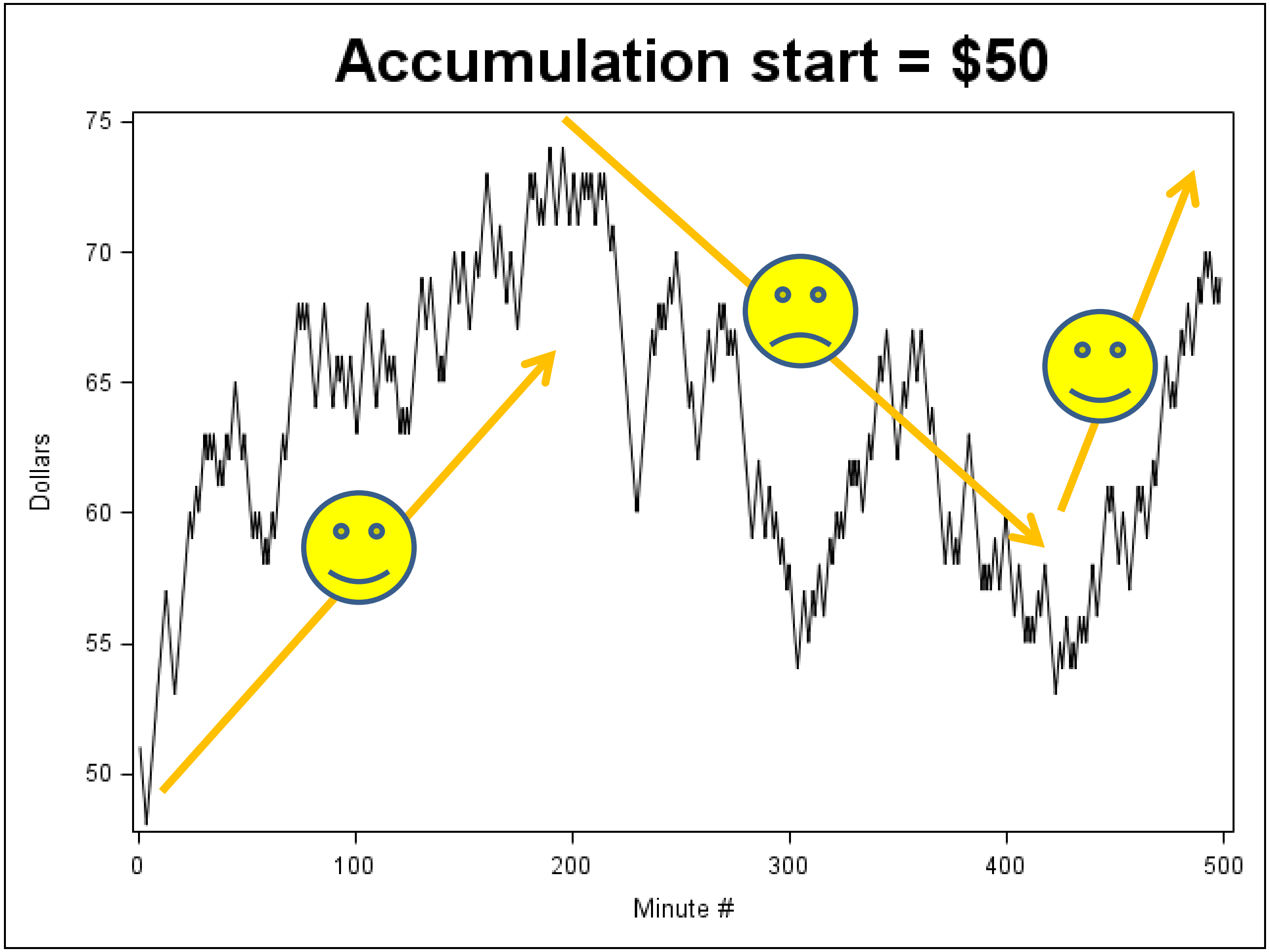
Unit Root Testing: History and Modern Applications

IAA September 2023

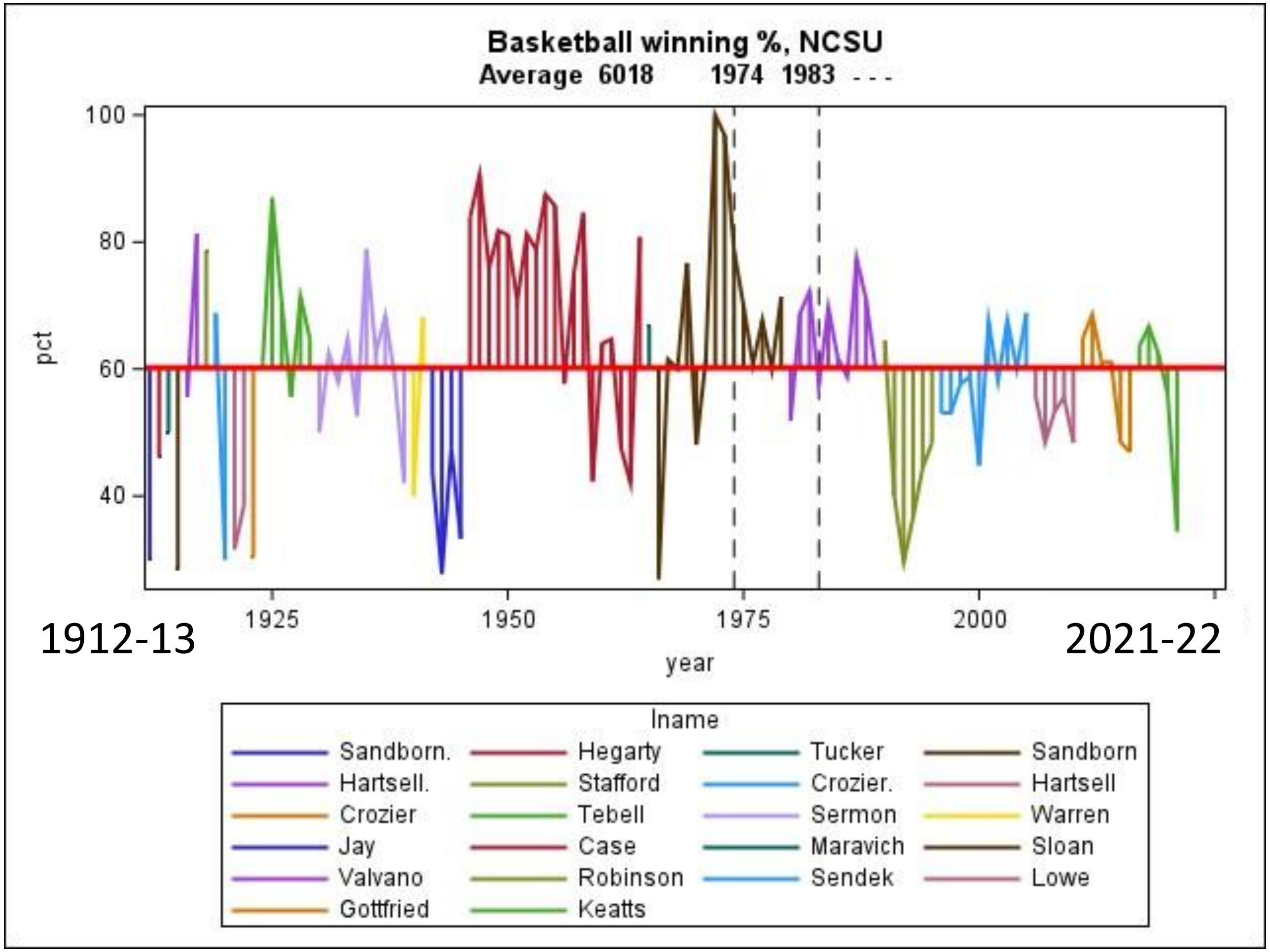
David A. Dickey – NC State University

Thanks to my sponsors:
State Retirement System & Social Security Administration

Stock price per minute for Unit Roots Inc.*

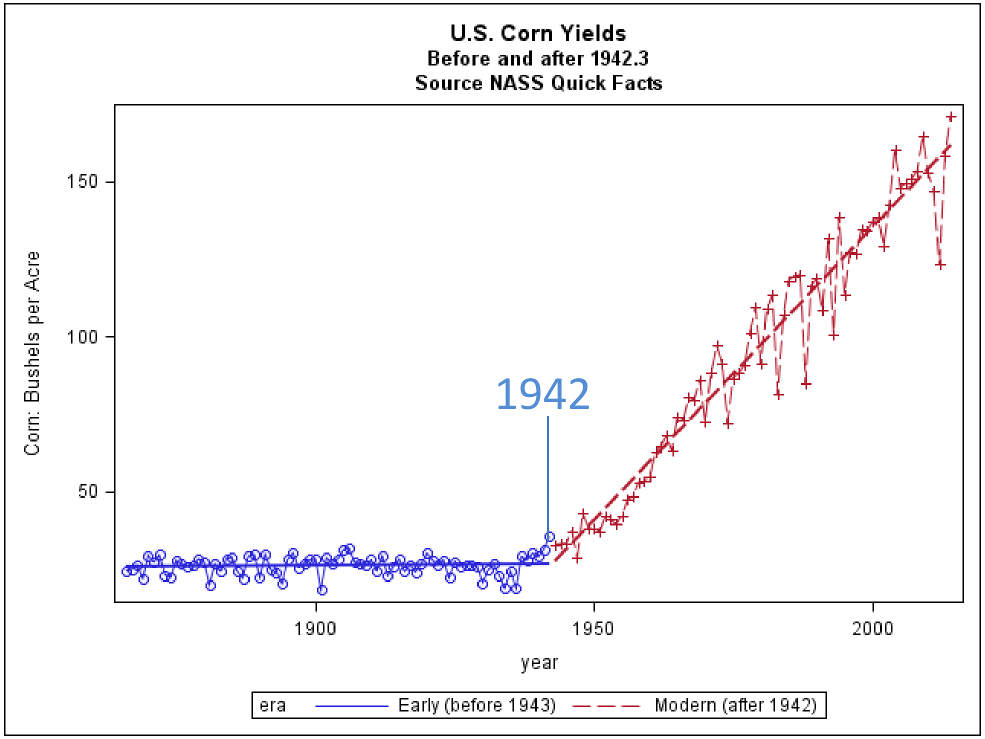


Winning Percentages



Corn Yields (BPA) Natl. Ag. Stat. Serv.

2014

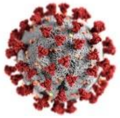


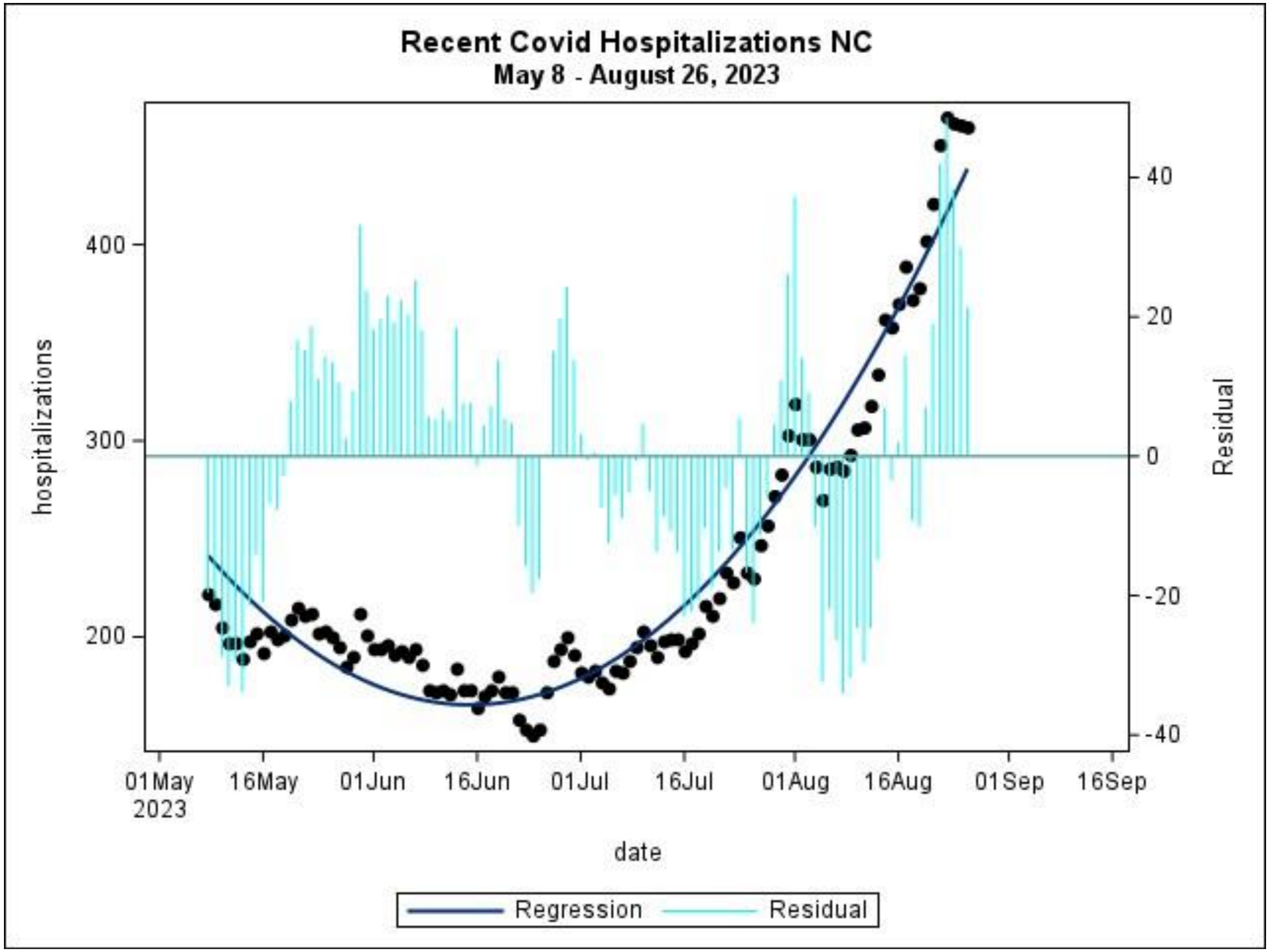
1866



Covid Hospitalizations

May 8, 2023– August 26 2023





Basic Model:

$$Y_t - f(t) = \rho(Y_{t-1} - f(t-1)) + e_t$$

$$Y_t - \mu \quad = \rho(Y_{t-1} - \mu) + e_t$$

$$H_0: \rho = 1$$

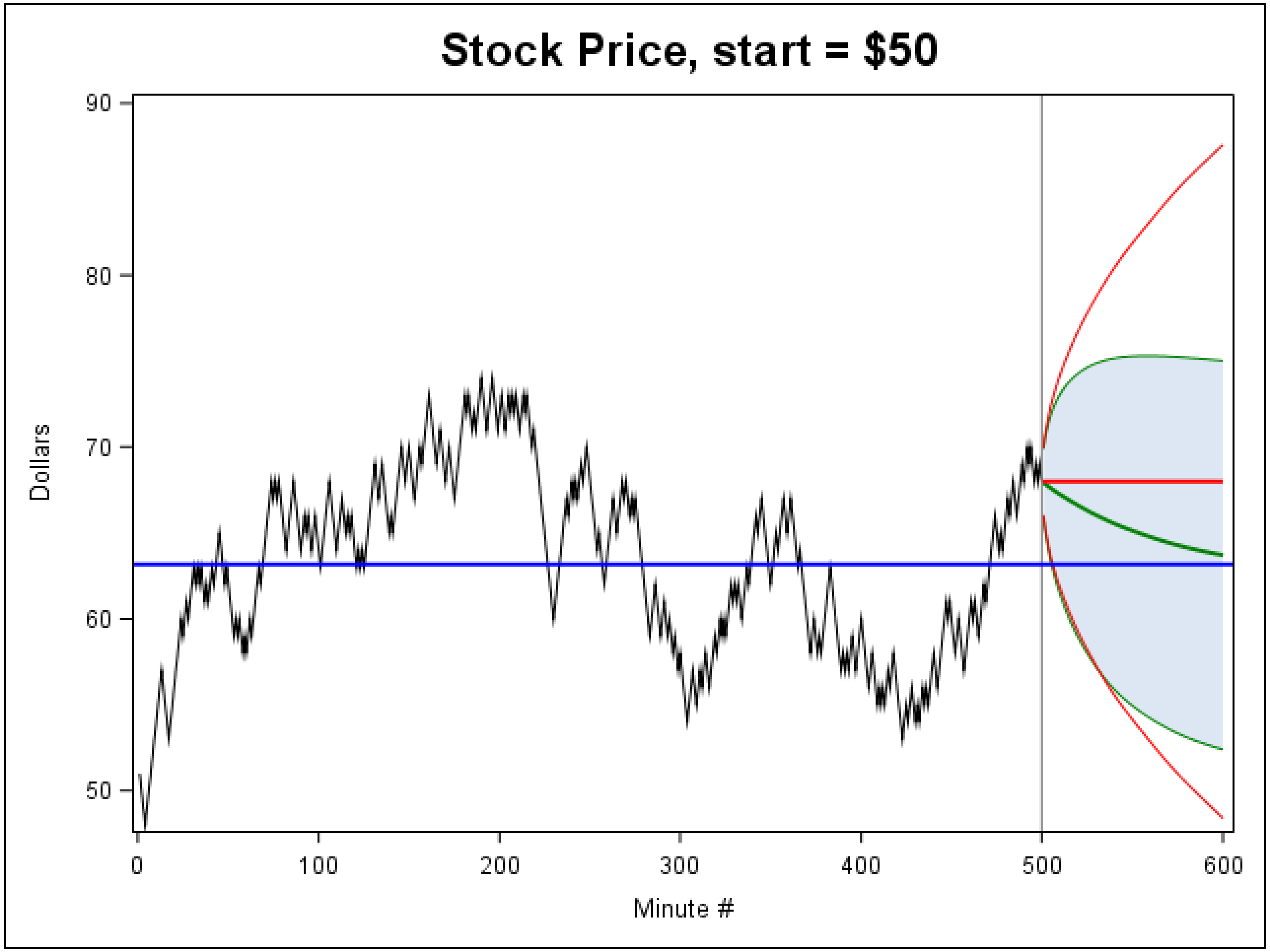
$$Y_t - \cancel{\mu} = Y_{t-1} - \cancel{\mu} + e_t$$

then $Y_t = Y_{t-1} + e_t$ (random walk)

Two forecasts:

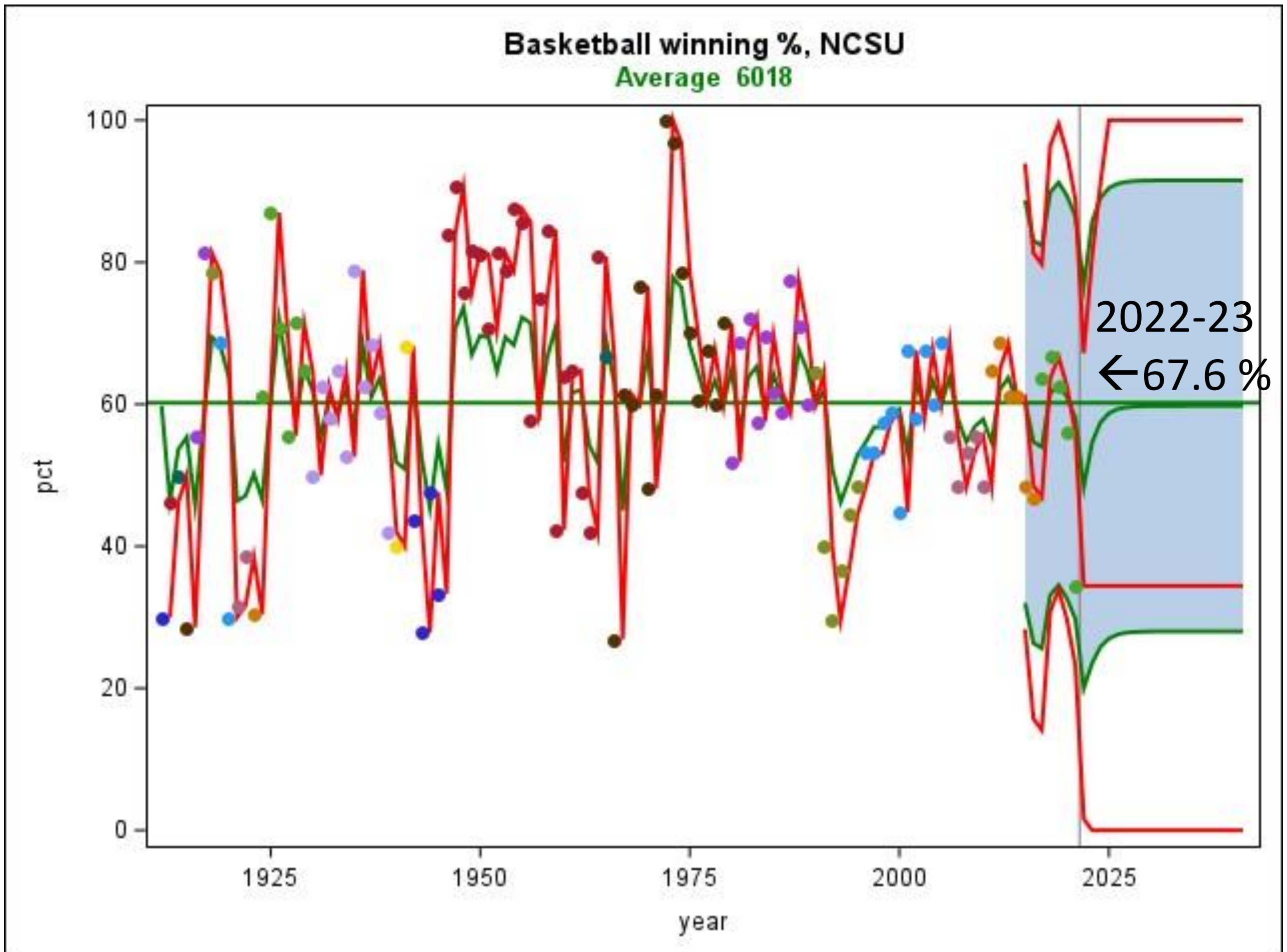
ρ estimated (0.9778)

$\rho = 1$



Winning  Proportions

Random Walk (ρ=1) and AR(1) (ρ=0.45)



~~Where are my clothes?~~

Basic Model, $f(t) = \mu$

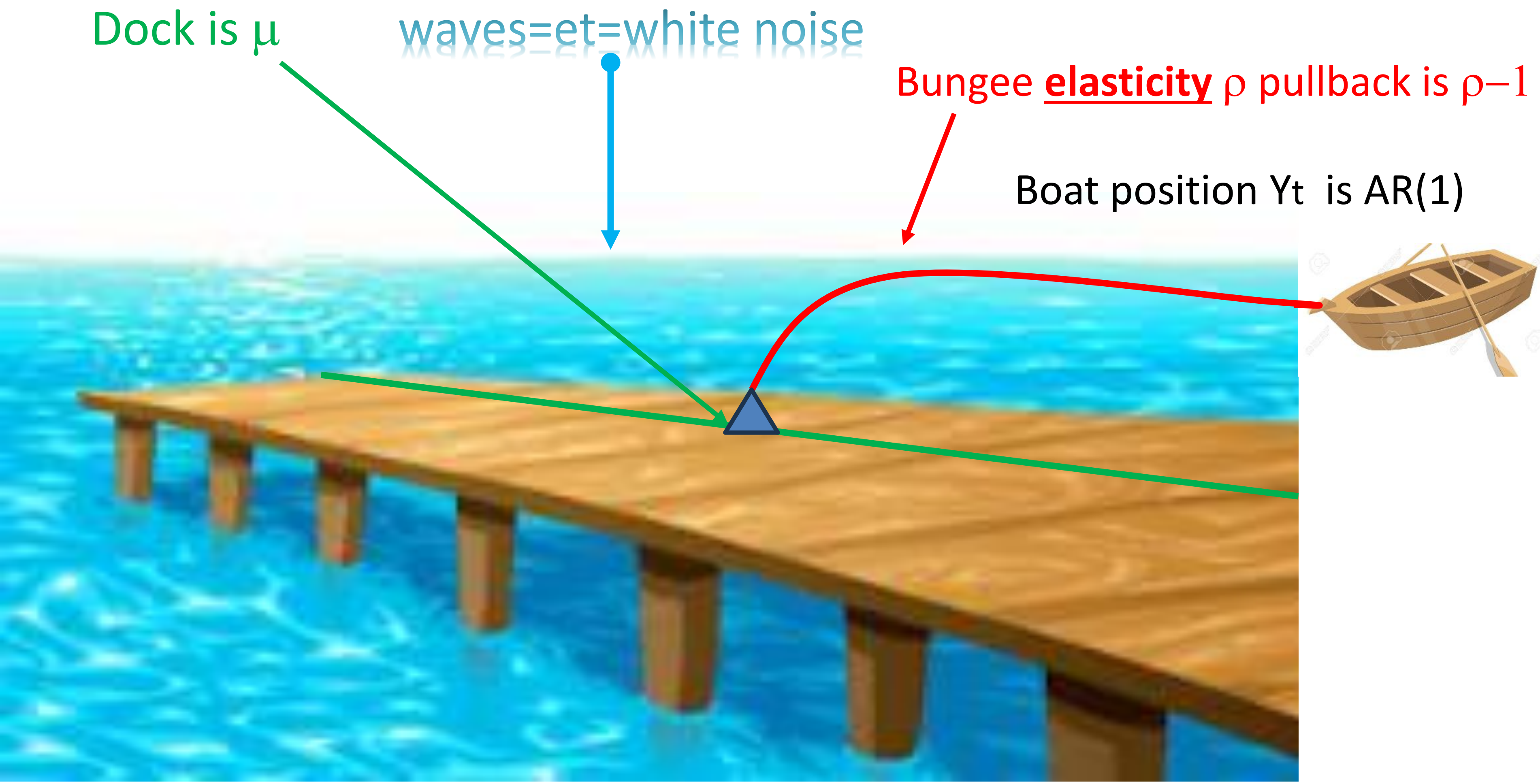
$$Y_t - \mu = \rho(Y_{t-1} - \mu) + e_t$$
$$-(Y_{t-1} - \mu) \quad - (Y_{t-1} - \mu)$$

$$\underbrace{Y_t - Y_{t-1}}_{\nabla Y_t} = (\rho - 1)(Y_{t-1} - \mu) + e_t$$

$$H_0: \rho - 1 = 0$$

Regress ∇Y_t on 1, $\underset{\text{t-stat}}{Y_{t-1}}$

$$Y_t - Y_{t-1} = (\rho - 1)(Y_{t-1} - \mu) + e_t$$



Basic Model:

$$Y_t - f(t) = \rho(Y_{t-1} - f(t-1)) + e_t$$

$$Y_t - \mu \quad = \rho(Y_{t-1} - \mu) + e_t$$

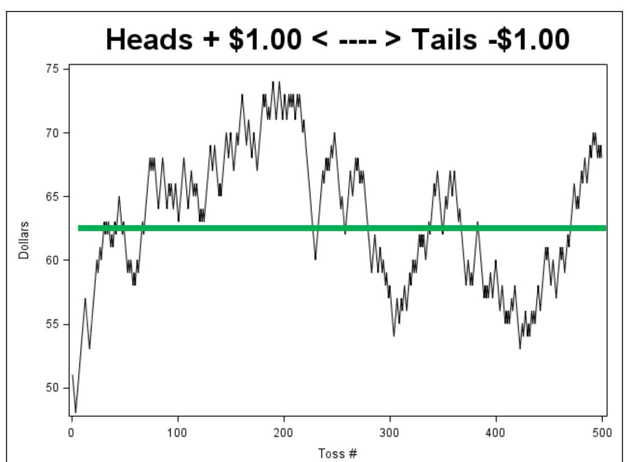
$$H_0: \rho = 1$$

$$Y_t - \cancel{\mu} = Y_{t-1} - \cancel{\mu} + e_t$$

then $Y_t = Y_{t-1} + e_t$ (random walk)

+ $1

- $1

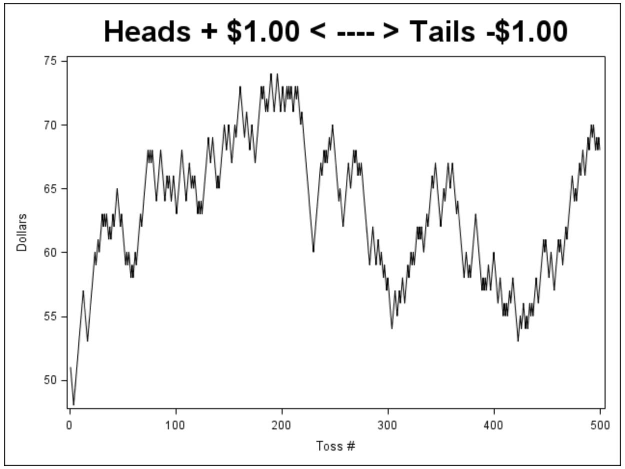


Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.32618	0.11908	2.74	0.0064
Lsum	1	-0.02219	0.00839	**-2.64**	**0.0084**

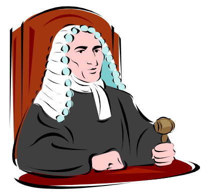
Reject H_0: $\rho=1$

But we KNOW that $\rho=1$????



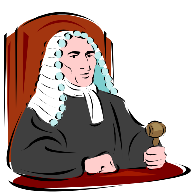
H_0:
H_1:

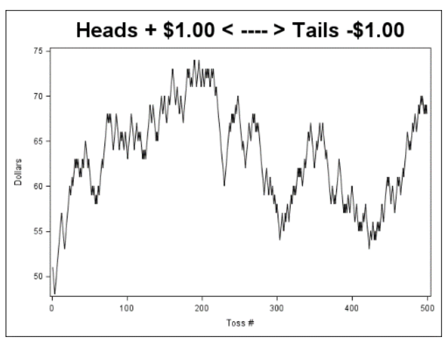
Testing a hypothesis



H_0: Innocence
H_1: Guilt

Beyond reasonable doubt
$P<0.05$





H_0: $\rho=1$
(random walk)
H_1: $|\rho|<1$
(stationary)

Truth: random walk.
(not “stationary”)

Conclusion: ($p=0.0084$) -> stationary

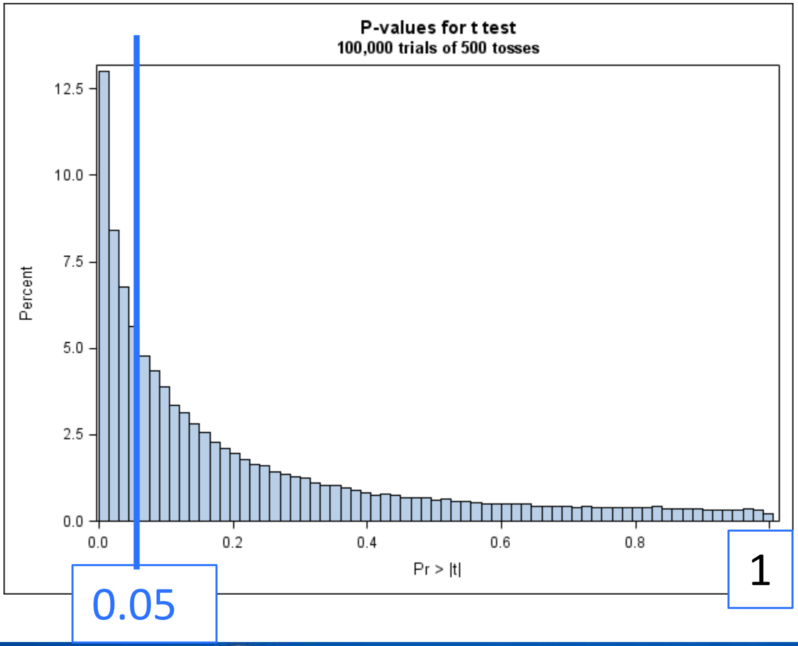
<u>Mistrial?</u>

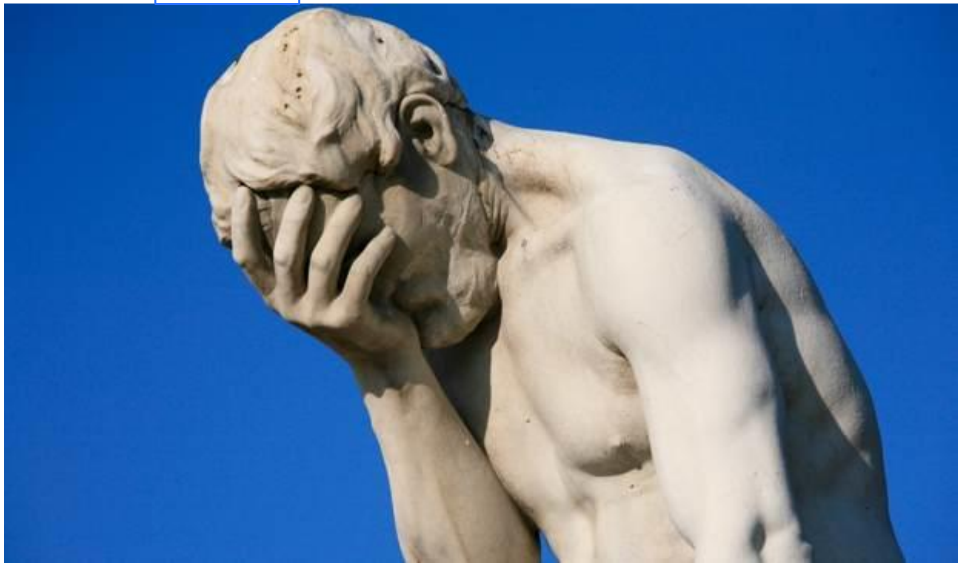
100,000 coin flip trials
of 500 tosses each

Reject H_0: $\rho=1$ in 30%

But we KNOW that $\rho=1$
(nominal $\alpha=0.05$)

100,000 p-values →





What's Wrong?

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + e_t$$

$$Y_t = \mu(1-\rho) + \rho Y_{t-1} + e_t$$

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + e_t$$

Linear model – yes!

$e_t \sim (0,\sigma^2)$ independently – yes!

Regressors fixed and known

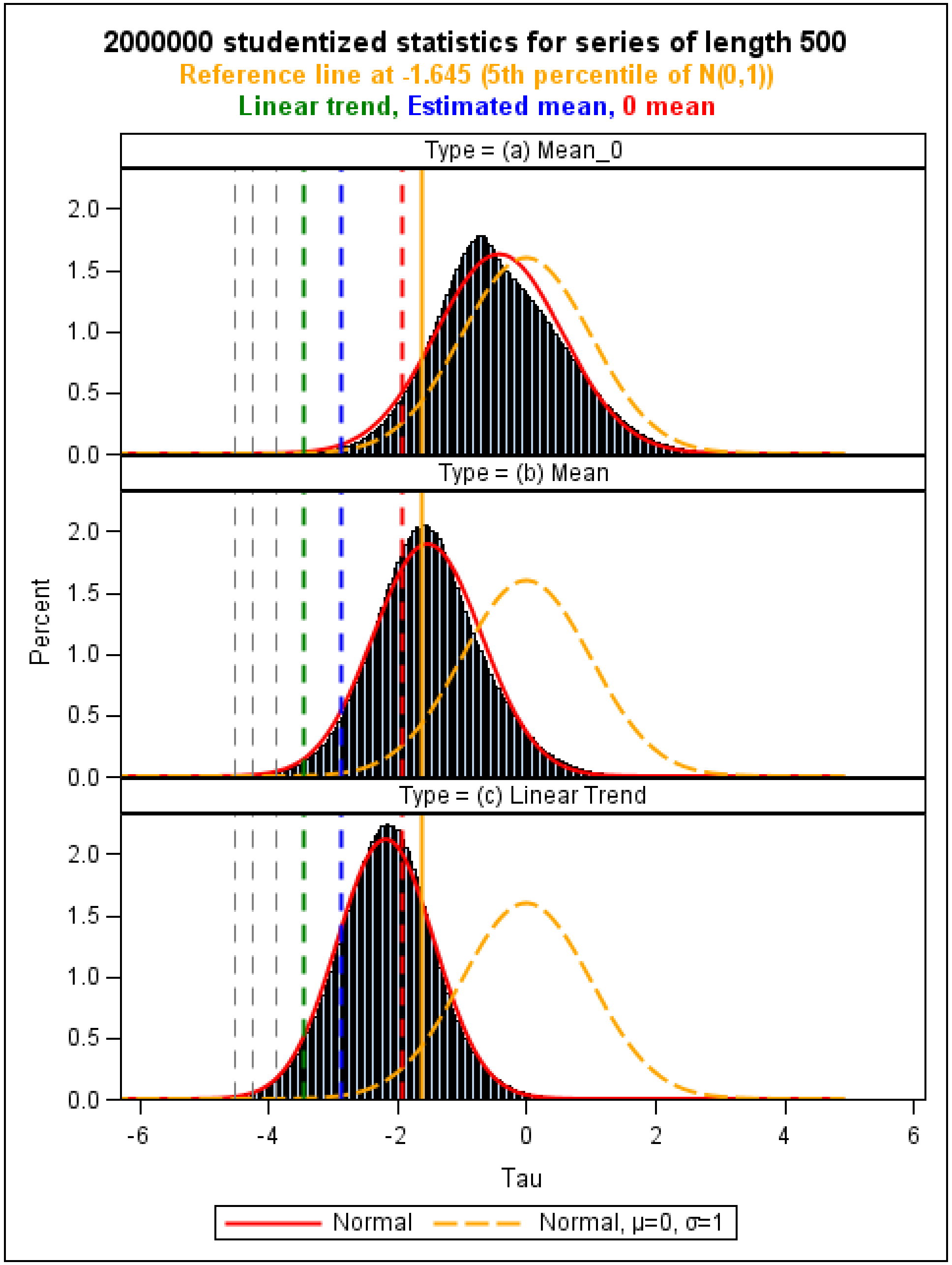
“I confess that in 1901
I said to my brother
Orville that man would
not fly for fifty years.”
(first flight Dec. 1903)

Wilbur
Wright

Lesson 1:

Jim Valvano NCSU basketball coach

Let's Take A Look



f(t)

Mean=0
← -1.95

Mean Estimated
← -2.86

Linear Trend
← -3.41

Lesson 1: Check your assumptions

```
PROC REG Data=cointoss;
  MODEL difference = Lsum;
```

COIN TOSS

		Parameter	Standard		
Variable	DF	Estimate	Error	t Value	Pr>|t|
Intercept	1	0.32618	0.11908	2.74	0.0064
Lsum	1	-0.02219	0.00839	**-2.64**	~~0.0084~~ **??**

Mean
Estimated
← -2.86

Commercial Products include SAS™ and others

```
PROC ARIMA Data=cointoss;
Identify var=sum stationarity=(Dickey=(0));
```

Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr<Rho	Tau	Pr<Tau
Zero Mean	0	-0.4367	0.5838	-0.28	0.5861
Single Mean	0	-11.0708	0.1030	**-2.64**	**0.0852 !!**
Trend	0	-11.8969	0.3173	-2.80	0.1996

™ SAS is the registered trademark of SAS Institute Cary NC

Mathematical Results:

Regress $Y_t - Y_{t-1}$ on Y_{t-1}

where $Y_t = \color{red}{1}\color{black}{Y_{t-1}} + e_t$

$$\boxed{n(\hat{\rho}-1) = \frac{\frac{1}{n}\sum Y_{t-1}e_t}{\frac{1}{n^2}\sum Y_{t-1}^2} = \frac{N}{D} \rightarrow \frac{\left(\sum_0^\infty \gamma_i Z_i\right)^2 - 1/2}{\sum_0^\infty \gamma_i^2 Z_i^2} = \frac{\int B(t)dB(t)}{\int B^2(t)dt}}$$

$$\gamma_i = \frac{2(-1)^{i+1}}{(2i-1)\pi}$$

B(t): Brownian Motion on [0,1]

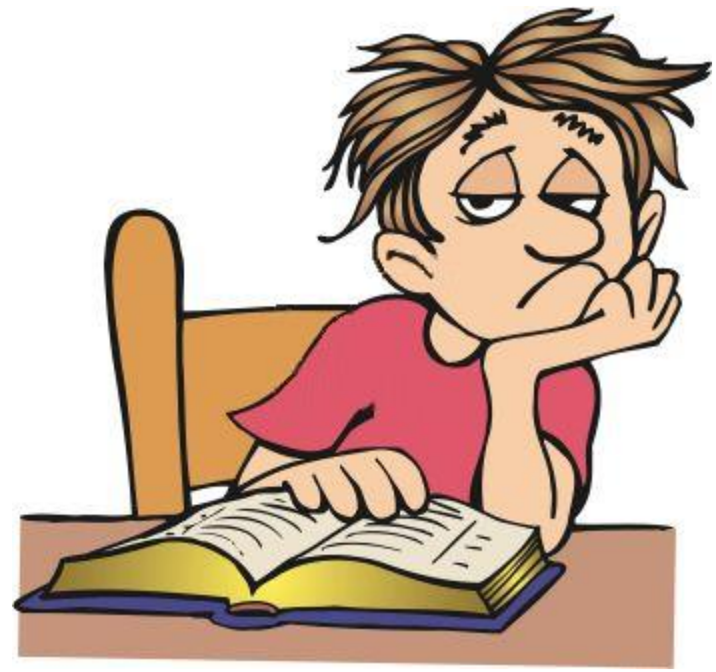
$Z_i \sim N(0,1)$, independently

WARNING:

Algebra like this has been known to cause headache, nausea, and lack of sleep, especially in MSA students.

Interesting formula!

Got more?



Part 2, the studentized test statistic (τ)

Regress $Y_t - Y_{t-1}$ on Y_{t-1}

$Y_t = 1Y_{t-1} + e_t$

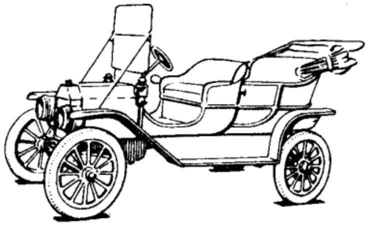
Mathematical Results:

$$\tau = \frac{\hat{\rho} - 1}{\sqrt{s^2 / \sum Y_{t-1}^2}} \Rightarrow \frac{\left(\sum_0^\infty \gamma_i Z_i\right)^2 - 1/2}{\sqrt{\sum_0^\infty \gamma_i^2 Z_i^2}} = \frac{\int B(t)\,dB(t)}{\sqrt{\int B^2(t)\,dt}}$$

Lesson 2:
Attitude matters

Whether you think you can, or you think you can't--you're right.

Henry Ford



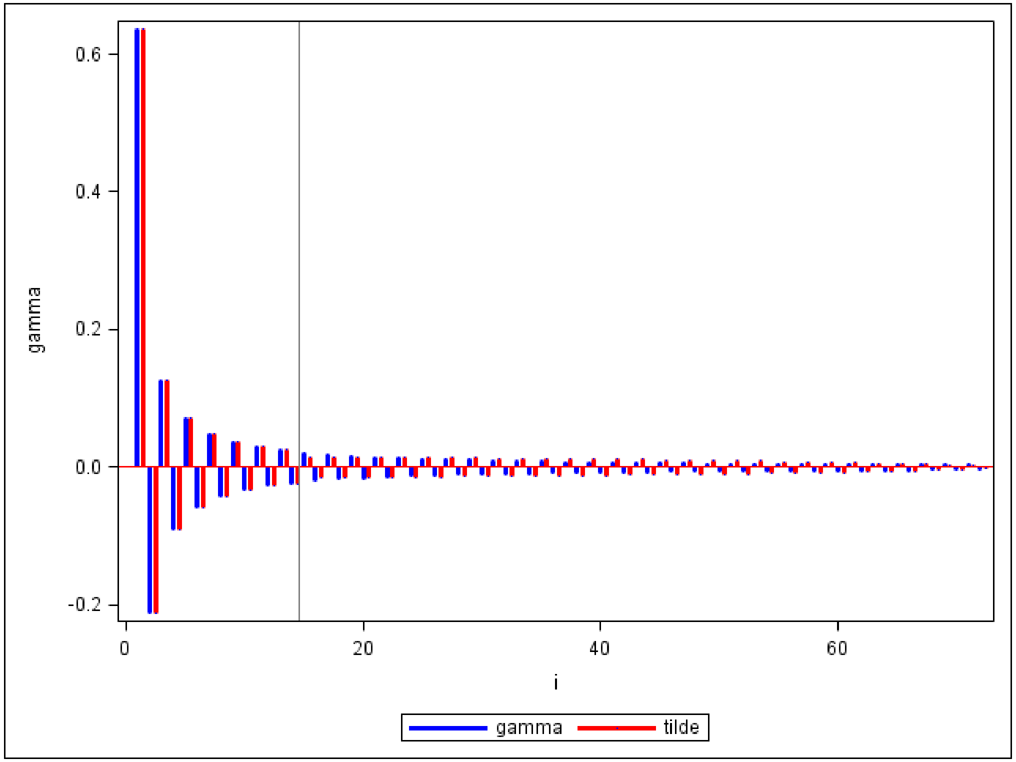
Numerator: $\left(\sum_0^\infty \gamma_i Z_i\right)^2 - 1/2 \approx \left(\sum_0^{72} \tilde{\gamma}_i Z_i\right)^2 - 1/2$

Denominator: $\sum_0^\infty \gamma_i^2 Z_i^2 \approx \sum_0^{72} \tilde{\gamma}_i^2 Z_i^2$

$$(1) \quad \tilde{\gamma}_i = \gamma_i = \frac{(-1)^{i+1} 2}{(2i-1)\pi}, i = 1,2,\cdots,14$$

$$(2) \quad \sum_0^\infty \gamma_i^2 = \sum_0^{72} \tilde{\gamma}_i^2$$

Lesson 3 :
Knowing a little
math can help



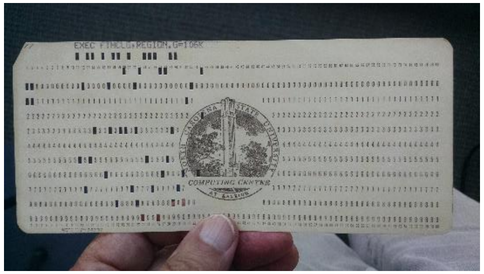
Winning idea:

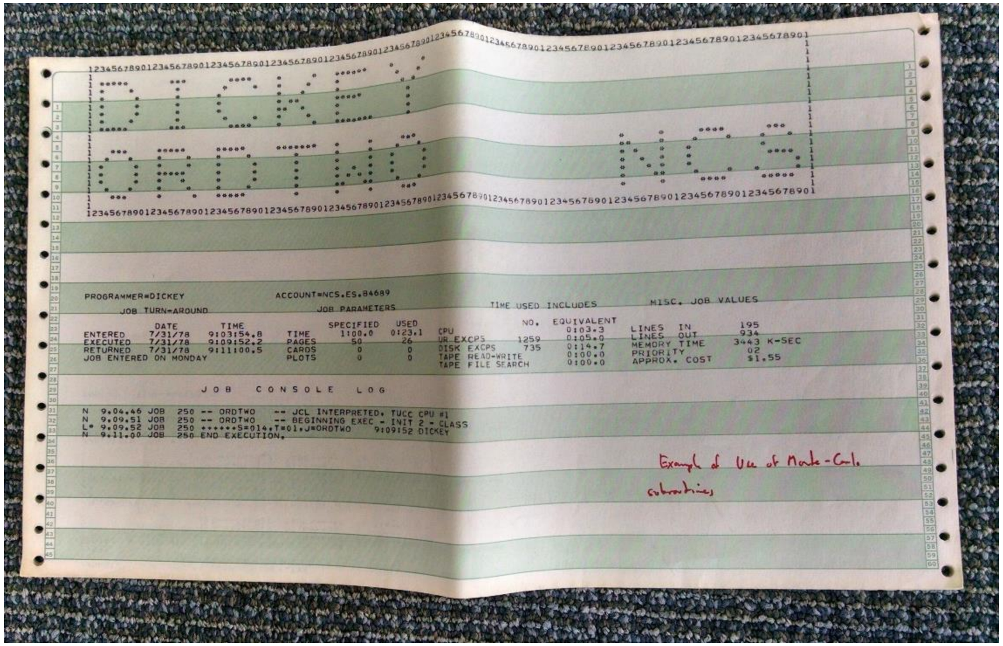
(1) Large simulations for finite sample size
(2) Simulate from limit approximation

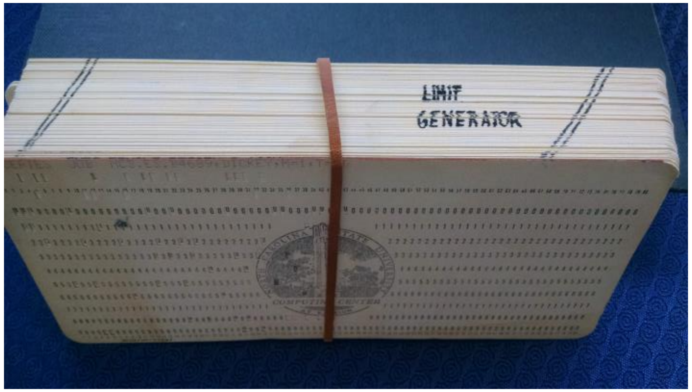
$$\boxed{\tau \to \frac{\left(\sum_0^m \tilde{\gamma}_i Z_i\right)^2 - 1/2}{\sqrt{\sum_0^m \tilde{\gamma}_i^2 Z_i^2}}}$$

Jimmy V
@NCSU

Wright glider Oct. 24, 1902


EXEC FTHCLG,REGION.G=106K
COMPUTING CENTRE


DICKEY
ORDTWO
NCS
PROGRAMMER=DICKEY
ACCOUNT=NCS.ES.84689
JOB TURN-AROUND
JOB PARAMETERS
TIME USED INCLUDES
MISC. JOB VALUES
DATE TIME SPECIFIED USED NO. EQUIVALENT
ENTERED 7/31/78 9:03:54.8 TIME 1:00.0 0:23.1 CPU 0:03.3 LINES IN 195
EXECUTED 7/31/78 9:09:52.2 PAGES 50 26 UR EXCPS 1259 0:05.0 LINES OUT 934
RETURNED 7/31/78 9:11:00.5 CARDS 0 0 DISK EXCPS 735 0:14.7 MEMORY TIME 3443 K-SEC
JOB ENTERED ON MONDAY PLOTS 0 0 TAPE READ-WRITE 0:00.0 PRIORITY 02
TAPE FILE SEARCH 0:00.0 APPROX. COST $1.55
J O B C O N S O L E L O G
N 9.04.46 JOB 250 -- ORDTWO -- JCL INTERPRETED. TUCC CPU #1
N 9.09.51 JOB 250 -- ORDTWO -- BEGINNING EXEC - INIT 2 - CLASS
L* 9.09.52 JOB 250 ******S=014,T=01,J=ORDTWO 9:09:52 DICKEY
N 9.11.00 JOB 250 END EXECUTION.
Example of Use of Monte-Carlo
subroutines


LIMIT
GENERATOR

n	P1	P5	P10	P25	P50	mean	P75	P90	P95	P99
					f(t)=0					
25	-2.67	-1.96	-1.61	-1.07	-0.47	-0.40	0.25	0.93	1.34	2.14
50	-2.61	-1.94	-1.61	-1.08	-0.49	-0.41	0.23	0.91	1.31	2.08
100	-2.59	-1.94	-1.62	-1.09	-0.49	-0.42	0.23	0.90	1.30	2.04
250	-2.58	-1.94	-1.62	-1.09	-0.50	-0.42	0.22	0.89	1.29	2.02
500	-2.57	-1.94	-1.61	-1.09	-0.50	-0.42	0.22	0.89	1.29	2.02
1000	-2.57	-1.94	-1.62	-1.09	-0.50	-0.42	0.22	0.89	1.28	2.02
					f(t)=μ					
25	-3.74	-2.99	-2.64	-2.09	-1.53	-1.52	-0.96	-0.37	0.00	0.72
50	-3.57	-2.92	-2.60	-2.09	-1.55	-1.53	-0.99	-0.40	-0.04	0.66
100	-3.50	-2.89	-2.58	-2.09	-1.56	-1.53	-1.00	-0.42	-0.06	0.64
250	-3.46	-2.87	-2.57	-2.09	-1.56	-1.53	-1.01	-0.43	-0.07	0.61
500	-3.44	-2.87	-2.57	-2.09	-1.56	-1.53	-1.01	-0.44	-0.08	0.61
1000	-3.44	-2.87	-2.57	-2.09	-1.57	-1.53	-1.01	-0.44	-0.08	0.61
					f(t)=α+βt					
25	-4.40	-3.61	-3.24	-2.69	-2.14	-2.17	-1.63	-1.14	-0.82	-0.17
50	-4.16	-3.51	-3.18	-2.68	-2.16	-2.18	-1.67	-1.20	-0.88	-0.25
100	-4.05	-3.46	-3.15	-2.67	-2.17	-2.18	-1.69	-1.22	-0.91	-0.29
250	-4.00	-3.43	-3.14	-2.67	-2.18	-2.18	-1.70	-1.23	-0.93	-0.31
500	-3.98	-3.42	-3.13	-2.67	-2.18	-2.18	-1.70	-1.24	-0.94	-0.32
1000	-3.97	-3.41	-3.13	-2.67	-2.18	-2.18	-1.70	-1.24	-0.94	-0.32

(2 million runs each. n=1000 took < 3 min. in SAS)

Winning   Percentages

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Tau	Pr < Tau
Zero Mean	0	-1.39	0.1519
	1	-1.11	0.2408
	2	-0.77	0.3798
Single Mean	0	-6.50	<.0001
	1	-5.64	<.0001
	2	-4.33	0.0007
Trend	0	-6.46	<.0001
	1	-5.61	<.0001
	2	-4.32	0.0044

Reject
H0: Unit Root

Definition: Power

The probability of rejecting a null hypothesis when you should (Pr{reject})

Modified Definition:

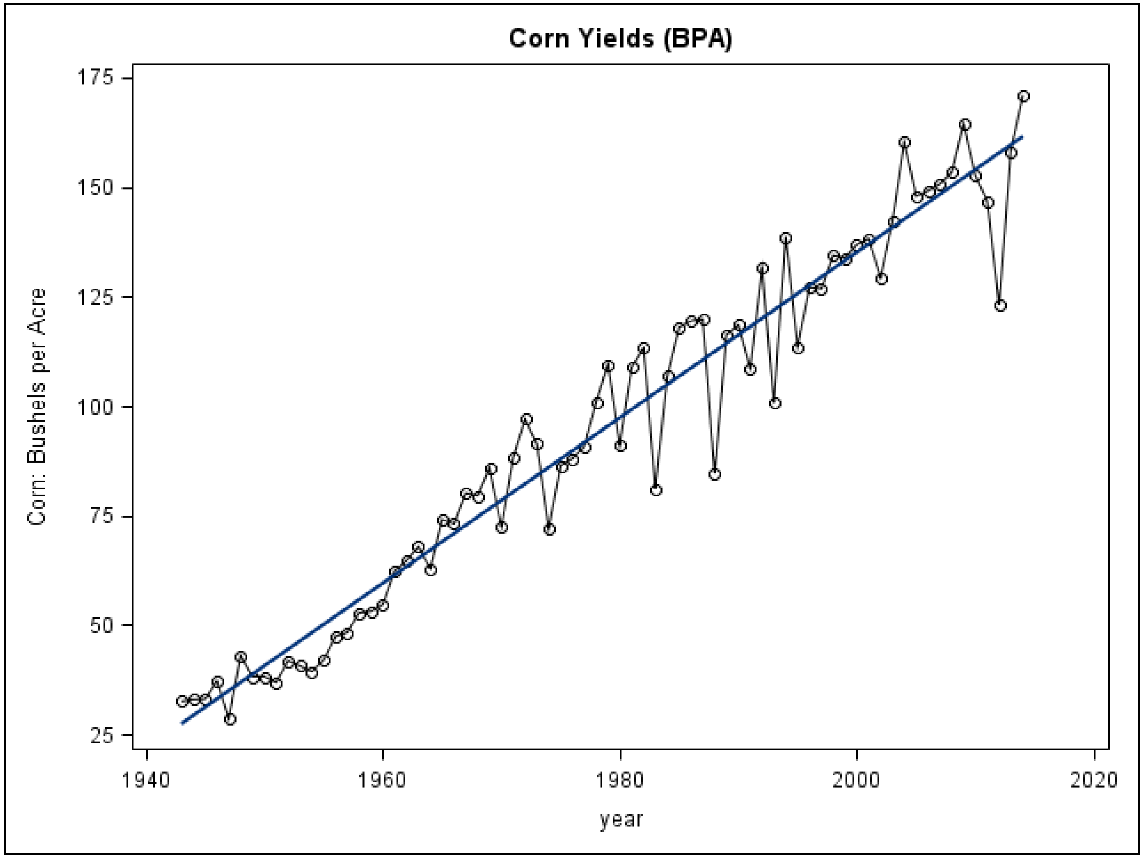
Parzen Power

Pr{reject}*Pr{use}

Dr. Emanuel Parzen, Distinguished Professor Emeritus of Statistics, Texas A&M University (1929 – 2016)

Basic Model 2: $f(t) = \alpha+\beta t$

$$Y_t - (\alpha+\beta t) = \rho(Y_{t-1} - (\alpha+\beta(t-1))) + e_t$$

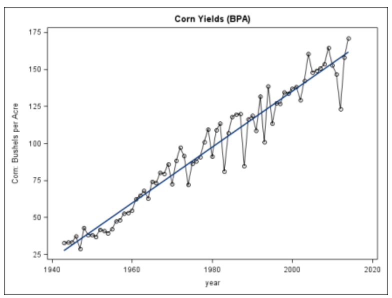


(1943 - 2014)

$$Y_t - (\alpha+\beta t) = \rho(Y_{t-1} - (\alpha+\beta(t-1))) + e_t$$
$$-(Y_{t-1} - (\alpha+\beta(t-1))) \qquad -(Y_{t-1} - (\alpha+\beta(t-1)))$$
$$Y_t - Y_{t-1} - \beta = (\rho-1)(Y_{t-1} - (\alpha+\beta t - \beta)) + e_t$$

$$Y_t - Y_{t-1} = \underline{(\beta - (\rho-1)(\alpha-\beta))} + \underline{(\rho-1)\beta}\, t + (\rho-1)Y_{t-1} + e_t$$

$$\beta_0 \qquad\qquad \beta_1 t$$



H_0: $\rho-1=0$

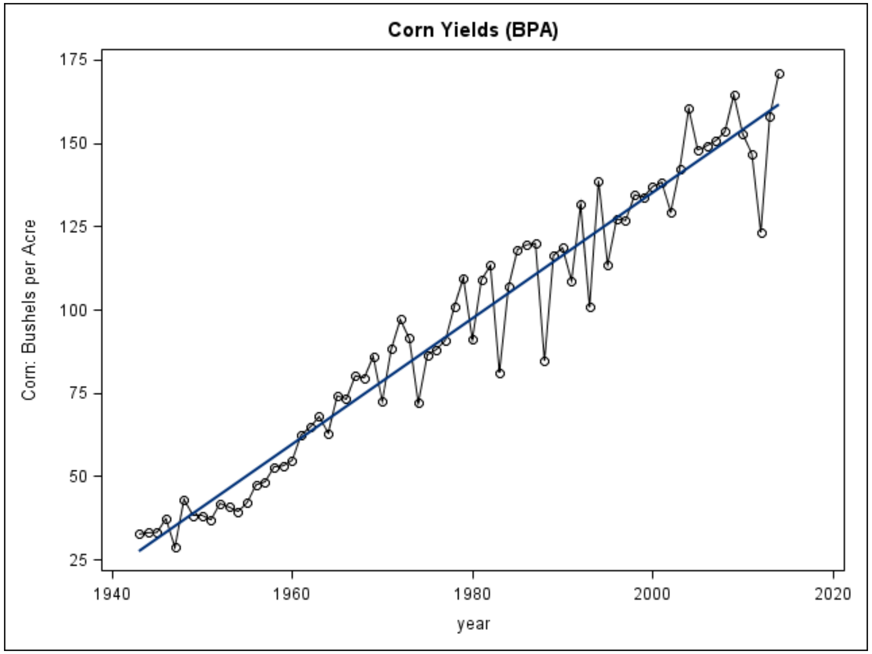
Regress $Y_t - Y_{t-1}$ on 1, t, Y_{t-1}

t-stat

$$Y_t - (\alpha + \beta t) \quad = \quad \rho(Y_{t-1} - (\alpha + \beta(t-1))) + e_t$$

If $H_0: \rho = 1$ is true then $Y_t = Y_{t-1} + \beta + e_t$

random walk with drift β



Corn yields, after 1942

Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr<Rho	Tau	Pr<Tau
Zero Mean	0	0.7047	0.8515	0.61	0.8454
Single Mean	0	-3.6170	0.5739	-1.23	0.6559
Trend	0	-74.8033	0.0002	**-8.65**	**<.0001**

Autocorrelation Check of Residuals
(from linear trend plus white noise)

To Lag	Chi-Square	DF	Pr > ChiSq	------------Autocorrelations-----------					
6	4.27	6	**0.6403**	-0.053	-0.042	-0.060	-0.022	0.204	-0.061
12	10.82	12	**0.5445**	-0.017	-0.087	-0.053	0.126	-0.205	-0.083
18	15.54	18	**0.6247**	0.029	0.020	-0.019	-0.200	0.041	-0.079
24	22.43	24	**0.5535**	0.188	-0.076	-0.109	0.006	0.076	0.090

Lesson 4: Just because a series is trending up does not mean you need to difference!

Lesson 4: Just because a series is trending up <u>does not</u> mean you need to difference!

Powered flight
Kill Devil Hills NC, Dec. 1903

If we all worked on the assumption that what is accepted as true is really true, there would be little hope of advance. (Orville Wright)

Corn yields, before 1943

Dickey-Fuller Unit Root Tests

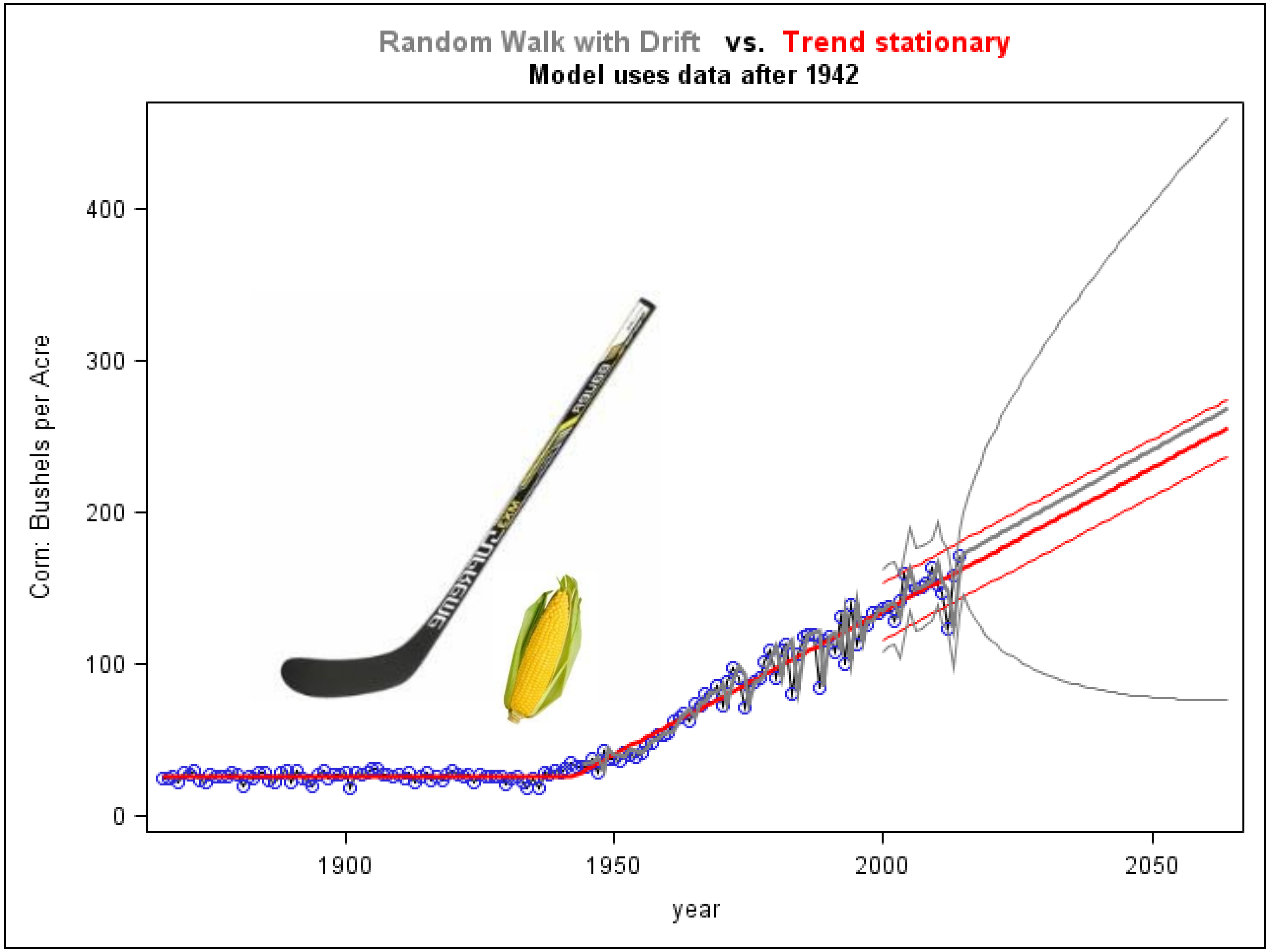
Type	Lags	Rho	Pr < Rho	Tau	Pr<Tau
Zero Mean	0	-0.4836	0.5712	-0.35	0.5569
Single Mean	0	-67.4004	0.0007	**-7.26**	**0.0001**
Trend	0	-67.4900	0.0002	**-7.23**	**<.0001**

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
MU	**26.17532**	0.36683	71.35	<.0001

Autocorrelation Check of Residuals

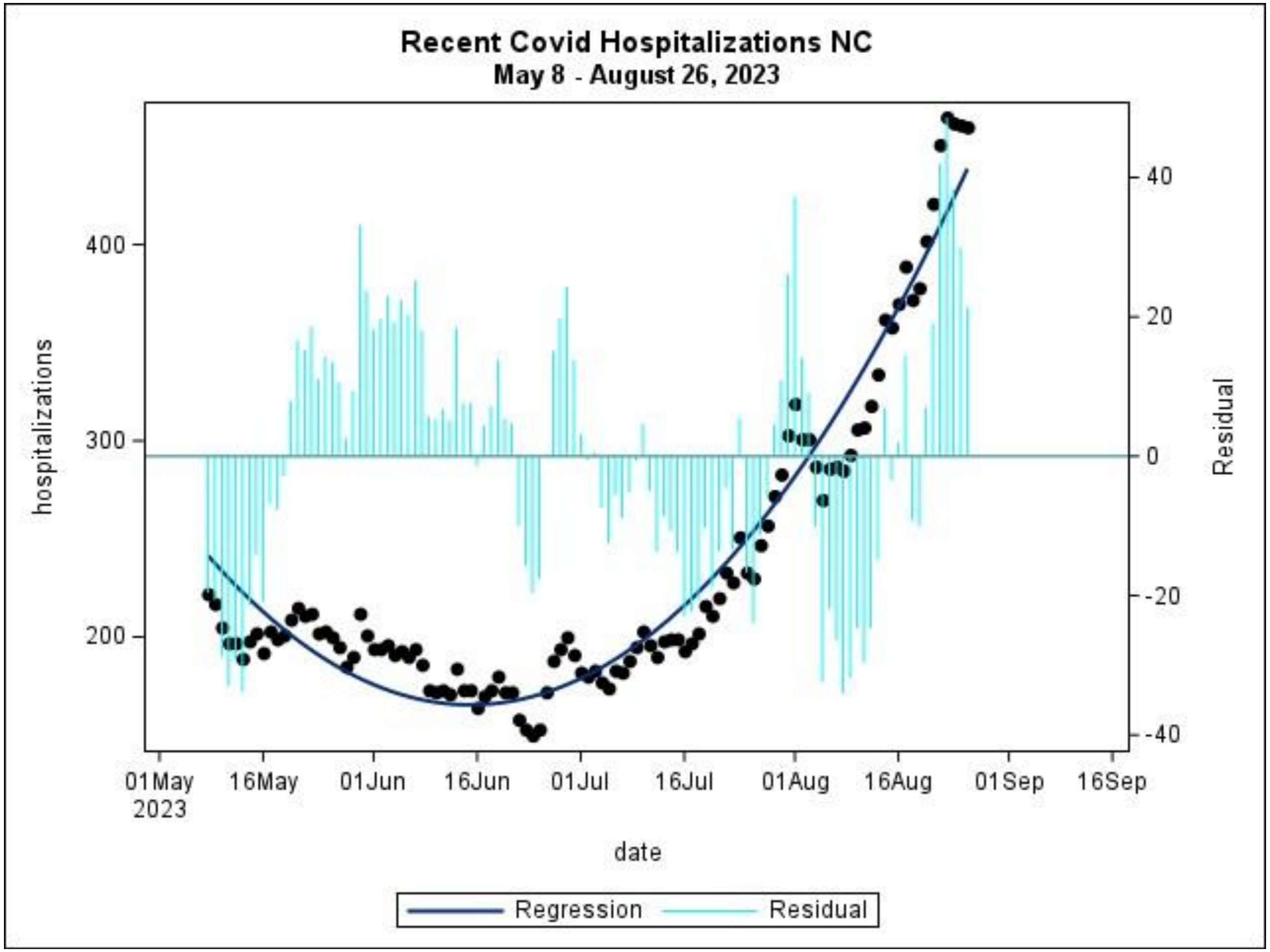
To Lag	Chi-Square	DF	Pr > ChiSq	------------	------------	Autocorrelations	------------	------------	------------
6	3.25	6	**0.7764**	0.100	0.070	0.034	0.029	-0.111	-0.098
12	7.07	12	**0.8532**	0.037	-0.111	-0.096	-0.067	-0.103	-0.064
18	8.76	18	**0.9651**	-0.033	0.014	-0.049	-0.018	0.102	-0.048
24	14.92	24	**0.9232**	-0.019	0.120	0.105	-0.134	0.041	-0.103


Random Walk with Drift vs. Trend stationary
Model uses data after 1942
Corn: Bushels per Acre
0
100
200
300
400
1900
1950
2000
2050
year

Covid Hospitalizations

May 8, 2023– August 26 2023

Quadratic residuals stationary??



```
proc arima data=covid;
identify var=Hospitalizations
stationarity=(dickey);
```

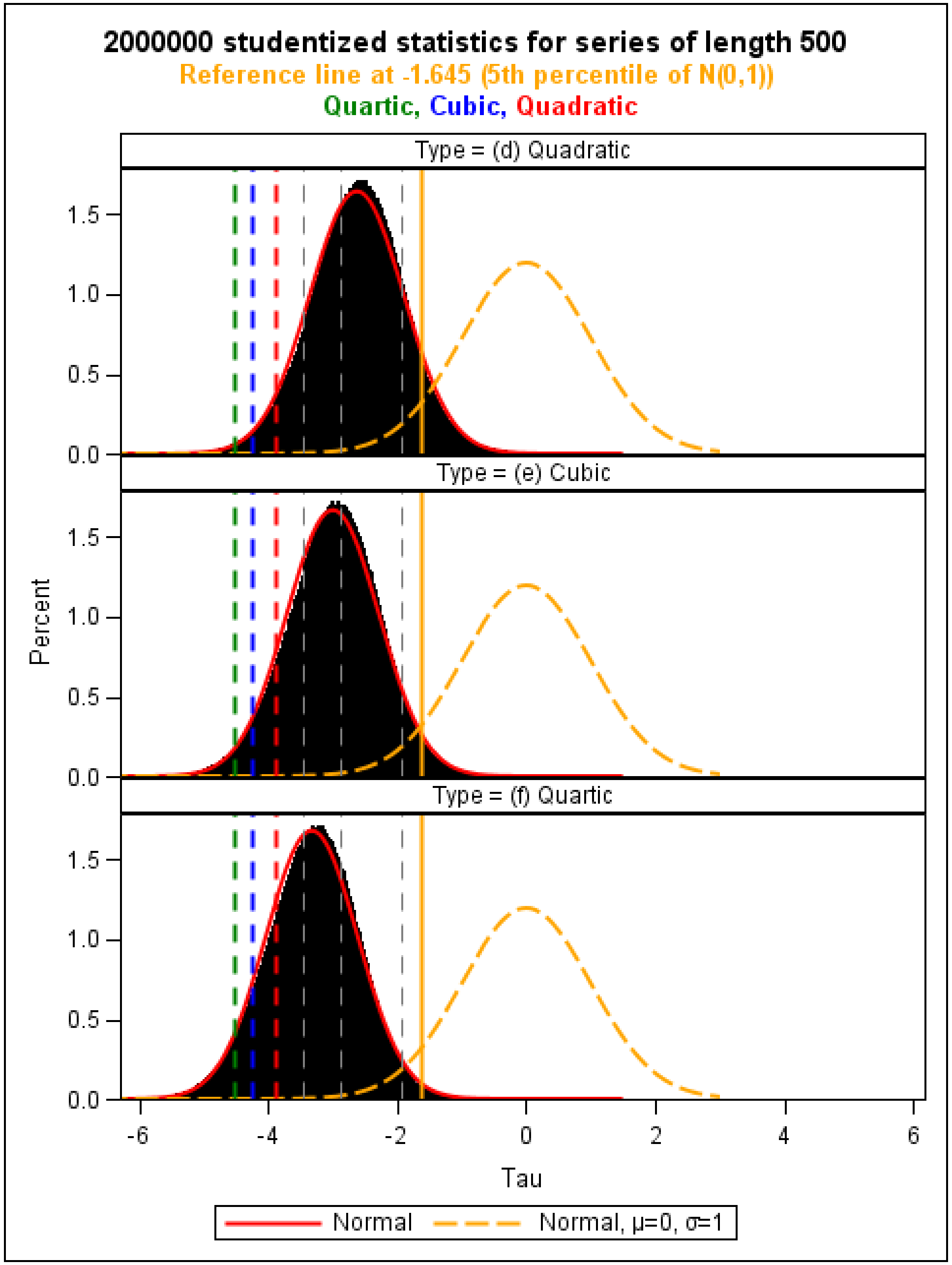
Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Tau	Pr < Tau
Zero Mean	0	2.87	0.9990
	1	2.37	0.9957
Single Mean	0	2.49	0.9999
	1	3.28	0.9997
Trend	0	-0.13	0.9937
	1	-0.25	0.9912

Conclusion: Neither 0, mean, nor trend can make residuals stationary!

How about quadratic?

Let's Take A Look



f(t)

Quadratic
← -3.84

Cubic
← -4.21

Quartic
← -4.53

Regression slope

No mean (intercept) $\frac{\sum X_t Y_t}{\sum X_t^2} = \frac{N}{D}$

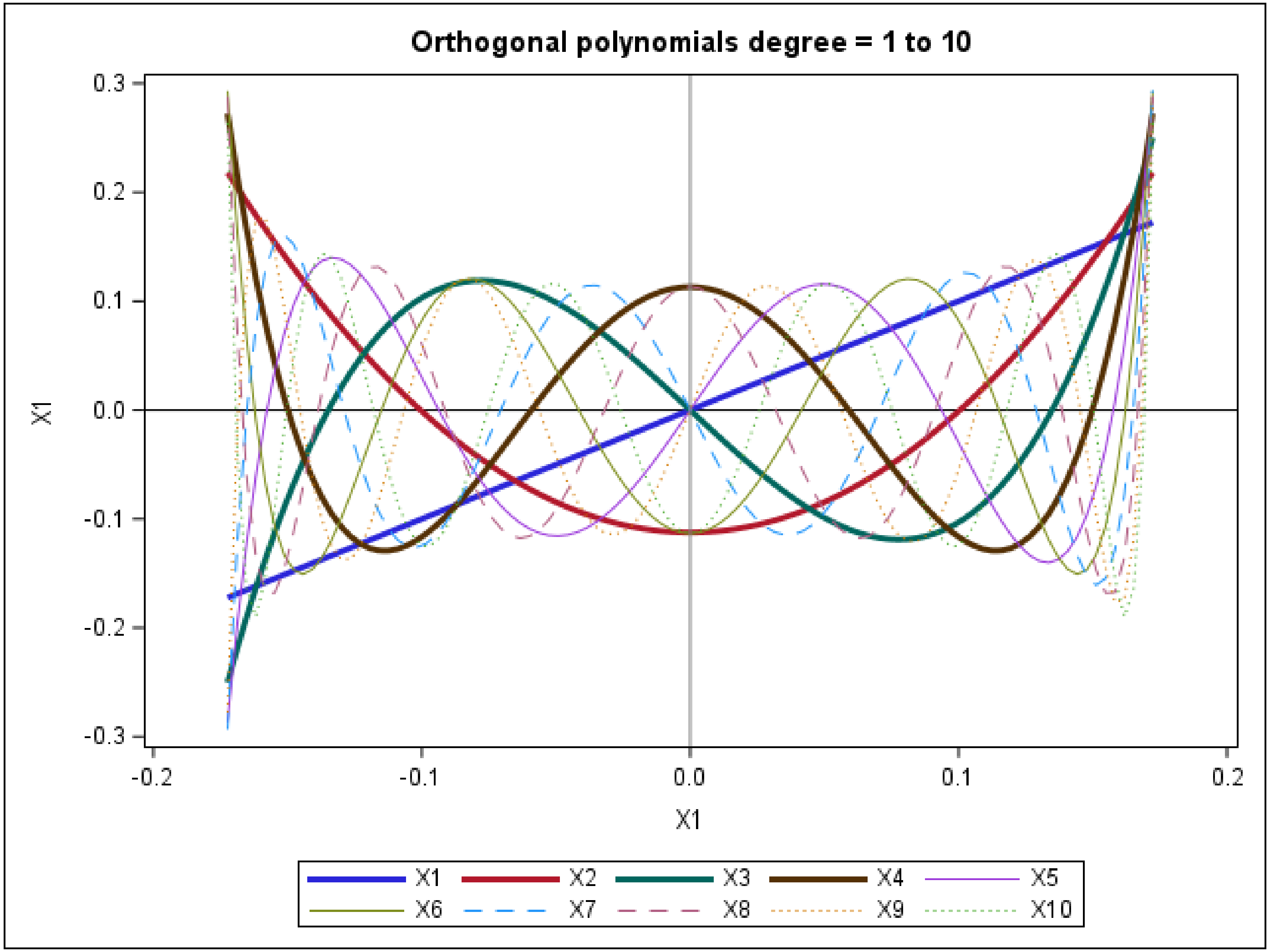
With mean $\frac{\sum(X_t - \bar{X})(Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2} = \frac{\sum X_t Y_t - n\bar{X}\bar{Y}}{\sum X_t^2 - n\bar{X}^2} = \frac{N - N_1}{D - D_1}$

	int	linear	quadratic
"Orthogonal"	1	(t-3)	$(t-3)^2-2$

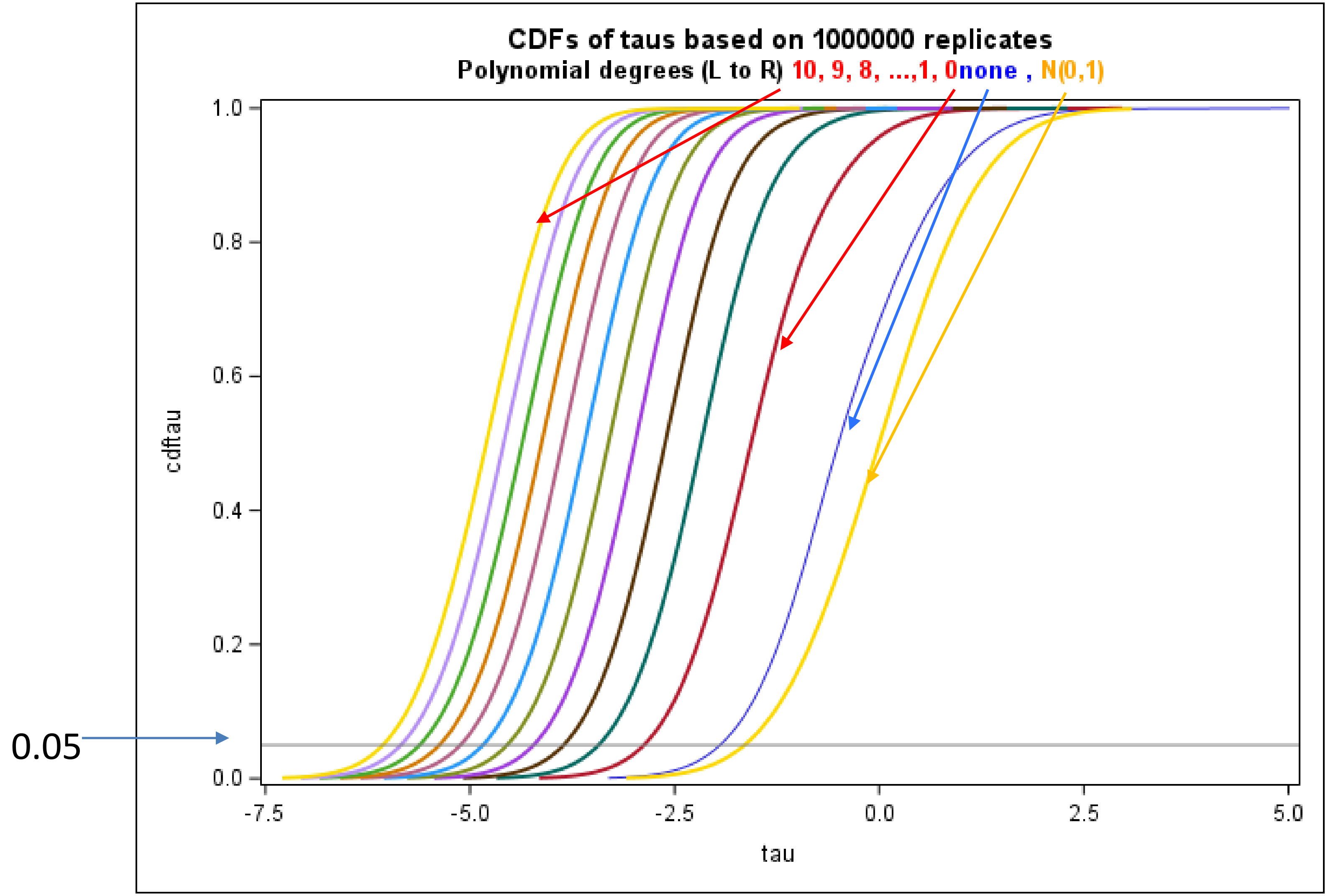
$$X = \begin{bmatrix} 1 & -2 & 2 \\ 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & -1 \\ 1 & 2 & 2 \end{bmatrix} \quad X'X = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 14 \end{bmatrix}$$

Slope = $\frac{N - N_1 - N_2 - N_3}{D - D_1 - D_2 - D_3}$

Use ORTHONORMAL polynomials for large simulations



One million series of length n=100
Empirical CDFs



Step 1: Fit quadratic and get residuals rQ

```
proc reg data=covid;  model Hospitalizations =  t t2;
output out=out1 residual=R predicted=p; run;
```

Step 2: Regress R(t)-R(t-1) on LagR = R(t-1)

Parameter Estimates

Variable	DF	Parameter Estimate	t Value	Pr > |t|
Intercept	1	0.34490	0.38	0.7036
LagR	1	-0.13972	-2.84	0.0054

Step 3: Use the right cutoff number

Quadratic ← -3.84

-3.84 < -2.84

Conclusion: H0:unit roots (is, is not) rejected

residuals (are, are not) shown to be stationary.

residuals (are, are not) shown to be nonstationary.

What Now? Difference!

$$\text{model } H_t = (at^2+bt+c)+Z_t$$

f(t) f(t-1)

$$(at^2+bt+c)-(a(t^2-2t+1)^2+b(t-1)+c)=2at+b-a$$

$$\text{Model } H_t-H_{t-1} = (\underline{b-a)+(2a)t} + Z_t-Z_{t-1}$$

```
proc arima data=covid;
identify var=Hospitalizations(1) crosscor=(t)
stationarity=(adf=4);
```

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Tau	Pr < Tau
Trend	0	-9.65	<.0001

Trend test OK for differences H_t-H_{t-1}

Z_t-Z_{t-1} is stationary! →

Finishing - model the differences!

```
estimate input=(t) p=(4) ml;
```

Maximum Likelihood Estimation

Approx

Parameter Estimate t Value Pr > |t| Lag Variable

MU -3.74522 -2.54 0.0112 0 hospitalizations AR1,1 -0.26081 -2.76
0.0058 4 hospitalizations NUM1 0.1076 4.64 <.0001 0 t

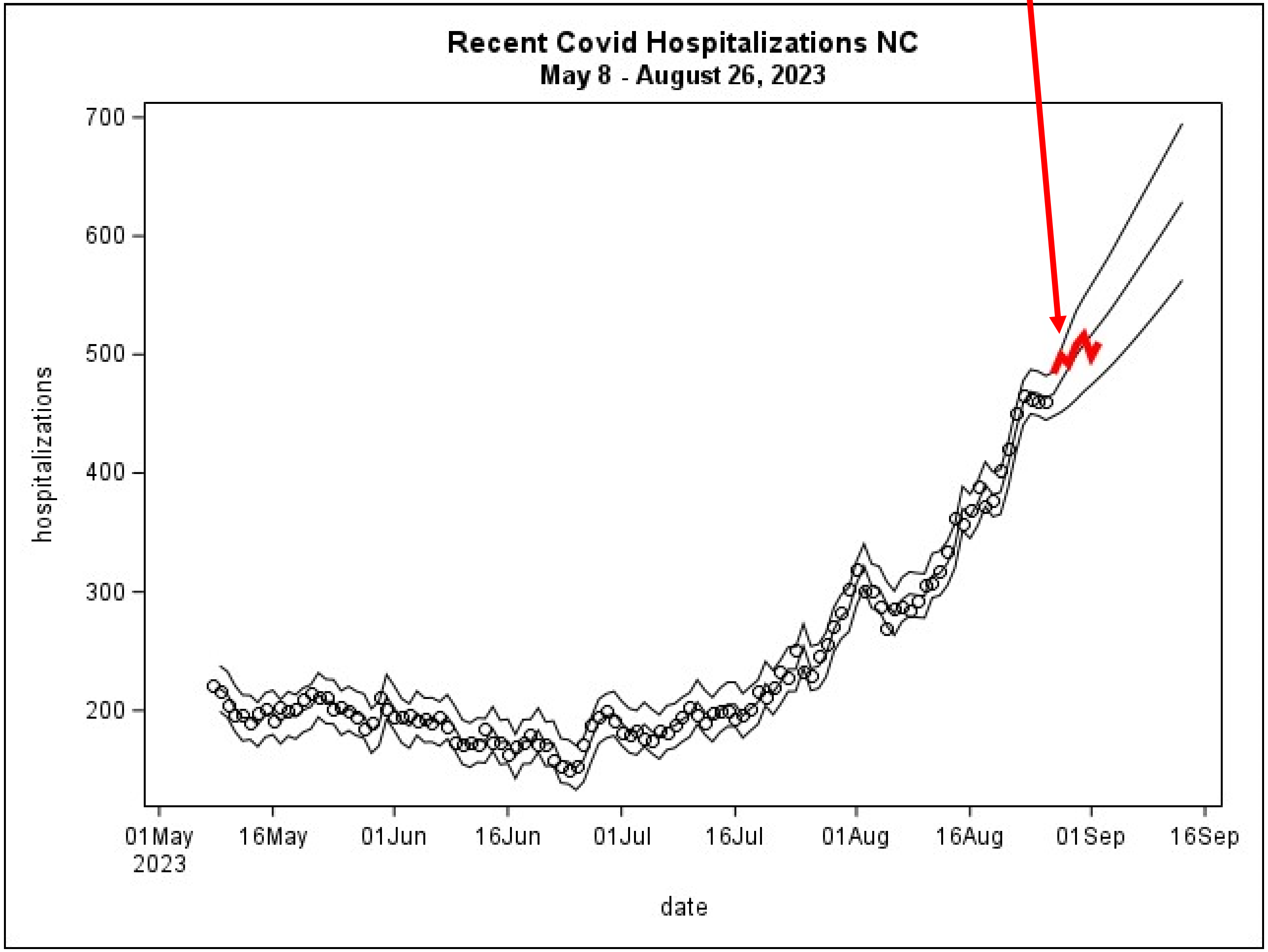
Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	4.89	5	0.4292	0.039	-0.060	0.054	0.008	-0.170	-0.069
12	9.02	11	0.6204	-0.013	0.017	-0.052	0.075	-0.049	-0.148
18	16.15	17	0.5134	0.125	0.060	-0.100	0.010	-0.102	-0.121
24	22.37	23	0.4980	-0.044	-0.038	0.153	0.059	0.076	0.091

Conclusion linear plus AR(4) fits (stationary) differences well.

```
forecast lead=18 out=out2
id=date;
```

New data released last week



Higher Order Models

stationary:

$$Y_t - \mu = 1.3(Y_{t-1} - \mu) - .4(Y_{t-2} - \mu) + e_t$$

$$\nabla Y_t = \boxed{-0.1}(Y_{t-1} - \mu) + .4(\nabla Y_{t-1}) + e_t$$

$$m^2 - 1.3m + 0.4 = (m - .5)(m - .8) = 0$$

"characteristic eqn."
roots 0.5, 0.8 (< 1)

note: –(1-.5)(1-.8) = –0.1

nonstationary

$$Y_t - \mu = 1.3(Y_{t-1} - \mu) - .3(Y_{t-2} - \mu) + e_t$$

$$\nabla Y_t = \boxed{0.0}(Y_{t-1} - \mu) + .3(\nabla Y_{t-1}) + e_t \,, \quad \nabla Y_t = .3(\nabla Y_{t-1}) + e_t$$

$$m^2 - 1.3m + 0.3 = (m - .3)(\underline{m - 1})$$

$$\text{"}unit\ root!\text{"}$$

Exit

Tests

These coefficients → normal!

Regress:

| ← → |

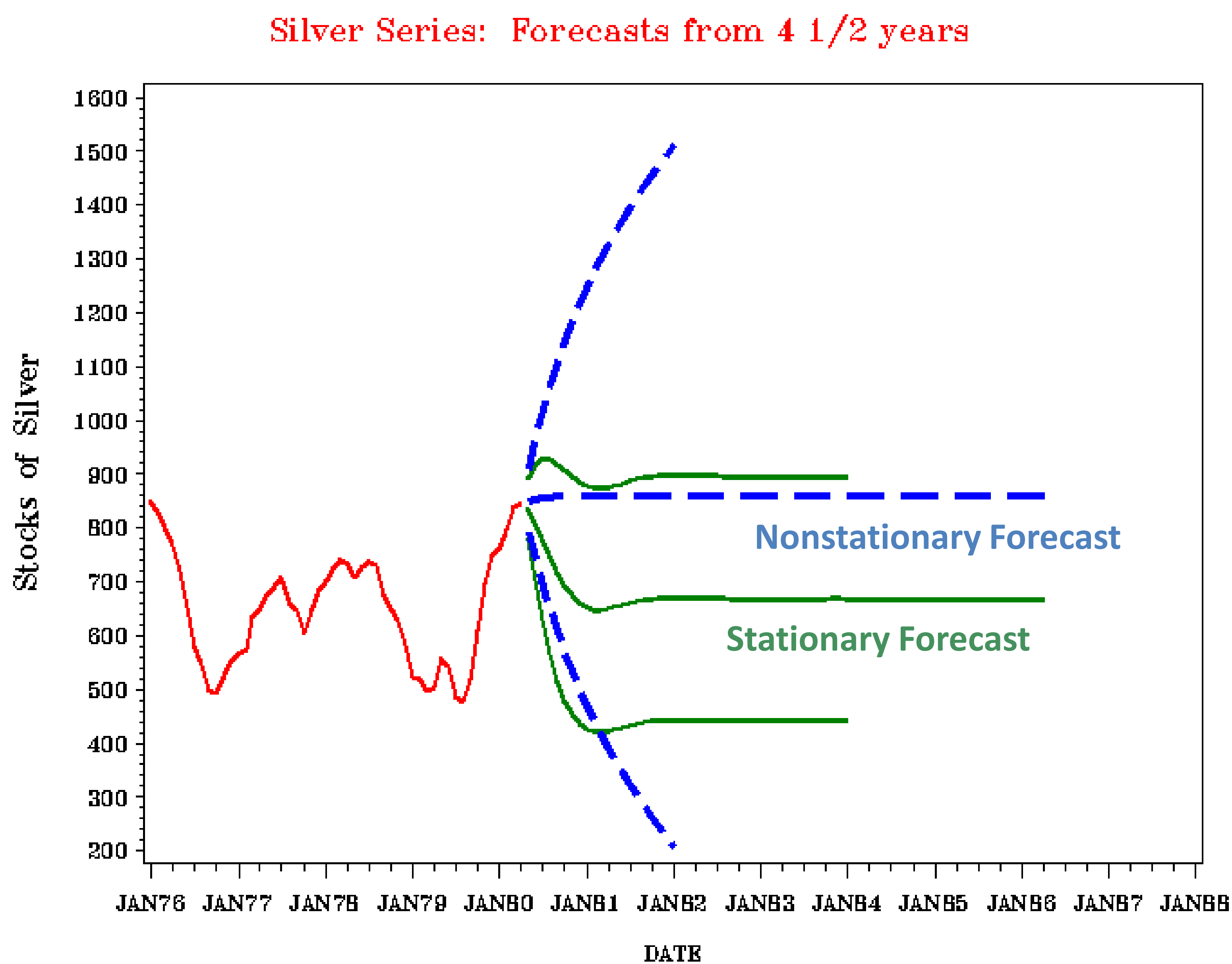
∇Y_t on (1, t) Y_{t-1} $\nabla Y_{t-1}, \nabla Y_{t-2}, \cdots, \nabla Y_{t-p+1}$

(augmenting terms)

ρ-1 ←augmenting affects limit distn.

(τ) ← " does not affect " "

Exit


Silver Series: Forecasts from 4 1/2 years
Stocks of Silver
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
JAN76
JAN77
JAN78
JAN79
JAN80
JAN81
JAN82
JAN83
JAN84
JAN85
JAN86
JAN87
JAN88
DATE
Nonstationary Forecast
Stationary Forecast

- Is AR(2) sufficient ?   test vs. AR(5).
- proc reg; model D = Y1 D1-D4; test D2=0, D3=0, D4=0;
-

Source	df	Coeff.	t	Pr>|t|
Intercept	1	121.03	3.09	0.0035
Y_{t-1}	1	-0.188	-3.07	0.0038
$Y_{t-1}-Y_{t-2}$	1	0.639	4.59	0.0001
$Y_{t-2}-Y_{t-3}$	1	0.050	0.30	0.7691
$Y_{t-3}-Y_{t-4}$	1	0.000	0.00	0.9985
$Y_{t-4}-Y_{t-5}$	1	0.263	1.72	0.0924

$F_{41}^{3} = 1152 / 871 = 1.32$ Pr>F = 0.2803

X

Fit AR(2) and do unit root test

Method 1: OLS output and tabled critical value (-2.86)

proc reg; model D = Y1 D1;

Source	df	Coeff.	t	Pr>|t|	
Intercept	1	75.581	2.762	0.0082	X
Y_{t-1}	1	-0.117	-2.776	0.0038	X
$Y_{t-1}-Y_{t-2}$	1	0.671	6.211	0.0001	☺

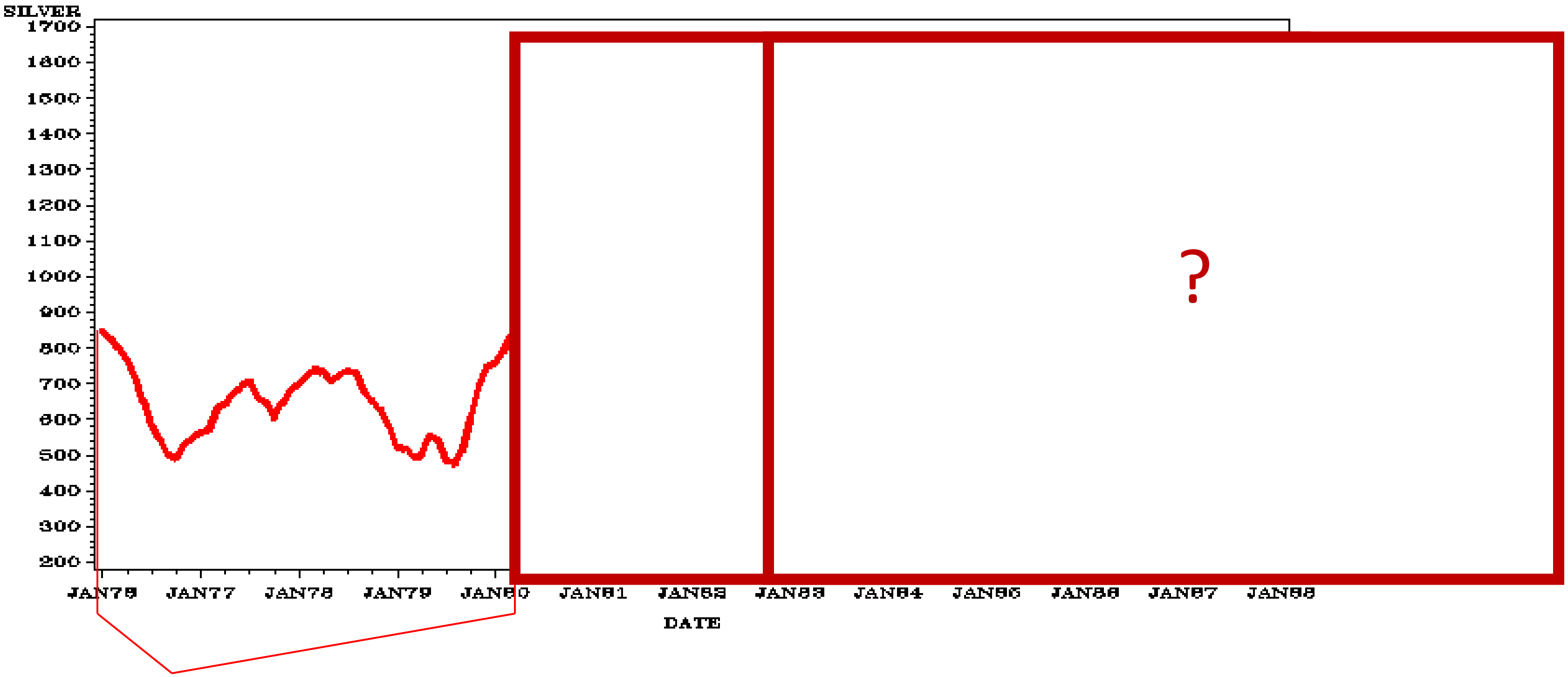
Method 2: OLS output and corrected p-values

proc arima; identify var=silver stationarity = (adf=(1));

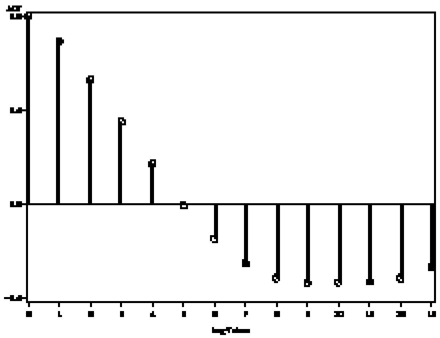
Augmented Dickey-Fuller Unit Root Tests

Type	Lags	t	Prob<t	
Zero Mean	1	-0.2803	0.5800	
Single Mean	1	-2.7757	0.0689	☺
Trend	1	-2.6294	0.2697	

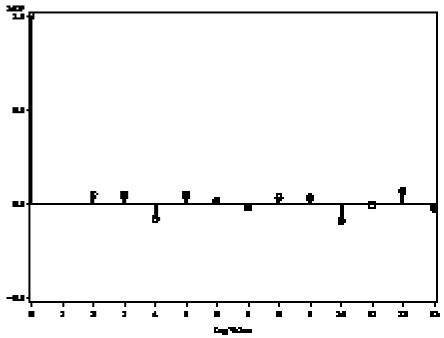
Silver Series – the rest of the story!



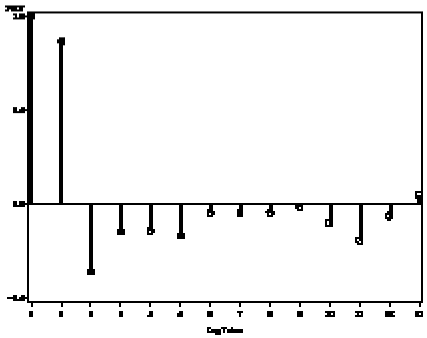
ACF



IACF



PACF



Your talk seems better now Grandpa!

Disclaimer: No babies ingested alcohol in the making of this slide.
(I personally made sure the bottle was empty).

Exit

Detecting overdifferencing

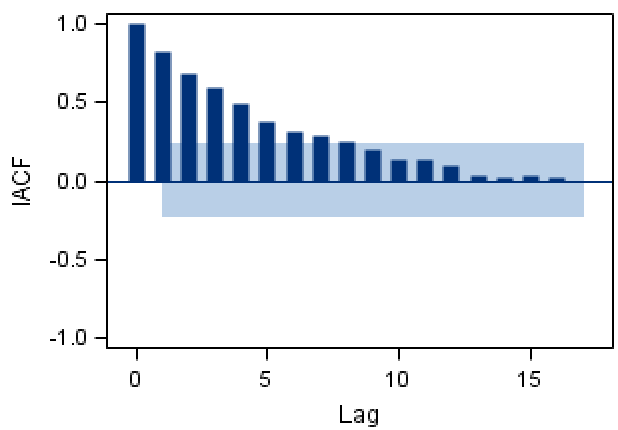
Generated Data → $Y_t = \alpha + \beta t + e_t \qquad e_t \sim independent\ \ N(0, \sigma^2)$

$Y_{t-1} = \alpha + \beta(t-1) + e_{t-1}$

Differences → $\nabla Y_t = \beta + e_t - \theta e_{t-1} \qquad \theta = 1$ (“Non-invertible moving average”)

Chang (1993) Moving average unit root (e.g. θ=1) → slow decay in IACF (Inverse AutoCorrelation Function)

Inverse Autocorrelations → (of differences)



Chang, M. C., Dickey, D. A. , (1993) "Recognizing Overdifferenced Time Series," Journal of Time Series Analysis , 15, 1-8.

$$Y_t-\mu = \rho(Y_{t-1}-\mu) + e_t - \theta e_{t-1}$$

Autoregressive Moving Average

Model $(|\theta|<1)$

(1) Inverse Autocorrelation Function (IACF) is autocorrelation function of

$$Y_t-\mu = \theta(Y_{t-1}-\mu) + e_t - \rho e_{t-1}$$

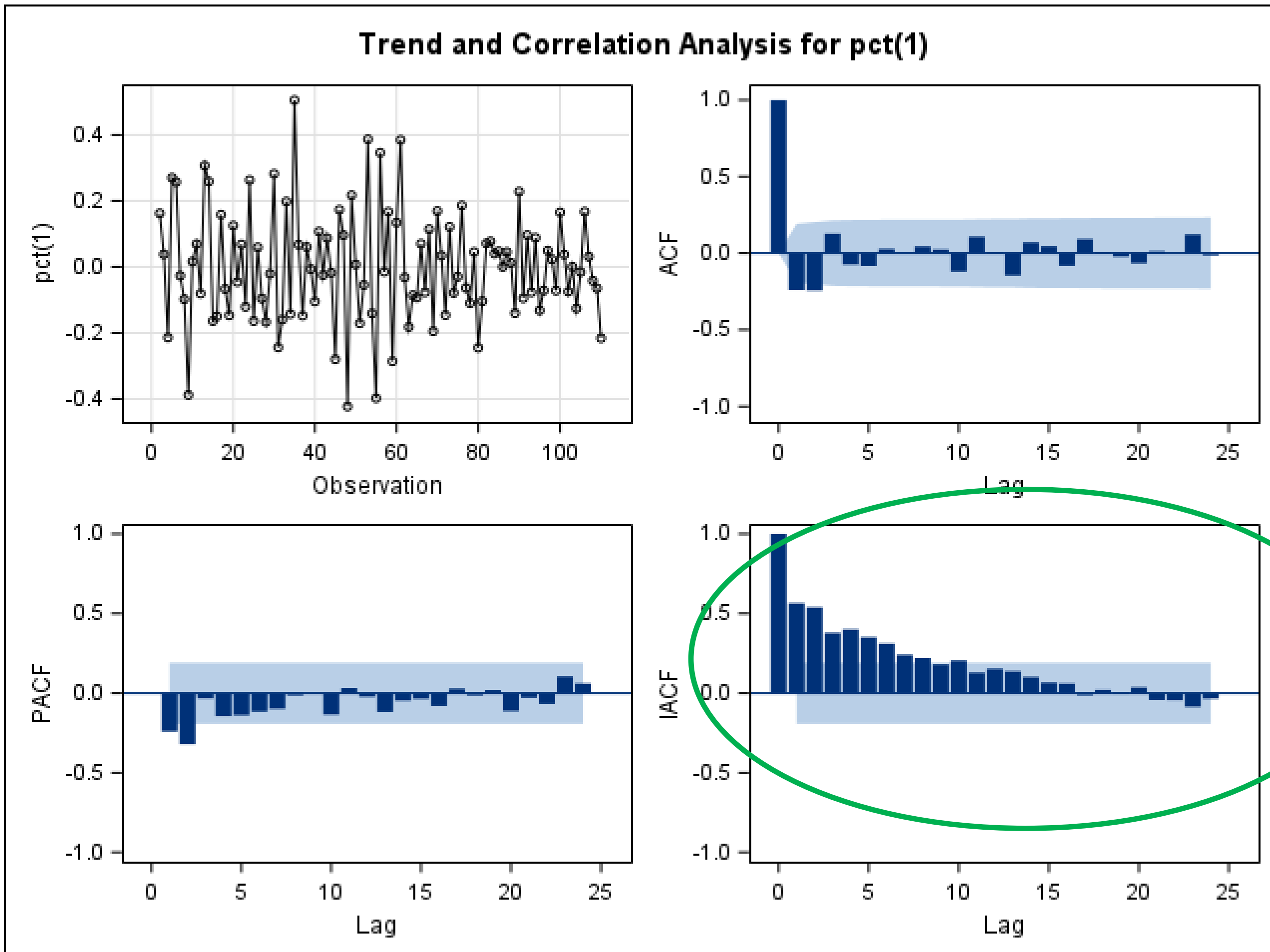
(2) Said E. Said*: With sufficient data and lagged differences, just treat this as autoregressive to test H_0: $\rho=1$.

* Said S.E. and Dickey, D. A. (1984). "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika 71, 599-607.

Winning

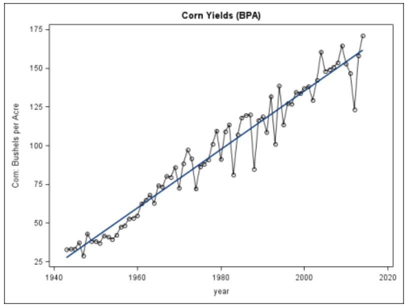
Percentages

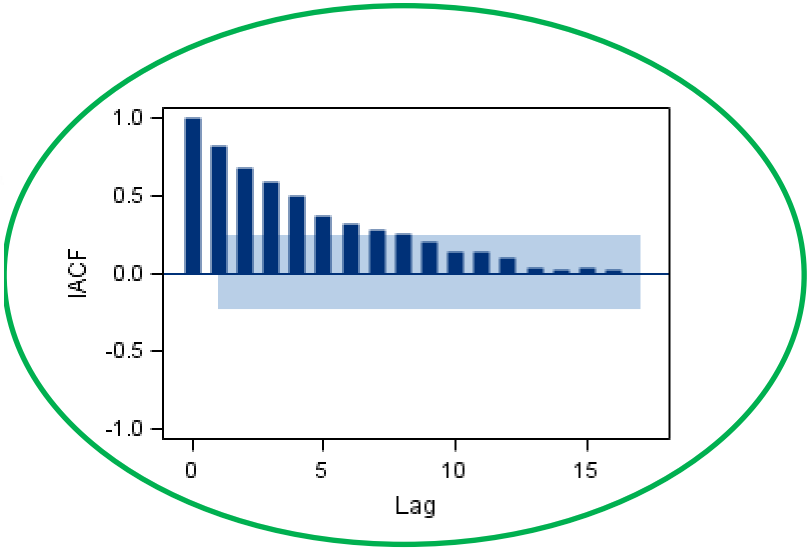
Differences



IACF

Post 1942
Corn Yields
(differenced)





IACF

(Time permitting)

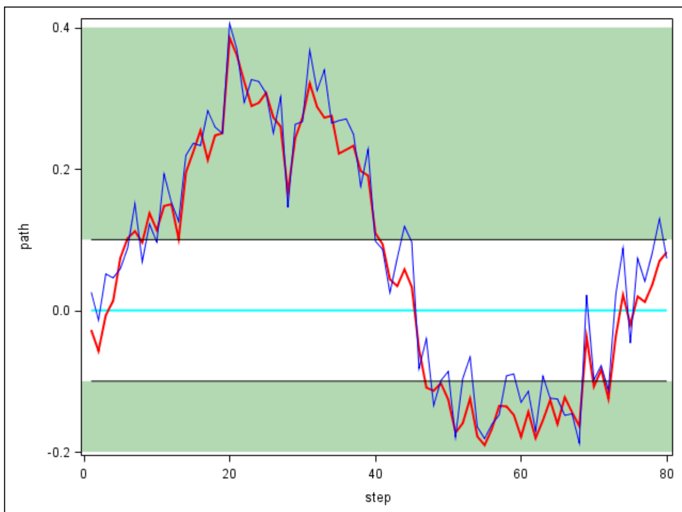
Cointegration analogy:
Drunk man walking a dog

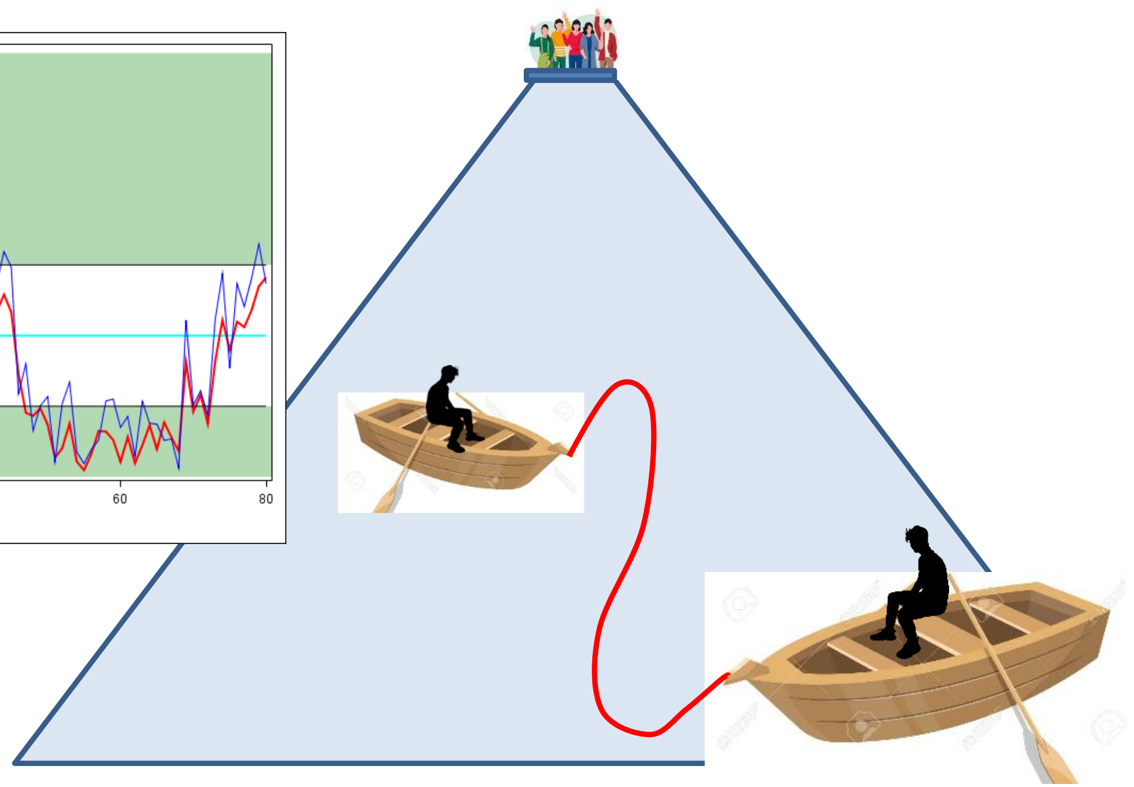
Man – red path

Man -> unit root process
Dog-> unit root process
Difference (man-dog) stationary!

Dog – blue path

Sidewalk





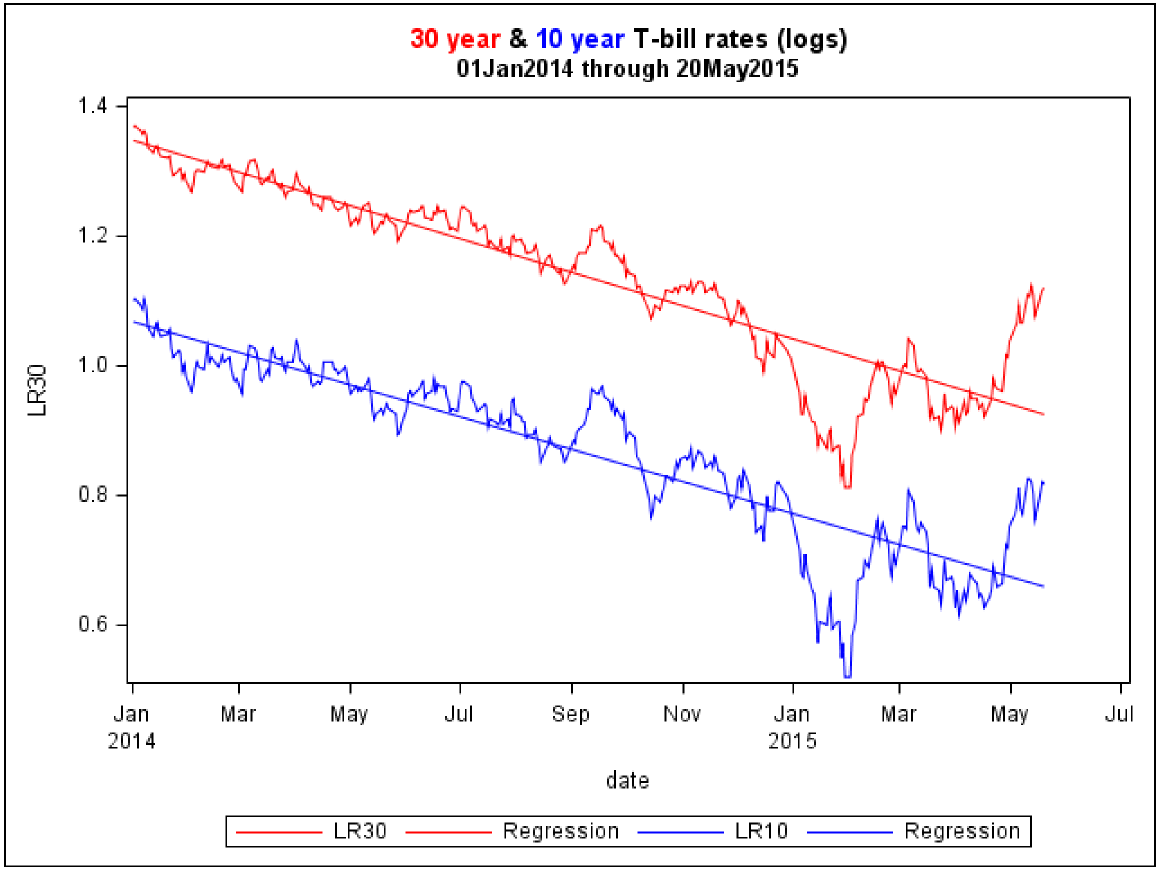
Exit

Data:

10 year t-bill yields

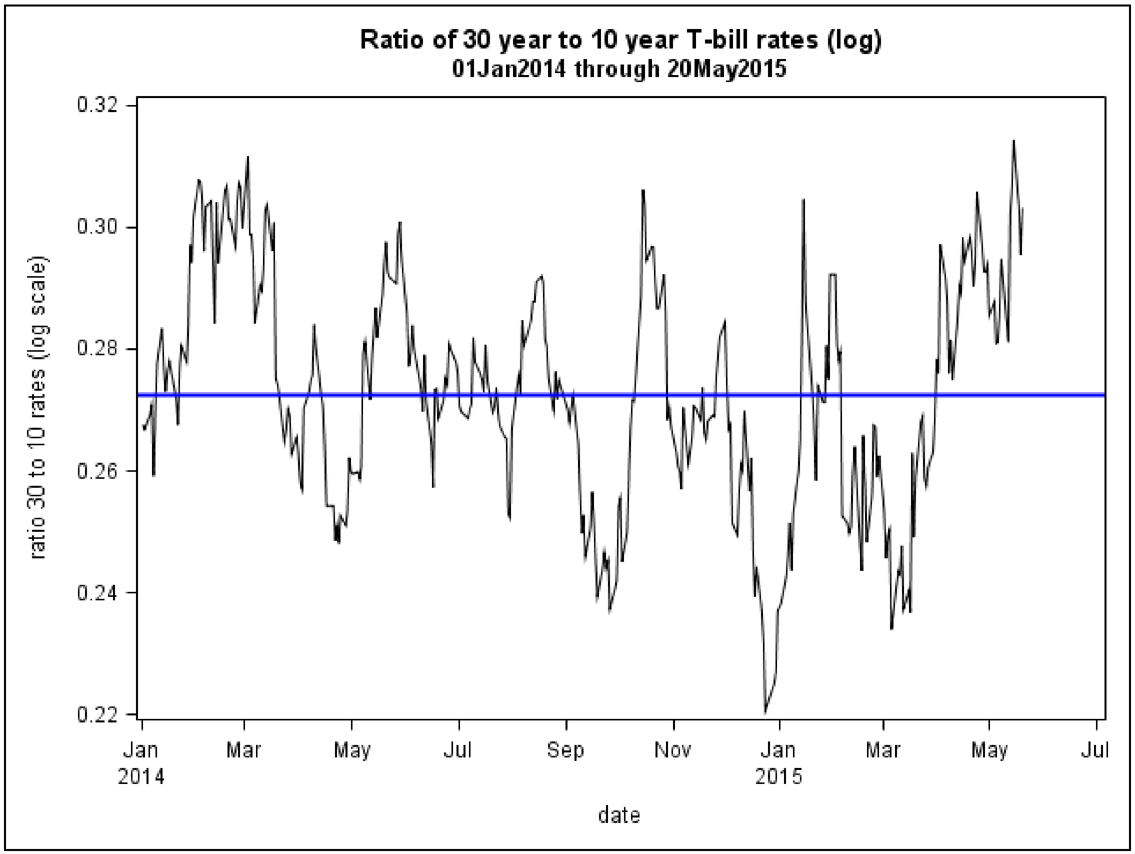
30 year t-bill yields

(in logarithms)



Log of Ratio

$$\text{Log of } \frac{\text{30 year t-bill yields}}{\text{10 year t-bill yields}} = \log(30\text{ yr.}) - \log(10\text{ yr.})$$



Check #1 both series are unit root types - yes

(10 year t-bills)

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-0.4461	0.5815	-1.05	0.2635
	1	-0.4438	0.5820	-1.09	0.2491
	2	-0.4283	0.5854	-1.07	0.2573
Single Mean	0	-5.4449	0.3922	-1.96	**0.3065**
	1	-5.2415	0.4106	-1.95	**0.3106**
	2	-5.0587	0.4278	-1.89	**0.3351**
Trend	0	-16.6981	0.1292	-2.63	**0.2652**
	1	-15.9949	0.1482	-2.51	**0.3207**
	2	-16.1441	0.1440	-2.47	**0.3421**

Can't
Reject
Unit Root

(30 year t-bills)

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-0.2652	0.6222	-1.09	0.2487
	1	-0.2662	0.6220	-1.13	0.2346
	2	-0.2595	0.6235	-1.08	0.2525
Single Mean	0	-3.5711	0.5870	-1.74	**0.4082**
	1	-3.5006	0.5953	-1.76	**0.4013**
	2	-3.5035	0.5950	-1.72	**0.4194**
Trend	0	-6.1699	0.7287	-1.24	**0.9002**
	1	-5.5451	0.7783	-1.12	**0.9231**
	2	-6.0531	0.7381	-1.17	**0.9140**

Can't
Reject
Unit Root

Check #2 Need more than one difference? – no

Tests on differenced data:

(10 year t-bills)

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-359.430	0.0001	-19.34	<.0001
	1	-369.931	0.0001	-13.56	<.0001
	2	-321.229	0.0001	-10.38	<.0001
Single Mean	0	-360.041	0.0001	**-19.35**	**<.0001**
	1	-371.989	0.0001	**-13.57**	**<.0001**
	2	-324.886	0.0001	**-10.39**	**<.0001**
Trend	0	-360.919	0.0001	-19.37	<.0001
	1	-374.663	0.0001	-13.60	<.0001
	2	-328.854	0.0001	-10.43	<.0001

<u>Reject</u>
Unit roots in differenced data

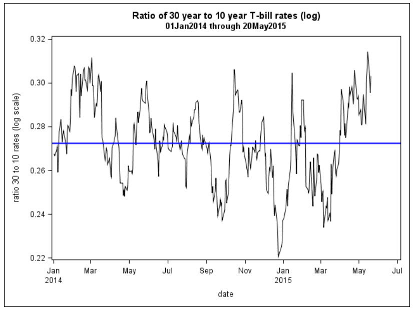
(30 year t-bills)

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	-355.741	0.0001	-19.13	<.0001
	1	-342.063	0.0001	-13.04	<.0001
	2	-288.960	0.0001	-10.00	<.0001
Single Mean	0	-356.585	0.0001	**-19.15**	**<.0001**
	1	-344.588	0.0001	**-13.06**	**<.0001**
	2	-293.062	0.0001	**-10.02**	**<.0001**
Trend	0	-358.485	0.0001	-19.23	<.0001
	1	-350.087	0.0001	-13.15	<.0001
	2	-301.163	0.0001	-10.11	<.0001

<u>Reject</u>
Unit roots in differenced data

Two unit root processes X_t and Y_t are cointegrated if some linear combination ($aX_t + bY_t$, e.g. $1X_t - 1Y_t$) is stationary.



We have 2 unit root processes

Are they "cointegrated?"

Can we use a = 1 and b = -1 ?

Is log(30 yr. rate) – log(10 yr. rate)

(= log(30 yr. rate/10 yr. rate)) stationary?

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	0	0.0028	0.6832	0.01	0.6844
	1	0.0243	0.6882	0.05	0.6991
	2	0.0324	0.6900	0.07	0.7060
	3	0.0204	0.6872	0.04	0.6970
Single Mean	0	-27.6337	0.0017	**-3.69**	**0.0048**
	1	-24.9002	0.0033	**-3.42**	**0.0113**
	2	-22.8252	0.0054	**-3.19**	**0.0216**
	3	-26.3394	0.0024	**-3.36**	**0.0136**
Trend	0	-28.5986	0.0100	-3.71	0.0227
	1	-25.8595	0.0186	-3.44	0.0486
	2	-23.7128	0.0298	-3.20	0.0866
	3	-27.6531	0.0124	-3.37	0.0574

Decision: Series are cointegrated!

Which row of the table should we have used?
How many lagged differences do we need?

Regresss

$$\Delta Y_t \quad on \quad 1, Y_{t-1}, \underbrace{\Delta Y_{t-1}, \Delta Y_{t-2}, \cdots, \Delta Y_{t-22}}_{\text{usual distributions (normal, t, F)}}$$

H_0: No lagged differences needed

$$F^{22}_{322} = 1.16 \qquad \Pr > F = 0.2807 > 0.05$$

Conclusion: Can use top (0 lagged differences) row

Single Mean	0	-27.6337	0.0017	**-3.69**	**0.0048**

Final model:

log(30 yr. rate/10 yr. rate) is stationary with estimated mean 0.2805 and autoregressive order 1 structure with ρ=0.92

The <u>full</u> cointegration story:

How do we know that the stationary combination of Y_t and X_t is $S_t = 1(Y_t) -1(X_t)$ rather than $S_t = 2(Y_t) -7(X_t)$ or something else?

(1) Estimate a and b in $S_t = a(Y_t) +b(X_t)$
(2) Find the effect on the unit root test of estimating a,b

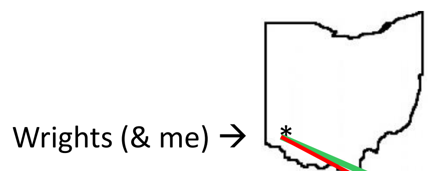
Solution: Engle and Granger (and others)

Result: Nobel Prize in Economics (2003)

One Final Wilbur Wright Quote:

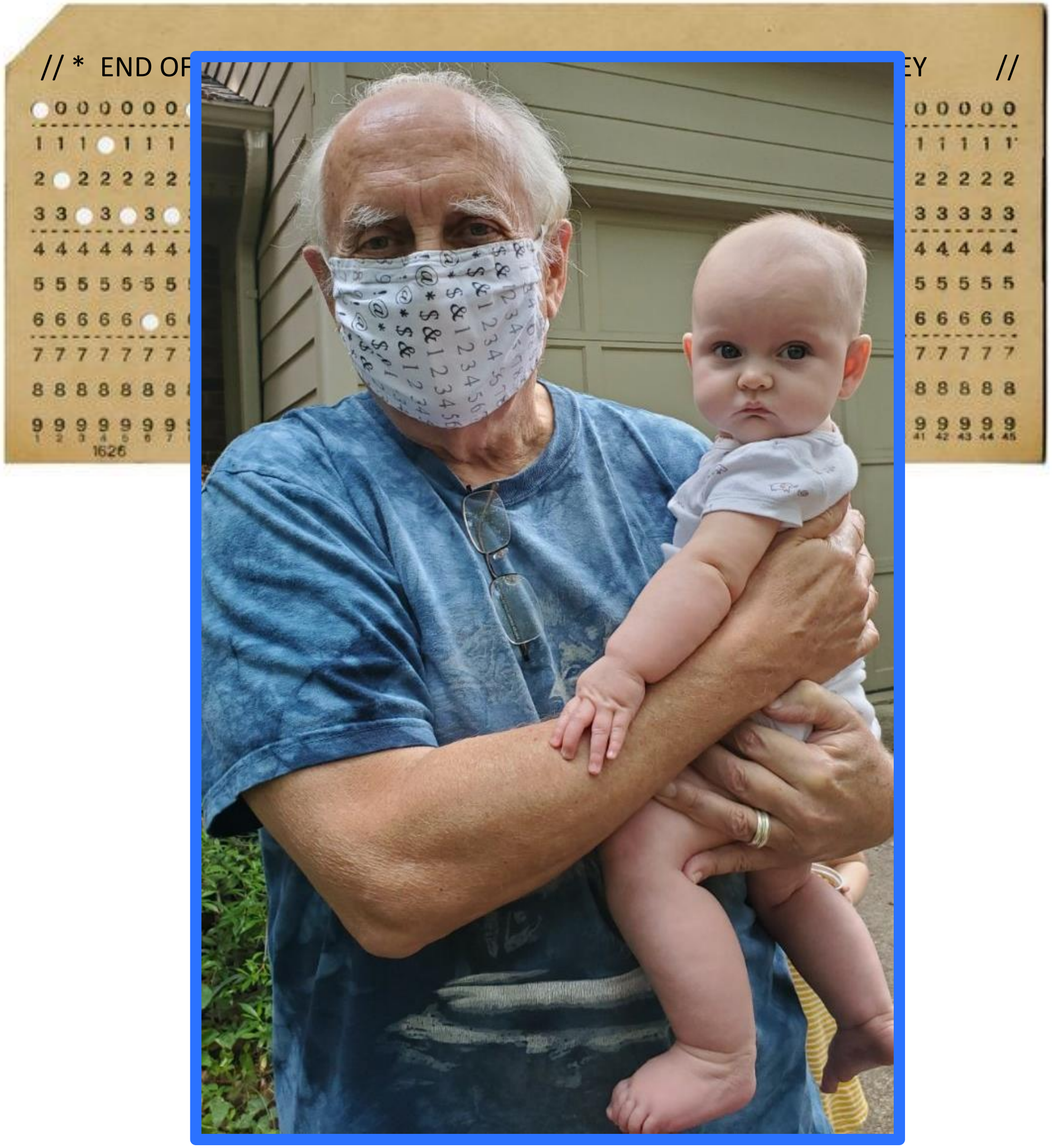
"If I were giving a young man advice as to how he might succeed in life, I would say to him, pick out a good father and mother, and begin life in Ohio."

--Wilbur Wright, 1910



… and one from Clark Gable:

"I'm just a lucky slob from Ohio who happened to be in the right place at the right time."



Lesson 8: Don’t take yourself too seriously 🙂