Lecture notes on forecasting

Robert Nau
Fuqua School of Business
Duke University

http://people.duke.edu/~rnau/forecasting.htm

Forecasting with adjustments for inflation and seasonality

- Deflation with price indices
- Seasonal decomposition
- Time series forecasting models for seasonal data
  - Averaging and smoothing combined with seasonal adjustment
  - Winters seasonal exponential smoothing model
- Side-by-side model comparisons in Statgraphics
  - Example: U.S. auto sales from 1970 to 1993

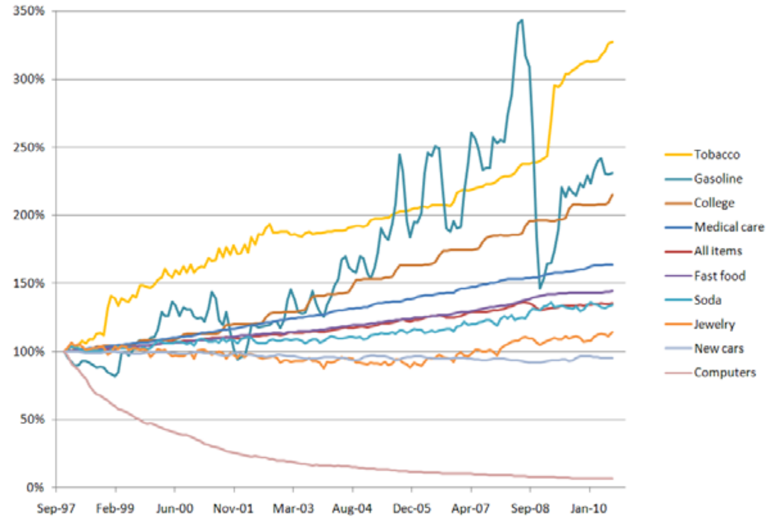
Modeling the effect of inflation

- Depending on the industry and the time horizon, price inflation can be a significant component of nominal growth
- If it is, there are two reasons why you might want to explicitly model its effect:
  - To measure *real* growth and estimate its dependence on other real factors (e.g. relative prices, market size, etc.)
  - To remove much of the trend and stabilize variance before fitting a model

Adjusting for inflation

- To “deflate” a variable, you *divide it by an appropriate price index variable*
- Depending on your objective you may want to use a *specific* index or a *general* one
  - To convert from nominal dollars to units of *real goods and services*, you want a *product-specific* index.
  - To convert from nominal dollars into equivalent “market baskets” of consumer goods, a *general* price index may suffice (or may be all you have!)
  - You can choose an arbitrary base year by scaling the price index to a value of 1.0 in the desired year.

Many indices to choose from!



Source: U.S. Bureau of Economic Analysis National Income and Product Accounts (NIPA)
Price indices for personal consumption expenditures. Available at http://www.economagic.com/nipa.htm

Putting inflation back in

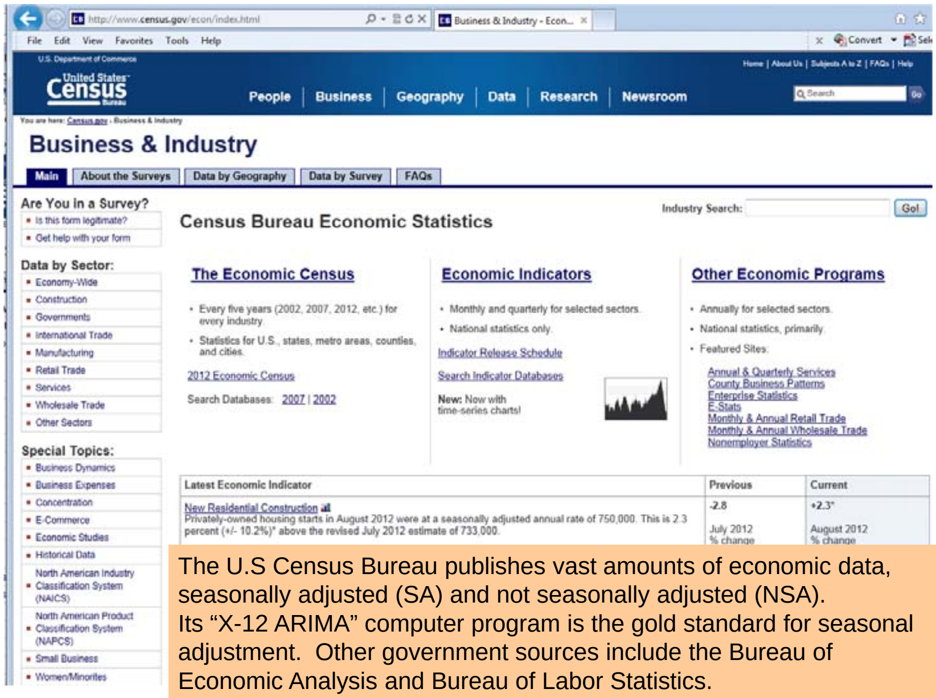
- To “re-inflate” forecasts of a *deflated* series (when appropriate), you multiply the forecasts and confidence limits by a *forecast* of the price index
- For short-term forecasts this isn’t much of an issue
- For longer-term forecasts, a price index forecast can be obtained from extrapolation of recent price trends or from financial markets (“CPI futures”) or expert consensus

When to log, when to deflate?

- *Deflation* should be used when you are interested in knowing the forecast in “real” terms and/or if the inflation rate is expected to change
- *Logging* is sufficient if you just want a forecast in “nominal” terms and inflation is expected to remain constant—*inflation just gets lumped with other sources of compound growth in the model*.
- Logging also ensures that forecasts and confidence limits have *positive values*, even in the presence of downward trends and/or high volatility.
- If inflation has been minimal and/or there is little overall trend or change in volatility, neither may be necessary

Seasonality

- A *repeating, periodic pattern* in the data that is keyed to the *calendar* or the *clock*
- Not the same as “cyclicality”, i.e., business cycle effects, which do *not* have a predictable periodicity
- Most regular seasonal patterns have an *annual* period (12 months or 4 quarters), but other possibilities exist:
  - Day-of-week effects (period=5, 6, or 7 days)
  - End-of-quarter effects (period = 3 months)
  - Hour-of-the-day effects



Business & Industry

Are You in a Survey?
- Is this form legitimate?
- Get help with your form

Data by Sector:
- Economy-Wide
- Construction
- Governments
- International Trade
- Manufacturing
- Retail Trade
- Services
- Wholesale Trade
- Other Sectors

Special Topics:
- Business Dynamics
- Business Expenses
- Concentration
- E-Commerce
- Economic Studies
- Historical Data
- North American Industry Classification System (NAICS)
- North American Product Classification System (NAPCS)
- Small Business
- Women/Minorities

Census Bureau Economic Statistics

The Economic Census
- Every five years (2002, 2007, 2012, etc.) for every industry.
- Statistics for U.S., states, metro areas, counties, and cities.

2012 Economic Census

Search Databases: 2007 | 2002

Economic Indicators
- Monthly and quarterly for selected sectors.
- National statistics only.

Indicator Release Schedule

Search Indicator Databases

New: Now with time-series charts!

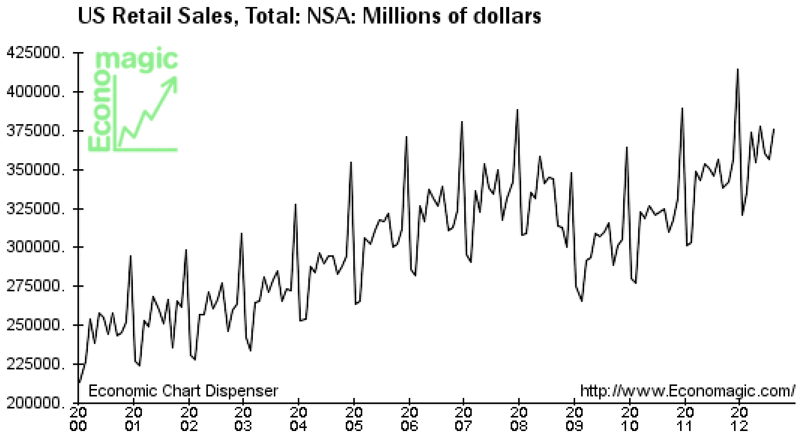
Other Economic Programs
- Annually for selected sectors.
- National statistics, primarily.
- Featured Sites:

Annual & Quarterly Services
County Business Patterns
Enterprise Statistics
E-Stats
Monthly & Annual Retail Trade
Monthly & Annual Wholesale Trade
Nonemployer Statistics

Latest Economic Indicator	Previous	Current
New Residential Construction Privately-owned housing starts in August 2012 were at a seasonally adjusted annual rate of 750,000. This is 2.3 percent (+/- 10.2%)* above the revised July 2012 estimate of 733,000.	-2.8 July 2012 % change	+2.3* August 2012 % change

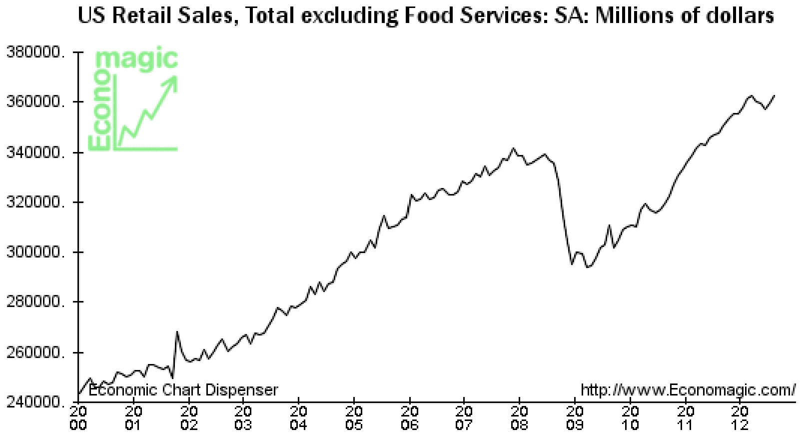
The U.S Census Bureau publishes vast amounts of economic data, seasonally adjusted (SA) and not seasonally adjusted (NSA).
Its “X-12 ARIMA” computer program is the gold standard for seasonal adjustment. Other government sources include the Bureau of Economic Analysis and Bureau of Labor Statistics.

Some examples of seasonal data, with and without seasonal adjustment



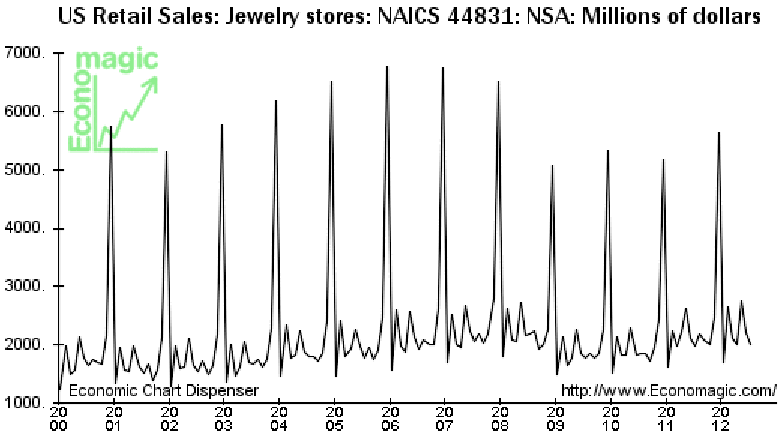
Total U.S. retail sales data has a dramatic seasonal pattern with a large Christmas-shopping "binge" in December and "hangover" in January-February.

US total retail sales (SA)



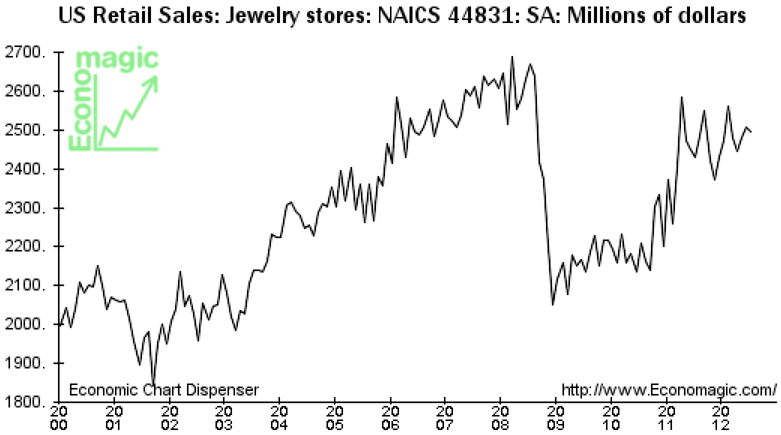
Seasonal adjustment highlights the sheer magnitude of the crash of 2008.

Jewelry (NSA)



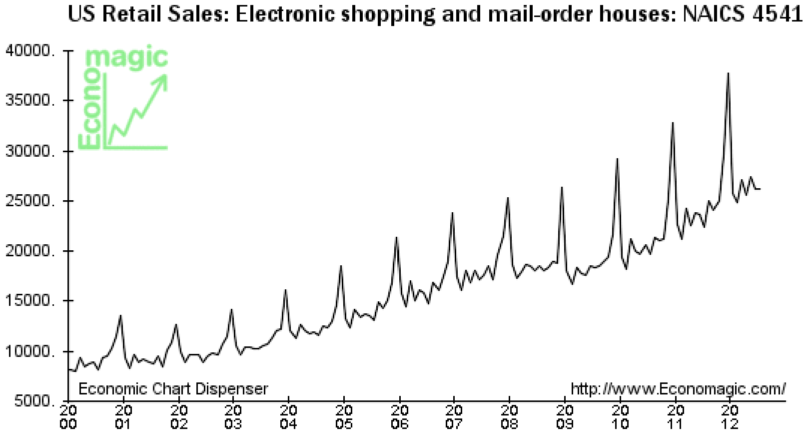
Jewelry sales have a particularly large spike in December…

Jewelry (SA)



…and had an even more dramatic crash in 2008 in seasonally adjusted terms.

Electronic shopping & mail order (NSA)



US Retail Sales: Electronic shopping and mail-order houses: NAICS 4541

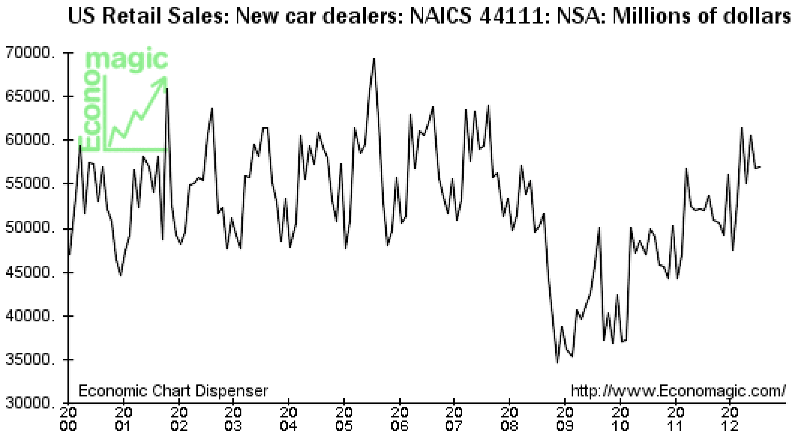
Electronic and mail order shopping also have a dramatic seasonal pattern…

Electronic shopping & mail order (NSA)

US Retail Sales: Electronic shopping and mail order houses: NAICS 4541

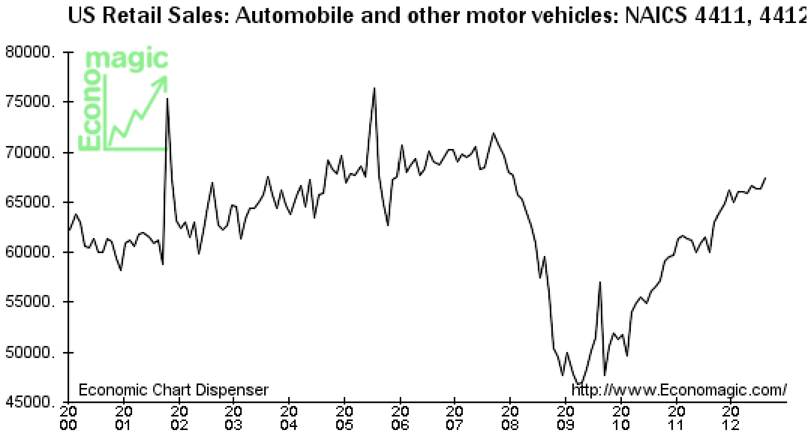
…but did *not* undergo a very big drop in 2008.

New cars (NSA)



The seasonal pattern of auto sales is more irregular (due to changes in timing of promotions and new models), and the long-term trend is relatively modest.

New cars (SA)



Seasonal adjustment reveals a gradual upward trend from 2000 to 2008, followed by a drop to far below the level 10 years earlier.

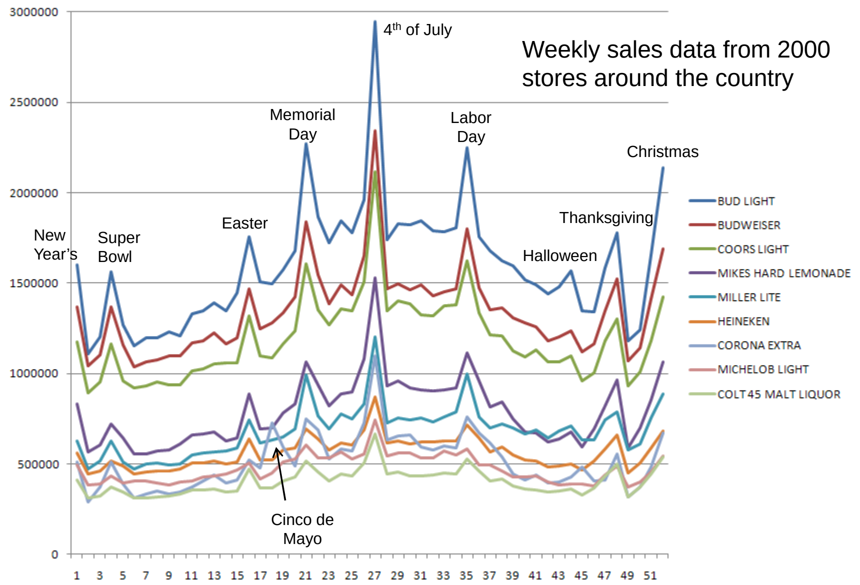
Seasonal patterns are complex, because the calendar is not rational

- Retail activity is geared to the business week, but months and years don’t have whole numbers of weeks
- A given month does not always have the name number of trading days or weekends
- Christmas shopping season has an enormous impact, and Christmas day can fall on any day of the week.
- Some major holidays (e.g., Easter) are “moveable feasts” that do not occur on the same calendar dates each year

Quarterly vs. monthly vs. weekly

- Quarterly data is easiest to handle: 4 quarters in a year, 3 months in a quarter, trading day adjustments have only minor effects.
- Monthly data is more complicated: 12 months in a year, but *not* 4 weeks in a month; trading day adjustments may be important.
- Weekly data requires special handling because a year is not exactly 52 weeks.

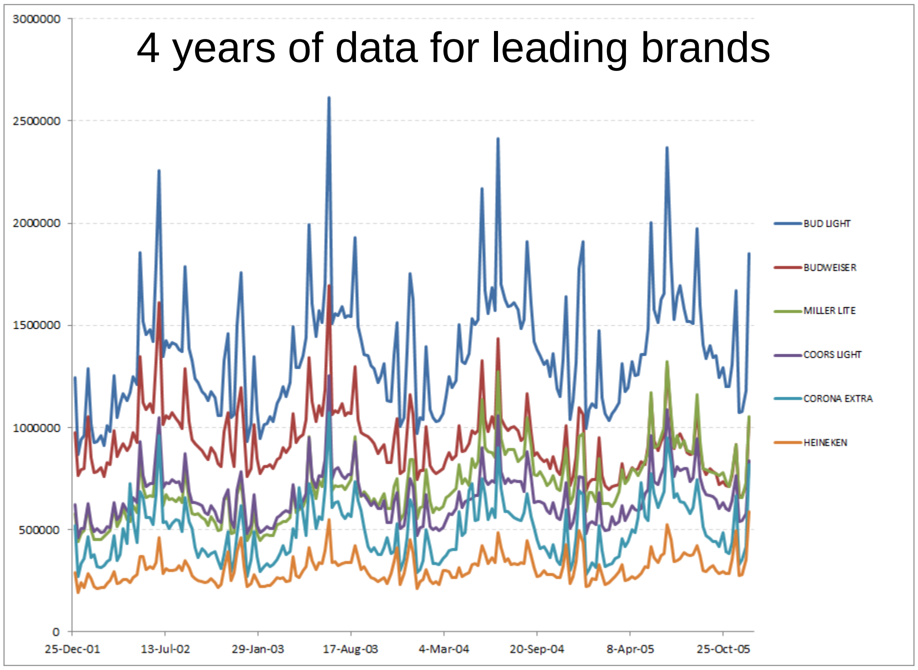
The beer calendar



Weekly sales data from 2000 stores around the country

Week 1 = week beginning Dec 30, 2002 **Week 52 = week beginning Dec 22, 2003**

4 years of data for leading brands



Multiplicative seasonality

Most natural seasonal patterns are *multiplicative:*

- Seasonal variations are roughly constant in *percentage* terms.
- Seasonal swings therefore get larger or smaller in *absolute* magnitude as the average level of the series rises or falls due to long-term trends and/or business cycle effects.
- Errors also tend to be consistent in size in % terms, in which case it's better to compare *Mean Absolute Percentage Error* (MAPE), rather than Root-Mean-Squared Error (RMSE), between models.

Additive seasonality

An *additive* seasonal pattern has *constant-amplitude* seasonal swings in the presence of trends and cycles.

- A *log transformation* converts a multiplicative pattern to an additive one, so if your model includes a log transformation, use additive rather than multiplicative seasonal adjustment.
- If the historical data sample has little trend and seasonal variations are not large in relative terms, additive and multiplicative adjustment yield very similar results.
- Additive seasonal patterns can also be fitted by using regression models with seasonal dummy variables.

Seasonal adjustment

Additive or multiplicative adjustment compensates for the anticipated effects of seasonality.

Two uses for seasonal adjustment:

- As an analysis tool that provides a different “view” of the data with seasonality removed
- As a component of a forecasting model in which a non-seasonal model is fitted to seasonally adjusted data

Seasonal Indices

- A seasonal index represents the expected percentage of “normal” in a given month or quarter.
- If January’s index is 89, this means that January’s value is expected to be 89% of normal, where “normal” is defined by the monthly average for the whole surrounding year.
- In this case, January’s *seasonally adjusted* value would be the actual value divided by 0.89

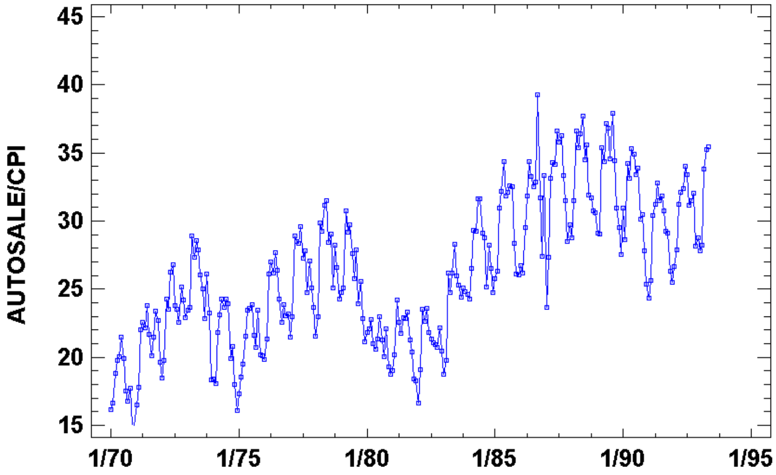
Seasonal indices, continued

- When the seasonal indices are assumed to be *stable* over time, they can be estimated by the "ratio to moving average" (RMA) method, illustrated on the following slides.
- This is also the method commonly used in forecasting software such as Statgraphics, and it can be easily implemented in spreadsheet models:
  - http://people.duke.edu/~rnau/411outbd.htm
  - http://www.exceluser.com/excel_dashboards/seasonality-sales.htm

---

- *Time-varying* seasonal indices can also be estimated with the Census Bureau's X-12 ARIMA program or Winters' seasonal exponential smoothing model.

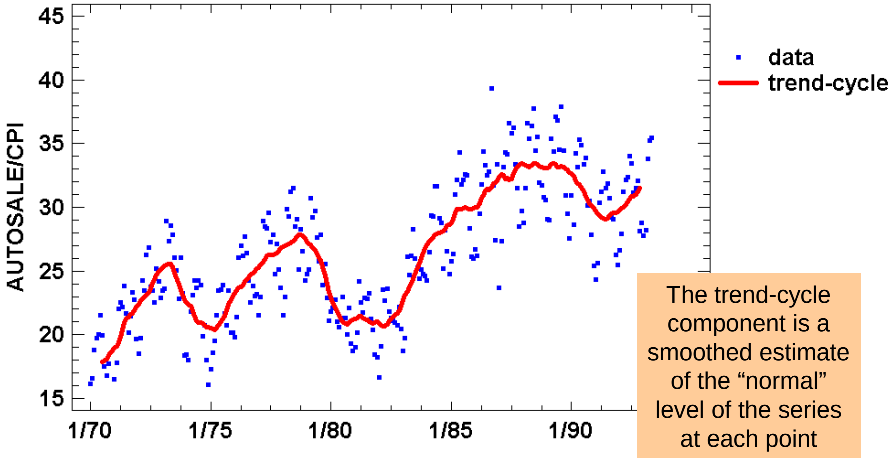
Example: U.S. total retail auto sales 1970-1993, deflated by CPI



Seasonal adjustment by RMA method: step 1

A “trend-cycle” component is estimated by a one-year-wide *centered moving average*

Trend-Cycle Component Plot for AUTOSALE/CPI

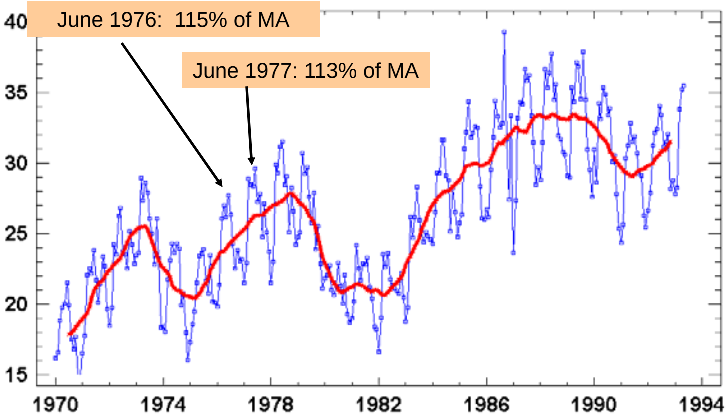


The trend-cycle component is a smoothed estimate of the “normal” level of the series at each point

Seasonal adjustment by RMA method: step 2

Compute the ratio of each observation to the value of the moving average at the same point in time

Trend-Cycle Component Plot for AUTOSALE/CPI



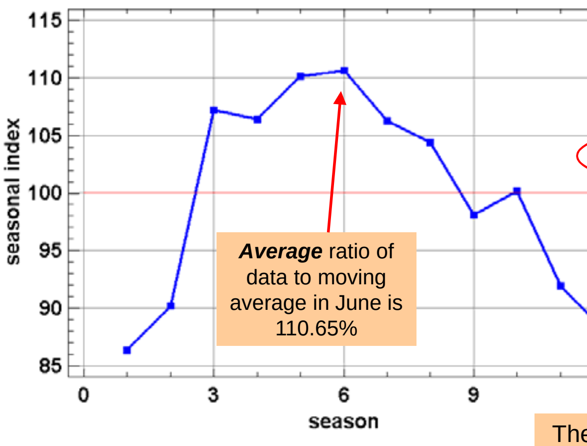
June 1976: 115% of MA

June 1977: 113% of MA

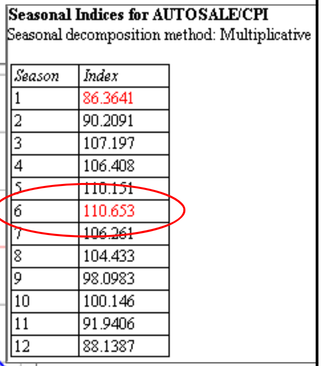
Seasonal adjustment by RMA method: step 3

Seasonal indices are estimated by averaging the ratio of original series to the moving average *by season* (month)

Seasonal Index Plot for AUTOSALE/CPI



Average ratio of data to moving average in June is 110.65%



Seasonal Indices for AUTOSALE/CPI

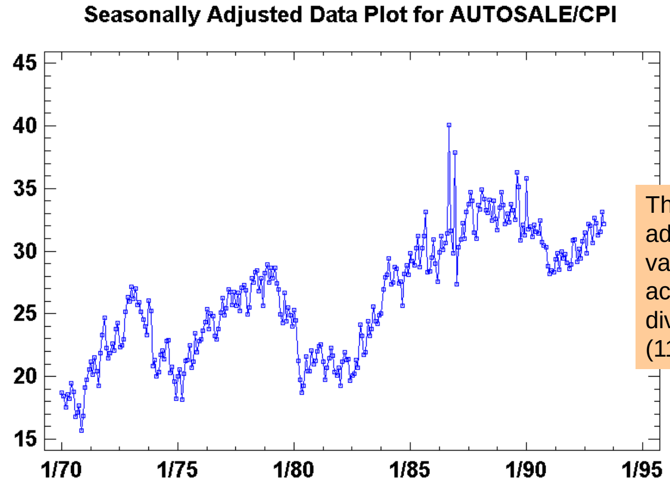
Seasonal decomposition method: Multiplicative

Season	*Index*
1	86.3641
2	90.2091
3	107.197
4	106.408
5	110.151
6	110.653
7	106.261
8	104.433
9	98.0983
10	100.146
11	91.9406
12	88.1387

The indices are renormalized if necessary so that their average is exactly 100%

Seasonal adjustment by RMA method: step 4

Seasonally adjusted series is the *original series divided by the seasonal indices*



The seasonally adjusted June values are the actual values divided by 1.1065 (110.65%)

Census Dept’s X-12 ARIMA seasonal adjustment procedure

- More bells and whistles: adjusts for trading days and uses forward and backward forecasting to avoid data loss at ends of series
- It begins by automatically fitting an ARIMA* model with regression variables to adjust for trading days, trends, etc., and uses it to forecast both forward and backward.
- Short-term tapered moving averages are then used to estimate time-varying seasonal indices on the extended data.
- Can be downloaded from this web site:

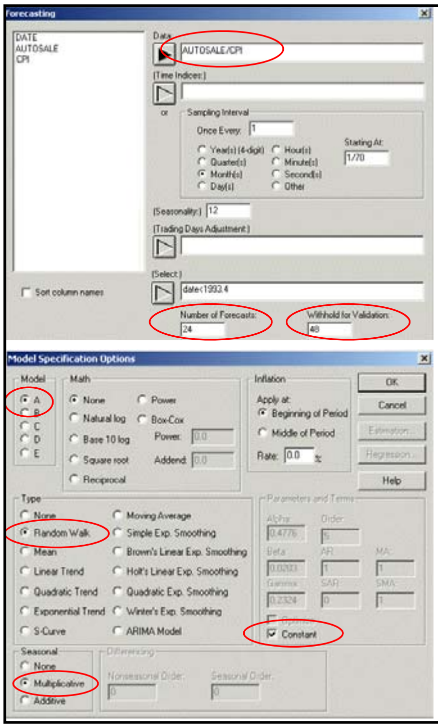
https://www.census.gov/srd/www/x12a/

Forecasting with seasonal adjustment

1. Seasonally adjust the data
2. Forecast the seasonally adjusted series using a nonseasonal model (e.g., random walk, linear trend, or exponential smoothing)
3. “Re-seasonalize” the forecasts *and confidence limits* by multiplying by the appropriate seasonal indices

- Statgraphics does this automatically when seasonal adjustment is used as a model option.
- See http://people.duke.edu/~rnau/411avg.htm for a discussion of averaging and smoothing models.

User-specified forecasting procedure in Statgraphics



Data: AUTOSALE/CPI

(Time Indices:)

Sampling Interval: Once Every: 1 ● Month(s) Starting At: 1/70

(Seasonality:) 12

(Trading Days Adjustment:)

(Select:) date<1993.4

Number of Forecasts: 24

Withhold for Validation: 48

Model: ● A

Math: ● None

Inflation: Apply at: ● Beginning of Period Rate: 0.0 %

Type: ● Random Walk

Parameters and Terms: Alpha: 0.4776 Order: 5 Beta: 0.0203 AR: 1 MA: 1 Gamma: 0.2324 SAR: 0 SMA: 1 ☑ Constant

Seasonal: ● Multiplicative

Differencing: Nonseasonal Order: 0 Seasonal Order: 0

Up to 5 models (code-named A-B-C-D-E) can be specified using various combinations of data transformations, seasonal adjustment, and model types.

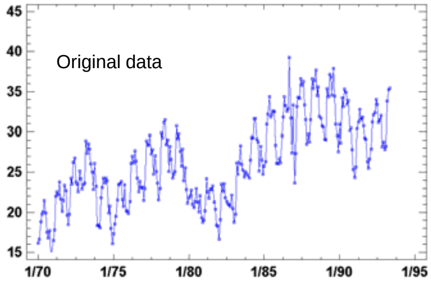
Here Model A is specified as random-walk-with-(constant)-drift together with multiplicative seasonal adjustment.

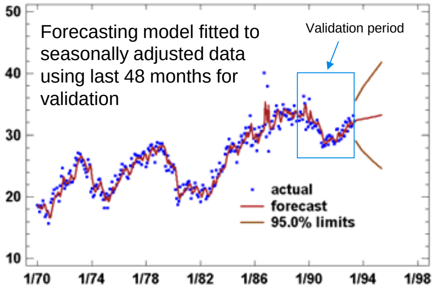
Output includes tables with side-by-side comparisons of error statistics for estimation and validation periods.

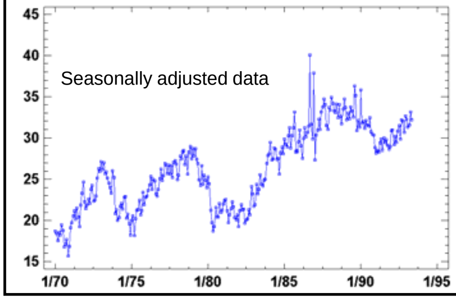
- The following pages show the results of fitting 4 different models to deflated auto sales, in combination with multiplicative seasonal adjustment:
  - A. Random walk
  - B. Simple moving average
  - C. Simple exponential smoothing
  - D. Holt's linear exponential smoothing

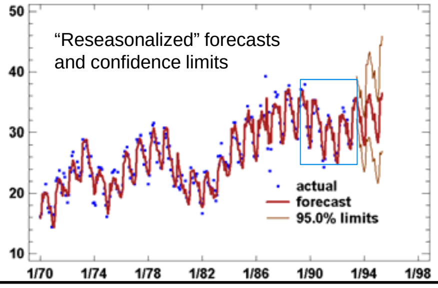
- Data from January 1970 to May 1993 (281 observations) was used.
- The last 48 months were held out from parameter estimation and used for out-of-sample validation.
- Forecasts were generated for the next 24 months.

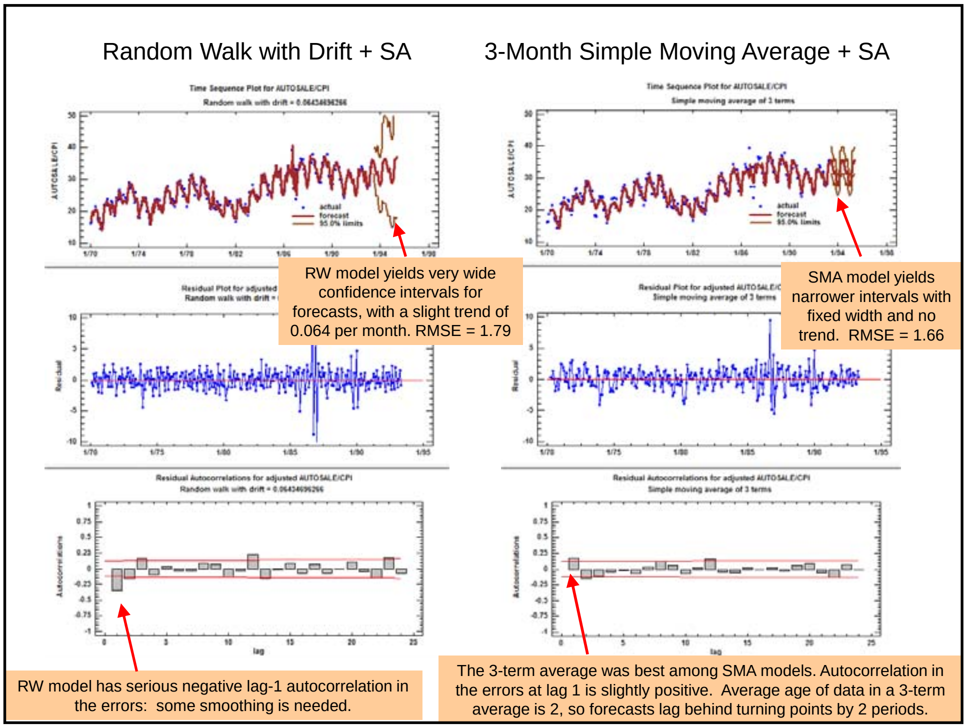
Forecasting with seasonal adjustment











Random Walk with Drift + SA

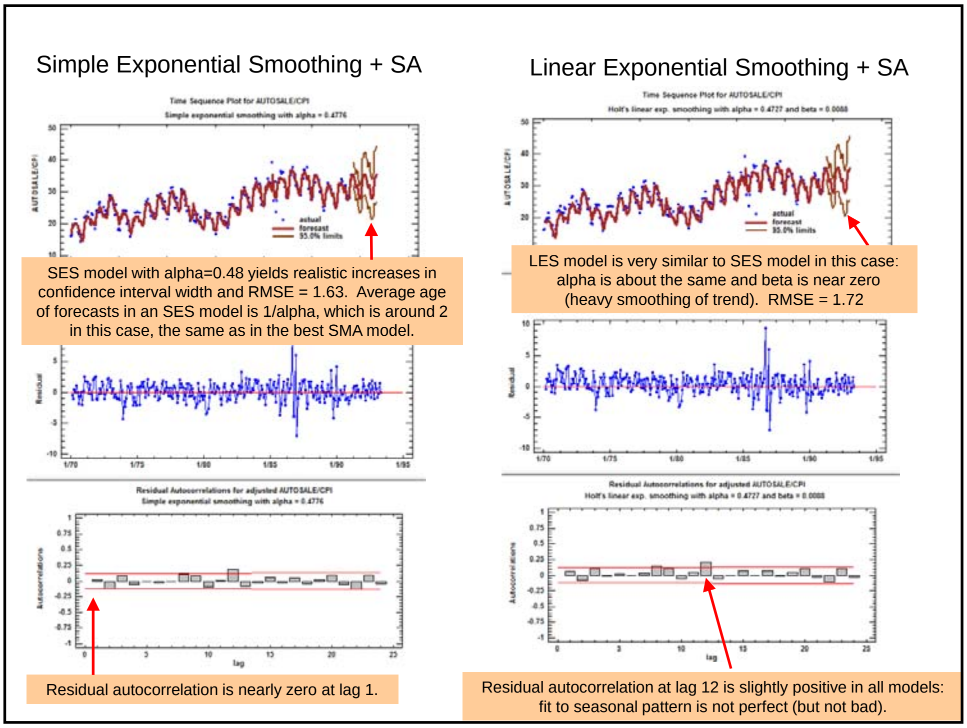
3-Month Simple Moving Average + SA

RW model yields very wide confidence intervals for forecasts, with a slight trend of 0.064 per month. RMSE = 1.79

SMA model yields narrower intervals with fixed width and no trend. RMSE = 1.66

RW model has serious negative lag-1 autocorrelation in the errors: some smoothing is needed.

The 3-term average was best among SMA models. Autocorrelation in the errors at lag 1 is slightly positive. Average age of data in a 3-term average is 2, so forecasts lag behind turning points by 2 periods.



Simple Exponential Smoothing + SA

Linear Exponential Smoothing + SA

SES model with alpha=0.48 yields realistic increases in confidence interval width and RMSE = 1.63. Average age of forecasts in an SES model is 1/alpha, which is around 2 in this case, the same as in the best SMA model.

LES model is very similar to SES model in this case: alpha is about the same and beta is near zero (heavy smoothing of trend). RMSE = 1.72

Residual autocorrelation is nearly zero at lag 1.

Residual autocorrelation at lag 12 is slightly positive in all models: fit to seasonal pattern is not perfect (but not bad).

Winters' Seasonal Smoothing

- The logic of Holt's LES model can be extended to recursively estimate *time-varying seasonal indices* as well as level and trend.
- Let L_t, T_t, and S_t denote the estimated level, trend, and seasonal index at period t.
- Let s denote the number of periods in a season.
- Let α, β, and γ denote *separate smoothing constants** for level, trend, and seasonality

*numbers between 0 and 1: smaller values → more smoothing

Winters' model formula for updated level

1. Updated level L_t is an interpolation between the *seasonally adjusted* value of the most recent data point and the previous forecast of the level:

$$L_t = \underbrace{\alpha \frac{Y_t}{S_{t-s}}}_{\text{Seasonally adjusted value of } Y_t} + (1-\alpha)\underbrace{(L_{t-1} + T_{t-1})}_{\text{Forecast of } L_t \text{ made at period } t\text{-1 is current estimated level plus current estimate trend}}$$

Winters’ model formula for updated trend

2. Updated trend T_t is an interpolation between the change in the estimated level and the previous estimate of the trend:

$$T_t \;=\; \beta\underbrace{(L_t - L_{t-1})}_{\text{Just-observed change in the level}} + (1-\beta)\underbrace{T_{t-1}}_{\text{Previous trend estimate}}$$

Winters’ model formula for updated seasonal index

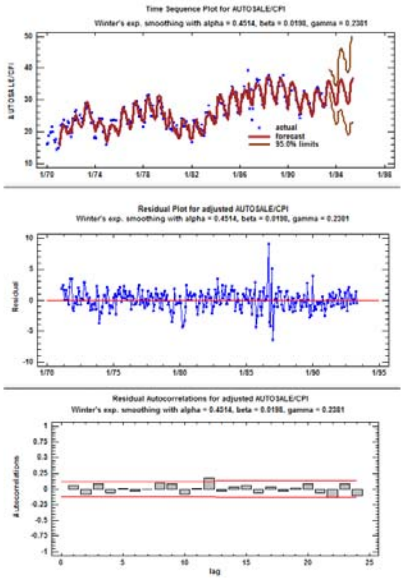
3. Updated seasonal index S_t is an interpolation between the ratio of the data point to the estimated level and the previous estimate of the seasonal index:

$$S_t \;=\; \gamma\underbrace{\frac{Y_t}{L_t}}_{\text{“Ratio to moving average” of current data point}} + (1-\gamma)\underbrace{S_{t-s}}_{\text{Last estimate of seasonal index in the same season}}$$

Winters' model formula for forecasts

4. *k*-step ahead forecast from period *t:*

$$\hat{Y}_{t+k} \;=\; \underbrace{(L_t + kT_t)}_{\substack{\text{Extrapolation of}\\ \text{level and trend}\\ \text{from period } t}}\,\underbrace{S_{t-s+k}}_{\substack{\text{Most recent estimate of}\\ \text{the seasonal index for } k^{th}\\ \text{period in the future}}}$$



Estimation of Winters model yields alpha and beta similar to LES model (0.45 and 0.02, respectively).

Gamma significantly above zero (0.24) indicates that a slowly changing seasonal pattern was detected.

RMSE = 1.72 (same as LES model).

However, note that confidence intervals for forecasts widen more rapidly for this model than for SES and LES.

It assumes a more uncertain future due to its allowance for changes in the seasonal pattern.

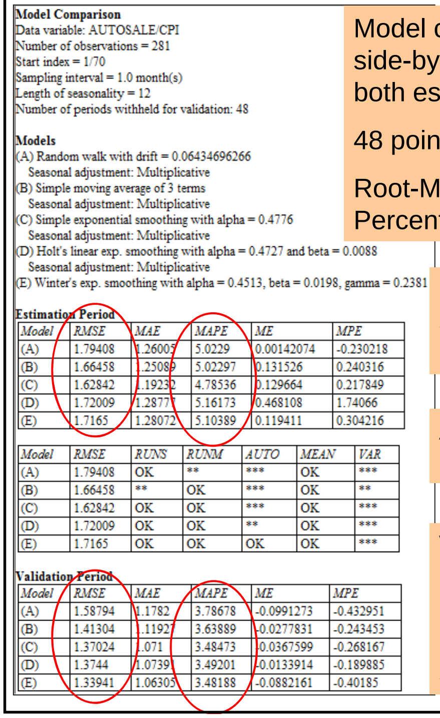
A very long history was used here for purposes of demonstration. An alternative way to deal with structural changes over time would be to use less history.

Winters’ model estimation issues

- Estimation of Winters’ model is tricky, and not all software does it well: sometimes you get crazy results.
- There are three separate smoothing constants to be jointly estimated by nonlinear least squares (α, β, γ).
- Initialization is also tricky, especially for the seasonal indices.
- Confidence intervals sometimes come out extremely wide because the model “lacks confidence in itself.”

Winters’ model in practice

- The Winters model is popular in “automatic forecasting” software, because it has a little of everything (level, trend, seasonality).
- Often it works very well, but difficulties in initialization & estimation can lead to strange results in some cases.
- It responds to recent changes in the seasonal pattern as well as the trend, but with some danger of unstable long-term trend projections.



Model Comparison
Data variable: AUTOSALE/CPI
Number of observations = 281
Start index = 1/70
Sampling interval = 1.0 month(s)
Length of seasonality = 12
Number of periods withheld for validation: 48

Models
(A) Random walk with drift = 0.06434696266
Seasonal adjustment: Multiplicative
(B) Simple moving average of 3 terms
Seasonal adjustment: Multiplicative
(C) Simple exponential smoothing with alpha = 0.4776
Seasonal adjustment: Multiplicative
(D) Holt's linear exp. smoothing with alpha = 0.4727 and beta = 0.0088
Seasonal adjustment: Multiplicative
(E) Winter's exp. smoothing with alpha = 0.4513, beta = 0.0198, gamma = 0.2381

Estimation Period

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*
(A)	1.79408	1.26005	5.0229	0.00142074	-0.230218
(B)	1.66458	1.25089	5.02297	0.131526	0.240316
(C)	1.62842	1.19232	4.78536	0.129664	0.217849
(D)	1.72009	1.28777	5.16173	0.468108	1.74066
(E)	1.7165	1.28072	5.10389	0.119411	0.304216

Model	*RMSE*	*RUNS*	*RUNM*	*AUTO*	*MEAN*	*VAR*
(A)	1.79408	OK	**	***	OK	***
(B)	1.66458	**	OK	***	OK	**
(C)	1.62842	OK	OK	***	OK	***
(D)	1.72009	OK	OK	**	OK	***
(E)	1.7165	OK	OK	OK	OK	***

Validation Period

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*
(A)	1.58794	1.1782	3.78678	-0.0991273	-0.432951
(B)	1.41304	1.11927	3.63889	-0.0277831	-0.243453
(C)	1.37024	1.071	3.48473	-0.0367599	-0.268167
(D)	1.3744	1.07391	3.49201	-0.0133914	-0.189885
(E)	1.33941	1.06305	3.48188	-0.0882161	-0.40185

Model comparison table from Statgraphics allows side-by-side comparisons of error measures in both estimation and validation periods.

48 points were held out for validation here.

Root-Mean-Squared error and Mean Absolute Percentage error are most important.

SES model is the winner by a slight margin in the estimation period on both RMSE and MAPE.

All three exponential smoothing models are pretty close in the validation period.

What's the bottom line? Smaller errors are better, but reasonableness of model assumptions (with respect to issues such as trend, volatility, and constancy of seasonal pattern) are also important.

Take-aways

- *Seasonal adjustment* (estimating and adjusting for a stable seasonal pattern) is useful both for description & prediction.
- Seasonal adjustment is traditionally performed by the *"ratio-to-moving-average" method* (with bells and whistles).
- Non-seasonal time series forecasting models (random walk, simple and linear moving averages, etc.) can be extended to seasonal data by combining with seasonal adjustment.
- Winter's model does everything at once and allows for time-varying seasonal indices, but it's a "black box": you don't see any details of the seasonal adjustement process
- Deflation (dividing by CPI) was used here to remove the inflationary component of trend & stabilize variance & seasonal pattern. Logging would have been an alternative.
- Further analysis of this same data, using ARIMA models, is here: http://people.duke.edu/~rnau/seasarim.htm