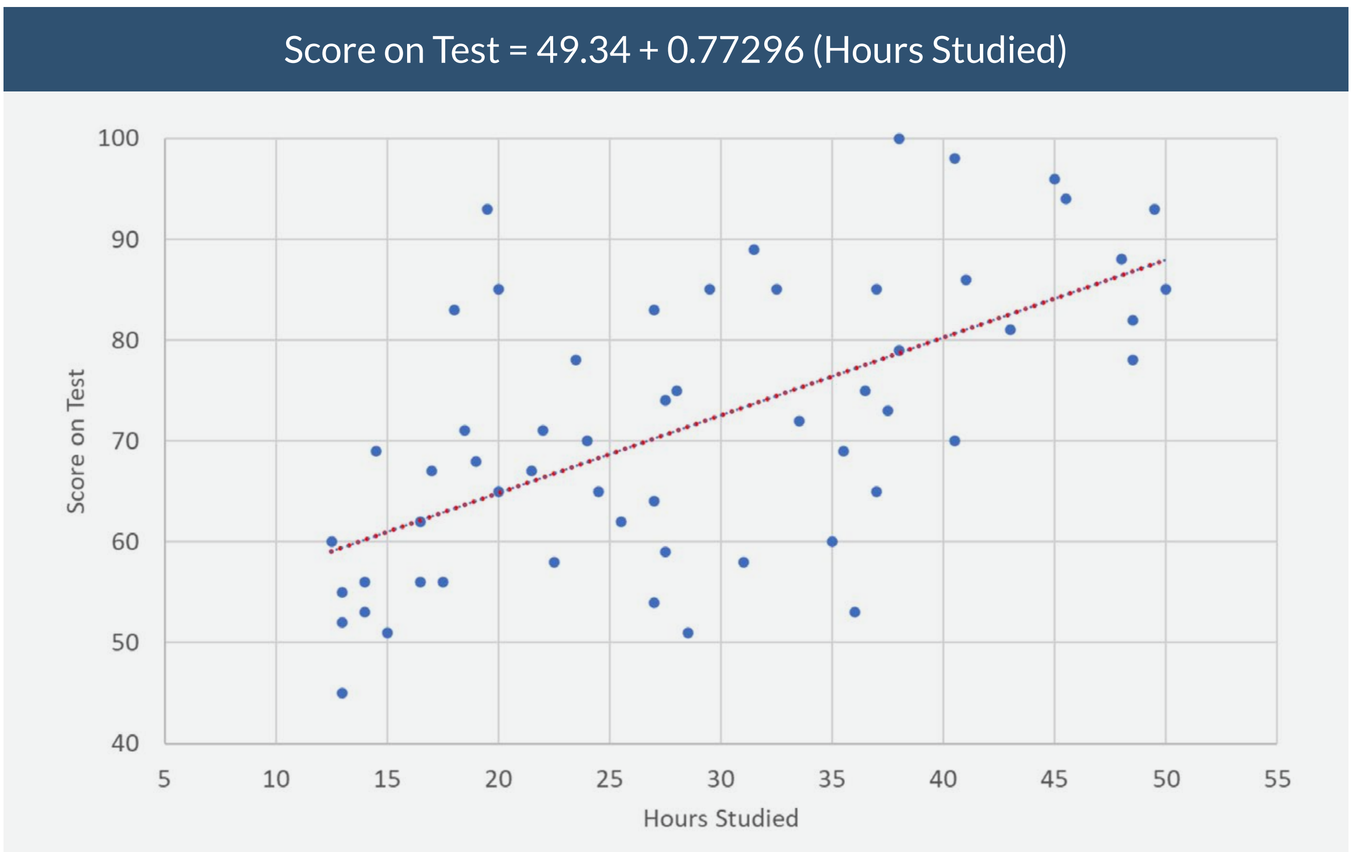
The resulting equation could be used to predict a student's test score based on the number of hours they study. In this example, the number of hours studied is the independent variable because it is being manipulated or controlled by the researcher, and the math test score is the dependent variable because it is being measured and is expected to change in response to changes in the independent variable.

In linear regression, the relationship between the dependent variable and the independent variables is assumed to be linear, which means that the relationship can be described by a straight line. In mathematics, the concept of a straight **line** and its equation is typically taught in middle school or early high school, usually in the context of algebra or geometry. In algebra, students are introduced to linear equations, which are equations that describe a straight line in the form y = mx + b, where m is the slope of the line and b is the y-intercept (the point where the line crosses the y-axis). Students learn how to graph a line using its equation, how to find the slope and y-intercept from the equation, and how to write an equation of a line given its slope and a point on the line.



The goal of linear regression is to find the line that best fits the data, which is accomplished by minimizing the sum of the squared differences or errors between the actual values of the dependent variable and the predicted values from the line. Linear regression is commonly used in various fields such as economics, social sciences, engineering, and business, to analyze and make predictions based on data. **Thus, linear regression can be used for forecasting, but it depends on the type of data and the nature of the relationship between the variables.** For example, if you have a set of historical sales data and you want to forecast future sales based on changes in advertising expenditure, linear regression may be an appropriate method if there is a linear relationship between sales and advertising expenditure, and if you expect that relationship to continue in the future. However, if the relationship between the variables is non-linear or if there are other factors that may affect the outcome variable in the future, other methods such as non-linear regression, time series analysis, or machine learning [artificial intelligence (AI)] techniques may be much more appropriate for forecasting.

Forecasting complex data such as stock prices is much more difficult because stock prices are affected by a wide range of factors, many of which are unpredictable and subject to rapid change. Some of the main reasons why forecasting stock prices is difficult include:

- **Complexity of the market:** The stock market is a complex system that is influenced by many variables, such as macro-economic, company management, geopolitical events, and investor sentiment. These variables interact with each other in complex unpredictable ways, making it difficult to identify the most crucial factors affecting stock prices at any time.
- **Randomness and volatility:** Stock prices are also subject to random fluctuations and can be highly volatile, especially in the short term. This means that even insignificant changes in market conditions or investor sentiment can cause large movements in stock prices.
- **Information asymmetry:** Investors have access to various levels of information about companies and market conditions, which can affect their decisions and lead to unpredictable price movements.
- **Behavioral factors:** Stock prices can also be influenced by psychological factors such as fear, greed, and herd behavior, which can cause investors to overreact or underreact to news and events.

Due to these and many other factors, forecasting stock prices with a high degree of accuracy is difficult, and even the most sophisticated models and techniques can have limited predictive power. However, we have made some breakthroughs in this area by studying one of the best financial quants of all-time, Jim Simons. Jim Simons is a mathematician, hedge fund manager, and philanthropist who is best known as the founder of Renaissance Technologies, a quantitative hedge fund that has generated exceptional returns using mathematical models and computer algorithms. Simons was born in 1938 in Massachusetts and earned a PhD in mathematics from the University of California, Berkeley in 1961. He worked as a codebreaker for the National Security Agency and as a mathematics professor at Harvard University and the Massachusetts Institute of Technology (MIT) before turning to finance. In 1982, Simons founded Renaissance Technologies, which used sophisticated mathematical models and data analysis techniques to identify profitable trading opportunities in financial markets. The firm's flagship Medallion Fund, which is only available to employees, has generated average annual returns of around 40% since its inception in 1988, making it one of the most successful hedge funds of all time.

One of the techniques Jim and his team uses is **high order polynomial curve fitting.** A high-order polynomial curve is a curve that is defined by a polynomial equation of a high degree, typically degree 3 or higher. The degree of a polynomial refers to the highest power of the independent variable (usually denoted as "x") in the equation. For example, a quadratic equation (degree 2) would be in the form of y = ax2 + bx + c, where a, b, and c are coefficients that determine the shape and position of the curve. A cubic equation (degree 3) would be in the form of y = ax3 + bx2 + cx + d, and so on for higher degrees. **High-order polynomial curves have complex and irregular shapes with multiple turning points, which makes them useful for modelling certain types of data such as experimental or empirical data [like economic or financial data].** However, they can also be prone to overfitting and may not generalize well to new data.

There is a solution to overfitting called GMDH. GMDH stands for Group Method of Data Handling. It is a machine learning algorithm that is used for regression and classification tasks. **The algorithm uses a self-organizing, adaptive model that can automatically select relevant features and determine the best model structure for a given data set.** GMDH was invented in the 1960s by a Ukrainian mathematician named Alexey Ivakhnenko, who was working at the Institute of Cybernetics in Kyiv Ukraine. Ivakhnenko developed the algorithm to optimize models without requiring any prior knowledge or assumptions about the underlying relationships between the variables. [We now call these neural networks.] GMDH has since been used in a wide range of applications, including engineering, biology, and environmental science. It is known for its ability to handle large and complex data sets, and for its high accuracy and robustness. However, like all machine learning algorithms, GMDH is not a one-size-fits-all solution, and its effectiveness depends on the specific problem and data set being analyzed. We have found polynomial curve fitting using the proper high order equation more useful than all technical techniques commonly used historically.

High-order polynomial curves have complex and irregular shapes with multiple turning points, which makes them useful for modelling certain types of data such as experimental or empirical data [like economic or financial data].

At Waterloo we use sophisticated polynomial curve fitting to forecast.

This is a mathematical technique using polynomial functions. A polynomial function is a mathematical function of the form:

$$f(x) = a_n xn + a_{n-1} x{n-1} + ... + a_1 x + a_0$$

where x is the independent variable, a_n, a_{n-1}, ..., a_1, a_0 are coefficients, and n is a non-negative integer. The highest power of x in the polynomial is called the degree of the polynomial.

Polynomial functions can take on many shapes, including curves, parabolas, or more complex curves. **The coefficients in the polynomial determine the shape of the curve, with higher degree polynomials producing more complex curves such as the ups and downs of the stock market.** To the right is the polynomial curve for the S&P 500 as of Sunday Feb 26th. The S&P 500 was trending upwards until December 2021 and spent most of 2022 falling until October. The curve then went back up until the beginning of February 2023. Since the beginning of February, it has been falling rapidly. This curve has an extraordinarily complex structure that is not discernible without some complex mathematical techniques. **However, after optimal complex curves are figured and drawn by the computer, one's eye can often see the pattern. Can you?**

The goal of polynomial curve fitting is to find a curve that best fits the data, so that it can be used to make good predictions. The process of polynomial curve fitting typically involves selecting a polynomial degree, which determines the order of the polynomial function that fits the data. Without GMDH or AI, we would not know the proper degree to use. **Higher degree polynomials more closely match the data points, but may also be more prone to overfitting, where the model is too complex and fails to generalize well to new incoming data.** A proper polynomial degree can be chosen using the least squares regression technique which minimizes the sum of the squared errors between the predicted values and the actual data points using Alexey Ivakhnenko' GMDH mathematical technique along with the powerful central processing units of today's computers. Given that there are an infinite number of polynomial curves, finding a curve with a good fit with small errors is a massive challenge.

The main advantage of GMDH over other polynomial curve fitting techniques is its ability to automatically select the optimal set of polynomial terms, which makes it a powerful tool for discovering complex polynomial relationships between variables. GMDH works by iteratively adding and deleting candidate polynomial terms and selecting the ones that minimize a given error metric, such as the sum of squared errors or the Akaike Information Criterion (AIC).* This approach helps to avoid overfitting and leads to more accurate models. GMDH can handle both linear and nonlinear relationships between variables, It is particularly useful when the underlying relationships between variables are unknown or complex data (Market and economic data are complex!) It can also handle noisy data (such as economic and financial time series.)

* The Akaike Information Criterion (AIC) is a mathematical formula that is used to compare statistical models and select the best one based on a trade-off between model complexity and goodness of fit. AIC was developed by the Japanese statistician Hirotugu Akaike in the 1970s, and it has become a widely used tool in model selection and hypothesis testing. The AIC is based on the principle of parsimony, which suggests that the simplest model that explains the data is usually the best. AIC is a useful tool for selecting a model that balances the trade-off between goodness of fit versus model complexity.

Higher degree polynomials more closely match the data points, but may also be more prone to overfitting, where the model is too complex and fails to generalize well to new incoming data.

Disclaimer

The commentary set forth herein represents the views of Waterloo Capital Management and its investment professionals at the time indicated and is subject to change without notice. The commentary set forth herein was prepared by Waterloo Capital Management based upon information that it believes to be reliable. Waterloo Capital Management expressly disclaims any responsibility to update the commentary set forth herein for any events occurring after the date indicated herein or otherwise. The commentary and other information set forth herein do not constitute an offer to sell, a solicitation to buy, or a recommendation for any security, nor do they constitute investment advice or an offer to provide investment advisory or other services by Waterloo Capital Management. The commentary and other information contained herein shall not be construed as financial or investment advice on any matter set forth herein, and Waterloo Capital Management expressly disclaims all liability in respect of any actions taken based on the commentary and information set forth herein. Hedge funds, private equity and other alternative investments involve a high degree of risk and can be illiquid due to restrictions on transfer and lack of a secondary trading market. They can be highly leveraged, speculative and volatile, and an investor could lose all or a substantial amount of an investment. Alternative investments may lack transparency as to share price, valuation and portfolio holdings. Complex tax structures often result in delayed tax reporting. Compared to registered mutual funds, hedge funds, private equity and other alternative investments are subject to less regulation and often charge higher fees. Alternative investment managers typically exercise broad investment discretion and may apply similar strategies across multiple investment vehicles, resulting in less diversification. Trading may occur outside the United States which may pose greater risks than trading on US exchanges and in US markets. Alternative Investments generally are offered through private placements of securities which are unregistered private placements and are available only to those investors who meet certain eligibility criteria. The above summary is not a complete list of the risks and other important disclosures involved in investing in alternative investments and is subject to the more complete disclosures contained in the Fund's confidential offering documents, which must be reviewed carefully prior to any investment.

WATERLOO™
CAPITAL

2700 Via Fortuna #100, Austin, TX 78746 • (512) 777-5900 • www.WaterlooCap.com

5555 Glenridge Connector, Suite 200 Atlanta, GA 30432 • (470) 552-5902 • www.WaterlooCap.com/IWS