum_{s=1}^{t} \sum_{j=1}^{s} u_s + t\Delta X_0 + X_0 \end{aligned}$$

3. Long Memory model (fractional integration) - $I(d)$ process

$$\begin{aligned} (1-L)^d X_t &= u_t \text{ or} \\ X_t &= \begin{cases} \sum_{j=0}^{\infty} \frac{(d)_j}{j!} u_{t-j} & |d| < \frac{1}{2} \\ \sum_{j=0}^{t} \frac{(d)_j}{j!} u_{t-j} + X_0 & d \geq \frac{1}{2} \end{cases} \end{aligned}$$

Effects of Trend

1. Observed behavior: divergence of process, no fixed mean, secular growth, explosive bubble, recurrence (visits every point in sample space)

2. Asymptotic form - standardized process (deterministic trend, semimartingale, Brownian motion, fractional Brownian motion): $f\left(\frac{t}{T}\right) \sim M(r)$ for $t = [Tr]$.

3. Changes in statistical theory and classical asymptotics (unit roots, cointegration, singularity of moment matrix limits due to common trends, degeneracy of limit theory, discontinuities in limit theory)

4. Importance of full trajectory + initialization

5. Prediction and prediction standard errors

6. Persistence of shocks, butterfly effects

Trend Extraction

1. Smoothing and Filtering

A. The Hodrick Prescott -Whittaker Filter: fit a trend to data $y^n = \{y_t\}_{t=1}^n$ by the smoother

$$\hat{f}_t = \arg\min_{f_t} \left\{ \underbrace{\sum_{t=1}^{n} (y_t - f_t)^2}_{\text{best least squares fit}} + \underbrace{\lambda \sum_{t=2}^{n} \left(\Delta^2 f_t\right)^2}_{\text{penalty for roughness}} \right\} = \hat{f}_t\left(y^n\right)$$

The fitted cycle is the residual

$$\hat{c}_t = y_t - \hat{f}_t$$

References

i Hodrick, R. J. and E. C. Prescott (1997), *J. Money, Credit and Banking, 29,* 1-16.

ii. Whittaker (1923). *Proc. Edinburgh Math. Assoc.* 78, 81-89..

Notes on the WHP Filter:

1. $\hat{f}_t$ depends on the full trajectory y^n - it smooths the data y^n.

2. As $\lambda \to \infty$, the penalty rises, $\hat{f}_t$ is smoother and eventually $\hat{f}_t = a + bt$ is linear

3. As $\lambda \to 0$, the penalty is less important (more roughness is allowed) until ultimately $\hat{f}_t = y_t$ and there is no smoothing.

4. $\lambda = 1600$ is often used in practical work with quarterly data

5. The solution satisfies the functional equation

$$\hat{f}_t = \frac{1}{\lambda L^{-2} (1-L)^4 + 1} y_t, \quad \hat{c}_t = \frac{\lambda L^{-2} (1-L)^4}{\lambda L^{-2} (1-L)^4 + 1} y_t$$

6. Observe that if $y_t = (1-L)^{-1} u_t$, so y_t is $I(1)$, then $\hat{c}_t = \frac{\lambda L^{-2}(1-L)^3}{\lambda L^{-2}(1-L)^4+1} u_t$ and $\hat{c}_t$ is apparently stationary.

7. Practical calculation of the WHP filter is usually by a numerical procedure.

B. **Band Pass Filtering**

(a) i. Ideal filter to extract the business cycle in the data is a bandpass filter that extracts components with periodic fluctuations in the business cycle frequency - say between 6-32 quarters.

ii. Baxter and King find the best approximant time domain filter corresponding to this (for frequencies greater than λ_0) is:

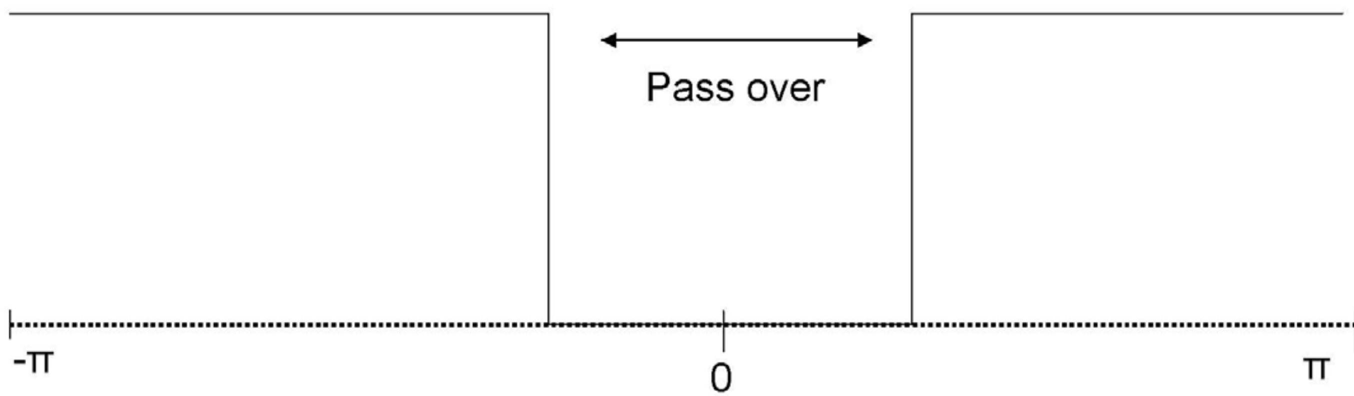
$$b(L) = \sum_{h=-K}^{K} b_h L^h, \quad \text{with } b_0 = \frac{\lambda_0}{\pi}, \ b_h = \frac{\sin(h\lambda_0)}{h\pi} \quad h = 1, 2, ..$$

References

i. Baxter and King (1999) *BRev. Econ. & Stat.* 81, 575-593.

ii. Corbae, Ouliaris & Phillips (2002). *Econometrica,* 70, 1067-1109..

iii Corbae & Ouliaris (2006) Ch. 6 in *Econometric theory and Practice* (ed. D. Corbae, S. Durlauf and B.Hansen) Cambridge.



An Ideal Band Pass Filter

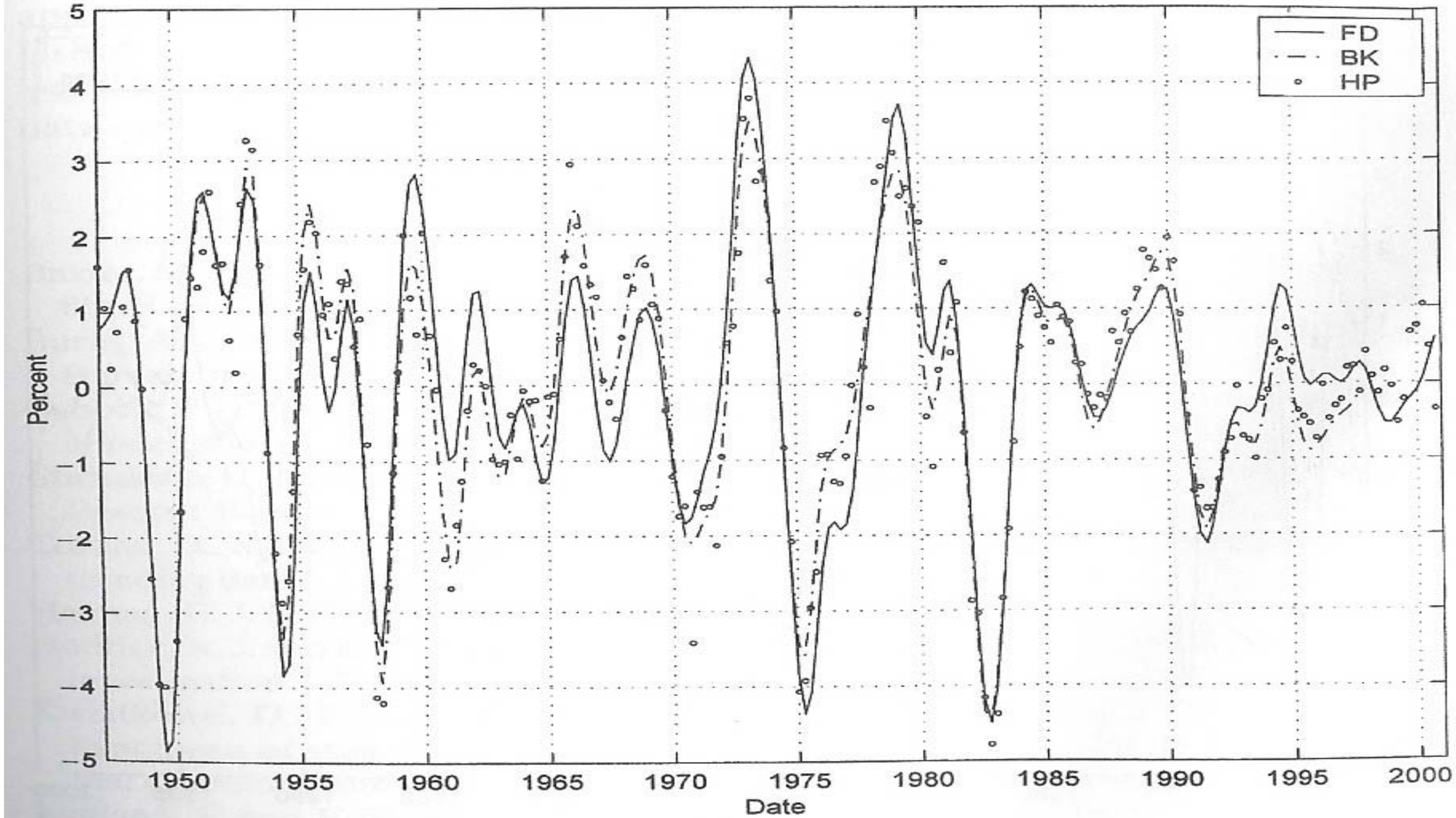


Figure 6.2. HP, BK, and FD filtered quarterly real GDP.

Business Cycles in Post War US GDP

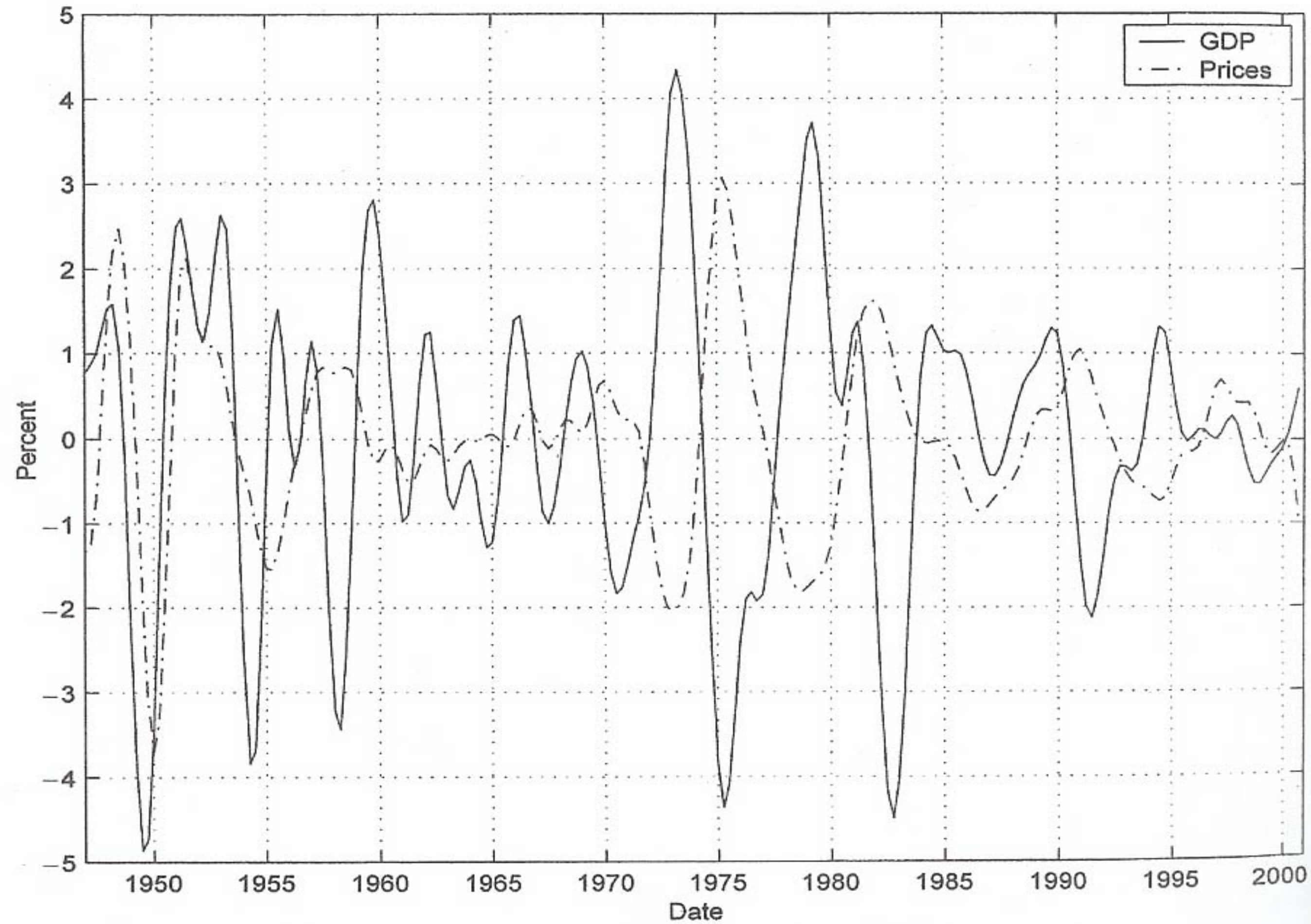


Figure 6.3. Cyclical component of real GDP and the price level.

Post War Cycles in US GDP and Prices

C. **Difference Filtering, Unit Root Determination, Quasi-Differencing**

$$\Delta X_t, \quad \Delta^2 X_t, \quad \Delta^m X_t, \quad (1-L)^d X_t, \quad (1-\theta_n L) X_t, \quad \theta_n = 1 + \frac{c}{k_n}$$

References

i. Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control.* Holden Day.

ii. Dickey D. and W. Fuller 1979, *Journal of the American Statistical Association* 74, 427–431.

iii. Dickey D. and W. Fuller 1981, *Econometrica* 49, 1057–1072.

iv. Phillips, P. C . B. (1987). *Econometrica,* 55, 277–302.

v. Phillips P. C. B. and W. Ploberger (1996) *Econometrica,* 64, 381-413.

2. Trend Extraction by Regression

Most Common Case of Time Polynomial Regression

$$
\begin{aligned}
X_t &= \beta_0 + \beta_1 t + ... + \beta_p t^p + u_t = \beta' x_t + u_t, \quad \text{say} \qquad (1) \\
\gamma_h &= E\left(u_t u_{t+h}\right), \qquad \sum_{h=-\infty}^{\infty} |\gamma_h| < \infty
\end{aligned}
$$

- Efficient time series regression is possible by least squares (OLS)
- **Grenander Rosenblatt Theorem**
  - OLS regression on (1) is asymptotically as efficient as GLS regression provided spectrum $f_u(\lambda)$ is continuous and nonzero at $\lambda = 0$.
  - Condition holds if $\sum_{h=-\infty}^{\infty} |\gamma_h| < \infty$, and $\sum_{h=-\infty}^{\infty} \gamma_h \neq 0$
- Asymptotic variance formula is

$$
\omega^2 \left(X'X\right)^{-1}, \quad \omega^2 = \sum_{h=-\infty}^{\infty} \gamma_h = \text{lrvar}\left(u_t\right) \qquad (2)
$$

Notes on Application of Grenander Rosenblatt Theorem

- Formula (2) for the asymptotic variance matrix holds in spite of the asymptotic singularity of $X'X$.

- The long run variance ω^2 can be estimated by the usual HAC estimator involving lag kernel methods, e.g.

$$\hat{\omega}^2 = \sum_{h=-M}^{M} k\left(\frac{h}{M}\right)\hat{\gamma}_h, \quad \frac{1}{M}+\frac{M}{n}\to 0, \quad k(\cdot) = \text{lag kernel (e.g. } k(x) = 1-|x|\text{)}$$

- Efficiency result extends to the case where x_t has a unit root and is strictly exogenous.

- Result fails when u_t has a root near unity or displays long memory. In these cases, $f_u(\lambda)$ is not continuous at the origin. Efficient estimation then involves dealing with the peak in the spectrum of $f_u(\lambda)$.

References on Trend Extraction by Regression

i. Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* Wiley

ii. Phillips, P. C. B. and J. Y. Park (1988), *Journal of the American Economic Association* 83, 111–115.

iii. Phillips, P.C.B. And C.C. Lee, (1996), In P.M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series: Essays in Memory of E.J. Hannan*, Springer–Verlag: New York.

iv. Canjels, N. And M. Watson (1997). *Review of Economics and Statistics, 79, 184-200.*

Figure 1: Relative efficiency in linear trend model with near-integrated errors

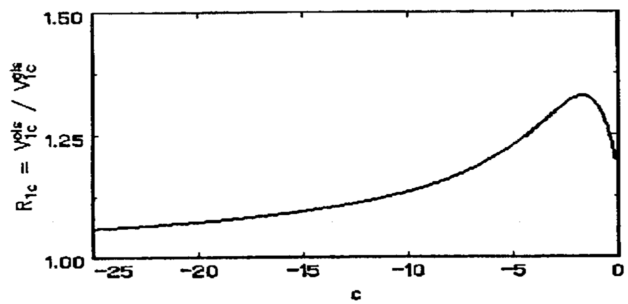


Figure 2: Relative efficiency in linear trend plus intercept model

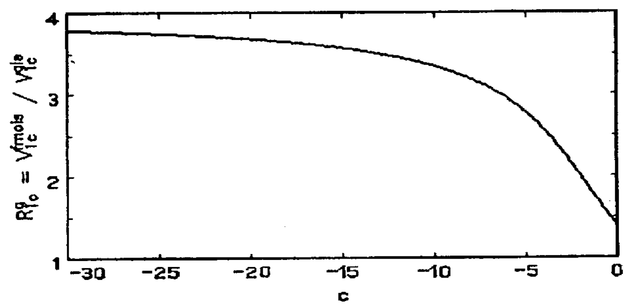


Figure 3: Relative efficiency in linear trend model with intercept and with c~c0T1/2

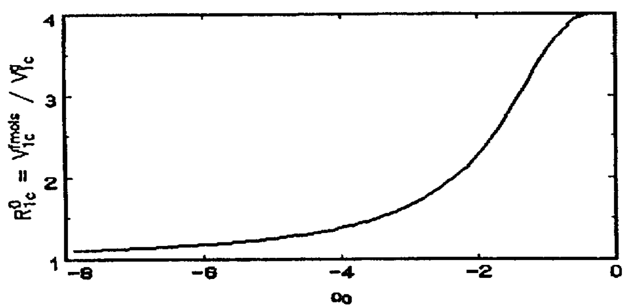
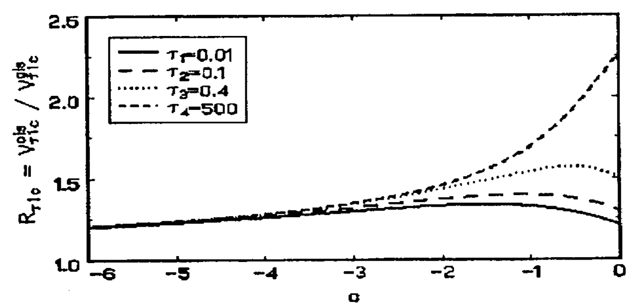


Figure 4: Efficiency curves for various τ



Relative Asymptotic Efficiency of OLS vs Quasi-Differencing + OLS in Deterministic Trend Regression

3. Nonparametric Trend Extraction

- **Sieve estimation, e.g. by polynomial regression approximation, spline smoothers such as**

$$\arg\min_f \left\{ \frac{1}{n} \sum_{t=1}^{n} \left(X_t - f\left(\frac{t}{n}\right)\right)^2 + \lambda \int \left(f''\right)^2 \right\}$$

- **Kernel regression**

$$\begin{aligned} X_t &= f\left(\frac{t}{n}\right) + u_t \\ \hat{f}(x) &= \frac{n^{-1}\sum_{t=1}^{n} X_t K_h\left(\frac{t}{n} - x\right)}{n^{-1}\sum_{t=1}^{n} K_h\left(\frac{t}{n} - x\right)} = \arg\min_f \sum_{t=1}^{n} \left(X_t - f\right)^2 K_h\left(\frac{t}{n} - x\right) \\ K_h(z) &= h^{-1} K\left(\frac{z}{h}\right),\ K(\cdot) = \text{ kernel function (e.g. } \frac{1}{\sqrt{2\pi}} e^{-z^2/2}),\ h = \text{bandwidth} \end{aligned}$$

- Local linear trend regression

$$\arg\min_{f_0, f_1} \sum_{t=1}^{n} \left(X_t - f_0 - f_1\left(\frac{t}{n} - x\right)\right)^2 K_h\left(\frac{t}{n} - x\right)$$

Asymptotics and Inference

- For kernel regression under regularity conditions and undersmoothing

$$\sqrt{nh}\left(\hat{f}(x) - f(x)\right) \sim N\left(0, \sigma_u^2 \int K(s)^2\, ds\right)$$

- When u_t is autocorrelated, such NP estimates are not asymptotically efficient - unlike parametric regression estimates. Refined procedures (like NP Cochrane-Orcutt transformations) help to improve efficiency and reduce the variance component σ_u^2 to σ_ε^2 where $u_t = C(L)\,\varepsilon_t$.

References on NP Regression + Efficiency

i. Xiao, Z. et. al. (2003) *J. American Statistical Association*, 98, 980-992.

ii. Su, L. and A. Ullah (2005) More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. Mimeo.

Asymptotic Variance involves the following limit for $x \in (0,1)$

$$
\begin{aligned}
n^{-1}\sum_{t=1}^{n} K_h\left(\frac{t}{n}-x\right) &= \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{2\pi}h}e^{-\frac{\left(\frac{t}{n}-x\right)^2}{2h^2}} \sim \int_0^1 \frac{1}{\sqrt{2\pi}h}e^{-\frac{(s-x)^2}{2h^2}}ds \\
&= \int_{-\frac{x}{h}}^{\frac{(1-x)}{h}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \rightarrow \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz = 1
\end{aligned}
$$

Model Choice, Order Determination and Automated Econometric Inference

- Model selection approaches - Bayesian, Information theoretic, Prequential, Likelihood inference
- Applications to: trend, order selection, differencing + unit roots, cointegration rank, parameter restrictions, Bayesian hyperparameters
- Automation in inference and prediction
- Nonparametric bandwidth selection, sieve order selection
- Data snooping
- Proximity theorems - how close can we get to the true model?
- Post Model Selection Inference

References

i. Schwarz, 1978. *Annals of Statistics* 6, 461–464.

ii. Vuong, Q. (1989). *Econometrica,* 57, 307-333.

iii. Phillips P. C. B. and W. Ploberger (1996) *Econometrica,* 64, 381-413.

iv. Phillips, P. C . B. (1996). *Econometrica,* 64, 763-812.

v. White, H. (2000). *Econometrica,* 68, 1097-1126.

vi. Ploberger, W. and P. C. B. Phillips (2003). *Econometrica,* 71, 627-673.

vii. Leeb, H. and B. M. Potscher (2005). *Econometric Theory*, 21, 21-59.

Model Selection - the Bayesian Approach

Assign prior probabilities to models and set up likelihoods and priors for individual models to explain data X^n:

$$\begin{aligned}
\text{Models} &: \quad M_j : j = 1, ..., J \\
\text{Prior Probabilities} &: \quad \pi_j : j = 1, ..., J \\
\text{Joint Probability: } P(M_j, X^n) &= P(M_j) P(X^n | M_j)) \\
&= P(X^n) P(M_j | X^n) \\
\text{Posterior Probability of Model:} P(M_j | X^n) &= \frac{P(M_j) P(X^n | M_j)}{P(X^n)} \\
&= \frac{\pi_j P(X^n | M_j)}{\sum_{k=1}^{J} \pi_k P(X^n | M_k)} \\
\text{Data Probability } P(X^n) &= \sum_{k=1}^{J} \pi_k P(X^n | M_k)
\end{aligned}$$

Selection Rule

- Choose model according to the rule that maximizes posterior probability of the model using $P(M_j|X^n) = \frac{P(M_j)P(X^n|M_j)}{P(X^n)}$

$$\hat{j} = \arg\max_j P(M_j|X^n) = \arg\max_j pdf(X^n|M_j)$$
$$\text{if prior probability } \pi_j = \frac{1}{j} \text{ is uniform across models}$$

- Requires evaluation of $P(X^n|M_j)$ or Bayes data density $pdf(X^n|M_j)$

Bayes Data Density

- Use Bayes Rule to extract data probability $P\left(X^n|M_j\right)$ for model M_j

$$P\left(X^n|M_j\right) = \underbrace{\int_{\Theta_j}}_{\substack{\Theta_j = \text{parameter space} \\ \text{for model } M_j}} \underbrace{\pi_{M_j}\left(\theta_j\right)}_{\substack{\text{prior density} \\ \text{for } \theta_j}} \underbrace{pdf_{M_j}\left(X^n|\theta_j\right)}_{\substack{\text{likelihood} \\ \text{for } \theta_j}} \underbrace{d\theta_j}_{\substack{\text{parameter} \\ \text{for model } M_j}}$$

Asymptotic Form of Data Density

- Let $\ell_n(\theta) = \log(pdf(X^n|\theta))$ be log likelihood. Then, under some general regularity conditions $\hat{\theta}$

$$
\begin{aligned}
pdf(X^n) &= \int_\Theta \pi(\theta)\, pdf(X^n|\theta)\, d\theta = \int_\Theta \pi(\theta)\, e^{\ell_n(\theta)} d\theta \\
&\sim \frac{(2\pi)^{k/2}\, \pi\left(\hat{\theta}\right) e^{\ell_n\left(\hat{\theta}\right)}}{\left|I_n\left(\hat{\theta}\right)\right|^{1/2}} \quad \text{PIC density, with} \begin{cases} \hat{\theta} = \text{MLE of } \theta \\ I_n\left(\hat{\theta}\right) = \text{information} \end{cases}
\end{aligned}
$$

- Log data density

$$
\begin{aligned}
\log(pdf(X^n)) &\sim \underbrace{\ell_n\left(\hat{\theta}\right)}_{\text{log likelihood}} - \underbrace{\frac{1}{2}\log\left|I_n\left(\hat{\theta}\right)\right|}_{\text{penalty involving sample information}} + \underbrace{O_{a.s.}(1)}_{\text{prior density is of smaller order}} \\
&= \text{penalized log likelihood}
\end{aligned}
$$

General Model Choice Rule – PIC Criterion:

$$\begin{aligned}\hat{j} &= \arg\max_j pdf\left(X^n|M_j\right) \\ &= \arg\max_j \left\{\ell_n^{M_j}\left(\hat{\theta}_j\right) - \frac{1}{2}\log\left|I_n^{M_j}\left(\hat{\theta}_j\right)\right|\right\}\end{aligned}$$

Stationary Case – BIC Order Criterion:

Sample information satisfies

$$\frac{1}{n}I_n\left(\hat{\theta}\right) = -\frac{1}{n}\frac{\partial^2\ell_n\left(\hat{\theta}\right)}{\partial\theta\partial\theta'} \to_{a.s.} I\left(\theta\right) = \text{limiting Fisher information}$$

so that the penalty term in the penalized likelihood

$$\frac{1}{2}\log\left|I_n\left(\hat{\theta}\right)\right| \sim \frac{1}{2}\log\left\{nI\left(\theta\right)\right\} = \frac{1}{2}\log\left(n^k\right) + \frac{1}{2}\log\left|I\left(\theta\right)\right| \sim \frac{k}{2}\log\left(n\right)$$

has the simple form

$$\frac{1}{2} \times \text{ Parameter Count } \times \log n$$

Proposition:

Models are not only unknown but inherently unknowable.

E. J. Hannan:

"Never any attainable true system generating the data."

Best to be hoped for —

"Such understanding of structure of system to be available that only a VERY RESTRICTED model class can be successfully used."

Proximity Theory

How close to true system can we come?

- **Quantify closeness**: KL distance, relying on

$$\left.\log\left(\frac{d\mathcal{G}}{dP_n^{\theta_n}}\right)\begin{array}{l}\leftarrow \text{ candidate data measure} \\ \leftarrow \text{ parametric measure}\end{array}\right\} = \text{relative likelihood}$$

- **Bounds?**: when parameters (θ_n) have to be estimated there is a bound on how close we can get to $P_n^{\theta_n}$

- **Factors:** bound depend on

— dimension of parameter space (curve of dimensionality)

— "information" in data

- **References:**

— Rissanen (1986, 1987); Ploberger & Phillips (1996,2003; *Econometrica*)

– Probability Framework –

- space: $(\Omega, \mathcal{F}, P)$, $\mathcal{F}_n$, $P_n = P|\mathcal{F}_n$
- data: $Y^n = (Y_t)_1^n$
- parameterized family: P_n^θ, $\theta \in \Theta$

$$\begin{aligned}
\theta_n^0 &= \arg\max_\theta \int \ln\left(\frac{dP_n^\theta}{dP_n}\right) dP_n \\
&= \arg\min KL(P_n, P_n^\theta)
\end{aligned}$$

– Popular Model Classes –

- **VARs + trends:** $\mathrm{Var}(p) + \mathrm{Tr}(t)$

$$y_t = J(L)y_{t-1} + d(t) + \varepsilon_t$$

- **Dynamic SEMs & Structural VARs**

$$By_t = J(L)y_{t-1} + d(t) + \varepsilon_t$$

- **RRRs & ECMs**

$$\begin{aligned} \Delta y_t &= \alpha\beta' y_{t-1} + \Phi(L)\Delta y_{t-1} + d(t) + \varepsilon_t \\ \Delta y_t &= \alpha^0 \beta^0(b)' y_{t-1} + \Phi(L)\Delta y_{t-1} + d(t) + \varepsilon_t \end{aligned}$$

- **BVAR's**

$$\Delta y_t = Ay_{t-1} + \Phi(L)\Delta y_{t-1} + d(t) + \varepsilon_t = Cx_t + \varepsilon_t$$

prior: $\pi(c) =_d N(\bar{c}, V_c)$, $V_c = V_c(\psi)$; hyperparameters: $\bar{c}, V_c = \mathrm{diag}(\lambda, \theta)$

– Why Reduce # Parameters? –

- **improve forecasting performance**

$$\text{VAR's} \begin{array}{l} \nearrow \ \text{RRR's} \\ \rightarrow \ \text{ECM's} \\ \searrow \ \text{BVAR's} \end{array}$$

- **help interpret results**

- **curse of dimensionality** (given n) can get

$$\frac{d\mathcal{G}^{M_1}}{dP^{\theta^0}} > \frac{d\mathcal{G}^{M_2}}{dP^{\theta^0}} \text{ for fitted } M_1,\ M_2$$

$$\text{when} \quad \#M_1 < \#M_2$$

even if P^{θ^0} has more parameters (and is closer in form to M_2)!!

- **small is beautiful**

— small models easy to adapt; big models hard to adapt - greater commitment to specification

– How to Choose Models –

- **Classical pretesting**

  — sequential tests

  — general to specific

  — specific to general

- **Bayesian**

  — posterior odds: $P(M_1)/P(M_2)$

  — Bayes factors: $dQ^{M_1}/dQ^{M_2} = \frac{pdf^1(X^n)}{pdf^2(X_n)}$

  — predictive odds (Geisser, Atkinson, Gelfand)

- **Prequential:** — sequential 1-period ahead forecast densities

$$\frac{\prod_{t=n_0+1}^{n} f_{M_1}(y_t|Y^{t-1}, \hat{\theta}_{t-1})}{\prod_{t=n_0+1}^{n} f_{M_2}(y_t|Y^{t-1}, \hat{\varphi}_{t-1})}$$

- **Information criteria:** stochastic complexity minimum description length

AIC, BIC, MDL, PIC

– Special Issues –

- **Models with hyperparameters**

$$y_t = \Pi(c)x_t + \varepsilon_t$$

— prior

$$\begin{aligned} c &=_d N(\bar{c}, V_c) \\ \bar{c} &= \bar{c}(\psi), V_c = V_c(\psi) \end{aligned}$$

— tightness hyperparameters ψ

- **No clear parameter count**

$$\begin{aligned} \# = \dim(c) &\ , \ V_c > 0 \\ \# = 0 &\ , \ V_c = 0 \ \ (c = c^0) \end{aligned}$$

- **continuum of choices** $[0, \#(c)]$
- **non nested models** — in VAR class (e.g., BVARs, RRRs)

Simple Illustration: Spurious Regression

$$\text{True DGP:} \quad y_t = y_{t-1} + u_t$$

$$\text{fitted model:} \quad y_t = \hat{b}t + \hat{u}_t$$

Limit behavior

$$\hat{b} \to_p 0$$

$$t(\hat{b}) \text{ divergent} \quad O_p(n^{1/2})$$

Conclusion

- deterministic trend proxies for unit root
- model shortcoming NOT statistical
- trends, I(1) data = powerful regressors
- can be "powerfully wrong" in forecasting

– Themes in Automated Modeling –

- **Role of Model**

— language to express regular features of data

Rissanen (1986) suggests goal is to

"remove untenable assumptions of data generation systems and 'true' parameters"

- **Primary task**

Dawid (1984)'s prequential approach

— "make sequential probability forecasts of future observations"

- **Modeling evolutionary mechanisms**

— data dependent $\begin{cases} \text{parameter count} \\ \text{initialization} \end{cases}$

LeCam & Yang (1990): "# parameters" depends on "# observations

– Use Model Selection –

for Parsimony & Practicality

- **Bayes factor** (LR)

$$\frac{pdf^0(X^n)}{pdf^1(X^n)} \begin{matrix} \geq \\ = \\ < \end{matrix} 1?$$

H_0: $pdf^0(X^n) = \int \pi_0(\theta) pdf(X^n|\theta) d\theta$

H_1: $pdf^1(X^n) = \int \pi_1(\psi) pdf(X^n|\psi) d\psi$

- **asymptotic form:**

$$\log(pdf^j(X^n) \sim \ell_n^j(\hat{\theta}_n^j) - \tfrac{1}{2} \log |I_n^j|$$

- **criterion:** choose model $M_{\widehat{j}}$ according to PIC criteria

$$\widehat{j} = \arg\max_j \left\{ \ell_n^j(\hat{\theta}_n^j) - \tfrac{1}{2} \log |I_n^j| \right\}$$

Application – Order Selection in Gaussian models

$$\text{AR}(k),\ \text{ARMA}(p,q),\ \text{Tr}(t)$$

- PIC $\arg\max_k \log|\Sigma_n| + \frac{1}{n}\log|I_n|$
- BIC $\arg\max_k \log|\Sigma_n| + \frac{k}{n}\log n$
- HQ $\arg\max_k \log|\Sigma_n| + \frac{k}{n}\log\log n$
- AIC $\arg\max_k \log|\Sigma_n| + \frac{2k}{n}$

- **PIC has greater penalty for trend**

$$\text{PIC:} \quad \log\left(\sum_{t=1}^{n} t^2\right) = \log n^3 + \text{const.} \sim 3\log n$$

$$\text{BIC:} \quad \log n$$

– Compare Predictive Odds –

- **Bayes predictive odds**

$$\frac{pdf^0(X^n_{n_0+1}|X^{n_0})}{pdf^1(X^n_{n_0+1}|X^{n_0})} \begin{matrix}\geq \\ = \\ <\end{matrix} 1 \ ?$$

$$pdf^j(X^n_{n_0+1}|X^{n_0}) = \frac{pdf^j(X^n)}{pdf^j(X^{n_0})}$$

- **Asymptotic form:** conditional PIC/PICF

$$\underbrace{\ell^j_n(\hat{\theta}^j_n)}_{\text{log likelihood}} \quad - \quad \underbrace{\tfrac{1}{2}\log(|I^j_n|/|I^j_{n_0}|)}_{\text{conditional penalty}}$$

- **Prequential form is** equivalent as $n, n_0 \to \infty$,

$$\frac{\hat{p}^0_{n,n_0}}{\hat{p}^1_{n,n_0}} = \frac{\displaystyle\prod_{t=n_0+1}^{n} f^0_t(\cdot|\hat{\theta}^0_{t-1}, X^{t-1})}{\displaystyle\prod_{t=n_0+1}^{n} f^1_t(\cdot|\hat{\theta}^1_{t-1}, X^{t-1})}$$

– VAR, RRR & BVAR Models –

• **Model VAR**(k, ℓ)

$$\begin{aligned}\Delta y_t &= A y_{t-1} + \sum_{i=0}^{k-1} \Phi_i \Delta y_{t-1-i} + \sum_{0}^{\ell} c_j t^j + \varepsilon_t \\ &= C x_t + \varepsilon_t,\ \ \varepsilon_t \equiv \mathrm{iid}(0, \Sigma)\end{aligned}$$

• **Model RRR**(r, k, ℓ)

$$A = \alpha\beta', \quad \beta' = [I_r, F] \text{ say}$$

• **Model BVAR**

prior $\quad \pi(c) \equiv N(\bar{c}, V_c),\ \ \bar{c} = \bar{c}(\psi)$

hyperparameters $\quad \psi,\ \ V_c = V_c(\psi)$

- **BVARM** — Minnesota priors

$$\begin{aligned}\bar{c} &= 0, 1 \text{ (main diagonal)} \\ \text{diag }(V_{c_i}) &= \begin{cases} (\lambda/a)^2, \ i = j & \text{own variable, lag} = a \\ \left(\frac{\lambda\theta\hat{\sigma}_i}{a\hat{\sigma}_j}\right)^2, \ i \neq j & \text{lag} = a \end{cases}\end{aligned}$$

- **BVAR – RBC** — Real business cycle model priors

  — Ingram & Whiteman (1996)

  — Schorfheide (2003)

– Automated Model Choice –

- **General form:** – selection criterion

$$PIC = \log|\hat{\Sigma}_n| + \tfrac{1}{n}\log(|I_n|/|I_{n_0}|)$$

- **VAR**(k,ℓ) **form**

$$I_n = \hat{\Sigma}_n^{-1} \otimes X'X$$

- **RRR**

$$I_n = \begin{bmatrix} \hat{\Sigma}_n^{-1} \otimes \hat{U}'\hat{U} & 0 \\ 0 & \hat{\alpha}_n'\hat{\Sigma}_n^{-1}\hat{\alpha}_n \otimes Y_{2,-1}'Y_{2,-1} \end{bmatrix} \begin{matrix} G \\ F \end{matrix}$$

model

$$\begin{aligned} \Delta y_t &= \alpha\beta' y_{t-1} + \Phi z_t + \varepsilon_t \\ &= G u_t \qquad\qquad\quad + \varepsilon_t \end{aligned} \qquad \text{stationary}$$

$$\beta' y_{t-1} = y_{1t-1} + F y_{2t-1} \qquad \text{nonstationary}$$

– BVAR Forms –

- **BVAR**

$$\text{prior} \qquad \pi(c) \equiv N(\bar{c}, V_c)$$

$$V_c = V_c(\psi)$$

$$\text{information} \quad I_{n,m} = \underset{\text{prior}}{V_c^{-1}} + \underset{\text{sample}}{\hat{\Sigma}_n^{-1} \otimes X'X}$$

- **BVARM case**

$$V_c = V_c(\lambda, \theta), \quad \lambda, \theta \text{ tightness parameters}$$

- **limits for tightness**

— $\lambda \to 0$ model: $\Delta y_t = c_d' d_t + \varepsilon_t$ only trend left

$$|I_{nm}|/I_{n_0m} \to \prod_{n_0+1}^{n} (1 + d_s'(D_{s-1}'D_{s-1})^{-1}ds$$
$$= |I_n|/|I_{n_0}|$$

— $\lambda \to \infty$ forecast error variance for model

$$|I_n|/|I_{n_0}| \to |I_n|/|I_{n_0}| \text{ for unrestricted VAR}$$

get continuum of models + penalties

– Optimized BVAR's –

- $(\hat{\lambda}, \hat{\theta}) = \underset{\lambda,\theta}{\arg\min} PIC^{BVARM(\lambda,\theta)}$
- optimal data determined values of hyperparameters
- makes use of BVAR's automatic

– Optimized RRR's –

- **Rule:** $(\hat{r},\hat{k},\hat{\ell}) = \arg\min_{r,k,\ell} PIC^{RRR(r,k,\ell)}$

$$\begin{aligned}\Delta y_t &= A y_{t-1} + \sum_{0}^{k-1} \Phi_i \Delta y_{t-1-i} + \sum_{0}^{\ell} c_j t^j + \varepsilon_t \\ A &= \underset{m\times r}{\alpha}\underset{r\times m}{\beta'}\end{aligned}$$

- **Consistent estimation of cointegrating rank** (Chao & Phillips, 1999: JOE)

$$\begin{aligned}\hat{r} &\to_p r \\ \hat{k} &\to_p k \\ \hat{\ell} &\to_p \ell\end{aligned}$$

in conjunction with lag order and trend order selection

- **Combine with MLE:** estimate cointegrating space + adjustment/factor loadings

$$\hat{\alpha}, \hat{\beta}$$

- **Compare the Classical Likelihood Ratio (LR) approach to testing** (Johansen, 1996)

  — not consistent unless size $\rightarrow 0$

  — vulnerable to initial settings of lag length and trend degree and inclusion of intercept

  — sequential testing procedures problematic - multiple routings

– Data Discarding and Lifetime of a Model –

- specify a recent history $[n_a, n_b]$ for calibration

- Permit range of initializations $\tau \in [n_0, n^0]$

  — n_0 = minimal information time

  — n^0 = latest possible initialization

- Data-determined τ:

$$\hat{\tau} = \underset{\tau \in [n_0, n^0]}{\arg\max} \frac{q_{n_b}(\cdot|\mathcal{F}_\tau^{n_a})}{q_{n_b}(\cdot|\mathcal{F}_{n_0}^{n_a})}$$

i.e.,

$$\tau = \max_\tau \left[q_{n_b}(\cdot|\mathcal{F}_\tau^{n_a}) = \frac{dQ_{n_b}}{dP_{n_b}} \bigg| \mathcal{F}_\tau^{n_a} \right]$$

maximize conditional Bayes data density $[n_a + 1, n_b]$ given $\mathcal{F}_\tau^{n_a}$

–Optimality Issues –

Can we do better in modelling the 'dgp'?

(Ploberger and Phillips, 1999,2003)

- Rissanen (1986, 1987): $\theta \in \Theta^k$ a.e.

$$\liminf_n \frac{E_\theta\{\log[f(Y^n; k, \theta)/g(Y^n)]\}}{(k/2)\log(n)} \geq 1$$

i.e.

— closest KL distance we can come on average to true density f is bounded below by "$(k/2)\log(n)$" as $n \to \infty$

except for

— negligible sets of θ ($\lambda\{...\} = 0$) —λ = Lebesgue measure

- Proof using $\sqrt{n}$ cgce, CLT for $\hat{\theta}_n$

– Extension to Cases of Random Information –

- for compact set $K \subset \Theta$

$$\lambda \left\{ \theta \in K : P_n^\theta \left[-\log \frac{d\mathcal{G}}{dP_n^\theta} \leq \frac{1}{2}(1-\varepsilon) \log |B_n| \right] \geq \alpha \right\} \to 0$$

$$\varepsilon, \alpha > 0 \quad \text{as } n \to \infty, \quad B_n = qv \text{ score} \sim I_n$$

- measure closeness to P_n^θ by $-\log(d\mathcal{G}/dP_n^\theta)$
- you can't come closer to P_n^θ than $\frac{1}{2}(1-\varepsilon)\log|B_n|$ with the probability as $n \to \infty$

Except for negligible sets with $\lambda(...) = 0$

- divine providence (know θ or parts of it)
- great guess
- prior information that reduces $\dim(\Theta)$

– Proximity of Bayes model & dgp –

- $$\log\left(\frac{dQ_n}{dP_n^\theta}\right) = \log c + \tfrac{1}{2}\underset{O_p(1)}{\underset{\shortparallel}{V_n' B_n^{-1} V_n}} + \tfrac{1}{2}\log|B_n| \text{ under } P_n^\theta$$
$$\sim -\tfrac{1}{2}\log|B_n| \text{ as } n \to \infty$$
comes arbitrarily close (up to $\varepsilon > 0$) to lower bound of approximation

- Cannot do better than Q_n (or $Q_n|\mathcal{F}_{n_0}$ if π improper) except on negligible θ-sets as $n \to \infty$

- justifies Bayes Q_n and classical predictive
$$\hat{P}_n = \Pi_{n_0}^n f_f(\cdot; \hat{\theta}_{t-1})$$
in sense that for an arbitrary empirical measure $\mathcal{G}_n$ we have
$$\log\left(\frac{d\mathcal{G}_n}{dP_n^\theta}\right) \geq_{\text{essentially}} \log\left(\frac{dQ_n}{dP_n^\theta}\right) \sim \frac{1}{2}\log|B_n|$$

– Example –

- **Gaussian linear model**

$$y_t = x_t'\theta + u_t \quad u_t \equiv \text{iid N}\left(0, \sigma^2\right)$$

- **Concentrated log likelihood & information**

$$\ell_n(\theta) = -\frac{1}{2}\sum(y_t - x_t'\theta)^2, \quad B_n = \sum_{t\leq n} x_t x_t'$$

- **Trend & stochastic regressor case**

$$x_t' = (1, t, W_1, ..., W_m, Z_1, ..., Z_p)\,, \quad W_t \equiv I\,(1)\,, Z_t \equiv I\,(0)$$

- **Asymptotic information content of data**

$$\frac{\log\det B_n}{2(\frac{1}{2}+\frac{3}{2}+m+\frac{p}{2})\log n} \to 1$$

– Implications –

- deterministic linear trend 'costs' (in terms of the distance between the empirical model and the DGP) *three times as much* as the lack of knowledge about the constant or the coefficient of a stationary variables!
- stochastic trend costs *twice as much!*
- higher order trends costs more.

– Prediction –

- **Optimal Predictor & arbitrary predictor**

$$\widehat{y}_t = E\left(y_t|\mathcal{F}_{t-1}\right) = x_t'\theta_0, \quad \overline{y}_t = \overline{y}_t(x_t, z^{t-1})$$

- **Associated empirical model $\mathcal{G}$ – from probability density**

$$\prod_{t\leq n} q(y_t|x_t, z^{t-1})$$

$$q_t(y_t|x_t, z^{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \overline{y}_t)^2}{2\sigma^2}\right)$$

- **Likelihood ratio of two models**

$$-\log\frac{dG}{dP_\theta} = \frac{1}{2\sigma^2}\sum_{t\leq n}\{(y_t - \overline{y}_t)^2 - (y_t - x_t'\theta_0)^2\} = \Delta_n$$

- **Ploberger - Phillips bounds**

$$\Delta_n \geq_{\text{essentially}} \frac{1}{2}\log\det B_n$$

– Implications for Prediction –

- **MSE of forecast bounds**

$$\sum_{t\le n}(y_t-\overline{y}_t)^2 \;\ge_{\text{essentially}}\; \sum_{t\le n}(y_t-x_t'\theta_0)^2+\frac{\sigma^2}{2}\log|B_n|$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\text{MSE}\,(\overline{y}_t) \qquad\qquad\qquad\qquad \text{MSE}\,(\widehat{y}_t)$$

— bound measures how close MSE is to that of optimal predictor!

— effect of trends on optimal prediction same as on dgp!

— distance depends on fitted model!

– Simulations –

- **Gaussian linear model**

$$y_t = x_t'\theta + u_t \quad u_t \equiv \text{iid N}(0,1)$$

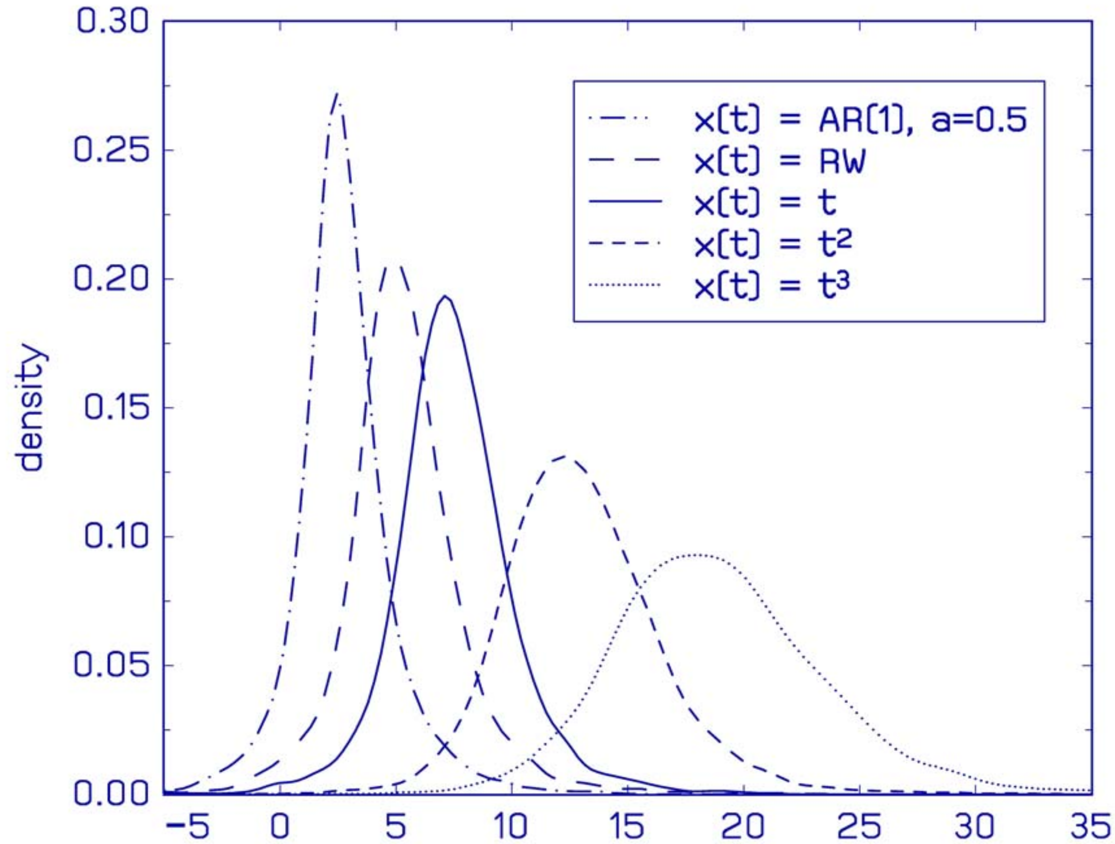
- **Regressors - stationary, unit root and deterministic trends**

$$x_t \equiv AR(1, \rho = 0.5), \; RW, \; t, t^2, t^3$$

- **Forecast Divergence**

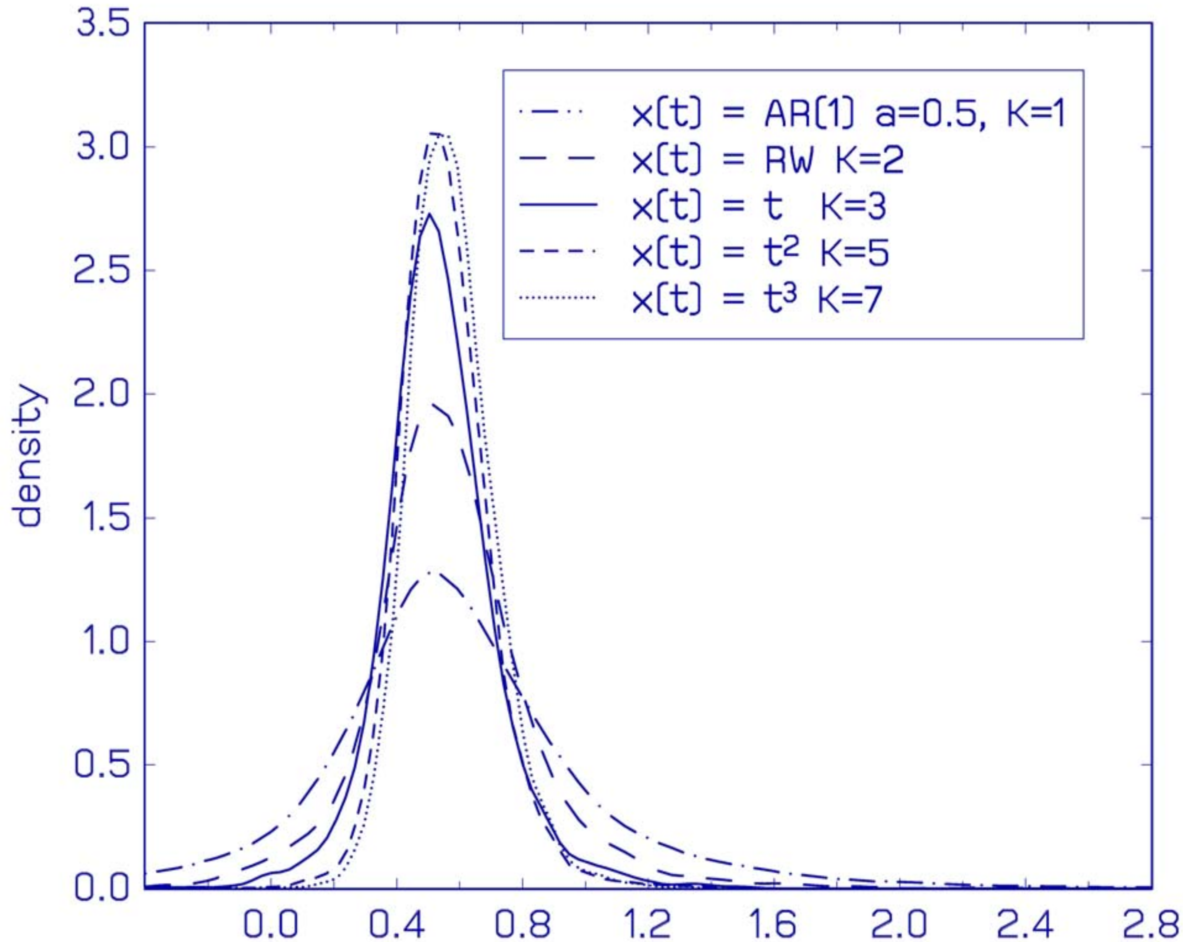
$$\Delta_n = \sum_{t\le n}\{(y_t - \hat{y}_t)^2 - \sum_{t\le n}(y_t - x_t'\theta_0)^2$$

- **Compute** $pdf\,(\Delta_n)\,,\; P\{\Delta_n > (1-\varepsilon)\,K\log n\}$ for $n = 10, ..., 100$ and $\varepsilon = 0.05$

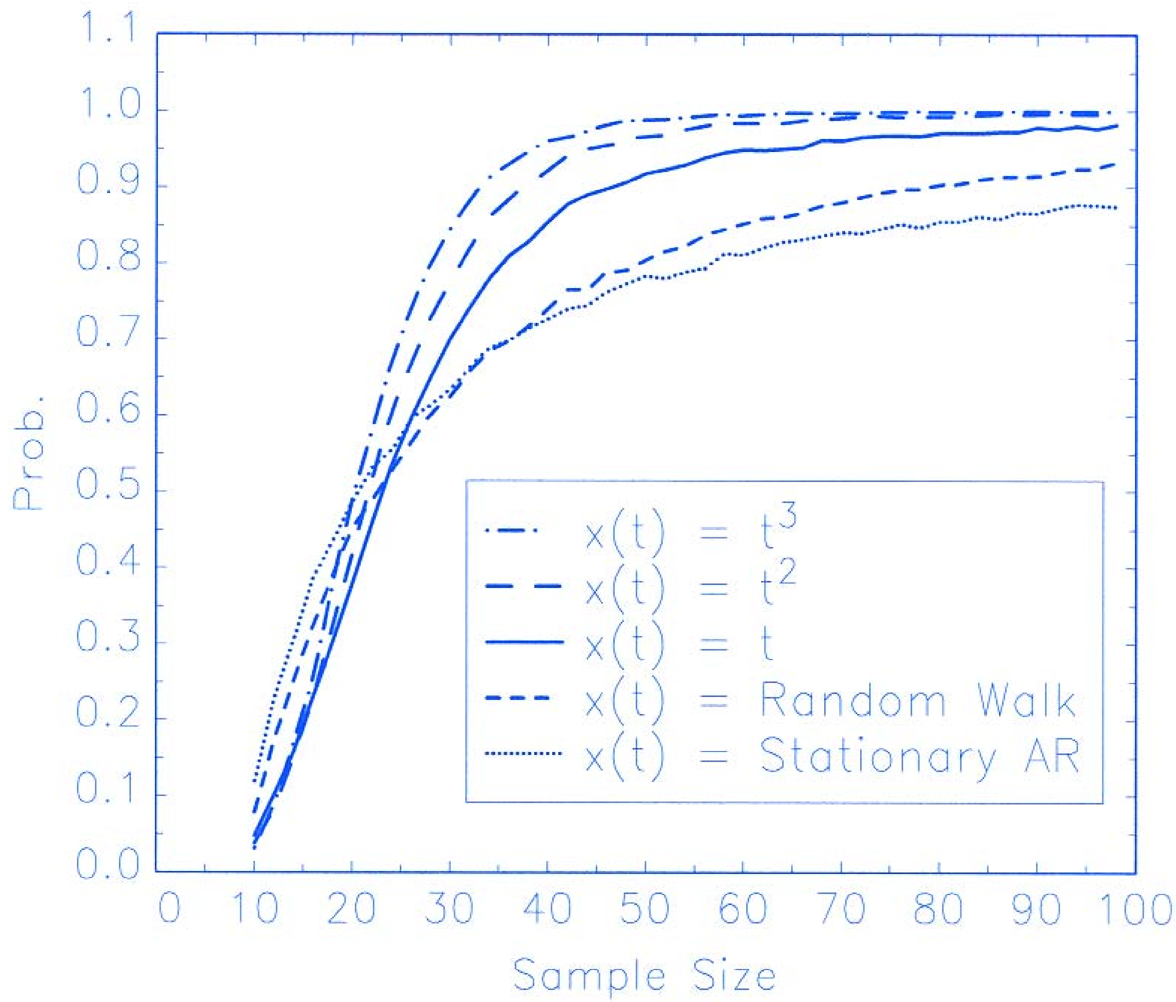


Probability Densities of Forecast Differential

$$\Delta_n = \sum_{t \leq n} (y_t - \hat{y}_t)' \, \Omega^{-1} \, (y_t - \hat{y}_t) - \sum_{t \leq n} (y_t - \tilde{y}_t)' \, \Omega^{-1} \, (y_t - \tilde{y}_t)$$



Probability densities of $\frac{\Delta_n}{K \log n}$



Simulation Estimates of $P\left\{\Delta_n \geq (1-\varepsilon)\log n\right\}$

– Automated Model Discovery –

Quo Vadis

- **General Approach**

  — data-based model determination - allows the data to choose

  — models evolve over time; PIC'ed by predictive odds criterion

  — has Bayesian, classical, prequential justifications

  — lag length, cointegrating rank, time trends, unit roots all determined automatically & adjusted period by period

  — order estimates all consistent, including cointegrating rank

  — can use in conventional time series tests, e.g. for causal effects

- **Methodology**

  — closer in philosophy to Rissanen (1986, 1987), West and Harrison (1986) & Dawid (1984) than to some common econometric methodologies

  — yields optimised BVAR($\widehat{\psi}$) and RRR($\widehat{r}, \widehat{k}, \widehat{l}$) models

  — finds 'Bayes model' model that is 'closest' to the true dgp and forecasts that are closest to optimal forecasts

- **Practical Experience**

  — ex post forecasting analyses in Phillips (1993, 1995, *J. Econometrics*) for US data and with Nelson & Plosser data;

  — ex ante forecasting experience in *Asia Pacific Economic Review* (1995-1999) for USA, Japan, Korea, Australia and New Zealand

  — comparisons with Fair Model on real GDP growth and inflation

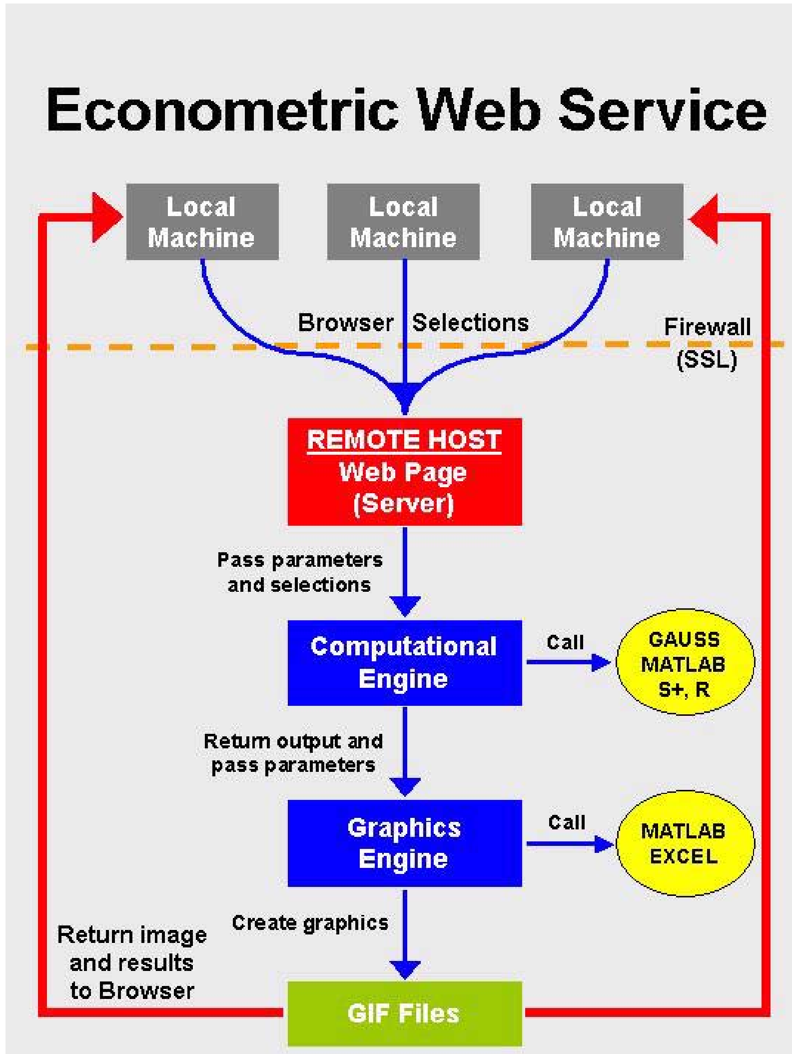
  — application to New Zealand with built-in policy analysis (effects of monetary policy changes and recession in US) Schiff & Phillips (1999, *NZEP*)

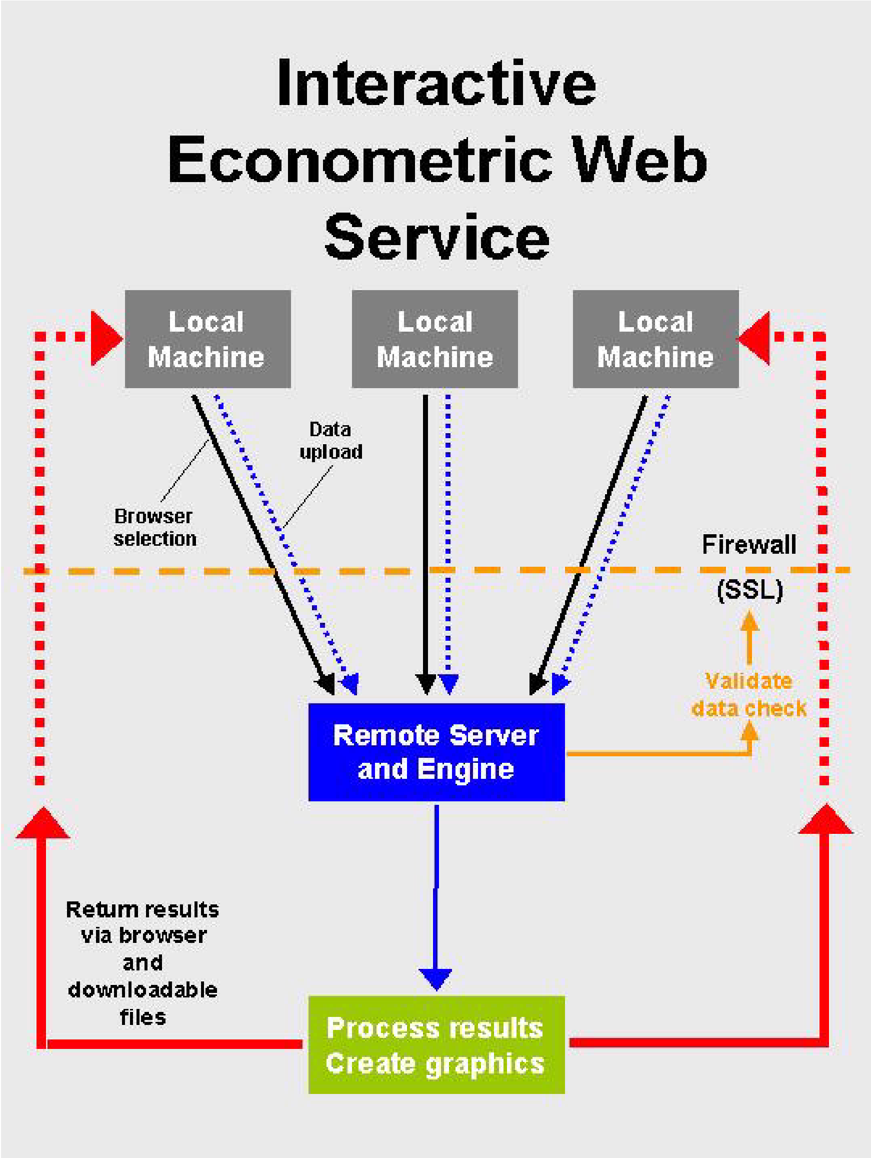
  — Web-based applications in New Zealand on Predicta website: http://covec.co.nz/

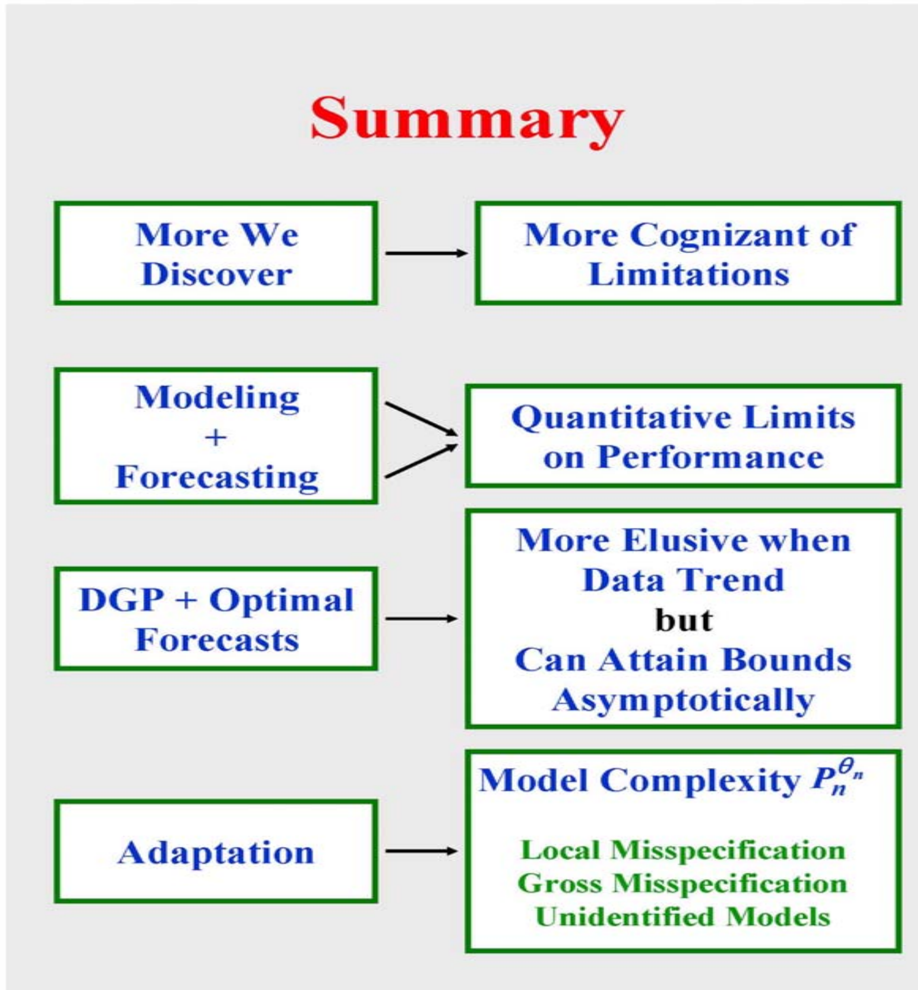
- **A New Research Goal: An Interactive Econometric Web Server**

  – real time econometric data & policy analysis to inform public economic debate

  – point, click, select series for modeling and forecasting & upload data for analysis.


Econometric Web Service
Local Machine
Local Machine
Local Machine
Browser Selections
Firewall (SSL)
REMOTE HOST
Web Page (Server)
Pass parameters and selections
Computational Engine
Call
GAUSS MATLAB S+, R
Return output and pass parameters
Graphics Engine
Call
MATLAB EXCEL
Create graphics
Return image and results to Browser
GIF Files


Interactive Econometric Web Service
Local Machine
Local Machine
Local Machine
Data upload
Browser selection
Firewall
(SSL)
Validate data check
Remote Server and Engine
Return results via browser and downloadable files
Process results Create graphics


Summary
More We Discover
More Cognizant of Limitations
Modeling + Forecasting
Quantitative Limits on Performance
DGP + Optimal Forecasts
More Elusive when Data Trend but Can Attain Bounds Asymptotically
Adaptation
Model Complexity $P_n^{\theta_n}$
Local Misspecification
Gross Misspecification
Unidentified Models


Sometimes
Total Model Failure
Compensation in Statistical
Properties & Asymptotics
e.g., Trend (Mis)Specification
Adjustments to Get Model on Track