Least Square Regressions and Model Building

by

Jacob Martin Vartuli-Schonberg

Submitted to the Board of Mathematics/Computer Science

in partial fulfillment of the requirements

for the degree of Bachelor of Arts

Purchase College

State University of New York

May 2015

Sponsor: Roger Tsai

Second Reader: Irina Shablinsky

Introduction:

There exists one common parameter in the gathering and extension of all disciplines of human thought. That parameter is observation. In order to establish patterns, one must first observe and collect data in systems being studied. Those in the business world make note of their revenue with respect to time. Astronomers make note of the luminosity of stars and analyze their light.

While the action of observation is important, it does not lead to a grand understanding of whatever system is being observed. You are limited to what *can* be observed. That limit can be defined, in most cases, as the bounds of the present and past. As far as I've observed, most disciplines aren't only concerned with what can be observed. They're also concerned with what might be observed.

The objective of this paper, and the consequent progression of this senior project, entails the study of one such method that allows the practitioners of science to transcend observation. To develop models of observed systems and to use those models to predict what else might come about as result of the laws and patterns of those systems, I employ methods of Mapping, Matrix Calculus and Optimization to push the boundary of observation.

In the conclusion of this paper, I will apply the theory of model development to various sets of real data including revenue data from Apple and Google and scientific data from the NIST.

Mapping:

In the course of attempting to understand the processes of model development, I initially employed the mathematical idea of a mapping to help wrap my mind around this relatively complex idea. In the following section, we will introduce the idea of a mapping which will be of great importance in understanding subsequent sections.

Point to point mapping:

Let us first consider a two dimensional plane, $P(x_{1,}x_2)$, whose orthogonal axes may be labeled arbitrarily. It should be obvious, but I will make note of it, that the intersection of the lines

$$x_1 = 0$$
$$x_2 = 0$$

is the point of origin and can be denoted by the vector

$$\vec{v} = \langle x_1, x_2 \rangle = \langle 0, 0 \rangle$$

To further extend our ability to traverse $P(x_{1,} x_2)$, we will allow any point in the plane to be defined as the intersection to the following lines.

$$x_1 = a$$
$$x_2 = b$$

For example

$$a = 1$$
$$b = 2$$

can be represented by the vector

$$\vec{v} = \langle 1, 2 \rangle$$

If we were to analyze a vector, we would obtain an arrow whose tail is the origin and whose head is the point $\langle a, b \rangle$. Something that I find beautiful is that if we were to allow for arbitrary a and b components,

$$a \in \mathbb{R}$$
$$b \in \mathbb{R}$$

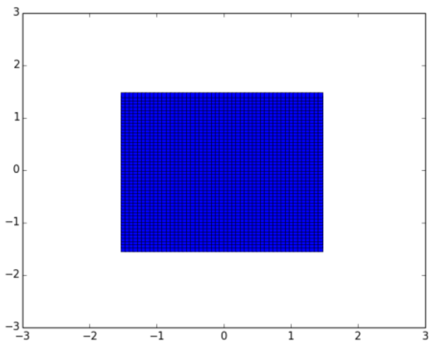
our vector would develop the entire plane. Keep this thought in mind as we proceed as its application to later sections lead to beautiful imagery.

Now that we've set up the framework of plane $P(x_{1,} x_2)$, let us consider another plane $T(y_1, y_2)$that is determined by the same principles we set in the previous section. However, points in plane T will correspond to the selection of points in plane P.

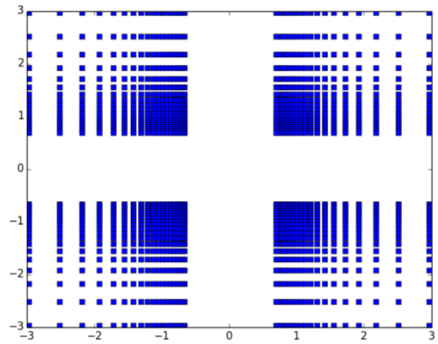
Consider the following example:

$$P(x_1, x_2) = \langle x_1, x_2 \rangle$$
$$T(y_1, y_2) = \langle y_1, y_2 \rangle = \left\langle \frac{1}{x_1}, \frac{1}{x_2} \right\rangle$$

This point mapping says the following. We have an independent space $P(x_1, x_2)$ from which we arbitrarily pick a point. This selection determines a point in our dependent space $T(y_1, y_2)$ which is obtained via the chosen component functions.



$P(x_1, x_2)$



$T(y_1, y_2)$

This mapping effectively turns the points of plane P inside out. As a variable point in $P\left(x_1, x_2\right)$ approaches the origin, the corresponding point in $T(y_1, y_2)$ will go towards infinity and conversely, if we were to have a variable point $P\left(x_1, x_2\right)$ approach infinity, our mapping $T(y_1, y_2)$ will increasingly transform that point closer to zero. There are a select few points that do not transform including:

$$P\{(-1,1),(-1,-1), (1,1),(1,-1)\}$$

I find these mappings quite interesting and warrant further study. While I have some idea what this particular point to point mapping looks like, I find myself wondering what other mysteries they hold. While these curiosities are furious, they are not the topic of this paper.

Point to Function Mapping:

Let us first consider planes $P(p_1, p_2)$ and $T(x, f(p_1, p_2, x))$. For simplicity we will initially introduce a linear polynomial whose coefficients are defined in $P(p_1, p_2)$ and exists in $T(x, f(p_1, p_2, x))$. We can represent this line in $T(x, f(p_1, p_2, x))$ algebraically in the following way.

$$f(p_1, p_2, x) = p_1(x) + p_2 \qquad\qquad p_1, p_2, x \in (-\infty, \infty)$$

However, for the purpose of furthering our understanding this concept and to remain in line with our previous sections we will rewrite this equation in Matrix form as follows.

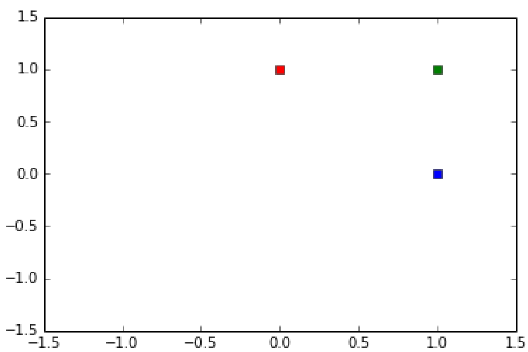
$$f(p_1, p_2, x) = \begin{bmatrix} 1 & 1 \end{bmatrix} \bullet \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \qquad p_1, p_2, x \in (-\infty, \infty)$$

In this form we have an independent vector, $\vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$, which, given mobility though a plane, $P(p_1, p_2)$, gives identity to our resultant function, $\vec{r} = \begin{bmatrix} x \\ f(p_1, p_2, x) \end{bmatrix}$. If we were to define vectors

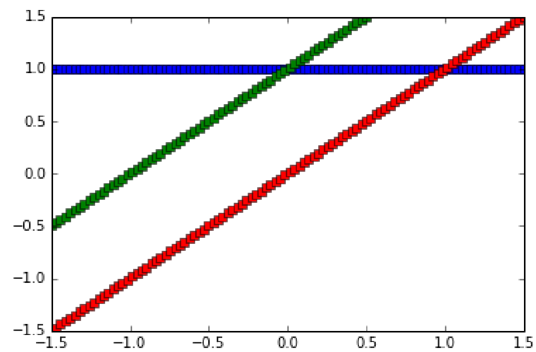
$$\vec{p}_r = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \vec{p}_g = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{and } \vec{p}_b = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{in } P(p_1, p_2)$$

we will develop the lines

$$\vec{r}_r = \begin{bmatrix} x \\ f(1,0,x) \end{bmatrix}, \ \vec{r}_g = \begin{bmatrix} x \\ f(1,1,x) \end{bmatrix} \text{and } \vec{r}_b = \begin{bmatrix} x \\ f(0,1,x) \end{bmatrix} \text{in } T(x, f(p_1, p_2, x)).$$

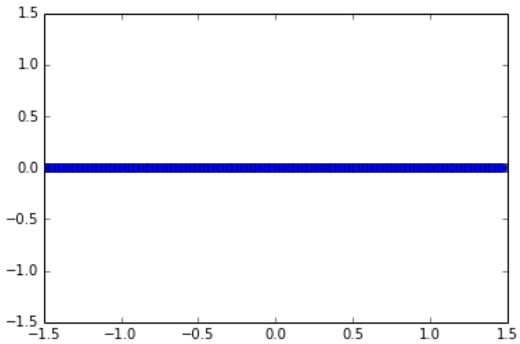


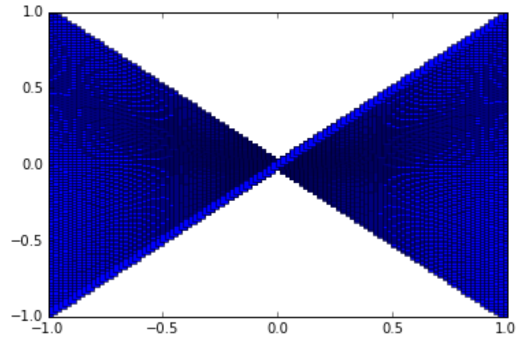
$P(p_1, p_2)$



$T(x, f(p_1, p_2, x))$

Using a variable vector $\vec{p}(p_1) = \begin{bmatrix} p_1 \\ 0 \end{bmatrix}$ where $p_1, x \in (-\infty, \infty)$, effectively maps an infinite number of functions, each of whichhas a slope dependent on the position of $\vec{p}(p_1)$. In algebraic terms, we have $f(x) = p_1 x + 0$. In the figure below we have $\vec{p}(p_1) = \begin{bmatrix} p_1 \\ 0 \end{bmatrix}$where $p_1 \in \left(-\frac{3}{2}, \frac{3}{2}\right)$ plotted in plane $P(p_1, p_2)$ and the associated lines plotted in $T(x, f(p_1, p_2, x))$.





In the previous examples we had our parameter $\vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$ and resultant vectors $\vec{r} = \begin{bmatrix} x \\ f(p_1, p_2, x) \end{bmatrix}$ where $p_1, p_2, x \in (-\infty, \infty)$ or $x \in (-\infty, \infty)$ and p_1 and p_2 be equal to some constant. We will now consider configuration which considers x to be a constant and allows $p_1, p_2 \in (-\infty, \infty)$.

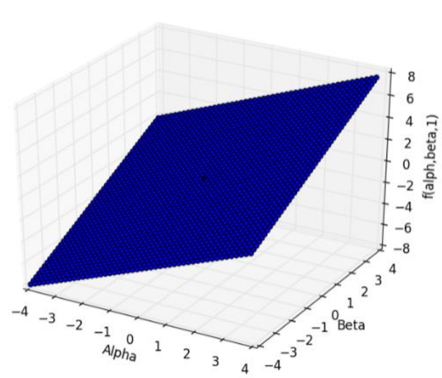
let:

$$x = x_1 = 1$$

$$p_1, p_2 \in (-\infty, \infty)$$

Hence:

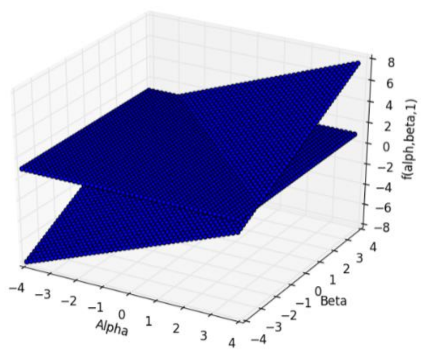
$$f(p_1, p_2, x) = \begin{bmatrix} 1 & 1 \end{bmatrix} \bullet \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = p_1 + p_2$$

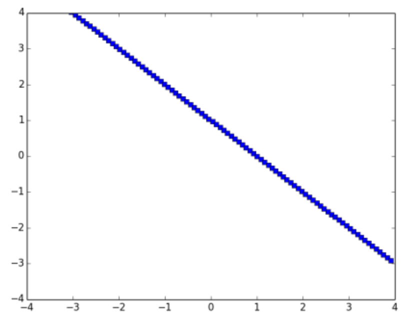


The figure represents all possible configurations of α and β and the resulting $f(p_1, p_2, x)$. If one was to receive a data point in $T(x, f(p_1, p_2, x))$ that reads:

$$\vec{r}_{x_1} = \begin{bmatrix} x_1 \\ f(p_1, p_2, x_1) \end{bmatrix} = \begin{bmatrix} 1 \\ p_1 + p_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The solution set for p_1 and p_2 will be the intersection of the above surface $\langle p_1, p_2, p_1 + p_2 \rangle$ and $\langle p_1, p_2, 1 \rangle$. And can be visualized in $P(p_1, p_2)$ as $\vec{p} = \begin{bmatrix} p_1 \\ 1 - p_1 \end{bmatrix}$





Now we will consider:

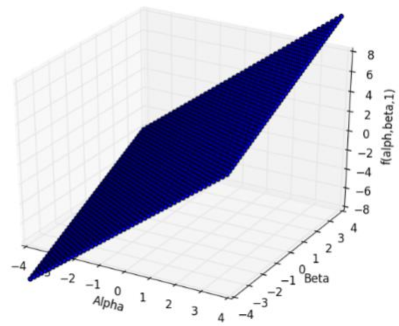
let:

$x = x_2 = 2$

$p_1, p_2 \in (-\infty, \infty)$

Hence:

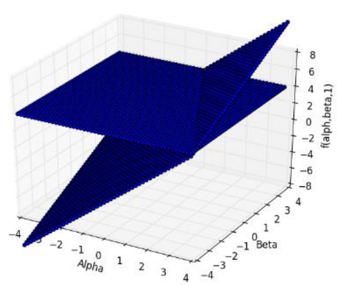
$$f(p_1, p_2, x) = \begin{bmatrix} 2 & 1 \end{bmatrix} \bullet \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = 2p_1 + p_2$$

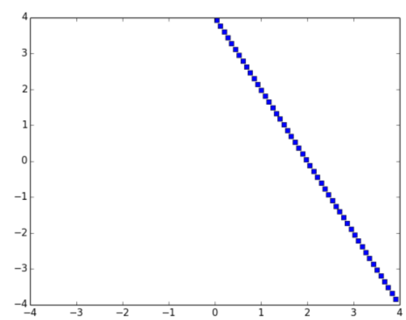


The figure represents all possible configurations of α and β and the resulting $f(p_1, p_2, x)$. If one was to receive a data point in $T(x, f(p_1, p_2, x))$ that reads:

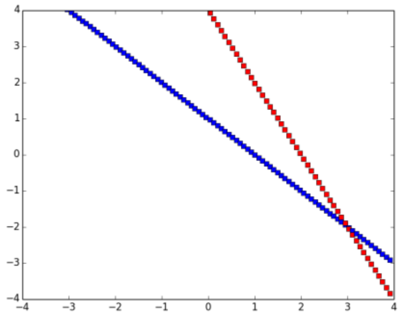
$$\vec{r}_{x_2} = \begin{bmatrix} x_2 \\ f\left(p_1, p_2, x_2\right) \end{bmatrix} = \begin{bmatrix} 2 \\ 2p_1 + p_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

The solution set for p_1 and p_2 will be the intersection of the above surface $\langle p_1, p_2, 2p_1 + p_2 \rangle$ and $\langle p_1, p_2, 4 \rangle$. And can be visualized in $P(p_1, p_2)$ as $\vec{p} = \begin{bmatrix} p_1 \\ 4 - 2p_1 \end{bmatrix}$





When we consider the solution sets of parameters for $r_{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $r_{x_2} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$



We see the intersection of these sets represents the parameters that are shared by $r_{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $r_{x_2} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$.

$$\vec{p}_{solution} = \begin{bmatrix} p_1 \\ 4-2p_1 \end{bmatrix} = \begin{bmatrix} p_1 \\ 1-p_1 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Hence the line that satisfies both points is $f(p_1, p_2, x) = \begin{bmatrix} x & 1 \end{bmatrix} \bullet \begin{bmatrix} 3 \\ -2 \end{bmatrix} = 3x - 2$

Let us now introduce a third point

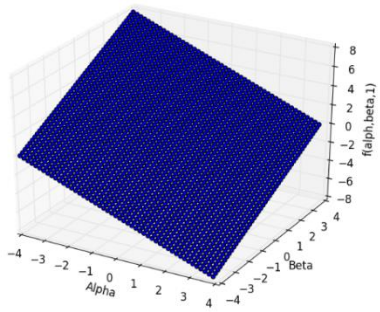
let:

$x = x_3 = -1$

$p_1, p_2 \in (-\infty, \infty)$

Hence:

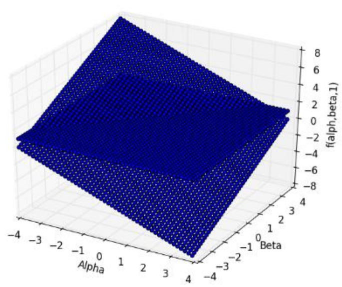
$$f(p_1, p_2, x) = \begin{bmatrix} -1 & 1 \end{bmatrix} \bullet \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = -p_1 + p_2$$

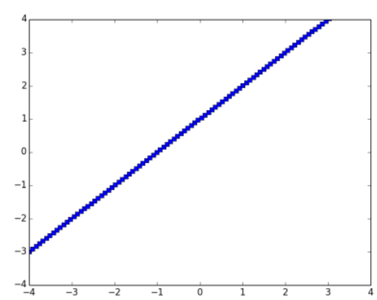


The figure represents all possible configurations of α and β and the resulting $f(p_1, p_2, x)$. If one was to receive a data point in $T(x, f(p_1, p_2, x))$ that reads:

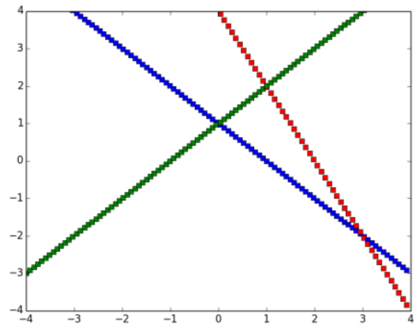
$$\vec{r}_{x_3} = \begin{bmatrix} x_3 \\ f\left(p_1, p_2, x_3\right) \end{bmatrix} = \begin{bmatrix} -1 \\ -p_1 + p_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

The solution set for p_1 and p_2 will be the intersection of the above surface $\langle p_1, p_2, -p_1 + p_2 \rangle$ and $\langle p_1, p_2, -1 \rangle$. And can be visualized in $P(p_1, p_2)$ as $\vec{p} = \begin{bmatrix} p_1 \\ 1 + p_1 \end{bmatrix}$





When we consider the solution sets of $r_{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $r_{x_2} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ and $r_{x_3} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ we get.



Notice that the parameter solution sets for $r_{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $r_{x_2} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ and $r_{x_3} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ are not intersecting. That is the collection of given points do not share a common parameter that satisfies all three points. Since there is no exact solution for the three given points, we must use a different measure of what parameters best describe the given data.

Least Square Regression:

Let us first introduce a plane $T(x_1, x_2)$. We are given a set of n points in $T(x_1, x_2)$ that are labeled with paired indices, x_{ij}. The ith index associates the component with its axis and the jth index associates the value with its pair. In $\mathbb{R}^2$, the scalars will be labeled and associated in the following manner:

$$(x_{11}, x_{21}),\ (x_{12}, x_{22}),\ \cdots,\ (x_{1i}, x_{2i}),\ \cdots,\ (x_{1n}, x_{2n})$$

This labeling may be extended to $\mathbb{R}^m$ and n points in the following manner:

$$(x_{11}, x_{21}, x_{31}, \dots, x_{i1}, \dots x_{m1}),\ (x_{12}, x_{22}, x_{32}, \dots, x_{i2}, \dots x_{m2}),\ \cdots,$$
$$(x_{1k}, x_{2k}, x_{3k}, \dots, x_{ik}, \dots x_{mk}),\ \cdots,\ (x_{1n}, x_{2n}, x_{3n}, \dots, x_{in}, \dots x_{mn})$$

Let us consider two sets of scalars labeled x_1 and x_2:

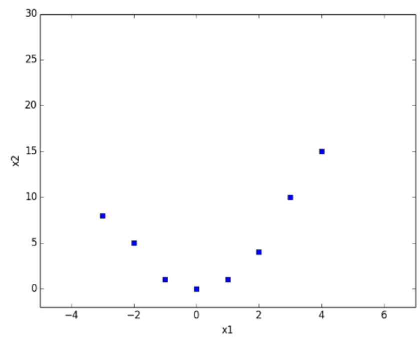
$$x_1 = \{x_{11}, x_{12}, x_{13}, \dots, x_{1n}\} = \{-3, -2, -1, 0, 1, 2, 3, 4, 5\}$$

$$x_2 = \{x_{21}, x_{22}, x_{23}, \dots, x_{2n}\} = \{8, 5, 1, 0, 1, 4, 10, 15, 24\}$$

Hence the point set is

$$M = \{(-3,8), (-2,5), (-1,1), (0,0), (1,1), (2,4), (3,10), (4,15), (5,24)\}$$

And can be visualized in plane $T(x_1, x_2)$ in the following manner:



Now that we have a collection of points M_n in $T(x_1, x_2)$ we seek to find a function that sufficiently describes the given points.

$$f(x_1) = x_2$$

In this particular case, our function in $\mathbb{R}^2$ may be represented as the polynomial

$$f(x_1) = p_1 x_1^{\,2} + p_2 x_1 + p_3$$

where x_1 and x_2 exist in plane $T(x_1, x_2)$ and parameters p_1, p_2 and $p_3 \in \mathbb{R}$ and may be represented in a space $P(p_1, p_2, p_3)$. Vectors in space $P(p_1, p_2, p_3)$, $\vec{p}$, determine the function mapping in $T(x_1, x_2)$. This can be represented in the matrix form as

$$y = Xp$$

where *y* is an *n* x 1 matrix whose rows represent x_2 data values, where *X* is a *n* x *i* matrix whose rows represent x_1 data values and columns represent each term of the selected function and *p* represents the coefficients of those terms. When *X* multiplies *p*, it creates in each of the resulting entries the function we wish to model. Considering the first associated x_{11} and x_{21} in our matrix equation, we get:

$$\left[f(x_{11}) \right] = \left[x_{11}^{\,2} \quad x_{11} \quad 1 \right] \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \left[p_1 x_{11}^{\,2} + p_2 x_{11} + p_3 \right]$$

Next we will consider the first and second x_1 and x_2.

$$\begin{bmatrix} f(x_{11}) \\ f(x_{12}) \end{bmatrix} = \begin{bmatrix} x_{11}^{2} & x_{11} & 1 \\ x_{12}^{2} & x_{12} & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} p_1 x_{11}^{2} + p_2 x_{11} + p_3 \\ p_1 x_{12}^{2} + p_2 x_{12} + p_3 \end{bmatrix}$$

Finally we will consider up to the n^th x_1 and x_2.

$$\begin{bmatrix} f(x_{11}) \\ f(x_{12}) \\ f(x_{13}) \\ \vdots \\ f(x_{1n}) \end{bmatrix} = \begin{bmatrix} x_{11}^{2} & x_{11} & 1 \\ x_{12}^{2} & x_{12} & 1 \\ x_{13}^{2} & x_{13} & 1 \\ \vdots & & \\ x_{1n}^{2} & x_{1n} & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} p_1 x_{11}^{2} + p_2 x_{11} + p_3 \\ p_1 x_{12}^{2} + p_2 x_{12} + p_3 \\ p_1 x_{13}^{2} + p_2 x_{13} + p_3 \\ \vdots \\ p_1 x_{1n}^{2} + p_2 x_{1n} + p_3 \end{bmatrix}$$

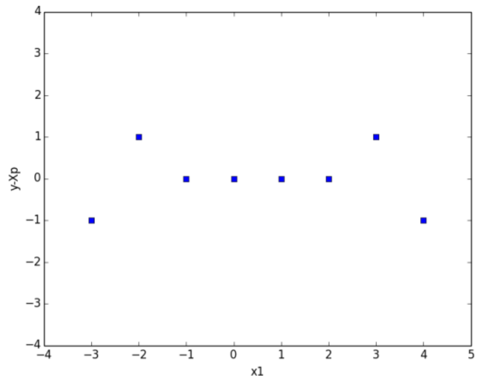
Because the x_1 and $f(x_1)$ are known values, the equation is asking what coefficient vector, $\vec{p}$, works to equate each corresponding index of *y* and *Xp*. In other words:

$$y - Xp = 0$$

However, we discovered in the previous section that unless our data set *M* is perfectly lined up with some function, the above equation cannot be solved perfectly. Hence, the equation becomes:

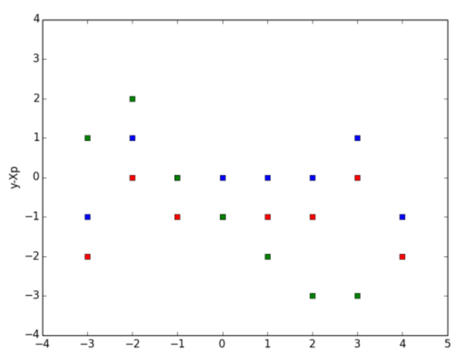
$$y - Xp \doteq 0$$

For example, let $p = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Therefore



$$y - Xp \doteq \begin{bmatrix} 8 \\ 5 \\ 1 \\ \vdots \\ 24 \end{bmatrix} - \begin{bmatrix} -3^2 & -3 & 1 \\ -2^2 & -2 & 1 \\ -1^2 & -1 & 1 \\ \vdots & & \\ 5^2 & 5 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 1 \\ \vdots \\ 24 \end{bmatrix} - \begin{bmatrix} -3^2 \\ -2^2 \\ -1^2 \\ \vdots \\ 5^2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ -1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

If we are to consider, $p_b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $p_r = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $p_g = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ the resulting differences are:



The sum of those differences for each vector $\vec{p}$,

$$\sum_{i=0}^{n} \left(y - Xp \right)_i$$

represent the total distortion of the data with respect to the selected function and its coefficients given by $\vec{p}$. In the previous example, we had a single vector $\vec{p}$ in $P(p_1, p_2, p_3)$ which produced, when applied to $\sum_{i=0}^{n}(y-Xp)_i$, gave us a value of the residuals at that point in $P(p_1, p_2, p_3)$. How would one consider all of the values of $\sum_{i=0}^{n}(y-Xp)_i$ in space $P(p_1, p_2, p_3)$ and find the minimal residual value?

Least square regression and Matrix Calculus:

Let us consider arbitrary sets

$$x_1 = \{x_{11}, x_{12}, \cdots, x_{1n}\}$$

$$x_2 = \{x_{21}, x_{22}, \cdots, x_{2n}\}$$

and an arbitrary polynomial.

$$x_2 = f(x_1) = p_i x_1{}^i + p_{i-1} x_1{}^{i-1} + p_{i-2} x_1{}^{i-2} + \ldots + p_2 x_1{}^2 + p_1 x_1 + p_0$$

From the above we will construct our residual vector, $y - Xp = 0$.

$$\begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} - \begin{bmatrix} x_{11}{}^i & x_{11}{}^{i-1} & \cdots & 1 \\ x_{12}{}^i & x_{12}{}^{i-1} & \cdots & 1 \\ \vdots & \vdots & & \\ x_{1n}{}^i & x_{1n}{}^{i-1} & \cdots & 1 \end{bmatrix} \begin{bmatrix} p_i \\ p_{i-1} \\ \vdots \\ p_0 \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In order to get a scalar value, we then dot the above with itself.

$$0 = (y - Xp)^T (y - Xp)$$

$$0 = (y^T - p^T X^T)(y - Xp)$$

$$0 = (y^T - p^T X^T) y - (y^T - p^T X^T) Xp$$

$$0 = yy^T - p^T X^T y - y^T Xp + p^T X^T Xp$$

$$0 = yy^T - 2y^T Xp + p^T X^T Xp$$

You'll notice that after expanding our equation we are left with three termed matrix equation, which resolves to a scalar, that describes the square residual for a particular X, y and p. In order to consider the residuals for all polynomials and to find the minimal value of the residuals, we must allow our vector p vary in its parameter space $P(p_0, p_1, \ldots, p_i)$ on $0 = yy^T - 2y^T Xp + p^T X^T Xp$.

$$\frac{\partial(0)}{\partial p} = \frac{\partial\left(yy^T - 2y^T Xp + p^T X^T Xp\right)}{p}$$

$$0 = \frac{\partial\left(yy^T\right)}{\partial p} - \frac{\partial\left(2y^T Xp\right)}{\partial p} + \frac{\partial\left(p^T X^T Xp\right)}{\partial p}$$

We must now consider what a matrix derivative is and prove what they evaluate to.

<u>The Matrix Derivative:</u>

In the previous section we encountered the matrix derivatives of three different types. Those being $\frac{\partial\left(yy^T\right)}{\partial p}$, $\frac{\partial\left(2y^T Xp\right)}{\partial p}$ and $\frac{\partial\left(p^T X^T Xp\right)}{\partial p}$. We must show that the kth element of the resulting matrix is the derivative of the original matrix product by p_k.

$$\frac{\partial(f(p))}{\partial p_k} = \begin{bmatrix} \frac{\partial(f(p))}{\partial p_1} \\ \frac{\partial(f(p))}{\partial p_2} \\ \vdots \\ \frac{\partial(f(p))}{\partial p_j} \\ \vdots \\ \frac{\partial(f(p))}{\partial p_i} \end{bmatrix}_k, \qquad k \in [1, i]$$

Case 1:

Let $yy^T = c$

$$\frac{\partial(c)}{\partial p} = 0$$

Case 2:

Let $2y^T X = b^T$

$$\left[\frac{\partial(b^T p)}{\partial p}\right]_k = [b]_k$$

$$\frac{\partial(b^T p)}{\partial p_k} = \frac{\partial(b_1 p_1 + b_2 p_2 + \ldots + b_k p_k + \ldots + b_i p_i)}{\partial p_k}$$

$$= \frac{\partial(b_1 p_1)}{\partial p_k} + \frac{\partial(b_2 p_2)}{\partial p_k} + \cdots + \frac{\partial(b_k p_k)}{\partial p_k} + \cdots + \frac{\partial(b_i p_i)}{\partial p_k}$$

$$= 0 + 0 + \cdots + b_k + \cdots + 0 = b_k = [b]_k$$

Therefore,

$$\frac{\partial(2y^T Xp)}{\partial p} = 2X^T y$$

Case 3:

Let $p^T X^T Xp = p^T Ap$

$$\left[\frac{\partial(p^T Ap)}{\partial p}\right]_k = [(A + A^T)p]_k$$

$$\frac{\partial(p^T Ap)}{\partial p_k} = \frac{\partial\left(\begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}\right)}{\partial p_k}$$

$$= \frac{\partial\left(\begin{bmatrix} \sum_{m=1}^{n} p_m a_{m1} & \sum_{m=1}^{n} p_m a_{m2} & \cdots & \sum_{m=1}^{n} p_m a_{mn} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}\right)}{\partial p_k}$$

$$= \frac{\partial\left(\sum_{l=1}^{n}\sum_{m=1}^{n} p_m a_{ml} p_l\right)}{\partial p_k}$$

$$= \frac{\partial\left(\sum_{l=k}\sum_{m=k} {p_k}^2 a_{kk} + \sum_{l\neq k}\sum_{m=k} p_k a_{kl} p_l + \sum_{l=k}\sum_{m\neq k} p_m a_{mk} p_k + \sum_{l\neq k}\sum_{m\neq k} p_m a_{ml} p_l\right)}{\partial p_k}$$

$$= \frac{\partial\left(\sum_{l=k}\sum_{m=k} {p_k}^2 a_{kk}\right)}{\partial p_k} + \frac{\partial\left(\sum_{l\neq k}\sum_{m=k} p_k a_{kl} p_l\right)}{\partial p_k} + \frac{\partial\left(\sum_{l=k}\sum_{m\neq k} p_m a_{mk} p_k\right)}{\partial p_k} + \frac{\partial\left(\sum_{l\neq k}\sum_{m\neq k} p_m a_{ml} p_l\right)}{\partial p_k}$$

$$= 2 p_k a_{kk} + \sum_{l\neq k} a_{kl} p_l + \sum_{m\neq k} p_m a_{mk} + 0$$

$$= \sum_{l\neq k} a_{kl} p_l + p_k a_{kk} + \sum_{m\neq k} p_m a_{mk} + p_k a_{kk}$$

$$= \sum_{l=1}^{n} a_{kl} p_l + \sum_{m=1}^{n} p_m a_{mk} = \sum_{i=1}^{n} a_{ki} p_i + \sum_{i=1}^{n} p_i a_{ik} = \sum_{i=1}^{n} \left(a_{ki} + a_{ik}\right) p_i$$

$$\left[\left(A + A^T\right)p\right]_k = \left[\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}\right)\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}\right]_k$$

$$= \left[\left(\begin{bmatrix} a_{11} + a_{11} & a_{12} + a_{21} & \cdots & a_{1n} + a_{n1} \\ a_{21} + a_{12} & a_{22} + a_{22} & \cdots & a_{2n} + a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{n1} + a_{1n} & a_{n2} + a_{2n} & \cdots & a_{nn} + a_{nn} \end{bmatrix}\right)\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}\right]_k$$

$$= \begin{bmatrix} \sum_{i=1}^{n} p_i \left(a_{1i} + a_{i1} \right) \\ \sum_{i=1}^{n} p_i \left(a_{2i} + a_{i2} \right) \\ \sum_{i=1}^{n} p_i \left(a_{3i} + a_{i3} \right) \\ \vdots \\ \sum_{i=1}^{n} p_i \left(a_{ni} + a_{in} \right) \end{bmatrix}_k = \sum_{i=1}^{n} \left(a_{ki} + a_{ik} \right) p_i$$

Therefore,

$$\frac{\partial \left(p^T X^T X p \right)}{\partial p} = \left(X^T X + \left(X^T X \right)^T \right) p = \left(2 X^T X \right) p$$

Given the above calculations, we can conclude that

$$0 = \frac{\partial \left(y y^T \right)}{\partial p} - \frac{\partial \left(2 y^T X p \right)}{\partial p} + \frac{\partial \left(p^T X^T X p \right)}{\partial p}$$

$$0 = 0 - 2 X^T y + 2 X^T X p$$

When solved for p, we get

$$\left(X^T X \right)^{-1} X^T y = p$$

which provides us with the parameters to the chosen equation that minimizes the sum of residuals squared.

<u>The Weighted Matrix:</u>

There may be time in which we would like to give certain data a higher importance, or weight, such that the resulting model will reflect that importance. How might we signify those weights? We can achieve this by introducing a matrix W

$$W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}$$

Such that when multiplied into our least square residual equations

$$0 = (y - Xp)^T W (y - Xp)$$

reduces or increases the size of the resulting squares. If we let $y - Xp = B$ where

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

We find

$$= \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix} \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

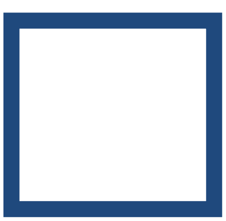
$$= \begin{pmatrix} b_1 w_1 & b_2 w_2 & \cdots & b_n w_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = b_1 w_1 b_1 + b_2 w_2 b_2 + \cdots + b_n w_n b_n$$

$$= b_1{}^2 w_1 + b_2{}^2 w_2 + \cdots b_n{}^2 w_n$$

As you can see the elements of matrix W multiply each of the values of the squares of the residuals. This can be visually represented, under the 3 cases of w_n, in the following way:

$w_n > 1$

$b_n{}^2$



$b_n{}^2 w_n$

$w_n = 1$

$w_n < 1$

When the original square, $b_n{}^2$, is multiplied by a weight $w_n > 1$, the resulting square $b_n{}^2 w_n$ will be larger. When the original square, $b_n{}^2$, is multiplied by a weight $w_n = 1$, the resulting square $b_n{}^2 w_n$ will be the same. When the original square, $b_n{}^2$, is multiplied by a weight $w_n < 1$, the resulting square $b_n{}^2 w_n$ will be smaller. The higher the value of w_n, the closer our selected line will have to get to that data point in order to minnimize the sum of residuals squared.

Reconsidering our residual sum of squares with matrix W we get,

$$
\begin{aligned}
0 &= (y - Xp)^T W (y - Xp) \\
0 &= (y^T - p^T X^T)(Wy - WXp) \\
0 &= y^T (Wy - WXp) - p^T X^T (Wy - WXp) \\
0 &= y^T Wy - 2y^T WXp + p^T X^T WXp
\end{aligned}
$$

and applying the results from the matrix calculus section we get,

$$
\frac{\partial(0)}{\partial p} = \frac{\partial(y^T Wy)}{\partial p} - \frac{\partial(2y^T WXp)}{\partial p} + \frac{\partial(p^T X^T WXp)}{\partial p}
$$

$$= 0 - 2X^T W^T y + \left(X^T WX + \left(X^T WX \right)^T \right) p$$

Note $W^T = W$

$$= 0 - 2X^T Wy + \left(2X^T WX \right) p$$

$$= \left(2X^T WX \right)^{-1} 2X^T Wy = p .$$

The form of the W matrix I find most useful is $w_n = m_1 e^{\frac{(f(x_{1n}, x_{2n}) - m_2)^2}{2m_3^2}}$. That is,

$$W = \begin{pmatrix} m_1 e^{\frac{(f(x_{11}, x_{21}) - m_2)^2}{2m_3^2}} & 0 & \cdots & 0 \\ 0 & m_1 e^{\frac{(f(x_{12}, x_{22}) - m_2)^2}{2m_3^2}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & m_1 e^{\frac{(f(x_{1n}, x_{2n}) - m_2)^2}{2m_3^2}} \end{pmatrix}$$

Where, $m = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}$ defines the shape of the underlying Gaussian curve and $f\left(x_{1n}, x_{2n}\right)$ gives the value we are to sample from the Gaussian curve for that w_n as a function of that weight's associated data point. When $f\left(x_{1n}, x_{2n}\right) = m_2$ our weight becomes $w_n = m_1$, which is an upper bound for our weights. For example, using $f\left(x_{1n}, x_{2n}\right) = \sqrt{\left(x_{1n} - x_{1k}\right)^2 + \left(x_{2n} - x_{2k}\right)^2}$ where $k \in [0, n]$ and $m = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ will weigh data close to $\left(x_{1k}, x_{2k}\right)$. What this is attempting to do is to use the data itself to sample a Gaussian curve for its weights. The motivation for using the Gaussian curve is that there is a known upper and lower bound to the equation and there is great malleability to the equation that allows for a vast amount of different orientations for different problems. That is, $m = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}$, gives us the degree of freedom to sample from many different curves. Determining how to form such a $f\left(x_{1n}, x_{2n}\right)$ would be interesting to consider further.

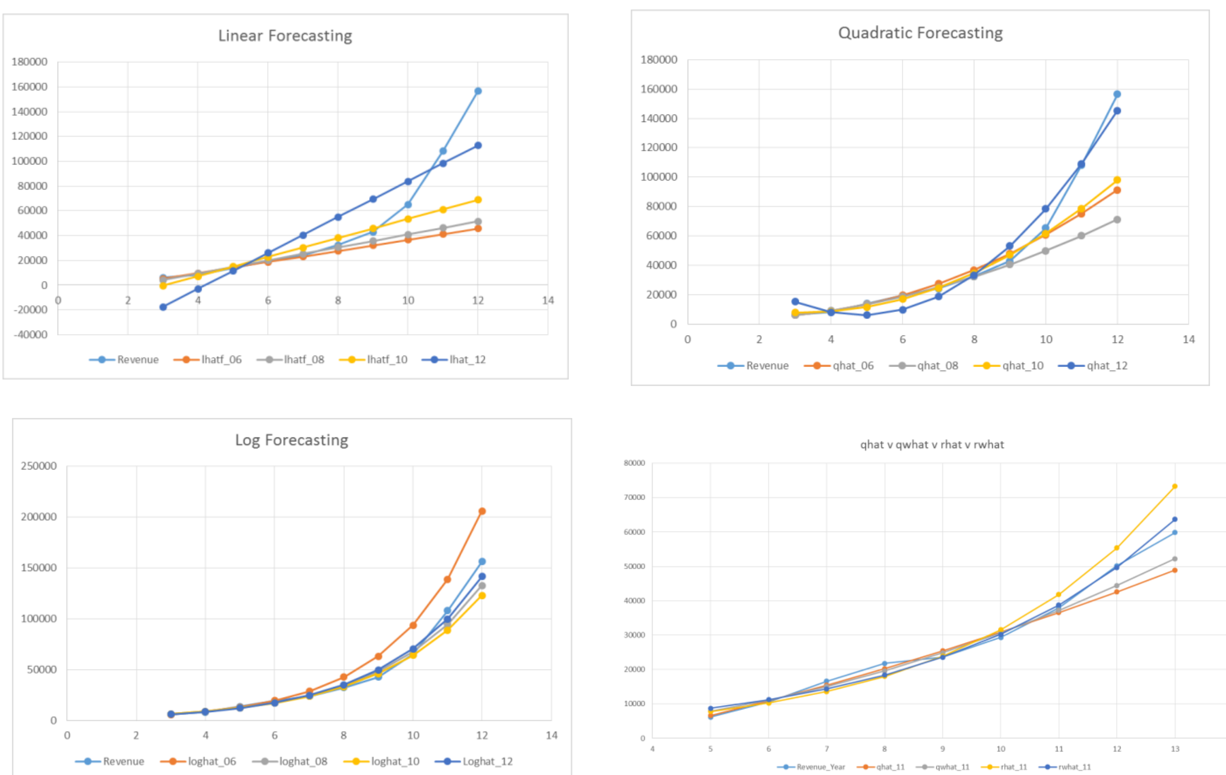
Experiments:

My first experiments exclusively dealt with developing models for Apple's revenue. Data was transferred from nasdaq.com in order to develop the matrices needed to develop a model of revenue as a function of time.

I then computed Linear (lhat_12), Quadratic (qhat_12) and Constant Percent Growth (loghat_12) models using the entirety of the data. The goal of these tests was to determine what model worked to minimize our residual errors most. As evident in the adjacent graphs, the model with the least residuals was the Constant Percent Growth Model. The percent error for the models are as follows:

Linear (50.88%), Quadratic (33.31%), Exp (19.93%)

Apple Revenue Model Forecasting



I next tackled the concepts of Model Forecasting. We can use varying data lengths (Revenue: 2003-2006, 2003-2008, 2003-2010, 2003-2012) to forecast the remaining years up to

2013 while comparing forecasted revenue with actual revenue. One can plainly see that as the length of historical revenue data increases, the lower the residual error is to the actual data.

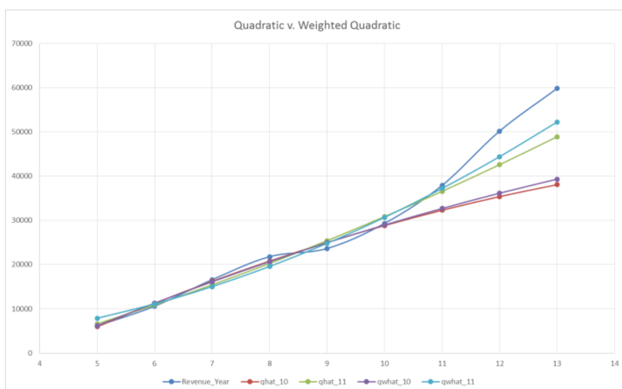
How does defining importance of certain data affect our residuals? This is what drove the experiments with Google's revenue data. Using an nxn exponential diagonal matrix (5), I was able to achieve a model weighting that took into account recent years data more so then data of past.

After computing the Quadratic models (qhat_10, qhat_11) and exponentially weighted Quadratic models (qwhat_10, qwhat_11) it was clear that there was a difference in the residual error to recent data. The percent error of Models are as follows:

qhat_10 = 28.51%, qwhat_10 = 26.87%,

qhat_11 = 13.42%, qwhat_11 = 9.5%

Google Revenue Model Forecasting



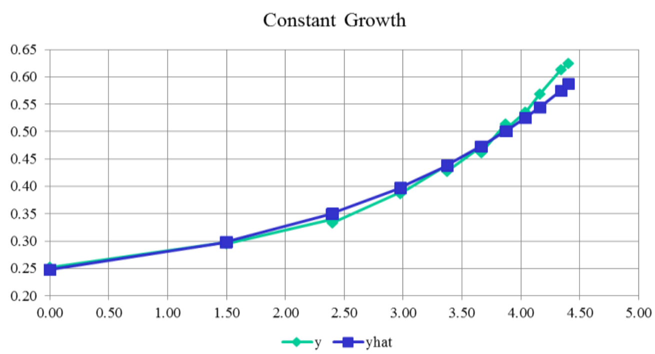
This effectively allows us to better predict where the company's revenue is heading as recent behavior will more likely be repeated. As you can see in figure Quadratic v. Weighted Quadratic, the weighted matrix works to pull our model closer to recent years.

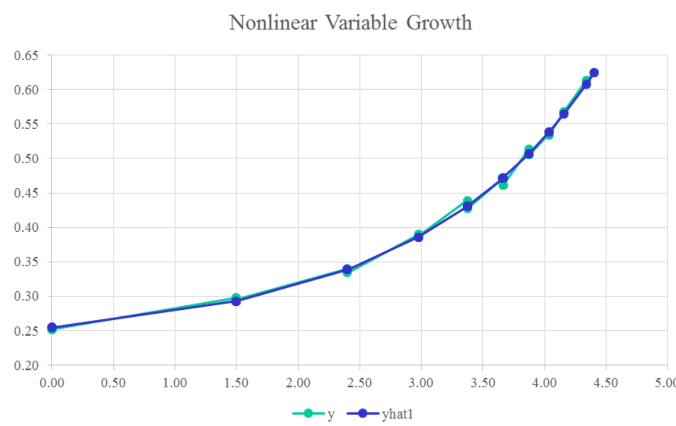
The final experiments conducted at this point were looking further into the Constant Percent Growth model. Specifically, what would happen if g, our constant growth, was allowed to vary. To test this curiosity we gathered observed data from NIST for quantum defects of Sulfur I.

First a standard Constant Percent Growth model was developed. We used this, figure Constant Growth, as a base to compare what variable growth could produce. Using excel tool, Solver, we were able to optimize the Nonlinear Variable Growth Model.

NIST Data Model

$$
\begin{aligned}
y &= f(x;\beta) + \epsilon \\
&= \beta_1 - \beta_2 x - \frac{\arctan\left[\frac{\beta_3}{x-\beta_4}\right]}{\pi} + \epsilon
\end{aligned}
$$





As evident by figure Nonlinear Variable Growth, the standard Constant Percent Growth model can be warped to create a far better fit by variable growth, in this case where g(x) = ax+b.

NIST originally used a physics based model (shown at the upper right corner of graph) to fit the data. We are able to fit well without knowing the physics model. The power of the empirical model is to be able to describe these processes without knowing the first principal.

<u>Simple Model of Light:</u>

As opposed to developing a model from given data, one may construct a model from initial axioms and compare to data taken later. The following is an attempt to mathematically describe the image of a large body by an observer.

First we must develop the local geometry of an object emitting photons. We will use a spherical model for following equations. A sphere may be defined in spherical coordinates in the following manner:

$$\bar{s}(\theta,\phi,\rho) = \langle \theta,\phi,\rho \rangle$$

$$\theta \in [n, n+2\pi]$$
$$\phi \in [n, n+\pi]$$
$$\rho \in (0, \infty)$$

Where θ is measured from the positive x axis, ϕ is measured from the positive z axis and ρ is measured as the distance from the origin to the point. When ρ is set to a constant, our vector $\bar{s}(\theta, \phi, c)$ represents a sphere of radius c. To inject this sphere into our model, we must convert our representation from a Spherical to Cartesian coordinate system. This can be achieved in the following way:

$$\bar{s}(s_1(\rho,\theta,\phi), s_2(\rho,\theta,\phi), s_3(\rho,\phi)) = \langle x, y, z \rangle$$

$$s_1(\rho,\theta,\phi) = \rho \sin\phi \cos\theta$$
$$s_2(\rho,\theta,\phi) = \rho \sin\phi \sin\theta$$
$$s_3(\rho,\phi) = \rho \cos\phi$$

Now that we've introduced our vector representation of a sphere to our $\langle x, y, z \rangle$ Cartesian space, we must give it the ability to be translated to positions beyond the point of origin. We can accomplish this by introducing a translation vector, $\bar{t} = (x_0, y_0, z_0)$, and adding it to our sphere object.

$$\bar{s} + \bar{t} = \langle \rho \sin\phi \cos\theta + x_0, \rho \sin\phi \sin\theta + y_0, \rho \cos\phi + z_0 \rangle$$

If we want to have this sphere in motion, we simply have the components of our translation vector, $\bar{t}$, be functions of time. In this paper, in order to avoid any confusion between the variable of time and the translation vector, we will call our time variable q. Hence, the above formula of a static sphere may be represented as a dynamic one by:

$$\bar{t}(q) = \langle f_1(q), f_2(q), f_3(q) \rangle$$

$$\bar{y}(q, \rho, \theta, \phi) = \bar{s} + \bar{t}(q) = \langle \rho \sin\phi \cos\theta + f_1(q), \rho \sin\phi \sin\theta + f_2(q), \rho \cos\phi + f_3(q) \rangle$$

It should be noted that if you want to have your sphere or ellipsoid in rotation, multiplying vector s by rotation matrix $R(q)$ will achieve this. Note the python program near the end. Now that we have a simple sphere in motion, let us calculate the lengths of the field of vectors that form when:

$$q \in (0,\infty)$$
$$\rho \in (0,\infty)$$
$$\theta \in [0,2\pi]$$
$$\phi \in [0,\pi]$$

$$\|\bar{y}(\theta,\phi,\rho,q)\| = \sqrt{\sum_{i=1}^{3}\left(f_i(q) + s_i(\rho,\theta,\phi)\right)^2}$$

$$= \sqrt{\rho^2 \sin^2\phi\cos^2\theta + 2\rho\sin\phi\cos\theta\cdot f_1(q) + f_1(q)^2 + \rho^2\sin^2\phi\sin^2\theta + 2\rho\sin\phi\sin\theta\cdot f_2(q) + f_2(q)^2 + \rho^2\cos^2\phi +}$$

$$= \sqrt{\rho^2\left(\sin^2\phi\cos^2\theta + \sin^2\phi\sin^2\theta + \cos^2\phi\right) + 2\rho\left(\sin\phi\cos\theta\cdot f_1(q) + \sin\phi\sin\theta\cdot f_2(q) + \cos\phi\cdot f_3(q)\right) + f_1(q)^2 + f_2}$$

$$= \sqrt{\rho^2\left(\sin^2\phi\left(\cos^2\theta + \sin^2\theta\right) + \cos^2\phi\right) + 2\rho\left(\sin\phi\cos\theta\cdot f_1(q) + \sin\phi\sin\theta\cdot f_2(q) + \cos\phi\cdot f_3(q)\right) + f_1(q)^2 + f_2(q)^2}$$

$$= \sqrt{\rho^2\left(\sin^2\phi(1) + \cos^2\phi\right) + 2\rho\left(\sin\phi\cos\theta\cdot f_1(q) + \sin\phi\sin\theta\cdot f_2(q) + \cos\phi\cdot f_3(q)\right) + f_1(q)^2 + f_2(q)^2 + f_3(q)^2}$$

$$= \sqrt{\rho^2 + 2\rho\left(\sin\phi\cos\theta\cdot f_1(q) + \sin\phi\sin\theta\cdot f_2(q) + \cos\phi\cdot f_3(q)\right) + f_1(q)^2 + f_2(q)^2 + f_3(q)^2}$$

As a digression, we can use our equation $\|\bar{y}\|(\theta,\phi)$ to determine the point on the sphere, of a constant ρ, that is closest to the observation point by allowing the variables θ , ϕ and q to on their intervals $[n, n+2\pi]$, $[n, n+\pi]$ and $[0,\infty)$ respectively.

$$\frac{\partial\left[\|\bar{y}(\theta,\phi,\rho,q)\|\right]}{\partial\theta} =$$

$$\frac{\rho\left(\sin\phi\cos\theta\cdot f_2(q) - \sin\phi\sin\theta\cdot f_1(q)\right)}{\sqrt{\rho^2 + 2\rho\left(\sin\phi\cos\theta\cdot f_1(q) + \sin\phi\sin\theta\cdot f_2(q) + \cos\phi\cdot f_3(q)\right) + f_1(q)^2 + f_2(q)^2 + f_3(q)^2}} = 0$$

$$\Rightarrow \min_\theta\left(\|\bar{y}(\theta,\phi,\rho,q)\|\right)$$

$$\frac{\partial\left[\left\|\bar{y}(\theta,\phi,\rho,q)\right\|\right]}{\partial\phi} =$$

$$\frac{\rho\left(\cos\phi\cos\theta\cdot f_1(q)+\cos\phi\sin\theta\cdot f_2(q)-\sin\phi\cdot f_3(q)\right)}{\sqrt{\rho^2+2\rho\left(\sin\phi\cos\theta\cdot f_1(q)+\sin\phi\sin\theta\cdot f_2(q)+\cos\phi\cdot f_3(q)\right)+f_1(q)^2+f_2(q)^2+f_3(q)^2}} = 0$$

$$\Rightarrow \min_{\phi}\left(\left\|\bar{y}(\theta,\phi,\rho,q)\right\|\right)$$

$$\frac{\partial\left[\left\|\bar{y}(\theta,\phi,\rho,q)\right\|\right]}{\partial q} =$$

$$\frac{\rho\left(\sin\phi\cos\theta\cdot\frac{\partial f_1(q)}{\partial q}+\sin\phi\sin\theta\cdot\frac{\partial f_2(q)}{\partial q}+\cos\phi\cdot\frac{\partial f_3(q)}{\partial q}\right)+2\sum_{i=1}^{3} f_i(q)\frac{\partial f_i(q)}{\partial q}}{\sqrt{\rho^2+2\rho\left(\sin\phi\cos\theta\cdot f_1(q)+\sin\phi\sin\theta\cdot f_2(q)+\cos\phi\cdot f_3(q)\right)+f_1(q)^2+f_2(q)^2+f_3(q)^2}} = 0$$

$$\Rightarrow \min_{q}\left(\left\|\bar{y}(\theta,\phi,\rho,q)\right\|\right)$$

We initially found the length of this field of vectors so that we could normalize them. Normalization works to reduce the length of all associated vectors to one while preserving their direction orientation. This can be achieved in the following manner.

$$\hat{y} = \frac{\bar{y}}{\|\bar{y}\|} = \frac{\bar{y}}{\sqrt{\sum_{n=1}^{3}\left(f_n(q)+s_n(\rho,\theta,\phi)\right)}}$$

$$\|\hat{y}\|(\theta,\phi) = \left\|\frac{\bar{y}}{\|\bar{y}\|}\right\| = \left\|\frac{\bar{y}}{\sqrt{\sum_{n=1}^{3}\left(f_n(q)+s_n(\rho,\theta,\phi)\right)}}\right\| = 1$$

It is reasonable to believe, due to the constant speed of photons, that two photons whose departure times from any surface are equal will have traveled an equal distance. It is also reasonable to believe that photons that are perceived have left the surface of the object at an angle such that its path will intercept an observation point. Now we want to represent these ideas in using our original equation $\bar{y}(\theta,\phi,\rho,q)$ in terms of $\hat{y}(\theta,\phi,\rho,q)$:

$$\bar{h}_y(\theta,\phi,\rho,q,l) = \left(\|\bar{y}\|(\theta,\phi,\rho,q)-l\right)\hat{y}(\theta,\phi,\rho,q)$$

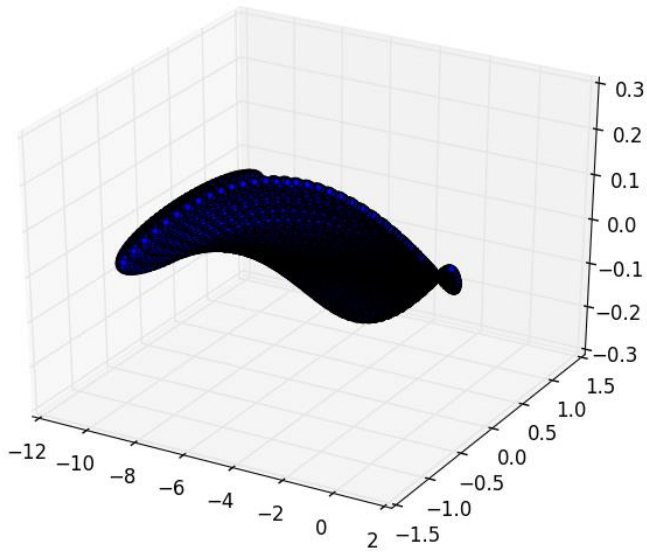
When $l=0$, we find that,

$$\bar{h}_y(\theta,\phi,\rho,q) = (\|\bar{y}\|(\theta,\phi,\rho,q) - 0)\hat{y}(\theta,\phi,\rho,q)$$

$$= \|\bar{y}\|(\theta,\phi,\rho,q)\hat{y}(\theta,\phi,\rho,q)$$

$$= \bar{y}(\theta,\phi,\rho,q)$$

Eq. 9 represents a translation of a unit l towards the point of origin, ie an observation point.



$$0 \le l \le \|y\|$$

To take measurements of the distortion of our object we must introduce a homothetic representation of the origin of our homothecy. We can achieve this by using the translation vector.

$$\bar{h}_t(q,l) = (\|t\|(q) - l)\hat{t}(q)$$

Our measure of distortion may be represented as,

$$Distortion = \|\bar{h}_y(\theta,\phi,\rho,q,l) - \bar{h}_t(q,l)\|$$

$$= \|(\|y\|(\theta,\phi,\rho,q) - l)\hat{y}(\theta,\phi,\rho,q) - (\|t\|(q) - l)\cdot\hat{t}(q)\|$$

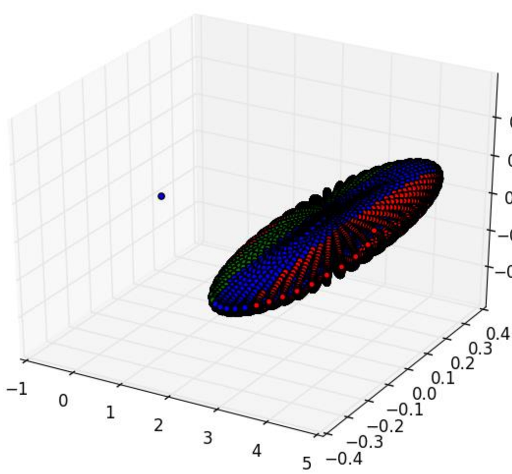
$$= \|\bar{y} - l\hat{y} - \bar{t} + l\hat{t}\|(\theta,\phi,\rho,q,l) = \|\bar{y} - \bar{t} + l(\hat{t} - \hat{y})\|(\theta,\phi,\rho,q,l)$$

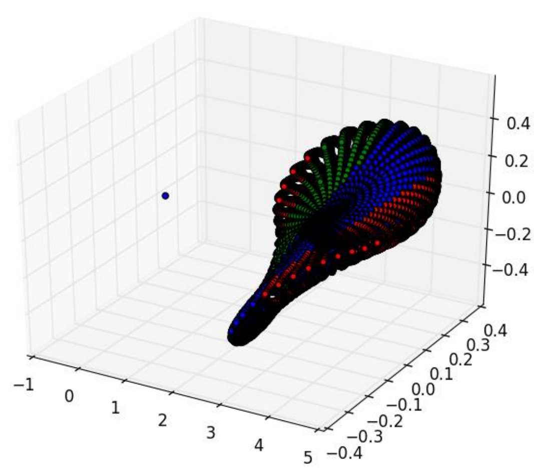
It should be noted that there are two points on our homothecy, those existing on a line that includes the origin of $\mathbf{R}^3$ and the origin of the local geometry of the sphere will experience no distortion with respect to the origin of the local geometry. That is if we begin with a sphere of radius ρ, Two points on any homothecy will be a distance ρ from the new origin for any l.

The following equation describes the apparent position of stars due to constant light speed.

Eq. $h_y(\theta,\phi,\rho,q,l) = s(\alpha,\beta,r) = \langle r\sin\alpha\cos\beta, r\sin\alpha\sin\beta, r\cos\alpha\rangle$

$$(\|\bar{y}\|(\theta,\phi,\rho,q) - l)\hat{y}(\theta,\phi,\rho,q) = s(\alpha,\beta,r)$$





The following is a python program using matplotlib.pyplot and numpy to generate these homothocys and tranformations.

```
import numpy as np

import math

import matplotlib.pyplot as plt

from mpl_toolkits.mplot3d import Axes3D


fig = plt.figure()

ax = fig.add_subplot(111, projection='3d')

##ax.set_xlim(-5,5)

##ax.set_ylim(-5,5)

##ax.set_zlim(-5,1)


        #ellipsoid parameters

a1 = .001

a2 = .05

a3 = .01


        #transform (T)

x1_0 = 3.5

x2_0 = 0

x3_0 = 0


        #view vector (v)
```

```
x1_1 = 0

x2_1 = 0

x3_1 = 0


    #T - v = B

x1_2 = x1_0 - x1_1

x2_2 = x2_0 - x2_1

x3_2 = x3_0 - x3_1


    #Space

angle = np.pi

rotangle = (.01)*np.pi  # rotang > 5*np.pi spirals

time_1 = 200

theta = np.linspace(0,2*angle,num=50)

phi = np.linspace(0,angle,num=50)

time = np.linspace(0,time_1,num=1500)


for i in range(0,1500):

    for j in range(0,50):

        for k in range(0,50):

                #Rotated object due to time

            x1_obj = a1*np.sin(phi[j])*np.sin(theta[k])+x1_2

            x2_obj = 
a2*np.sin(phi[j])*np.cos(theta[k])*np.cos(time[i]*(rotangle/time_1)) + 
a3*np.cos(phi[j])*np.sin(time[i]*(rotangle/time_1)) + x2_2
```

```
                    x3_obj = -
1*a2*np.sin(phi[j])*np.cos(theta[k])*np.sin(time[i]*(rotangle/time_1)) +
a3*np.cos(phi[j])*np.cos(time[i]*(rotangle/time_1)) + x3_2


                    n_obj = math.sqrt((x1_obj**2)+(x2_obj**2)+(x3_obj**2))


                    nx1_obj = x1_obj/n_obj

                    nx2_obj = x2_obj/n_obj

                    nx3_obj = x3_obj/n_obj


                    l_t = (n_obj/time_1)*time[i]


                    x1_hom = (n_obj - l_t)*nx1_obj

                    x2_hom = (n_obj - l_t)*nx2_obj

                    x3_hom = (n_obj - l_t)*nx3_obj


                    n_hom = math.sqrt((x1_hom**2)+(x2_hom**2)+(x3_hom**2))


                    if(n_hom > 0.975 and n_hom <= 1.0):


                        print '('+str(i)+', '+str(j)+', '+str(k)+', '+str(x1_obj)+',
'+str(x2_obj)+', '+str(x3_obj)+', '+str(n_hom)+')'


                        if(time[i]<10 and time[i]>=0):

                            ax.scatter(x1_obj,x2_obj,x3_obj,c='r')

                        elif(time[i]<20 and time[i]>=10):
```

```
        ax.scatter(x1_obj,x2_obj,x3_obj,c='g')
    elif(time[i]<30 and time[i]>=20):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='b')
    elif(time[i]<40 and time[i]>=30):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='r')
    elif(time[i]<50 and time[i]>=40):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='g')
    elif(time[i]<60 and time[i]>=50):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='b')
    elif(time[i]<70 and time[i]>=60):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='r')
    elif(time[i]<80 and time[i]>=70):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='g')
    elif(time[i]<90 and time[i]>=80):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='b')
    elif(time[i]<100 and time[i]>=90):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='r')
    elif(time[i]<110 and time[i]>=100):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='g')
    elif(time[i]<120 and time[i]>=110):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='b')
    elif(time[i]<130 and time[i]>=120):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='r')
    elif(time[i]<140 and time[i]>=130):
        ax.scatter(x1_obj,x2_obj,x3_obj,c='g')
```

```
            elif(time[i]<150 and time[i]>=140):
                ax.scatter(x1_obj,x2_obj,x3_obj,c='b')
            elif(time[i]<160 and time[i]>=150):
                ax.scatter(x1_obj,x2_obj,x3_obj,c='r')
            elif(time[i]<170 and time[i]>=160):
                ax.scatter(x1_obj,x2_obj,x3_obj,c='g')
            elif(time[i]<180 and time[i]>=170):
                ax.scatter(x1_obj,x2_obj,x3_obj,c='b')
            elif(time[i]<190 and time[i]>=180):
                ax.scatter(x1_obj,x2_obj,x3_obj,c='r')
            elif(time[i]<200 and time[i]>=190):
                ax.scatter(x1_obj,x2_obj,x3_obj,c='g')


ax.scatter(0,0,0)
plt.show()
```

Conclusion:

Model building is an incredibly useful tool in any concentration of study and a beautiful system to consider. It's been an incredible experience to consider spaces beyond that of the three dimensions. While the topic is broad, there are certain forms of equations that are not solvable using this method, including non-linear variable growth. Even in my own research, I've found that the concepts and skills I've gathered in the pursuit of understanding least square regressions has helped me greatly. I hope to continue as I have and look deeper into reality with this tool.

References:

Tsai, R. (2014, Spring). *Topics in Advanced Mathematics*. Lectures conducted from State University of New York at Purchase, NY.

Empirical Model Building: Data, Models, and Reality, by James Thompson, Wiley series in probability and statistics, ISBN 978-0470467039, 2011

Applied Regression Analysis A Research Tool, , by John Ralings, Sastry Pantula, David Dickey, SPringer Science and Business Mdeia, 2006