Earnings Projections Using Linear Algebra

Samantha Armstrong
Sponsor: Michael Crumley

Fundamentals of Forecasting

Forecasting is the process of making predictions of future values based on past performance, quantifying the cost of making errors in those predictions, and the confidence one should place in them. Some common subjects of forecasting are:

- Monthly sales
- Stock market performance
- Home values
- Interest rates
- Economic growth rates
- Unemployment rates

Some Considerations when Forecasting

- Does my model reflect the fundamentals of the thing to be forecasted?
- What type of business is it? What types of patterns should I expect to see in the data?
- Is the data of a periodic nature?
- What is the cost of being wrong in my forecast? How wrong can I afford to be? How wrong should I expect to be?
- Has my model been artificially constructed to fit the data? Have I considered that the quantity may be, to some degree, fundamentally unpredictable?

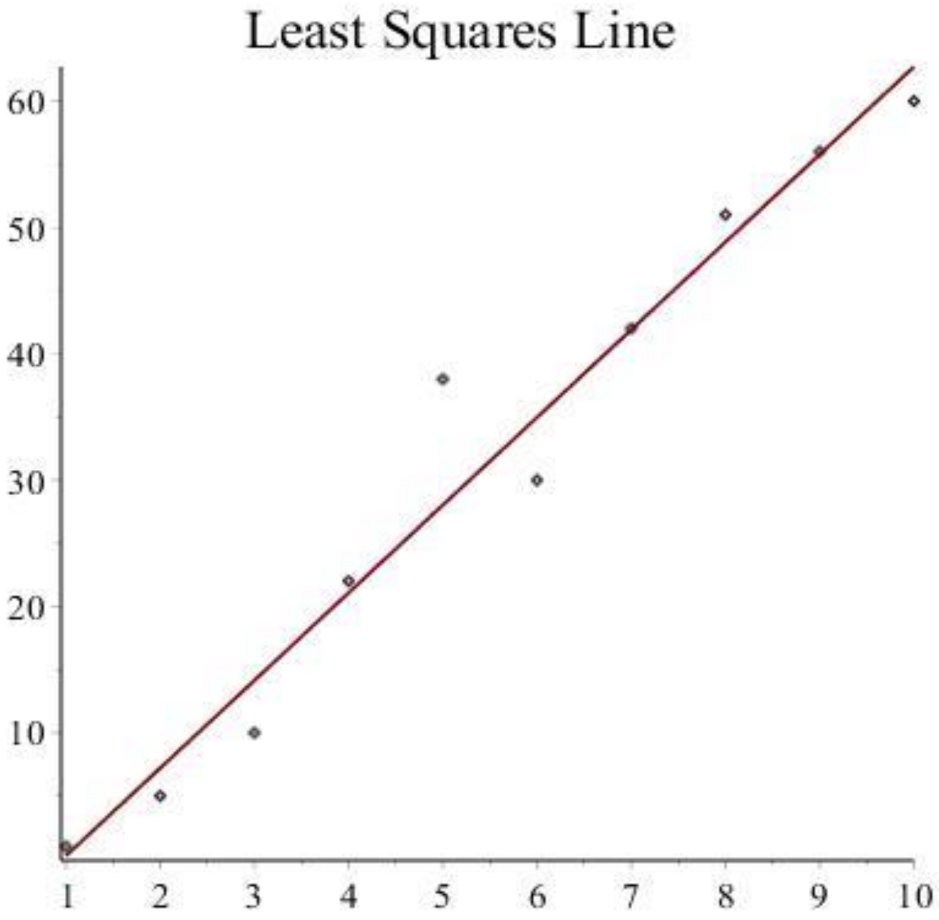
The Least Squares Line

Definition. For a set of points $(x_1, y_1), \ldots, (x_n, y_n)$, the **least squares line** $\hat{y} = b_1 x + b_0$ is the line which minimizes the squared deviations of the data to the line. It is computed as

$$b_1 = \frac{\mathbf{x} \cdot \mathbf{y} - n\bar{\mathbf{x}}\bar{\mathbf{y}}}{\mathbf{x} \cdot \mathbf{x} - n(\bar{\mathbf{x}})^2}$$

$$b_0 = \bar{\mathbf{y}} - b_1 \bar{\mathbf{x}}$$

The least squares line is very often the <u>best</u> model for data which is expected to be <u>linear</u>.

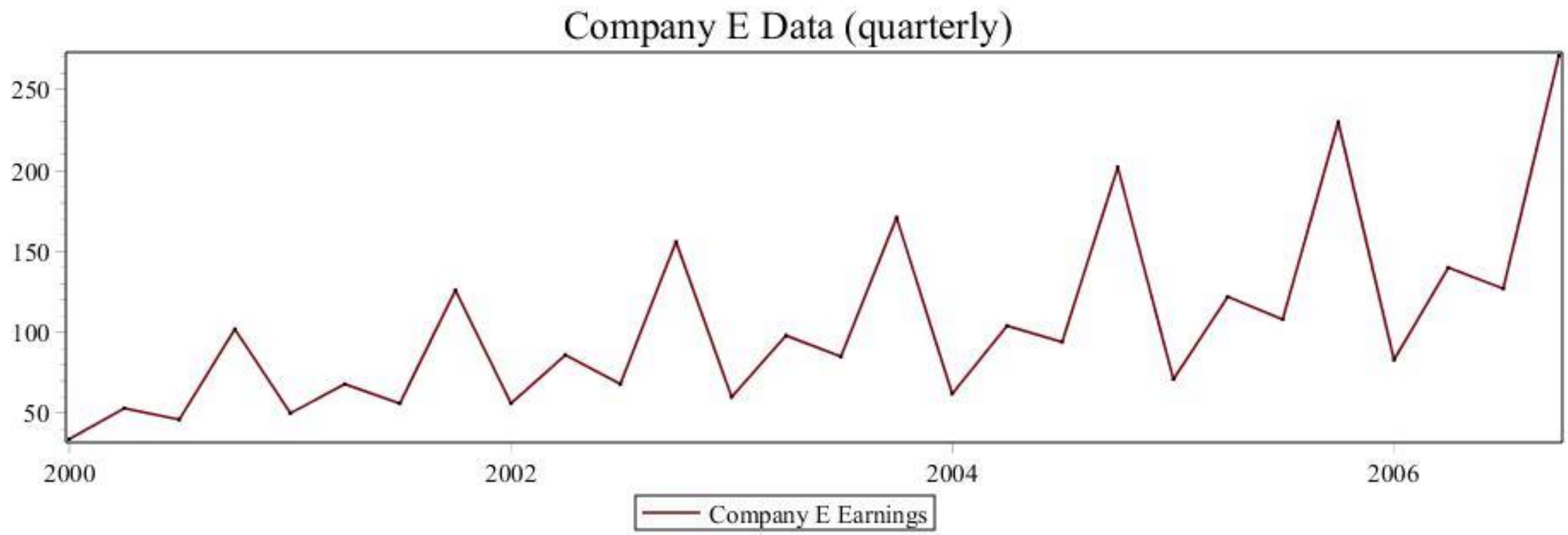


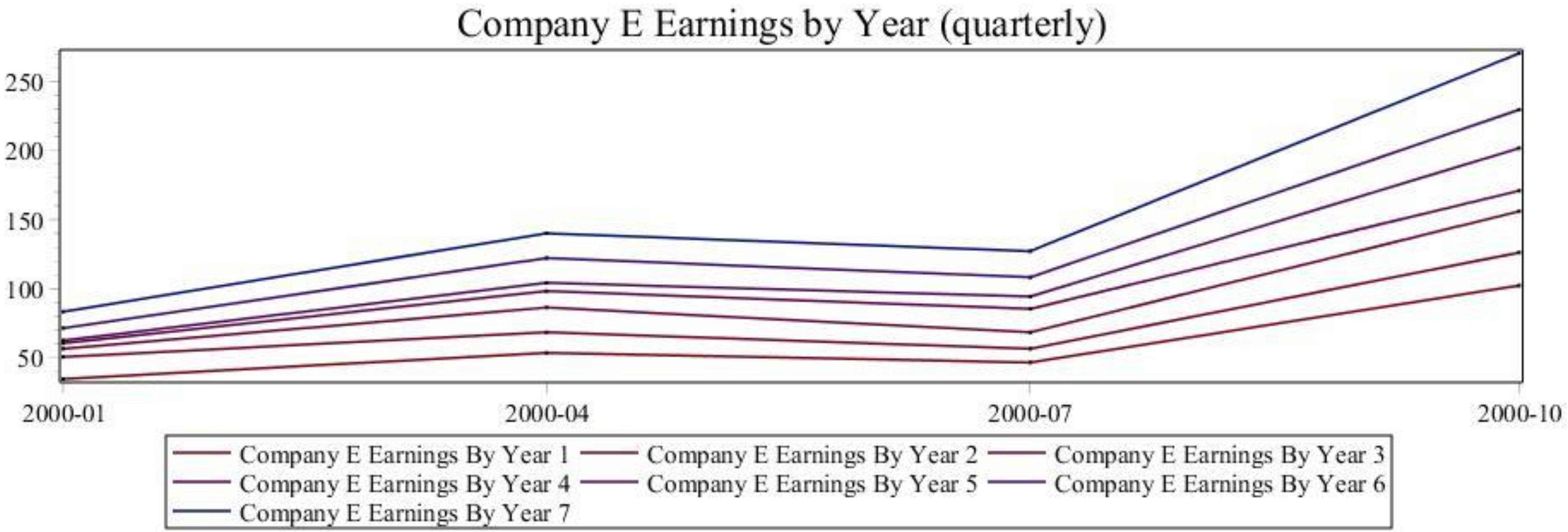
$$y = \frac{229}{33}x - \frac{20}{3}$$

Our Model: Linear trend adjusted by periodic indices

The data we consider is either monthly or quarterly data reported over several years. We denote this data by:

$$X_t = \text{value of the reported data}$$
$$\text{at } t \text{ months/quarters}$$


Company E Data (quarterly)
250
200
150
100
50
2000
2002
2004
2006
Company E Earnings


Company E Earnings by Year (quarterly)
250
200
150
100
50
2000-01
2000-04
2000-07
2000-10
Company E Earnings By Year 1
Company E Earnings By Year 2
Company E Earnings By Year 3
Company E Earnings By Year 4
Company E Earnings By Year 5
Company E Earnings By Year 6
Company E Earnings By Year 7

Major Assumptions of the Model

- The long term, year-to-year average, is a linear function L_t
- The data exhibits a periodic trend year-to-year
- The periodic effect of each month/quarter is to multiply the long term average by an appropriate constant, called the **index** of each month/quarter, unique to each month/quarter

In other words:

$$
\begin{aligned}
\widehat{X}_t &= \text{forecasted value at time } t \\
L_t &= \text{linear long term average} \\
I_t &= \text{index of month/quarter corresponding to time } t
\end{aligned}
$$

$$\widehat{X}_t = L_t I_t$$

Double Moving Average

Definition. The **double moving average** of the quarterly data X_t is

$$X_t^{**} = \frac{X_{t-2} + 2X_{t-1} + 2X_t + 2X_{t+1} + X_{t+2}}{8}$$

For monthly data it is

$$X_t^{**} = \frac{X_{t-6} + 2\sum_{k=-5..5} X_{t+k} + X_{t+6}}{24}$$

Notice: These averages are taken over an entire year.
So, the effect of monthly/quarterly variations should cancel each other out.

Company E Sales and Double Moving Average
250
200
150
100
50
2000
2002
2004
2006
Double Moving Average
Company E Sales

Indices

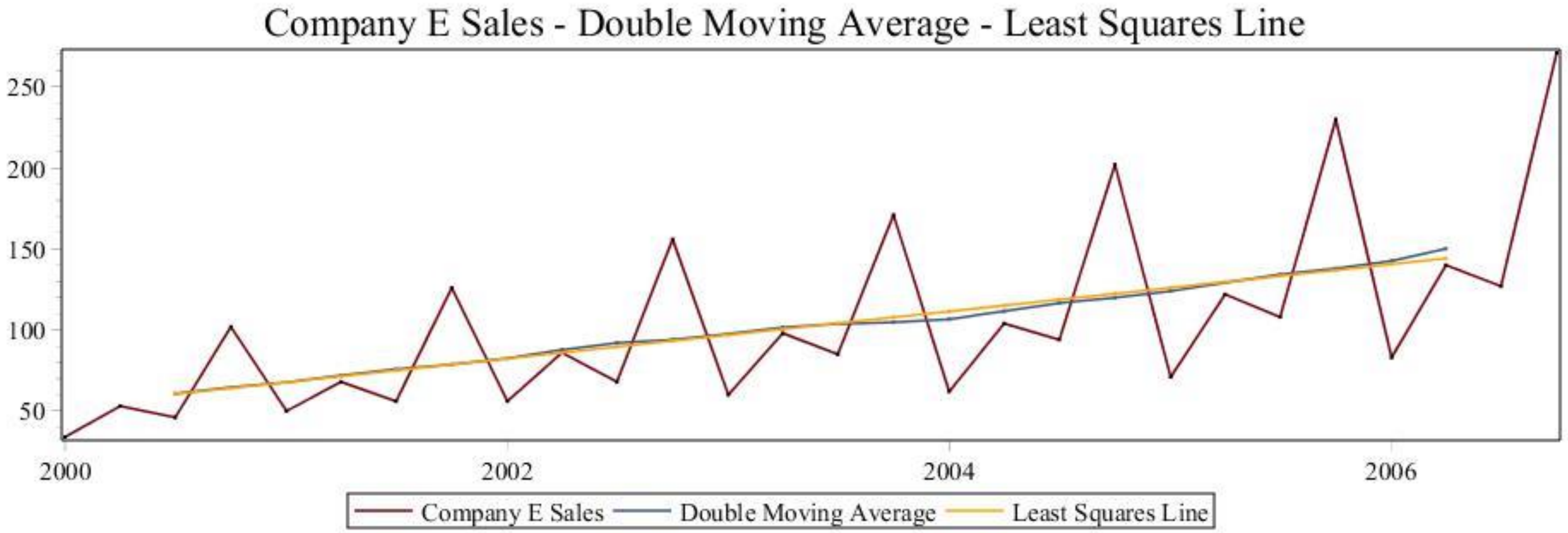
Definition. The **index** corresponding to month or quarter k, denoted I_k, is gotten by taking the median of X_t/X_t^{**} for all available data for month or quarter k:

$$I_k = \text{median}\left(X_k/X_k^{**}, X_{k+4}/X_{k+4}^{**}, X_{k+8}/X_{k+8}^{**}, \ldots\right)$$

Company E Indices

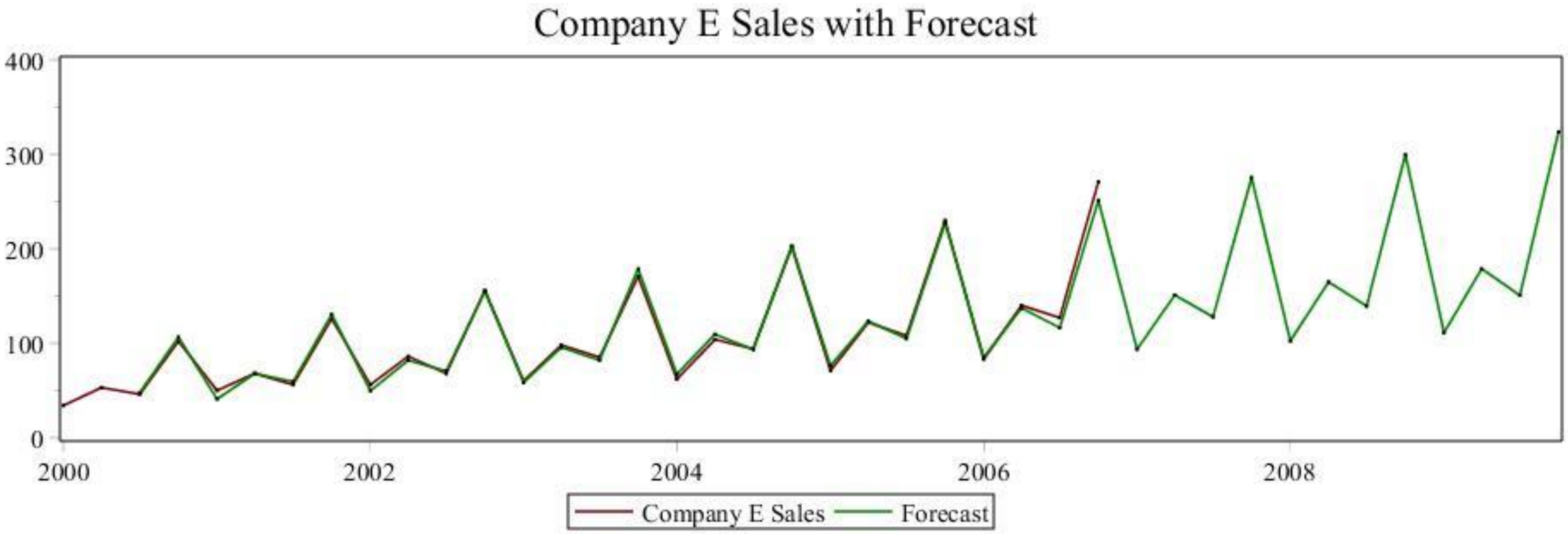
	Q1	Q2	Q3	Q4
2000	X	X	.757	1.578
2001	.738	.944	.739	1.600
2002	.679	.980	.739	1.659
2003	.615	.964	.819	1.632
2004	.581	.932	.806	1.683
2005	.572	.943	.804	1.667
2006	.582	.932	X	X
Index	.602	.951	.789	1.658

Definition. The **linearized long term average**, de-noted L_t, is the least squares line for the double moving average X_t^{**}.



What we’ve all been waiting for…

Definition. Our **forecast** $\widehat{X}_t$ for time t is $\widehat{X}_t = L_t I_t$

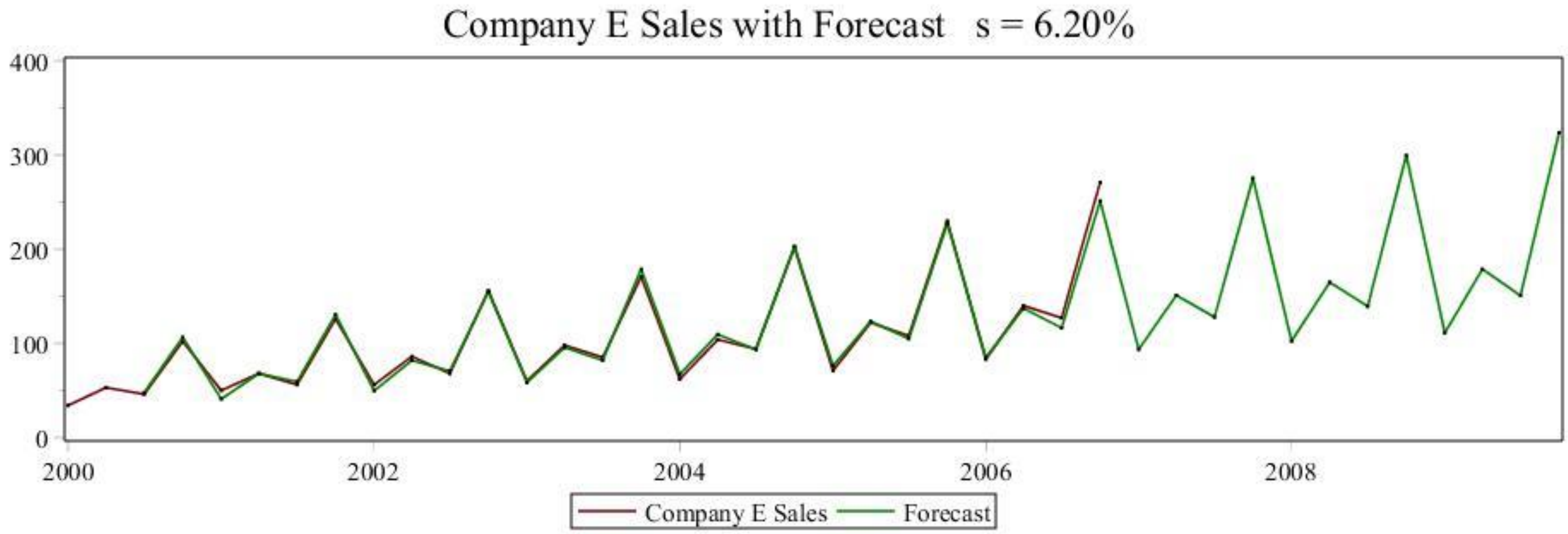


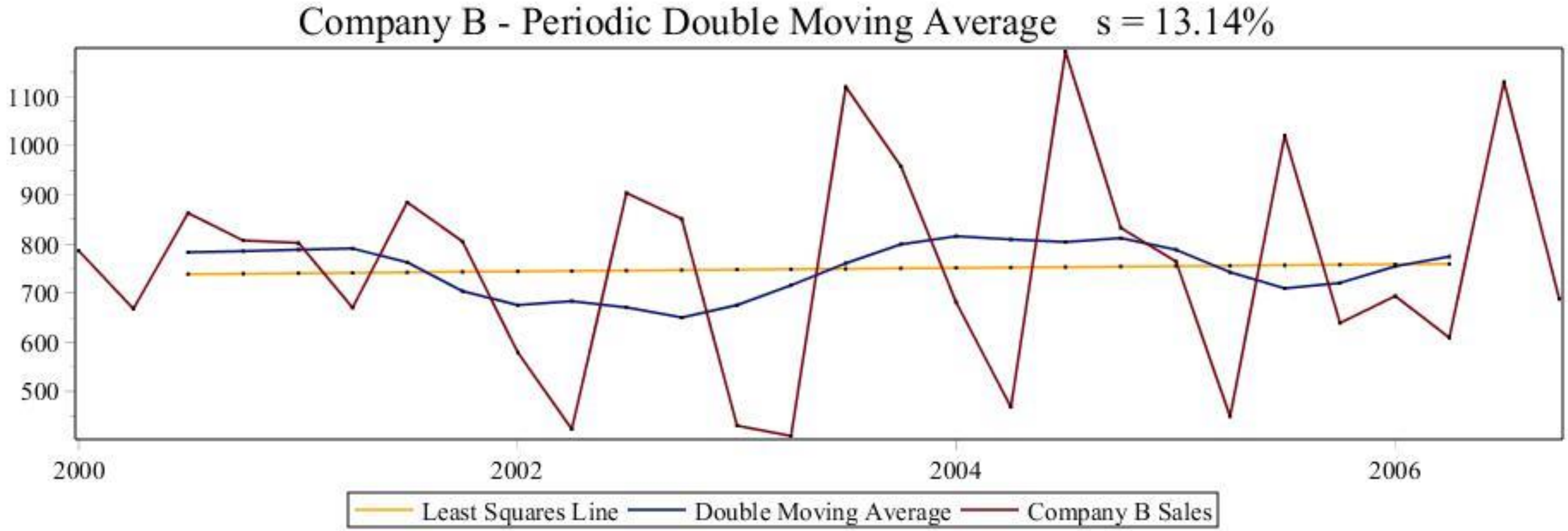
Evaluating the Goodness of the Model

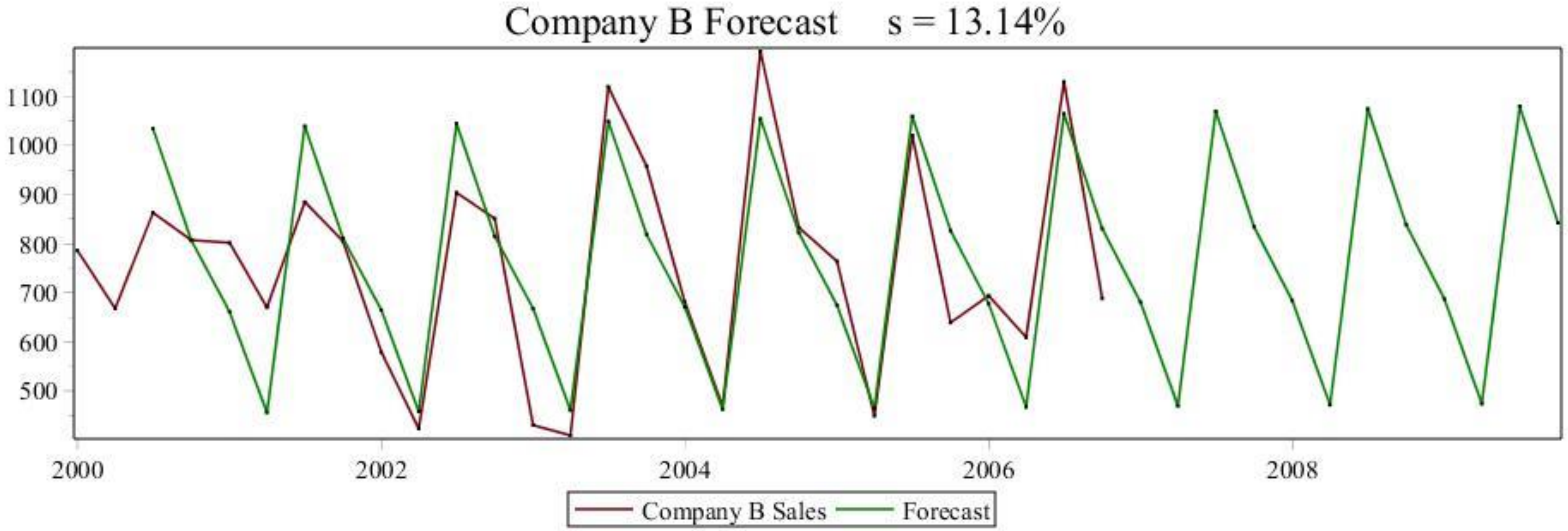
We have decided to measure the goodness of the model by the **average percentage difference** between the observed and predicted values over the last available year:

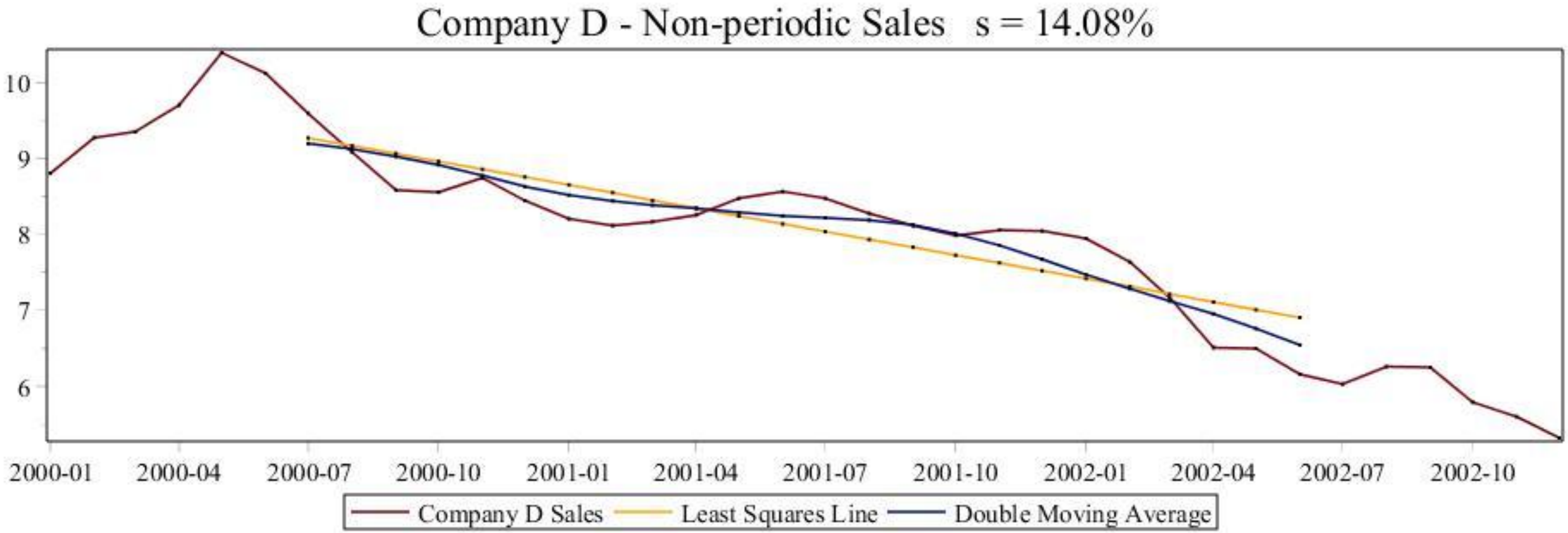
$$s = \frac{1}{n} \sum_{\text{last year}} \left| \frac{\widehat{X}_i - X_i}{X_i} \right|$$

*** Important Note: only previous years were used to construct the model $\widehat{X}_t$

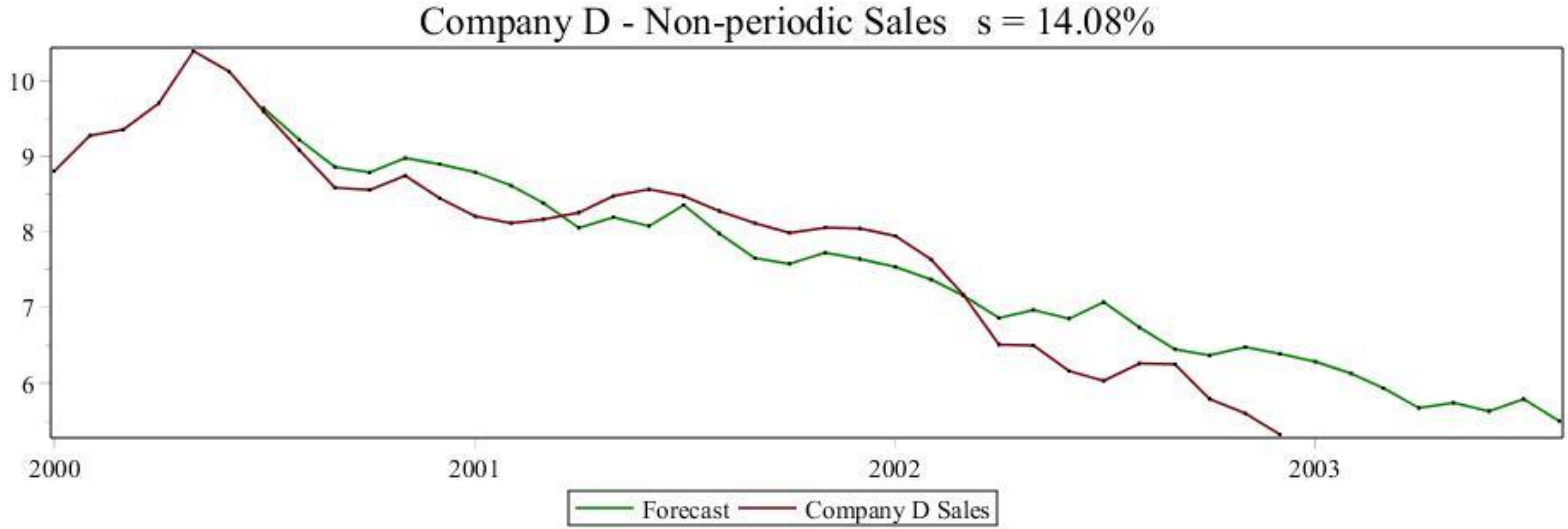

Company E Sales with Forecast s = 6.20%
400
300
200
100
0
2000
2002
2004
2006
2008
Company E Sales
Forecast


Company B - Periodic Double Moving Average s = 13.14%
1100
1000
900
800
700
600
500
2000
2002
2004
2006
Least Squares Line
Double Moving Average
Company B Sales


Company B Forecast s = 13.14%
1100
1000
900
800
700
600
500
2000
2002
2004
2006
2008
Company B Sales
Forecast

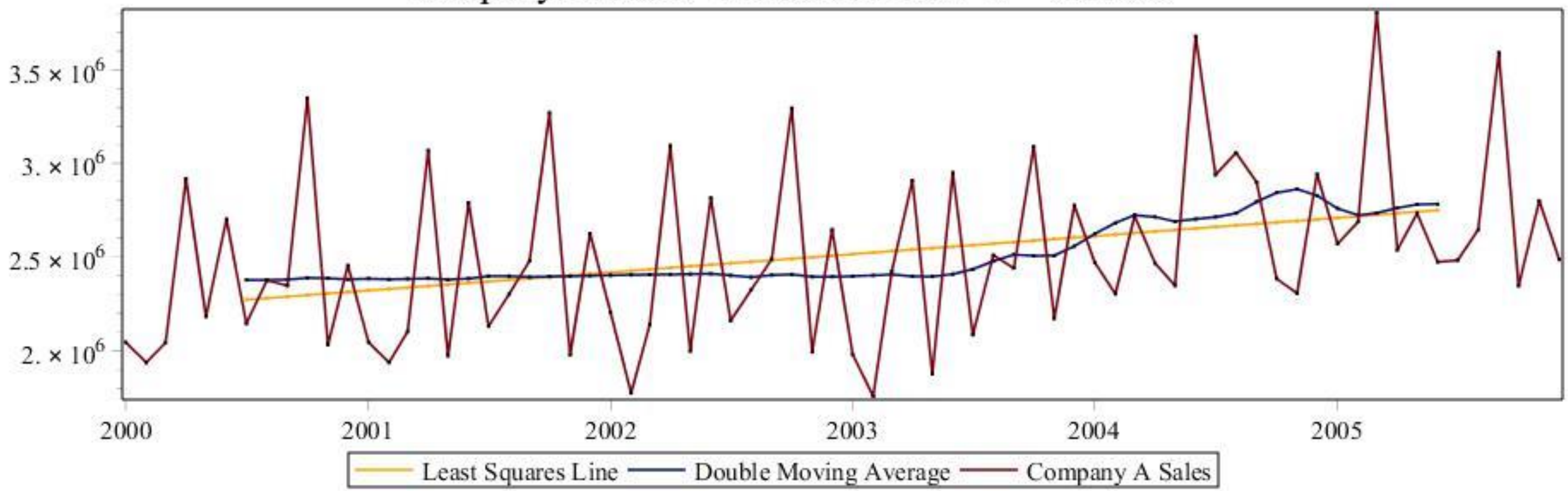


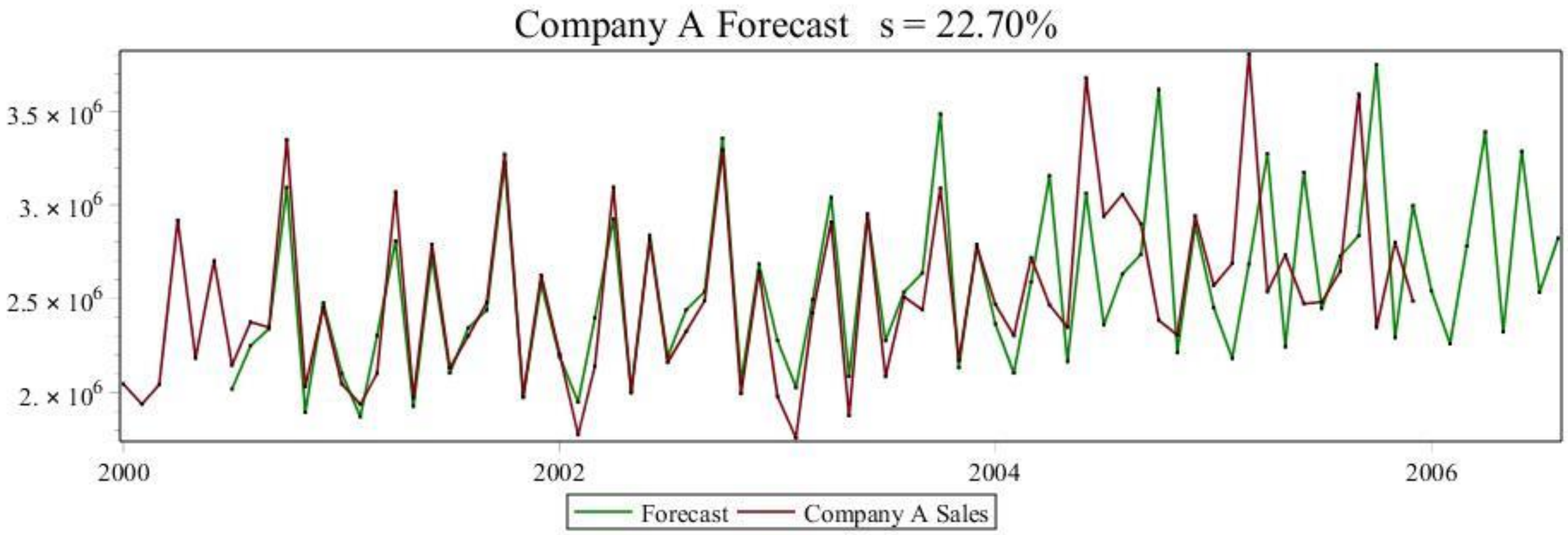
Company D - Non-periodic Sales s = 14.08%
10
9
8
7
6
2000-01
2000-04
2000-07
2000-10
2001-01
2001-04
2001-07
2001-10
2002-01
2002-04
2002-07
2002-10
Company D Sales
Least Squares Line
Double Moving Average



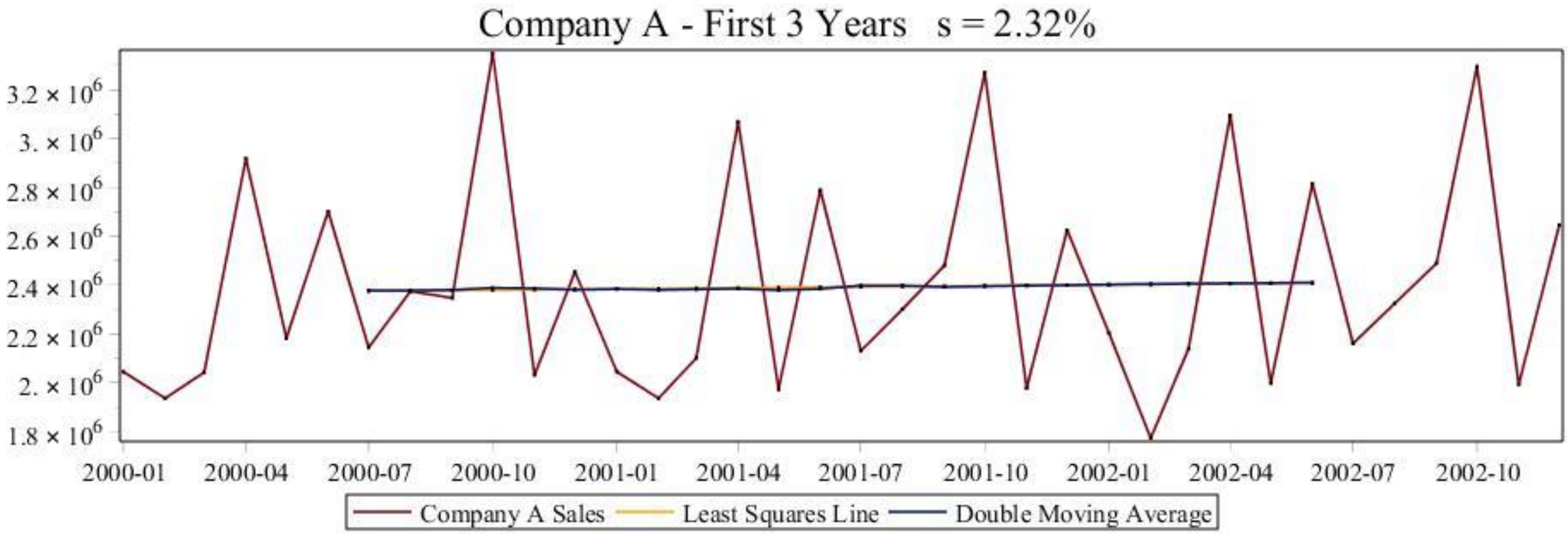
Company D - Non-periodic Sales s = 14.08%
10
9
8
7
6
2000
2001
2002
2003
Forecast
Company D Sales

Company A Sales - Periodic at first   s = 22.70%

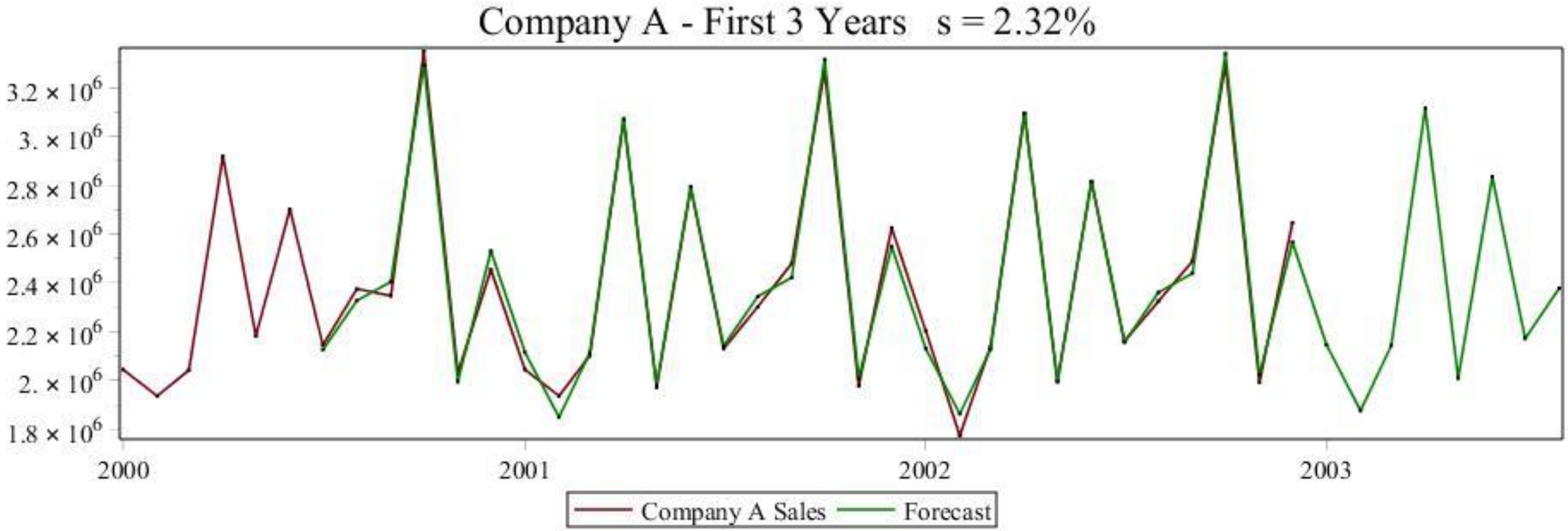




Company A Forecast s = 22.70%
3.5×10^6
$3. \times 10^6$
2.5×10^6
$2. \times 10^6$
2000
2002
2004
2006
Forecast
Company A Sales



Company A - First 3 Years s = 2.32%
3.2 × 10^6
3. × 10^6
2.8 × 10^6
2.6 × 10^6
2.4 × 10^6
2.2 × 10^6
2. × 10^6
1.8 × 10^6
2000-01
2000-04
2000-07
2000-10
2001-01
2001-04
2001-07
2001-10
2002-01
2002-04
2002-07
2002-10
Company A Sales
Least Squares Line
Double Moving Average



Company A - First 3 Years s = 2.32%
3.2 × 10^6
3. × 10^6
2.8 × 10^6
2.6 × 10^6
2.4 × 10^6
2.2 × 10^6
2. × 10^6
1.8 × 10^6
2000
2001
2002
2003
Company A Sales
Forecast

Final Remarks

Many objections could be raised to this model, depending on the particular characteristics of the data:

- Is it fair to assume the long term average is *linear*? Many financial quantities instead obey an *exponential* curve.
- Is X_t^{**} the best way to capture the long term average?
- Is it fair to assume the periodicity is *constant*? Is the effect of periodicity in fact *additive* rather than multiplicative, i.e. should we instead seek a model of the form $\widehat{X}_t = L_T + I_t$?
- Is average percentage difference the best measure of goodness? Is *absolute* instead of percentage difference a better measure of the cost of being wrong?

Thank You!

Thank you to Dr. Mike Crumley for his assistance with this project!

Reference:

Newbold and Boss, (1994). Introductory Business and Economic Forecasting (2nd ed.). Cincinatti: Southwestern Publishing Co.