TIME SERIES ANALYSIS LECTURE 1

datascience@berkeley

Introduction to Time Series Analysis

How Is Univariate Time Series Analysis (TSA) Different From Classical Linear Regression Analysis?

- **Focus:**
  - In univariate TSA, we focus on one series.
    - Example: the monthly unemployment rates in the United States between 1950 and 2014.
  - In regression analysis, we focus on using a set of independent variables to explain the dependent variable. Put differently, we are trying to characterize the conditional mean function of the dependent variable, conditional on the independent variables.
    - Example: Student and family characteristics are used to "explain" student test scores.
- **Dependency among observations**
  - In univariate TSA, the observations are very likely to be dependent of each other. In fact, it is the dependency structure of the series that we want to characterize.
  - In CLR, the observations within each of the variables are assumed to be uncorrelated with each other. (Remember the random sampling assumption (i.e., iid) when we studied the Classical Linear Regression models?)
- **Data:**
  - In univariate TSA, the data consist only of **T** observations of the same variable; the dataset takes the form of a **Tx1** vector.
  - In classical linear regression analysis, the data consist of **n** observations of the dependent variable and a set of **k** independent variables, each of which also consists of **n** observations. Therefore, the dataset takes the form of a **n x k** matrix.

Objectives and Applications of Time Series Analysis

Time series analysis has many applications:

1. Capture the key pattern observed in the data, such as the long-term trend in climate, global average surface temperature in the last 120 years, or an annual growth rate in national health care expenditure in the United States.
2. Predict future values of a variable of interest, such as house price index, consumer spending, population size, sales, and exchange rates (e.g., USD/EUR).
3. Separate (or filter) "signals" from "noise," such as removing the seasonal component when projecting the long-term trend of red wine sales.
4. Test hypotheses using historical data, such as the hypothesis related to global warning.
5. Simulation: The water level of a reservoir depends heavily on daily water input to the system. If the input is modeled using a time series model, once the model is estimated, it can be used to "simulate" a large number of independent sequences of daily inputs. Combining the size and mode of operation of a particular reservoir with the simulated inputs, one can determine under which input conditions and at which timing the reservoir will run out of water.

Time Series Analysis Plays an Important Role in Forecasting

Time series analysis plays an important role in forecasting :

1. Governments forecast tax revenue, retail sales, unemployment claims, etc.
2. Companies forecast sales, consumers' demand, expenditure, and labor cost for strategic planning.
3. National school enrollments projection, which can be used for facilities usage and faculty recruitment planning.
4. Natural gas suppliers use TSA to forecast demand to determine the number of orders to place from the offshore fields.
5. Airlines forecast future capacity for fleet-expansion decision.
6. Retail stores, such as Walmart, Target, and Home Depot, forecast demand during holiday seasons to plan for inventory level and staffing.

Example 1: Government Budget and Key Economic Indictor Projections

Deficits Projected in CBO's Baseline

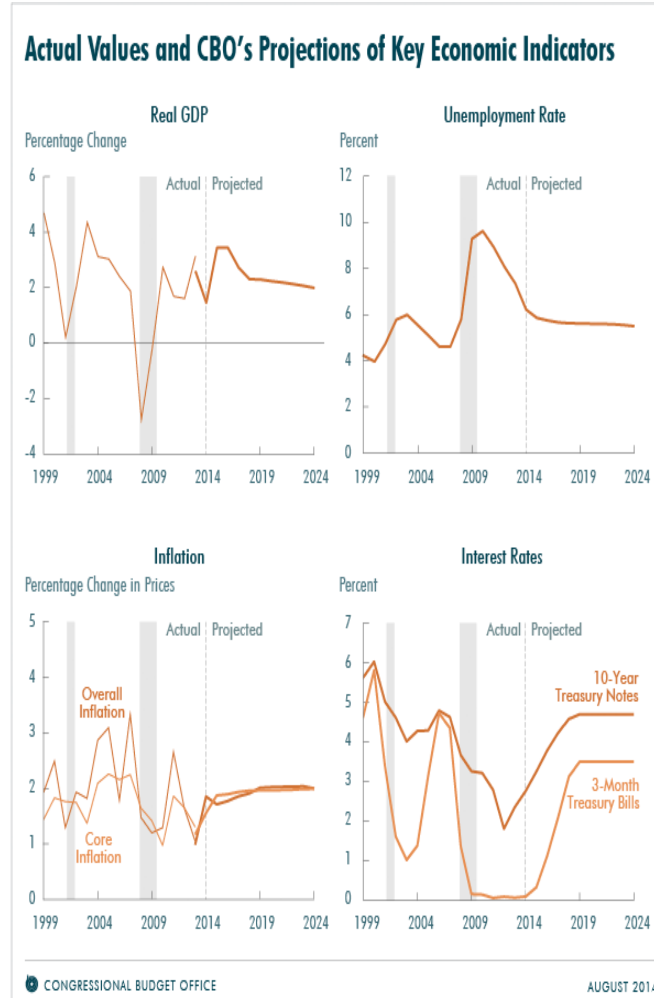
	Actual, 2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total 2015-2019	Total 2015-2024
Revenues	2,775	3,006	3,281	3,423	3,605	3,748	3,908	4,083	4,257	4,446	4,644	4,850	17,965	40,243
Outlays	3,455	3,512	3,750	3,979	4,135	4,308	4,569	4,820	5,076	5,391	5,601	5,810	20,741	47,439
Total Deficit	**-680**	**-506**	**-469**	**-556**	**-530**	**-560**	**-661**	**-737**	**-820**	**-946**	**-957**	**-960**	**-2,777**	**-7,196**
Net Interest	221	231	251	287	340	412	492	566	627	687	746	799	1,782	5,207
Primary Deficit[a]	-459	-275	-218	-269	-190	-148	-169	-170	-193	-259	-211	-161	-995	-1,989
Memorandum (As a percentage of GDP):														
Total Deficit	-4.1	-2.9	-2.6	-2.9	-2.7	-2.7	-3.0	-3.3	-3.5	-3.8	-3.7	-3.6	-2.8	-3.2
Primary Deficit[a]	-2.8	-1.6	-1.2	-1.4	-1.0	-0.7	-0.8	-0.8	-0.8	-1.1	-0.8	-0.6	-1.0	-0.9
Debt Held by the Public at the End of the Year	72.0	74.4	74.0	73.6	73.0	72.8	73.1	73.6	74.3	75.4	76.4	77.2	n.a.	n.a.

Source: Congressional Budget Office.

Note: GDP = gross domestic product; n.a. = not applicable.

a. Excludes net interest.

Source: https://www.cbo.gov/publication/45653



Example 2: Companies Forecast Sales

Nestle forecasts 2015 sales near low end of long-term target

http://www.bloomberg.com/news/articles/2015-02-19/nestle-reports-slowest-annual-sales-growth-in-five-years

---

Walmart reported decent holiday quarter results but its forecasts suggest it knows it has to raise its game a few notches.
http://fortune.com/2015/02/19/walmarts-holiday-results/

Caterpillar warns of 2015 sales hit from falling oil price … cut its 2015 profit outlook and warned the plunge in oil prices would hurt its energy equipment business.
http://www.reuters.com/article/2015/01/27/us-caterpillar-results-idUSKBN0L01E420150127

Joseph Hinrichs, Ford Motor Co.'s president of the Americas, said today the company is predicting industrywide U.S. auto sales will top 17 million next year, the most since 2001.
http://www.bloomberg.com/news/articles/2014-09-25/ford-s-hinrichs-sees-auto-sales-at-14-year-high-in-2015

Example 3: NCES Projections of Education Statistics

- National Center for Education Statistics makes projections for enrollment, graduates, teachers, and expenditures in both public and private schools.
- Multiple methods, such as single and double exponential smoothing and linear regressions, were used.

Source: http://nces.ed.gov/pubs2013/2013008.pdf

An excerpt from NCSE Projections of Education Statistics to 2021, pp. 81:

When using single exponential smoothing for a time series, P_t, a smoothed series, $\hat{P}$, is computed recursively by evaluating

$$\hat{P}_t = \propto P_t + (1 - \propto)\hat{P}_{t-1}$$

where $0 < \propto \leq 1$ is the smoothing constant.

By repeated substitution, we can rewrite the equation as

$$\hat{P}_t = \propto \sum_{s=0}^{t-1} (1 - \propto)^s P_{t-s}$$

where time, s, goes from the first period in the time series, 0, to time period t-1.

The forecasts are constant for all years in the forecast period. The constant equals

$$\hat{P}_{T+k} = \hat{P}_T$$

where T is the last year of actual data and k is the kth year in the forecast period where $k > 0$.

These equations illustrate that the projection is a weighted average based on exponentially decreasing weights. For higher smoothing constants, weights for earlier observations decrease more rapidly than for lower smoothing constants.

For each of the approximately 1,200 single exponential smoothing equations in this edition of *Projections of Education Statistics*, a smoothing constant was individually chosen to minimize the sum of squared forecast errors for that equation. The smoothing constants used to produce the projections in this report ranged from 0.001 to 0.999.

Example 4: CMS Projection on National Health Expenditure

- The Centers for Medicare & Medicaid Services (CMS) produces short-term (11 years) projections of health care spending for categories in the National Health Expenditure Accounts annually.

- Its baseline projection relies on an econometric model estimated using time series data.

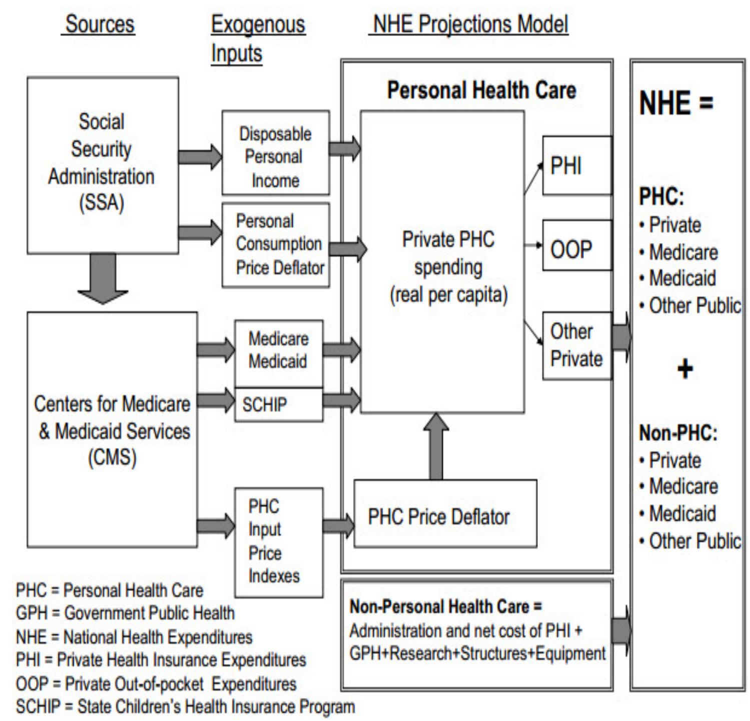
- As is common in many statistical and econometric applications, the empirical model is built on under theories (and in this case, economic theory and health economics literature).

- Importantly, it relies on other inputs (such as macroeconomic forecast) that are themselves projection from other econometric models.

a. Personal health care (PHC) spending

The Baseline NHE Projection Model for health spending by all private sources of funds is an econometric model that is estimated based on time series data from the historical National Health Expenditures. The structure and parameters of the model draw on standard economic theory and the health economics literature. The model is reestimated annually following the release of updated data for the NHEA. The fit and appropriateness of model specifications for individual series are reviewed at this time.

The diagram below provides a schematic view of the aggregate health sector within the Baseline NHE Projections Model and shows the linkages among the data sources, exogenous data, the personal health care (PHC) model, the non-PHC output, and the aggregate baseline NHE projections.



Source: Projections of National Health Expenditures (7/28/2011)

Berkeley SCHOOL OF INFORMATION