# Simulation of GARCH for Predictive Analytics from Real-Time Data

Statement of Work to Create a Business Prototype (Functional Proof of Concept & Working Model Based on Real Business Process)

## Introduction

Classical linear regression offers static optimization for (production, distribution, inventory) systems in common use. Periodic data (demand, forecasts, capacity) are used for traditional optimization. Elaborate computations generate company-wide plan for next month's (or year's) production schedule or inventory levels. Seldom such plans meet the optimal course of action or can adapt to uncertainty. Managers often introduce *ad hoc* adjustments [1]. Thus real-time optimization is of value in the evolution of digital supply chain management. The slow progress of the latter is compounded by an inadequate penetration of technologies to acquire real-time data and ignorance of advances in econometrics for analysis of real-time data. It has been estimated that some companies spend about 10% of their revenue on safety stock. Could they benefit from right-time supply chain management? Quite simply, increased information used effectively equals increased profit [2]. One highly touted but rarely implemented solution calls for inclusion of software agents in dynamic decision support systems. In addition to data acquisition and event monitoring, agents may heuristically take corrective action in local domains because it is not feasible to continuously re-run the planning algorithm for every fluctuation. Corporate profitability may improve from efficient management of supply chain risks. One approach to increase efficiency may be found in exploring the GARCH approach [3] for real-time optimization in the face of demand volatility and other forms of uncertainty.

Globalization has increased profits and introduced uncertainties that must be managed by organizations that are rapidly evolving as meta-national organizations. They procure from one part of the world, manufacture in another and distribute somewhere else. The complexity of the global consumer driven supply network introduces several unknowns and error terms that are traditionally ignored in models or aggregated to represent a homogeneous form, where, in reality, errors may be highly volatile. To be effective, real-time modeling cannot merely ignore but must account for these traditional assumptions yet cannot compromise its accuracy by accepting static (homoskedastic) representation of error terms or disregard volatility in dynamic analytics. Allowing for heteroskedastic behavior of error terms is essential to manage dynamic risks and improve accuracy of predictive models. Use of the GARCH tool has accomplished the latter and proven effective for predictive analytics in dealing with risk and volatility of financial markets where high volume of data (change in price of shares per minute) is available. With increased penetration of automatic identification technologies (RFID, UWB, sensors) we are able to track & trace objects through networks and geographies, thereby, yielding high volume of real-time object data in sharp contrast to sample data or periodic data points. Hence, automatic identification may enable the business world to fulfill its demand for accurate predictive decisions based on real-time data, patterns and trends for products with short life-cycle (electronics). It is in this regard that a successful adaptation of GARCH may generate tools for operational analytics and boost systemic efficiency of businesses where the real-world heteroskedastic behaviour is the norm. If integrated with agent based real-time knowledge discovery through semantic search engines, the combined decision support potential of such a system may handle any real-world challenge in near real-time (security, defense, emergency).

This proof of concept has modest goals. We wish to simulate the GARCH [3] technique in supply chain management and aim to compare the performance of CLRM vs GARCH in a simulated environment (or use real-world data for testing the model, if available). In the next step, we may combine this technique with business intelligence and data mining tools to generate a knowledge discovery based decision support framework that may be useful now, without waiting for the semantic infrastructure to evolve (as long as we have access to high volume real-time data from RFID or sensors networks to satisfy the requirements of the GARCH technique).

**Scope**

We aim to evaluate the pros and cons of CLRM vs GARCH for accuracy of forecasting and predictive modeling. In a simulated end-to-end business process, we will explore 2 scenarios: [a] CLRM using data points with latency before data is used for optimization and [b] simulated RFID data for real-time decision support. Business process model:

1. Procurement of raw materials
2. Manufacturing
3. Storage and transport of finished goods to manufacturer's distribution center (DC)
4. Transport of finished goods to from manufacturer's DC to distributor's DC
5. Transport from distributor's DC to retailers
6. Service options (returns to manufacturer)

| PHASE | **TASKS** | TIME |
|---|---|---|
| I | Upon approval of funding, gather pre-arranged team members to discuss phases, tasks, milestones. Ensure clarity of project and secure software tools (eViews, GoldSim, Arena). | 16 weeks |
| II | Create business process (supply network) simulation and assign parameters. Identify data generating nodes for CLRM (periodic data) and GARCH (RFID data) scenarios. | 16 weeks |
| III | Test CLRM model for forecasting with simulated data. | 16 weeks |
| IV | Test GARCH model for forecasting (same variables as in CLRM) using simulated real-time data (assumption: obtained from nodes identified in phase II). | 8 weeks |
| V | Compare CLRM vs GARCH model. Evaluate and optimize for improvements. Determine value of data and improved visibility. Couple improvements to scm planning and decision support. | 16 weeks |
| VI | Couple time series data and real-time predictive analytical tool with right-time data mining for improved decision support (better knowledge discovery). | **OPTIONAL** 24 weeks |

**References**


[1] R. Kalakota, J. Stallaert and A. C. Whinston. Implementing Real time Supply Chain Optimization Systems. Supply Chain Management (1995)

[2] J. Dyche. Real Time or Right Time - Explaining The Real Time Enterprise. CRM Guru (2003)

[3] S. Datta and C. Granger. Potential to Improve Forecasting Accuracy (2006) http://esd.mit.edu/WPS/esd-wp-2006-11.pdf ; http://esd.mit.edu/WPS/esd-wp-2006-10.pdf

**Can you risk not investing in GARCH simulation ?**

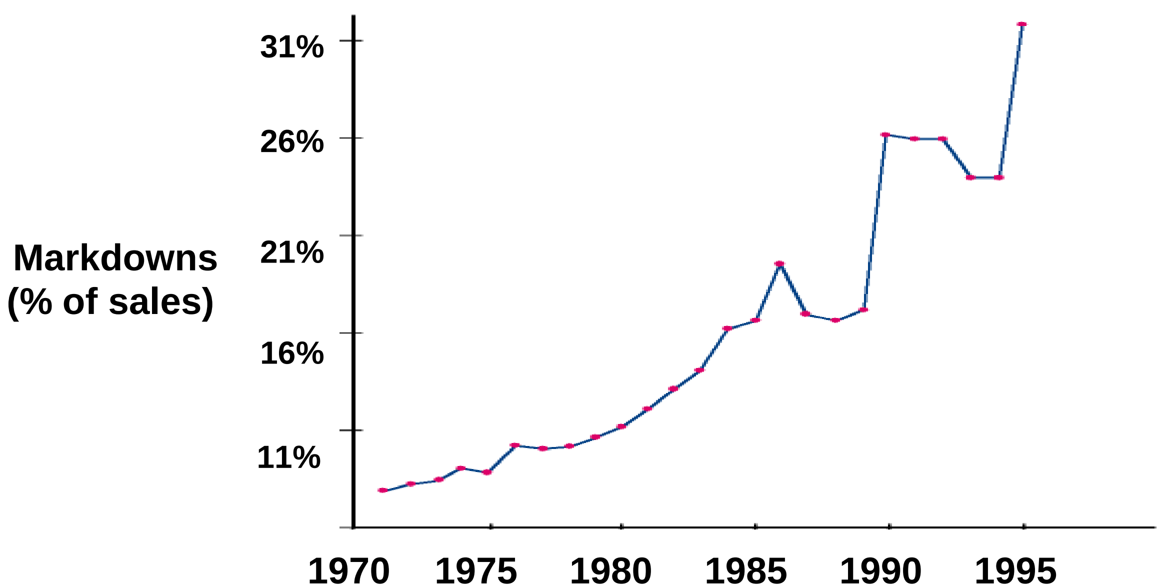
Dr Shoumen Datta, Research Director, Forum for Supply Chain Innovation, Massachusetts Institute of Technology shoumen@mit.edu
>

Port closures due to foot and mouth disease (FMD) and severe acute respiratory syndrome (SARS) or the disruption of raw material procurement (to produce rubber) due to civil war in Liberia are perhaps well-known examples of environmental and political risks that may impact the profitability of local and global businesses.

At 8:00 PM on 17 March 2000, lightning struck a power grid and caused a fire in a Philips semiconductor plant in Albuquerque, NM. The fire was soon extinguished but the plant ceased to operate. This factory supplied ASIC chips to Nokia and Ericsson. Were Nokia and Ericsson prepared to mitigate such supply risks? Nokia's systems detected shipment discrepancies within three days. The supplier (Philips) was pushed hard to offer new supply sources. Modular architecture enabled Nokia to use and adapt to new chip design by grabbing capacity elsewhere in the global ASIC market. Ericsson remained oblivious of the problem for weeks due to slow chain of information analysis (poor key performance indicators or KPI) and lack of contingent planning. When it realized the potential of the catastrophe, all the available capacity in the global market was already committed. It suffered a $400 million loss and quit phone manufacturing (Avoiding Supply-Chain Breakdown. MIT Sloan Management Review, Fall 2004).

On 27 April 2001, the San Jose Mercury News exposed how Cisco was stuck with stacks of chips, circuit boards and other components worth over $2.5 billion of inventory that it believed it won't be able to sell within the next year. Why did a revered industry leader like Cisco failed to detect a brewing multi-billion dollar inventory risk?

A study of retail clothing stores found that "a third of customers entering a store leave without buying. They can't find what they came to buy." A fact reflected by increased markdowns (inventory risk, price risk, customer service risk and capital efficiency risk). http://gsbwww.uchicago.edu/news/capideas/summer02/measuremanage.html




Risk in supply chain management originates from 2 key areas: supply and demand. Next, of equal importance, are environmental, political and process risks. Political and environmental risks may always remain amorphous and refractory to adequate quantification. However, the current definition of process risk is poorly differentiated due to lack of clarity among business school pundits and includes examples as diverse as healthcare management for employees as well as standardization of operational procedures. Process discrepancies between organizations increase risks especially when information and communication technologies are used as a medium of exchange (or in systems integration). Lack of a 'common' business vocabulary is currently addressed by adopting 'band-aid' solutions from RosettaNet or in the form of global process standards such as ISO 9000 or developing specialized languages such as e-business eXtended Mark-up Language (ebXML). It is this approach that stands to undergo a radical metamorphosis with the gradual emergence of ontological frameworks and the diffusion of the semantic web.

**How Paris, France lost the bid for Olympic 2012 & created a supply chain nightmare for textile retailers**


,

Opacity of data from supply chain nodes (supplier, distributor, transportation provider) will increase risk whereas transparency may reduce risk, if the data is analysed and its impact sufficiently understood to deploy risk mitigation steps, at the right time. Operational transparency at or within supply chain nodes may improve with the increase in object associated data acquisition that may be possible through pervasive adoption of automatic identification technologies (RFID, UWB, GPS, sensors). The use and analysis of this data in a model that captures the end-to-end business network (as well as links to other factors that may impact the function of a specific node) may help to reduce risk. It is in this context that a combinatorial use of MGARCH and VAR techniques may offer value hitherto unimaginable.

This proposal is relevant to those who are increasingly using “lean” principles and have global outsourcing practices which may compromise the visibility of the supply chains. Transparency of operations within the corporation (internal risk drivers) are as critical as data from business partners in “lean” and “global” operations to evaluate external risk drivers. In some cases, outlier events may be even more influential.

Businesses often introduce risk under two broad categories: [a] quantitative anomalies resulting from selection bias or principal component analysis and [b] qualitative effects stemming from pressures to enhance productivity, eliminate waste, remove duplication and minimize cost yet increase service levels to customers. Balancing these priorities require continuous risk mitigation strategies, real-time data and analytical tools. However, neither the data nor the tools to analyse such data (estimate risk) are adequate, at present. Often, risk is viewed as simplistic as merely the product of frequency and consequence. High-frequency but low-consequence event (currency exchange rates) are viewed as similar to a low-frequency but high-consequence event (sinking of a cargo ship laden with spare parts). In reality such apparently “similar risks” may have vastly different effects. Often sensational risks grab attention and beg for resource-consuming mitigation while risk managers tend to ignore the smaller risks that create the real friction in the supply chain.

With the increasingly complex business environment that is the hallmark of globalization, supply chain presents a myriad of factors that represent the complexity of supply-demand network risks. If accounted as parameters in traditional optimization equations, the sheer number of factors will exponentially increase the state space and as a result may grind the computation of the optimization algorithms to a pace that may become unacceptable for decision support systems in the management of supply chain adaptability.

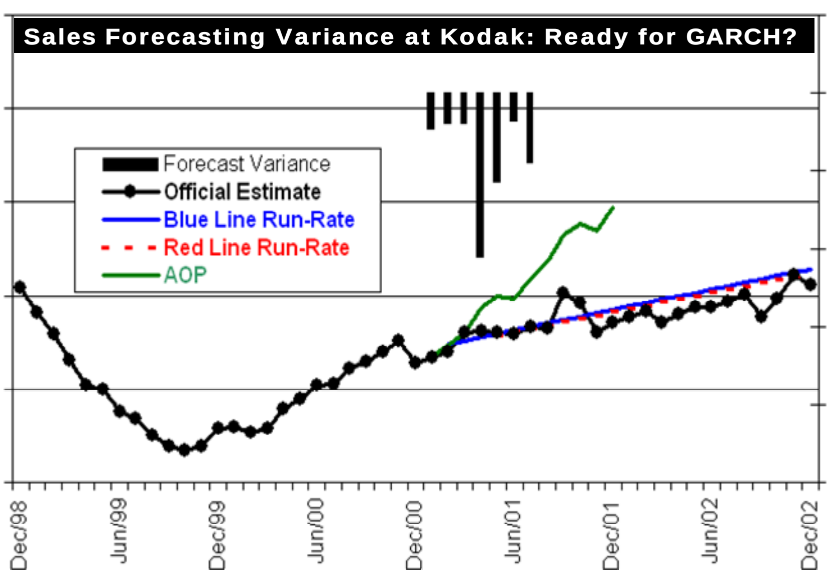
The GARCH model may be designed to take into account the details of the operational nodes (assuming we have data available from each of these nodes/processes). Recurring analysis performed in near real-time (assuming real-time data is available to the analytical engine) may offer results that may predict or detect risks in the operational model (supply chain) far in advance of what is possible at present. This is the expectation from this proposal. The validity of this proposal as a tool for risk analysis may be tested by simulating a MGARCH-VAR model of a real world business operation and running the simulation with real-time data (observed or simulated) to test the technique and the tool to be considered a risk mitigation strategy. Research to create such a simulation testing tool is currently under consideration.

Use of GARCH model in supply chain risk analysis may also help create a merger of fields and minds by integrating financial and physical supply chains. The financial supply chain, which drives financial settlement, takes over where the physical supply chain ends. Exporters want rapid payment while importers demand accurate data on goods received to better manage inventory and cash-flow to optimize working capital management. Thus, capital efficiency (the traditional domain of the CFO) depends on data and sharing of information (the traditional domain of the CTO) about cross-border movement of goods (customs and excise), transfer of title, risk mitigation and payment. Therefore, facilitation of the flow of (decisionable and actionable) information, across physical and financial supply chains has a direct impact on working capital optimization.

From a risk management perspective, the global supply chain is, therefore, a component of the CFO's responsibility. Adapting the GARCH model to serve as a tool in supply chain risk analysis may offer financial managers a familiar tool that may yield clues to supply chain risks. Such a tool is highly desirable for financial managers to improve capital efficiency which is threatened by a heightened regulatory emphasis that is driving further inefficiencies in the cash conversion cycle. Global security risks have triggered the 24-hour manifest rule by the US customs and border protection agencies that require importers to submit an electronic manifest of goods to verify the validity of a ship's cargo. Such requirements underscore how increased concerns around issues of security related to global supply chain activities (risk of a tampered shipping container) also have direct impact on the CFO.

Comprehensive solutions are necessary over the life of a transaction cycle that may integrate cash management, trade settlement, finance, logistics, supply nodes, procurement, demand projections, inventory, human resources, regulatory policy compliance and management of information across physical and financial supply chains. Creating one or more models that may work in synergy and integrate such real-world scenarios will be a challenge. The proven success of GARCH in finance and the potential to adapt GARCH for business operations (supply chains) may offer a synergistic multi-faceted tool for risk-adjusted supply chain management by acting as a bridge for some of the interdependent issues in business: finance, supply chain and management of risk.

How can you risk not to reap the rewards of the application potential of GARCH to improve your profitability?



# Risk in the Global Supply Chain

*Transparency is Key to Forecasting Risk and Risk Analysis is the Key Element in Security*

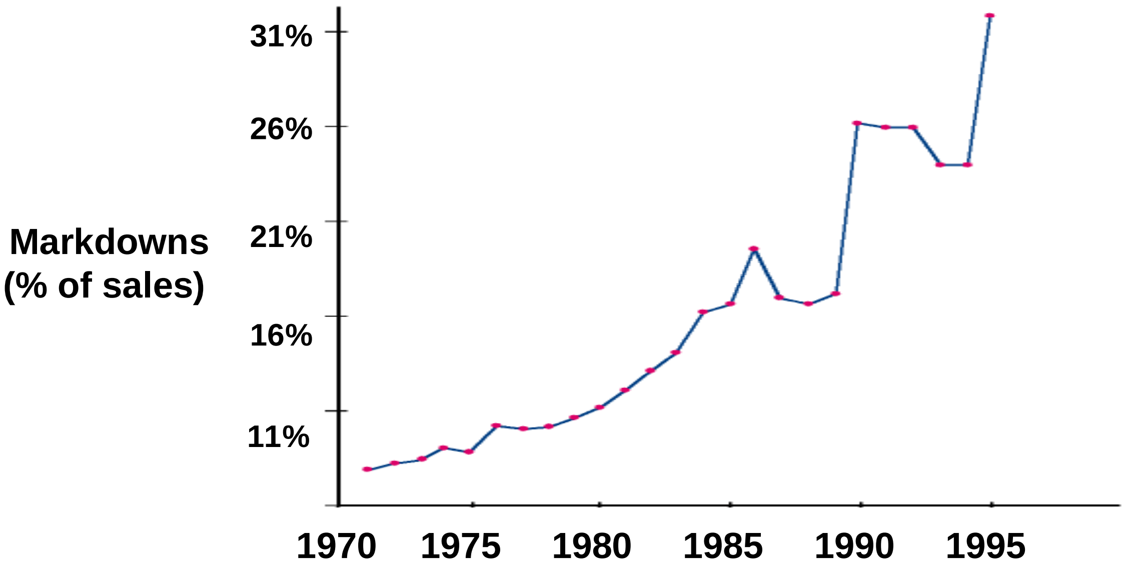
Dr Shoumen Palit Austin Datta, School of Engineering, MIT and Co-Founder & Research Director, MIT Forum for Supply Chain Innovation

At 8pm on 17 March 2000, lightning struck a power grid and caused a fire in a Philips semiconductor plant in Albuquerque, NM. The fire was soon extinguished but the plant ceased to operate. This factory supplied ASIC chips to Nokia and Ericsson. Were Nokia and Ericsson prepared to mitigate such supply chain risks? The answer is implicit in the fact that several years later we still have Nokia but Ericsson is now better known as Sony-Ericsson. Specifically, Nokia's systems detected shipment discrepancies within 3 days. The supplier (Philips) was pushed hard to offer new supply sources. Modular architecture enabled Nokia to use and adapt to new chip design by grabbing capacity elsewhere in the global ASIC market. Ericsson remained oblivious of the problem for weeks due to an inefficient chain of information analysis (poor key performance indicators or KPI) and lack of contingent planning. When it realized the potential of the catastrophe, all the available capacity in the global market was already committed. It suffered a $400 million loss of revenue and was forced to exit mobile phone manufacturing. Sony bought Ericsson, soon thereafter.
(Avoiding Supply-Chain Breakdown in MIT Sloan Management Review, Fall 2004).

Port closures due to foot and mouth disease (FMD) and severe acute respiratory syndrome (SARS) or the disruption of raw material procurement (to produce rubber) due to civil war in Liberia are common examples of environmental and political risks that may impact the profitability of local and global businesses. In addition, uncertainty has forged an even stronger alliance with security. If an explosive device containing radioactive material or bio-material (such as, anthrax) were to detonate when a container is off-loaded, the authorities may close all the nation's ports until every container on every site in the country is inspected. In October 2002, a "war game" that mimicked that scenario found that closing US ports for as few as 12 days created a 60-day container backlog and cost the economy roughly $58 billion.

On 27 April 2001, the San Jose Mercury News made headlines by exposing how Cisco was stuck with stacks of chips, circuit boards and other components worth over $2.5 billion of inventory that it believed it won't be able to sell within the next year. Why did a revered industry leader like Cisco failed to detect a brewing multi-billion dollar inventory risk? A study of retail clothing stores (see illustration) found that "a third of customers entering a store leave without buying. They can't find what they came to buy." A fact reflected by increased markdowns (inventory risk, price risk, customer service risk and capital efficiency risk).

Risk in supply chain management originates from two key areas: supply and demand. At the next level of equal importance are environmental, political, process and security risks. Political and environmental risks may always remain amorphous and refractory to adequate quantification. Security risks are even more volatile but on a far higher priority level. However, the current definition of process risk is poorly differentiated due to lack of clarity among business school pundits and includes examples as diverse as healthcare management for employees as well as standardization of operational procedures. Process discrepancies between organizations increase risks especially when information and communication technologies (ICT) are used as a medium of exchange (or in systems integration). Lack of a 'common' business vocabulary is currently addressed by adopting 'band-aid' solutions from RosettaNet or in the form of global process standards such as ISO 9000 or developing specialized languages such as e-business eXtended Mark-up Language (ebXML). It is this approach that stands to undergo a radical metamorphosis with the gradual emergence of enterprise-wide ontological frameworks and sufficient diffusion of the semantic web.




DeHoratius, N. http://gsbwww.uchicago.edu/news/capideas/summer02/measuremanage.html

Opacity of data from supply chain nodes (supplier, distributor, warehouse, transportation provider) will increase risk but transparency may reduce risk, if the data is analysed and its impact sufficiently understood to deploy risk mitigation steps, at the right time. Operational transparency at or within supply chain nodes may improve with the increase in object associated data acquisition that may be possible through pervasive adoption of automatic identification technologies (RFID, UWB, GPS, sensors). The use and analysis of this data in a model that captures the end-to-end business network (as well as links to other factors that may impact the function of a specific node) may help to reduce risk. It is in this context that a combinatorial use of MGARCH and VAR techniques may offer value hitherto unimaginable (Adapting Decisions, Optimizing Facts and predicting Figures by Shoumen Palit Austin Datta, in this volume) in the area of forecasting and analysis of risk. Forecasting and analysis of risk is particularly relevant to those who are increasingly using “lean” principles and have global outsourcing practices which may compromise the visibility of the supply chain. Transparency of operations within the corporation (internal risk drivers) are as critical as data from business partners in “lean” and “global” operations to evaluate external risk drivers. In some cases, outlier events may be more influential.

Businesses often introduce risk under two broad categories: [a] quantitative anomalies resulting from selection bias or principal component analysis and [b] qualitative effects stemming from pressures to enhance productivity, eliminate waste, remove duplication and minimize cost yet increase service levels to customers. Balancing these priorities require continuous risk mitigation strategies, real-time data and analytical tools. However, neither the data nor the tools to analyse such data (estimate risk) are adequate, at present. Often, risk is viewed as simplistic as merely the product of frequency and consequence. A high-frequency but low-consequence event (currency exchange rates) are viewed as similar to a low-frequency but high-consequence event (sinking of a cargo ship laden with spare parts). In reality such apparently “similar risks” may have vastly different effects. Often sensational risks grab attention and beg for resource-consuming mitigation while risk managers tend to ignore the smaller risks that create the real friction in the supply chain. With the increasingly complex business environment that is the hallmark of globalization, supply chain presents a myriad of factors that represent the complexity of supply-demand network risks. If accounted as parameters in traditional optimization equations, the sheer number of factors will exponentially increase the state space and as a result may grind the computation of the optimization algorithms to a pace that may become unacceptable for decision support systems in the management of supply chain adaptability.

Generalized autoregressive conditional heteroskedasticity or the GARCH model (Adapting Decisions, Optimizing Facts and predicting Figures by Shoumen Palit Austin Datta, in this volume) may be designed to take into account the details of the operational nodes (assuming we have data available from each of these nodes/processes). Recurring analysis performed in near real-time (assuming real-time data is available to the analytical engine) may offer results that may predict or detect risks in the operational model (supply chain) far in advance of what is possible at present. The validity of this proposal as a tool for risk analysis may be tested by simulating multivariate GARCH-VAR (vector auto regression) model of a real world business operation and running the simulation with real-time data (observed or simulated) to test the technique and the tool to be considered for risk analysis. The importance of research to create such a simulation testing tool cannot be overemphasized given the looming security threats from cross-border global commerce.

Use of GARCH model in supply chain risk analysis may also help create a merger of fields and minds by integrating financial and physical supply chains. The financial supply chain, which drives financial settlement, takes over where the physical supply chain ends. Exporters want rapid payment while importers demand accurate data on goods received to better manage inventory and cash-flow to optimize working capital management. Thus, capital efficiency (the traditional domain of the CFO) depends on data and sharing of information (the traditional domain of the CTO) about cross-border movement of goods (customs and excise), transfer of title, risk mitigation and payment. Therefore, facilitation of the flow of decisionable actionable information, across physical and financial supply chains has a direct impact on working capital.

From a risk management perspective, the global supply chain is, therefore, a component of the CFO's responsibility. Adapting the GARCH model to serve as a tool in supply chain risk analysis may offer financial managers a familiar tool that may yield clues to supply chain risks. Such a tool is highly desirable for financial managers to improve capital efficiency which is threatened by a heightened security and regulatory (Sarbanes-Oxley Act) emphasis that is driving further inefficiencies in the cash conversion cycle. Global security risks have triggered the 24-hour manifest rule by the US customs and border protection agencies that require importers to submit an electronic manifest of goods to verify the validity of a ship's cargo. Such requirements underscore how increased concerns around issues of security related to global supply chain activities (risk of a tampered shipping container) also have direct impact on the CFO.

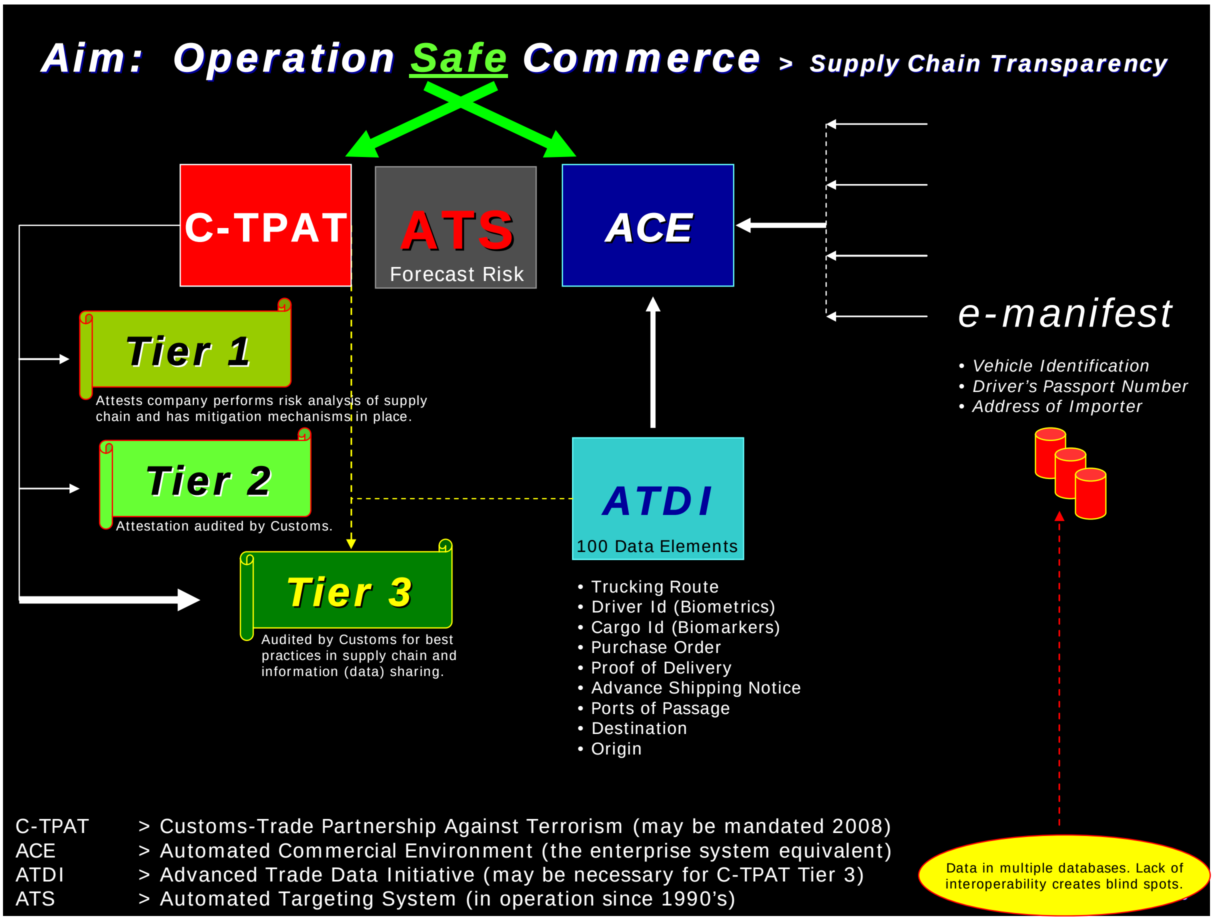
Hence, comprehensive solutions are necessary over the life of a transaction cycle that may integrate cash management, trade settlement, finance, logistics, supply nodes, procurement, demand projections, inventory, human resources, regulatory policy compliance and management of information across physical and financial supply chains. Creating one or more models that may work in synergy and integrate such real-world scenarios will be a challenge. The proven success of GARCH in finance and the potential to adapt GARCH for business operations (supply chains) may offer a synergistic multi-faceted tool for risk-adjusted supply chain management by acting as a bridge for some of the interdependent issues in business: finance, supply chain, security and management of risk.

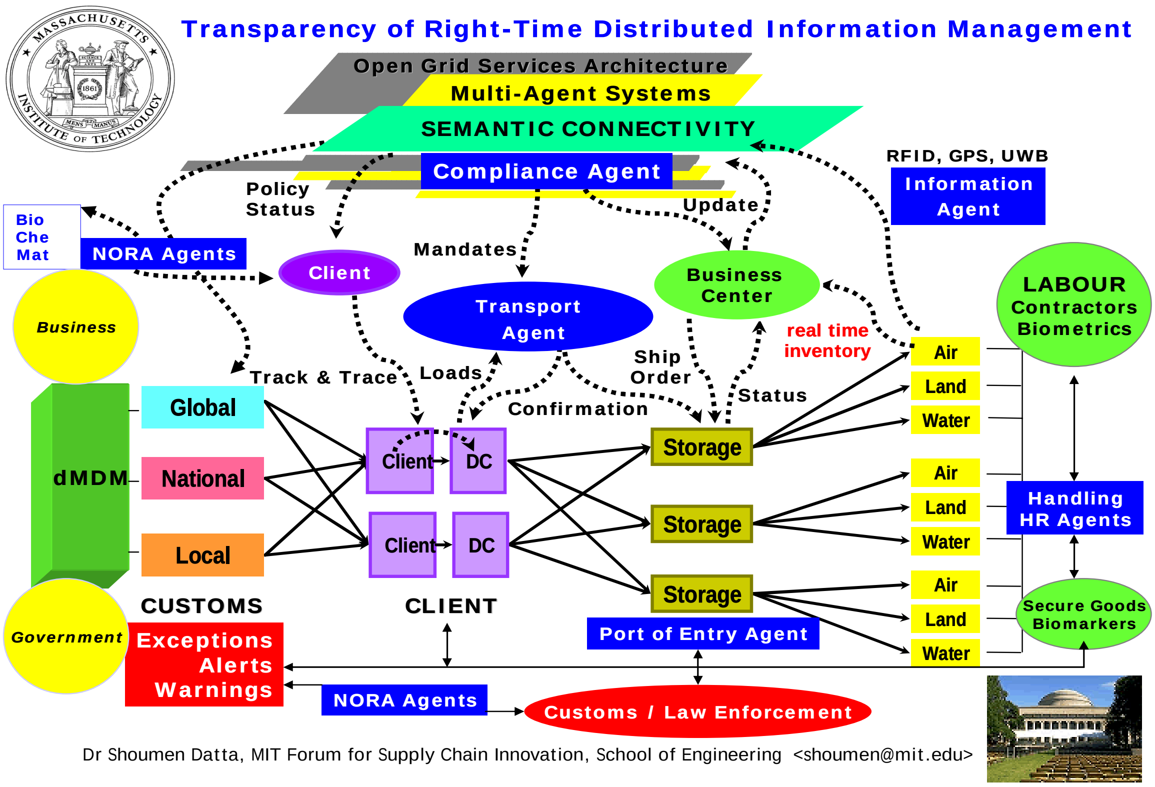
The inescapable focus on supply chain transparency may have been, thus far, a business process decision to optimize for profitability. Wal*Mart may be the poster child (or ogre) for its pursuit of transparency, in its preferred form, by driving the use of RFID tags on certain shipments. While this form of transparency offers limited business process visibility, it offers few systemic benefits to the supply chain partners. The cost for doing business with such behemoths simply increases because the cost of the technology (RFID) must be assumed by the supplier while Wal*Mart shareholders expect to reap the benefits. Data sharing or information exchange is still the exception and giants like Wal*Marts can dictate the terms of data sharing, if at all. However, change is looming on the horizon. It may soon appear in the form of mandates from the US Department of Homeland Security for the justifiable reason of national security (US). What if, even sooner, Japan demands real-time test results of each head of cattle from US in order to resume US beef exports without reservations? US tests 1% of the 35 million cattle slaughtered each year in the US for meat production.

Cost of doing business with and in the US may soon have to figure in the costs necessary to implement transparency. Businesses must share data with US Department of Homeland Security if their goods originate overseas. The model of data sharing may soon be adopted by other countries determined to counter terrorism. The move toward global supply chain transparency if not a matter of if but a question of when. The lack of systems interoperability and the inability to make intelligent decisions may create many more problems before it starts providing solutions. If even a tiny fraction of the 25,000 containers that arrive in US ports each day require inspection, then goods will face customs clearance delays.

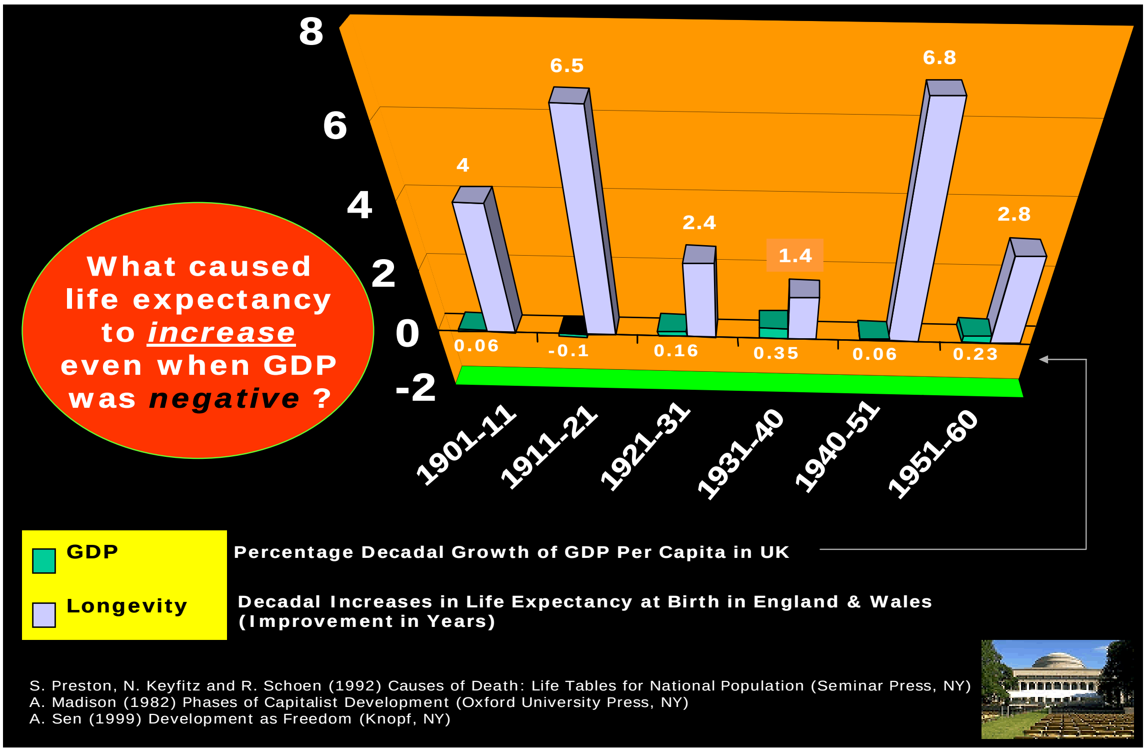
The illustration below outlines some of the pilot projects in progress in the US. It points toward a possible mandate by the US in the form of Customs-Trade Partnership Against Terrorism. To qualify for Tier 3 certification in C-TPAT, it will be necessary for businesses to share data through the Advanced Trade Data Initiative. Sharing business sensitive data will add layers of data security necessary to ensure safety from hackers in pursuit of competitive information or business intelligence. With data from ATDI, the customs "enterprise" system or Automated Commercial Environment (ACE) is expected to run analytical engines to spot anomalies, integrate biometric information about individuals involved in handling goods, perform non-obvious relationship analysis (NORA) and forecast risk associated with individual containers or shipments in general. Armed with this risk profile, customs may inspect cargo containers that exceed a threshold.

To achieve even a limited proficiency in this operation, it will be necessary to pursue convergence of intelligent data mining through Agents based on artificial intelligence coupled with tools such as GARCH plus the innovations from track & trace technologies including radio frequency identification, biometrics, sensors, software defined radio and GPS. This is a vision in dire need of interoperability between systems and a feasible yet meaningful convergence of several ideas.






The illustration above outlines the multitude of connections necessary for supply chain transparency. Analytical tools may process the acquired data to shed light on security risks. The underlying theme of sharing (in this case, data) may draw an important lesson from economics (illustration below) as outlined by Amartya Sen (1998 Nobel Prize for Economics).



# Simulation of GARCH for Predictive Analytics from Real-Time Data

Statement of Work to Create a Business Prototype (Functional Proof of Concept & Working Model Based on Real Business Process)

## Introduction

Classical linear regression offers static optimization for (production, distribution, inventory) systems in common use. Periodic data (demand, forecasts, capacity) are used for traditional optimization. Elaborate computations generate company-wide plan for next month's (or year's) production schedule or inventory levels. Seldom such plans meet the optimal course of action or can adapt to uncertainty. Managers often introduce *ad hoc* adjustments [1]. Thus real-time optimization is of value in the evolution of digital supply chain management. The slow progress of the latter is compounded by an inadequate penetration of technologies to acquire real-time data and ignorance of advances in econometrics for analysis of real-time data. It has been estimated that some companies spend about 10% of their revenue on safety stock. Could they benefit from right-time supply chain management? Quite simply, increased information used effectively equals increased profit [2]. One highly touted but rarely implemented solution calls for inclusion of software agents in dynamic decision support systems. In addition to data acquisition and event monitoring, agents may heuristically take corrective action in local domains because it is not feasible to continuously re-run the planning algorithm for every fluctuation. Corporate profitability may improve from efficient management of supply chain risks. One approach to increase efficiency may be found in exploring the GARCH approach [3] for real-time optimization in the face of demand volatility and other forms of uncertainty.

Globalization has increased profits and introduced uncertainties that must be managed by organizations that are rapidly evolving as meta-national organizations. They procure from one part of the world, manufacture in another and distribute somewhere else. The complexity of the global consumer driven supply network introduces several unknowns and error terms that are traditionally ignored in models or aggregated to represent a homogeneous form, where, in reality, errors may be highly volatile. To be effective, real-time modeling cannot merely ignore but must account for these traditional assumptions yet cannot compromise its accuracy by accepting static (homoskedastic) representation of error terms or disregard volatility in dynamic analytics. Allowing for heteroskedastic behavior of error terms is essential to manage dynamic risks and improve accuracy of predictive models. Use of the GARCH tool has accomplished the latter and proven effective for predictive analytics in dealing with risk and volatility of financial markets where high volume of data (change in price of shares per minute) is available. With increased penetration of automatic identification technologies (RFID, UWB, sensors) we are able to track & trace objects through networks and geographies, thereby, yielding high volume of real-time object data in sharp contrast to sample data or periodic data points. Hence, automatic identification may enable the business world to fulfill its demand for accurate predictive decisions based on real-time data, patterns and trends for products with short life-cycle (electronics). It is in this regard that a successful adaptation of GARCH may generate tools for operational analytics and boost systemic efficiency of businesses where the real-world heteroskedastic behaviour is the norm. If integrated with agent based real-time knowledge discovery through semantic search engines, the combined decision support potential of such a system may handle any real-world challenge in near real-time (security, defense, emergency).

This proof of concept has modest goals. We wish to simulate the GARCH [3] technique in supply chain management and aim to compare the performance of CLRM vs GARCH in a simulated environment (or use real-world data for testing the model, if available). In the next step, we may combine this technique with business intelligence and data mining tools to generate a knowledge discovery based decision support framework that may be useful now, without waiting for the semantic infrastructure to evolve (as long as we have access to high volume real-time data from RFID or sensors networks to satisfy the requirements of the GARCH technique).

**Scope**

We aim to evaluate the pros and cons of CLRM vs GARCH for accuracy of forecasting and predictive modeling. In a simulated end-to-end business process, we will explore 2 scenarios: [a] CLRM using data points with latency before data is used for optimization and [b] simulated RFID data for real-time decision support. Business process model:

1. Procurement of raw materials
2. Manufacturing
3. Storage and transport of finished goods to manufacturer's distribution center (DC)
4. Transport of finished goods to from manufacturer's DC to distributor's DC
5. Transport from distributor's DC to retailers
6. Service options (returns to manufacturer)

| PHASE | **TASKS** | TIME |
|---|---|---|
| I | Upon approval of funding, gather pre-arranged team members to discuss phases, tasks, milestones. Ensure clarity of project and secure software tools (eViews, GoldSim, Arena). | 16 weeks |
| II | Create business process (supply network) simulation and assign parameters. Identify data generating nodes for CLRM (periodic data) and GARCH (RFID data) scenarios. | 16 weeks |
| III | Test CLRM model for forecasting with simulated data. | 16 weeks |
| IV | Test GARCH model for forecasting (same variables as in CLRM) using simulated real-time data (assumption: obtained from nodes identified in phase II). | 8 weeks |
| V | Compare CLRM vs GARCH model. Evaluate and optimize for improvements. Determine value of data and improved visibility. Couple improvements to scm planning and decision support. | 16 weeks |
| VI | Couple time series data and real-time predictive analytical tool with right-time data mining for improved decision support (better knowledge discovery). | **OPTIONAL** 24 weeks |

**References**

[1] R. Kalakota, J. Stallaert and A. C. Whinston. Implementing Real time Supply Chain Optimization Systems. Supply Chain Management (1995)

[2] J. Dyche. Real Time or Right Time - Explaining The Real Time Enterprise. CRM Guru (2003)

[3] S. Datta and C. Granger. Potential to Improve Forecasting Accuracy (2006) http://esd.mit.edu/WPS/esd-wp-2006-11.pdf ; http://esd.mit.edu/WPS/esd-wp-2006-10.pdf

**Can you risk not investing in GARCH simulation ?**

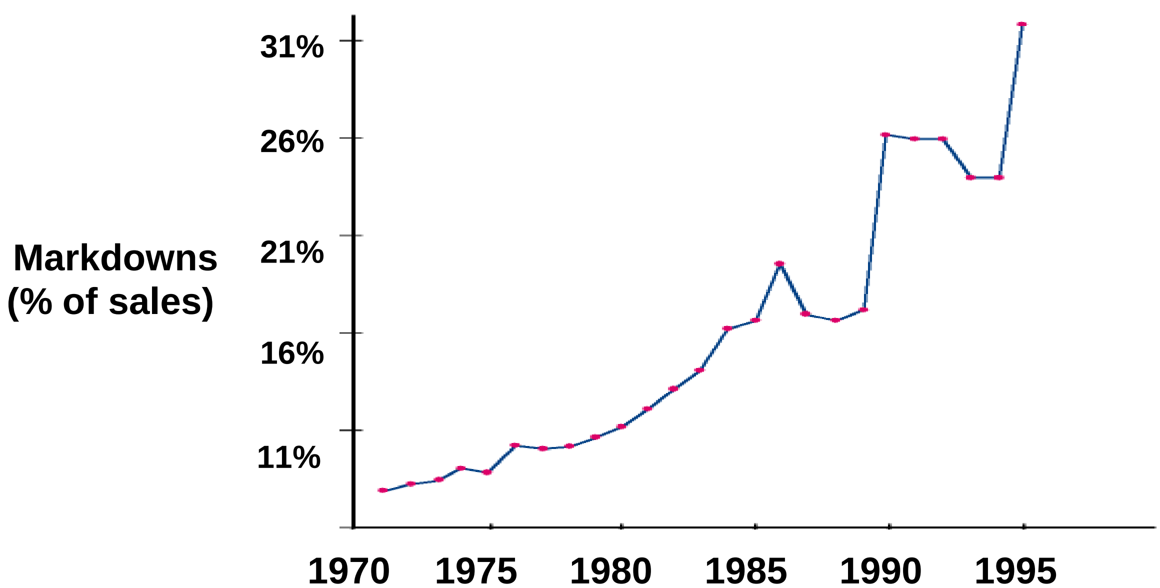
Dr Shoumen Datta, Research Director, Forum for Supply Chain Innovation, Massachusetts Institute of Technology shoumen@mit.edu
>

Port closures due to foot and mouth disease (FMD) and severe acute respiratory syndrome (SARS) or the disruption of raw material procurement (to produce rubber) due to civil war in Liberia are perhaps well-known examples of environmental and political risks that may impact the profitability of local and global businesses.

At 8:00 PM on 17 March 2000, lightning struck a power grid and caused a fire in a Philips semiconductor plant in Albuquerque, NM. The fire was soon extinguished but the plant ceased to operate. This factory supplied ASIC chips to Nokia and Ericsson. Were Nokia and Ericsson prepared to mitigate such supply risks? Nokia's systems detected shipment discrepancies within three days. The supplier (Philips) was pushed hard to offer new supply sources. Modular architecture enabled Nokia to use and adapt to new chip design by grabbing capacity elsewhere in the global ASIC market. Ericsson remained oblivious of the problem for weeks due to slow chain of information analysis (poor key performance indicators or KPI) and lack of contingent planning. When it realized the potential of the catastrophe, all the available capacity in the global market was already committed. It suffered a $400 million loss and quit phone manufacturing (Avoiding Supply-Chain Breakdown. MIT Sloan Management Review, Fall 2004).

On 27 April 2001, the San Jose Mercury News exposed how Cisco was stuck with stacks of chips, circuit boards and other components worth over $2.5 billion of inventory that it believed it won't be able to sell within the next year. Why did a revered industry leader like Cisco failed to detect a brewing multi-billion dollar inventory risk?

A study of retail clothing stores found that "a third of customers entering a store leave without buying. They can't find what they came to buy." A fact reflected by increased markdowns (inventory risk, price risk, customer service risk and capital efficiency risk). http://gsbwww.uchicago.edu/news/capideas/summer02/measuremanage.html




Risk in supply chain management originates from 2 key areas: supply and demand. Next, of equal importance, are environmental, political and process risks. Political and environmental risks may always remain amorphous and refractory to adequate quantification. However, the current definition of process risk is poorly differentiated due to lack of clarity among business school pundits and includes examples as diverse as healthcare management for employees as well as standardization of operational procedures. Process discrepancies between organizations increase risks especially when information and communication technologies are used as a medium of exchange (or in systems integration). Lack of a 'common' business vocabulary is currently addressed by adopting 'band-aid' solutions from RosettaNet or in the form of global process standards such as ISO 9000 or developing specialized languages such as e-business eXtended Mark-up Language (ebXML). It is this approach that stands to undergo a radical metamorphosis with the gradual emergence of ontological frameworks and the diffusion of the semantic web.

**How Paris, France lost the bid for Olympic 2012 & created a supply chain nightmare for textile retailers**


Opacity of data from supply chain nodes (supplier, distributor, transportation provider) will increase risk whereas transparency may reduce risk, if the data is analysed and its impact sufficiently understood to deploy risk mitigation steps, at the right time. Operational transparency at or within supply chain nodes may improve with the increase in object associated data acquisition that may be possible through pervasive adoption of automatic identification technologies (RFID, UWB, GPS, sensors). The use and analysis of this data in a model that captures the end-to-end business network (as well as links to other factors that may impact the function of a specific node) may help to reduce risk. It is in this context that a combinatorial use of MGARCH and VAR techniques may offer value hitherto unimaginable.

This proposal is relevant to those who are increasingly using “lean” principles and have global outsourcing practices which may compromise the visibility of the supply chains. Transparency of operations within the corporation (internal risk drivers) are as critical as data from business partners in “lean” and “global” operations to evaluate external risk drivers. In some cases, outlier events may be even more influential.

Businesses often introduce risk under two broad categories: [a] quantitative anomalies resulting from selection bias or principal component analysis and [b] qualitative effects stemming from pressures to enhance productivity, eliminate waste, remove duplication and minimize cost yet increase service levels to customers. Balancing these priorities require continuous risk mitigation strategies, real-time data and analytical tools. However, neither the data nor the tools to analyse such data (estimate risk) are adequate, at present. Often, risk is viewed as simplistic as merely the product of frequency and consequence. High-frequency but low-consequence event (currency exchange rates) are viewed as similar to a low-frequency but high-consequence event (sinking of a cargo ship laden with spare parts). In reality such apparently “similar risks” may have vastly different effects. Often sensational risks grab attention and beg for resource-consuming mitigation while risk managers tend to ignore the smaller risks that create the real friction in the supply chain.

With the increasingly complex business environment that is the hallmark of globalization, supply chain presents a myriad of factors that represent the complexity of supply-demand network risks. If accounted as parameters in traditional optimization equations, the sheer number of factors will exponentially increase the state space and as a result may grind the computation of the optimization algorithms to a pace that may become unacceptable for decision support systems in the management of supply chain adaptability.

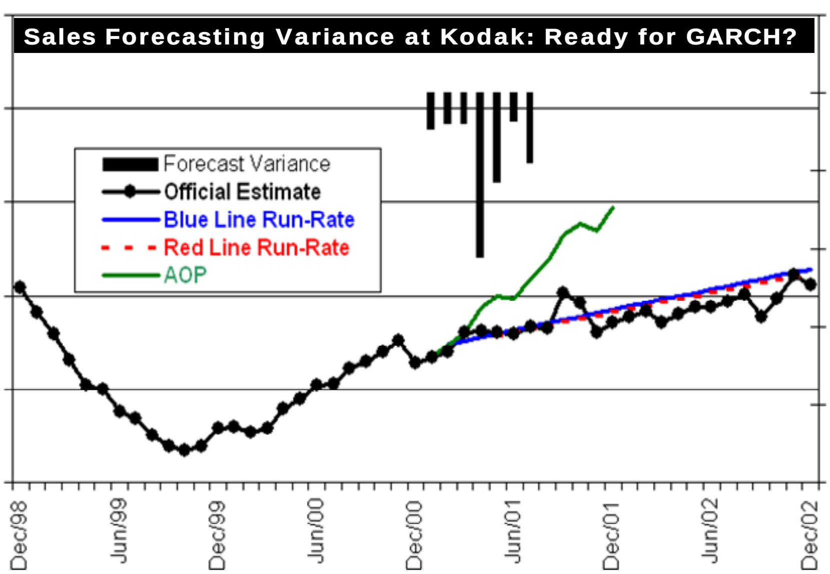
The GARCH model may be designed to take into account the details of the operational nodes (assuming we have data available from each of these nodes/processes). Recurring analysis performed in near real-time (assuming real-time data is available to the analytical engine) may offer results that may predict or detect risks in the operational model (supply chain) far in advance of what is possible at present. This is the expectation from this proposal. The validity of this proposal as a tool for risk analysis may be tested by simulating a MGARCH-VAR model of a real world business operation and running the simulation with real-time data (observed or simulated) to test the technique and the tool to be considered a risk mitigation strategy. Research to create such a simulation testing tool is currently under consideration.

Use of GARCH model in supply chain risk analysis may also help create a merger of fields and minds by integrating financial and physical supply chains. The financial supply chain, which drives financial settlement, takes over where the physical supply chain ends. Exporters want rapid payment while importers demand accurate data on goods received to better manage inventory and cash-flow to optimize working capital management. Thus, capital efficiency (the traditional domain of the CFO) depends on data and sharing of information (the traditional domain of the CTO) about cross-border movement of goods (customs and excise), transfer of title, risk mitigation and payment. Therefore, facilitation of the flow of (decisionable and actionable) information, across physical and financial supply chains has a direct impact on working capital optimization.

From a risk management perspective, the global supply chain is, therefore, a component of the CFO's responsibility. Adapting the GARCH model to serve as a tool in supply chain risk analysis may offer financial managers a familiar tool that may yield clues to supply chain risks. Such a tool is highly desirable for financial managers to improve capital efficiency which is threatened by a heightened regulatory emphasis that is driving further inefficiencies in the cash conversion cycle. Global security risks have triggered the 24-hour manifest rule by the US customs and border protection agencies that require importers to submit an electronic manifest of goods to verify the validity of a ship's cargo. Such requirements underscore how increased concerns around issues of security related to global supply chain activities (risk of a tampered shipping container) also have direct impact on the CFO.

Comprehensive solutions are necessary over the life of a transaction cycle that may integrate cash management, trade settlement, finance, logistics, supply nodes, procurement, demand projections, inventory, human resources, regulatory policy compliance and management of information across physical and financial supply chains. Creating one or more models that may work in synergy and integrate such real-world scenarios will be a challenge. The proven success of GARCH in finance and the potential to adapt GARCH for business operations (supply chains) may offer a synergistic multi-faceted tool for risk-adjusted supply chain management by acting as a bridge for some of the interdependent issues in business: finance, supply chain and management of risk.

How can you risk not to reap the rewards of the application potential of GARCH to improve your profitability?



*Generalized AutoRegressive Conditional Heteroskedasticity*

# Can GARCH Improve Accuracy ?

**Dr Shoumen Datta**

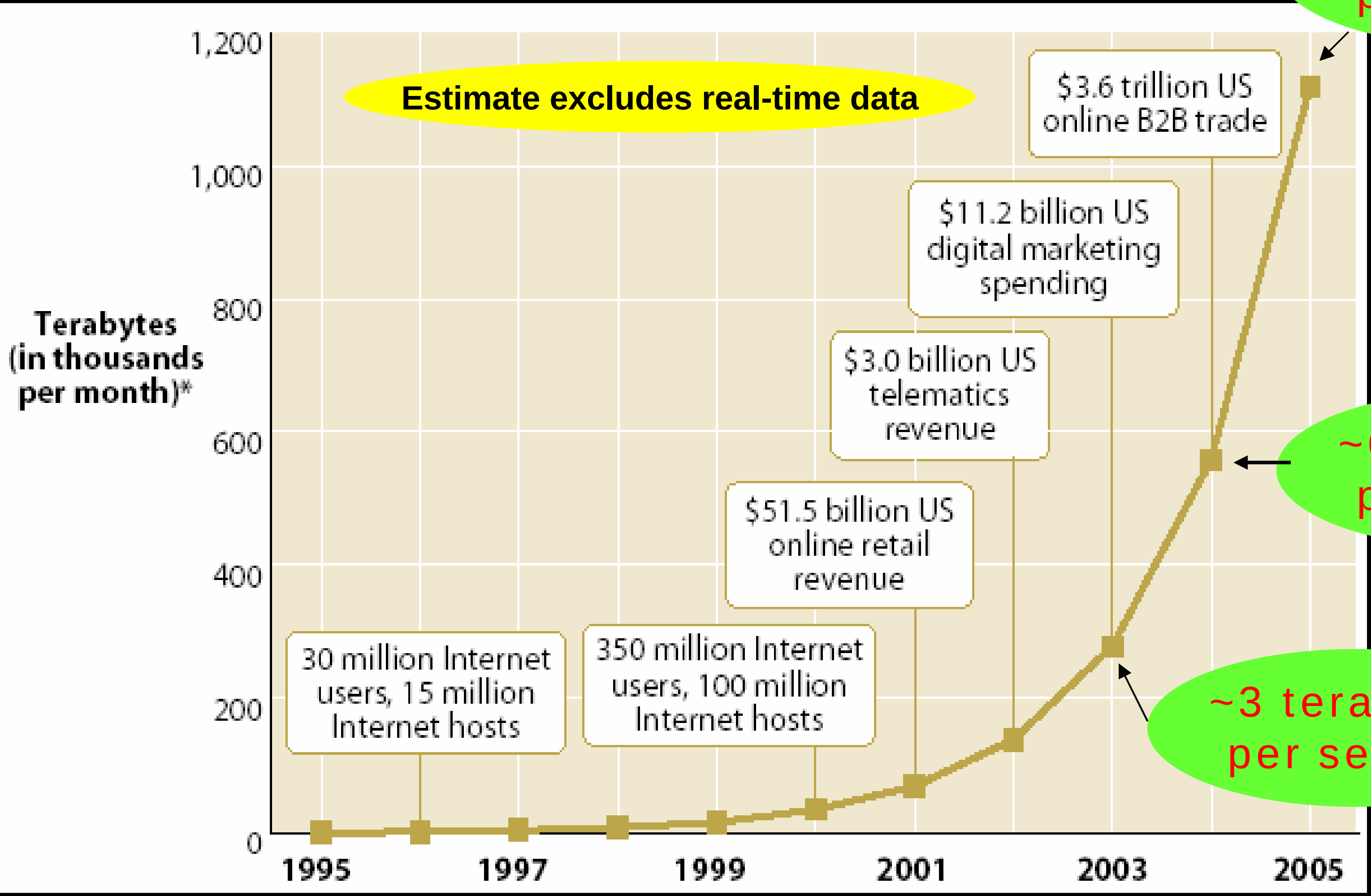
**Engineering Systems Division**
**School of Engineering**

**Massachusetts Institute of Technology**

# Why GARCH: Data vs Noise

~10 terabytes per second

2005

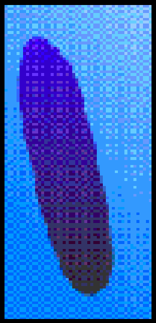



~6 terabytes per second

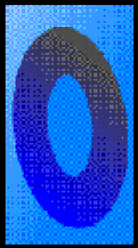
~3 terabytes per second

# What is this ?

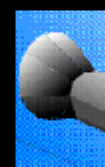
# What is this ?



# What is this ?

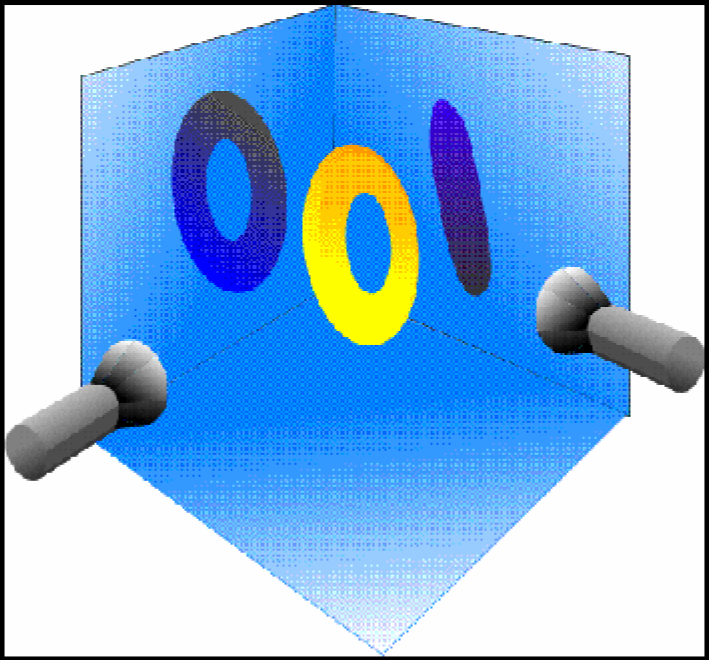


# What is this ?

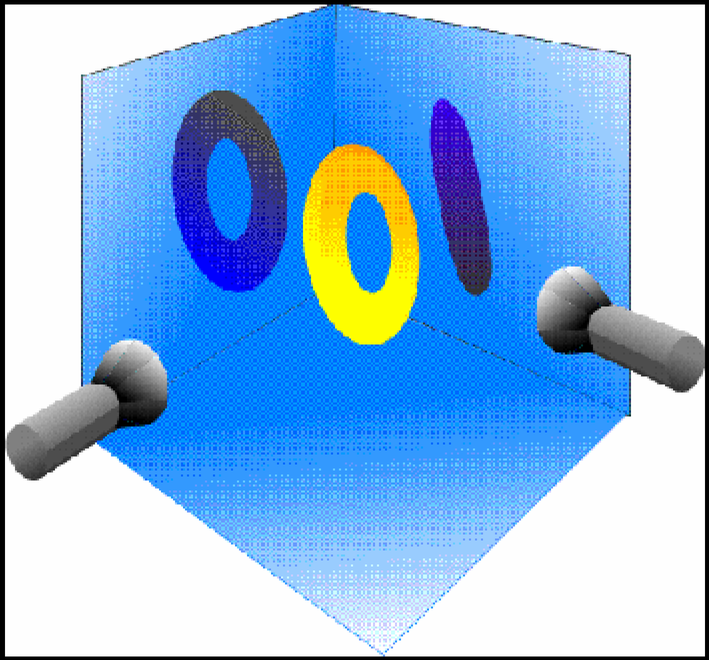


http://obelia.jde.aca.mmu.ac.uk/multivar/pca_graf.htm

# What is this ?



# Principal Component Analysis



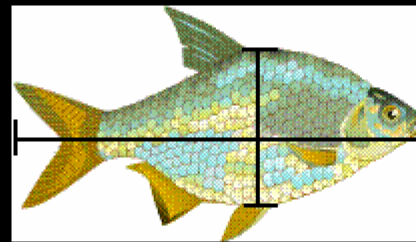
**The doughnut never changes shape even though the projections are quite different.**

# Principal Component Analysis

*Low Dimensional Structures from High Dimensional Samples*

The 3D golden eagle is projected onto 2D but can still be recognized as a eagle, because the image retains a significant amount of information.

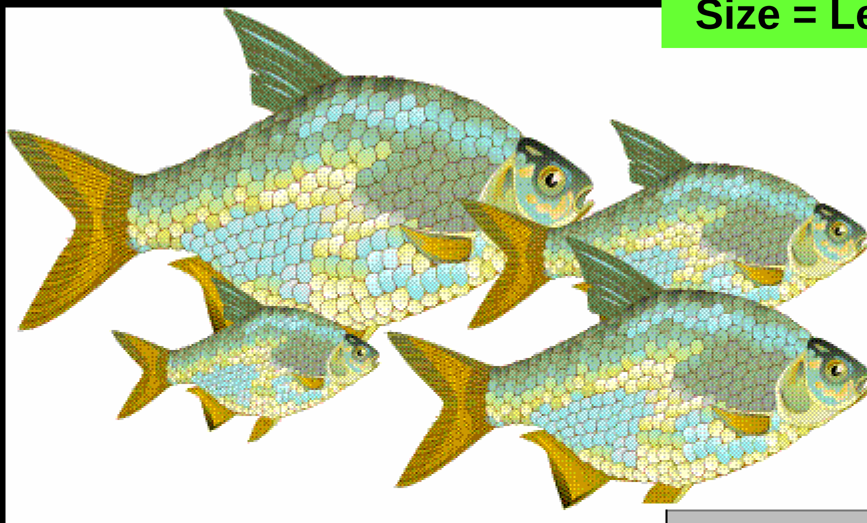
# Data vs Information: Systems introduce Artifacts and Inaccuracies

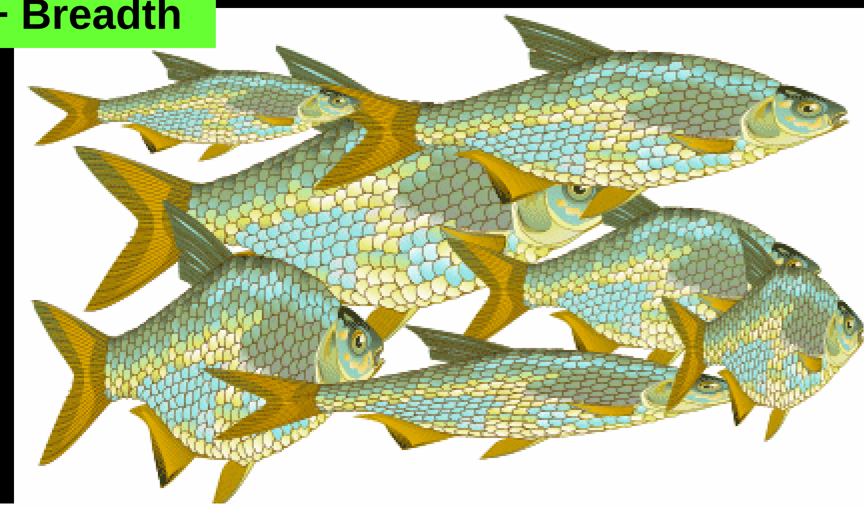


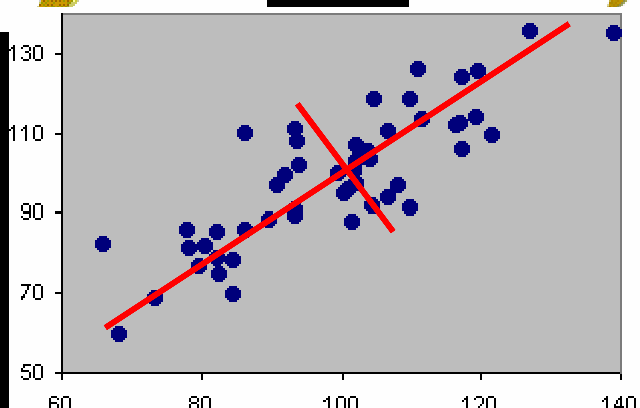
**Retain 87.5% of the information**

**Retain 62.5% of the information**

**Size = Length + Breadth**



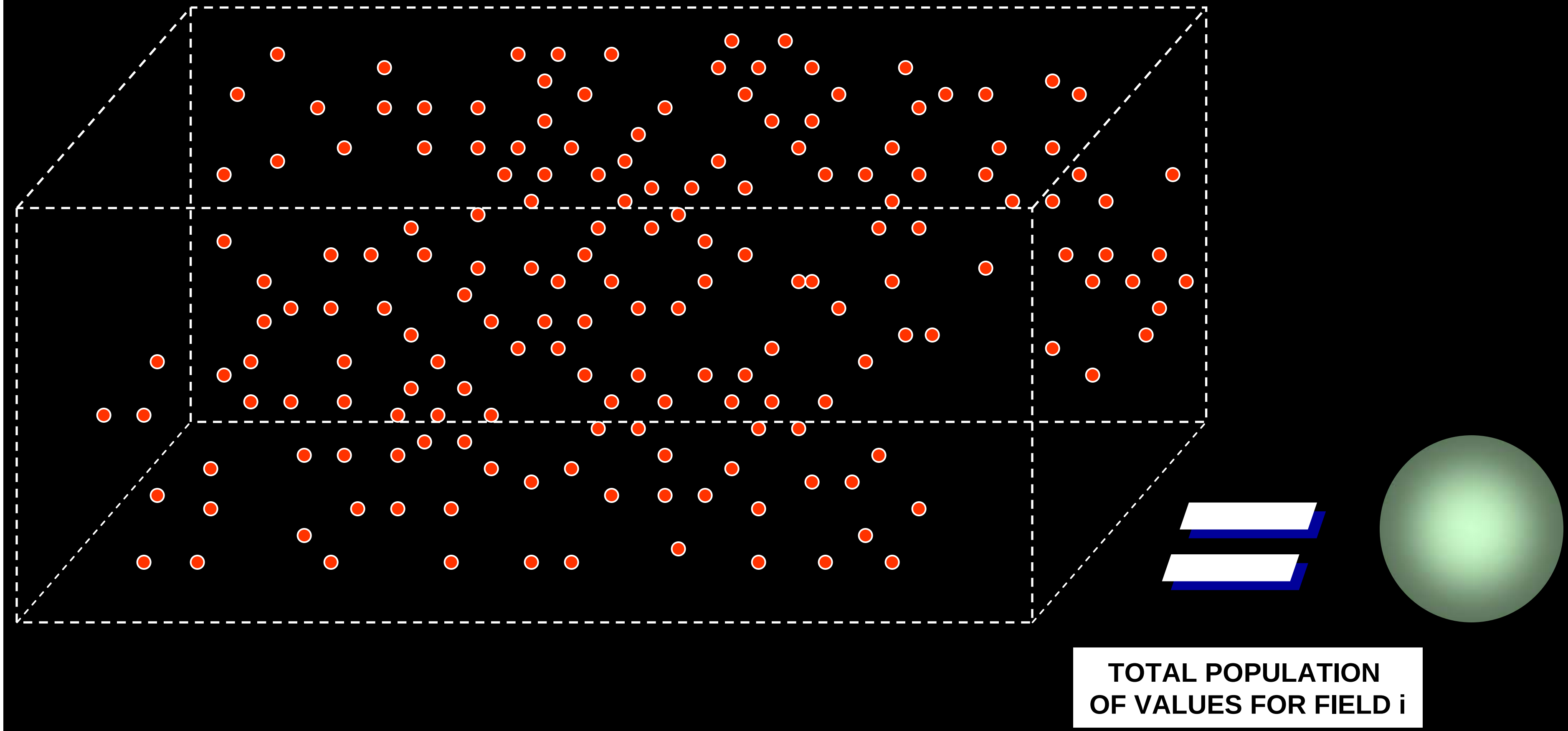


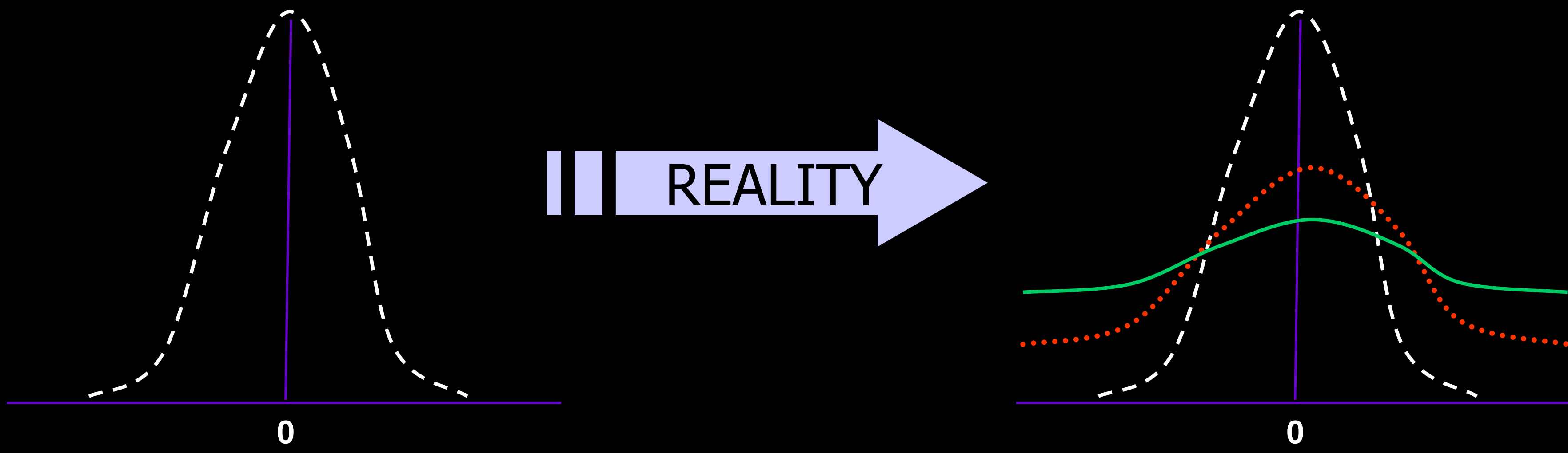



Dr Shoumen Datta, MIT <shoumen@mit.edu>

http://obelia.jde.aca.mmu.ac.uk/multivar/pca_graf.htm

# Reductionist Approach



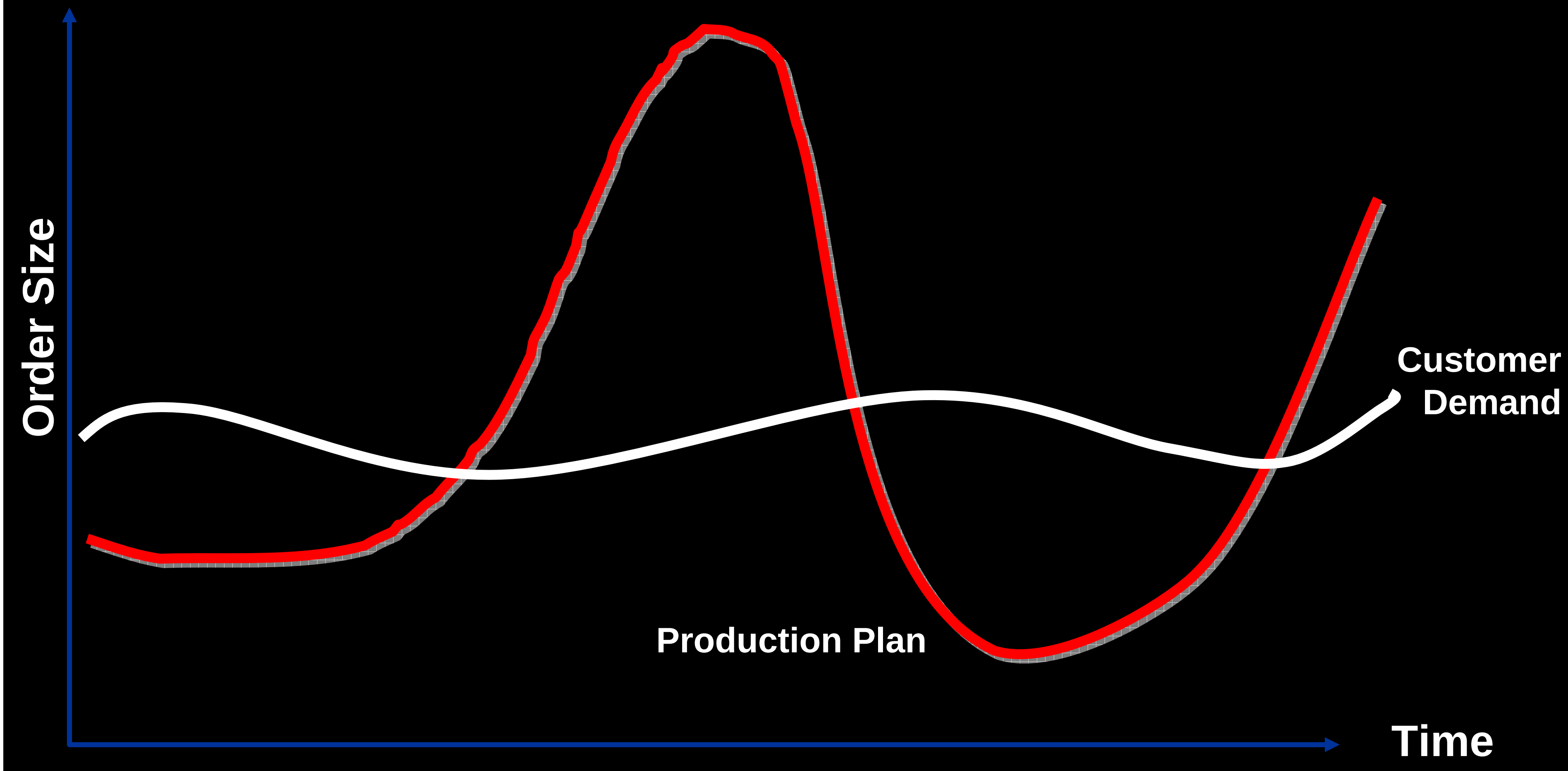

0
REALITY
0

# Information Asymmetry between Demand and Supply

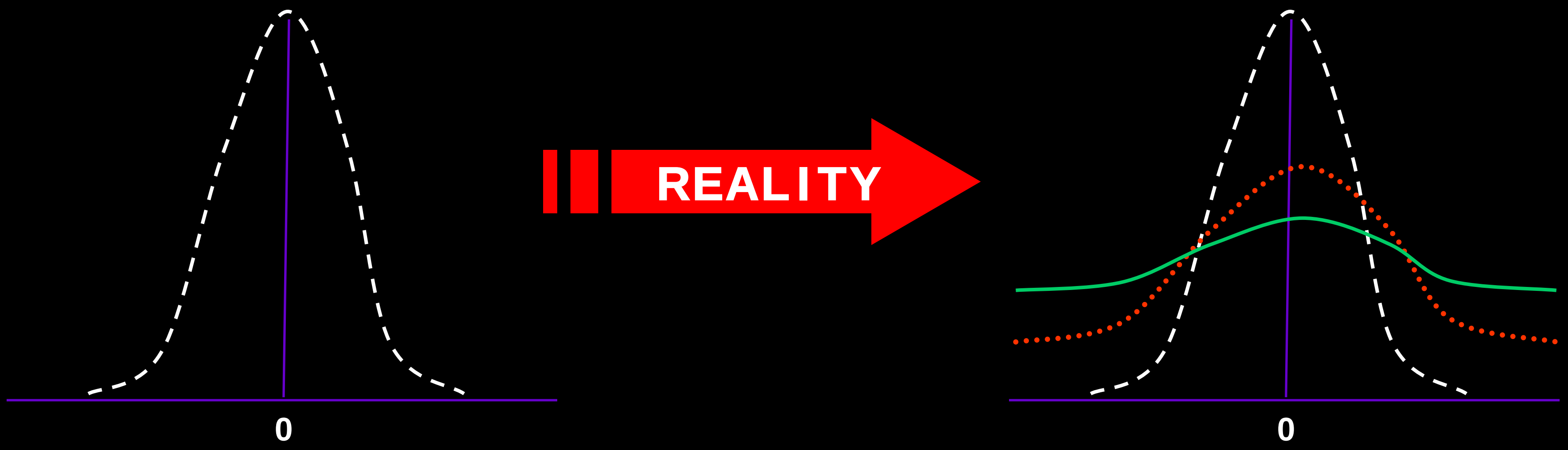
## The Bullwhip Effect


Source: Tom McGuffog, Electronic Commerce and Value Chain Management

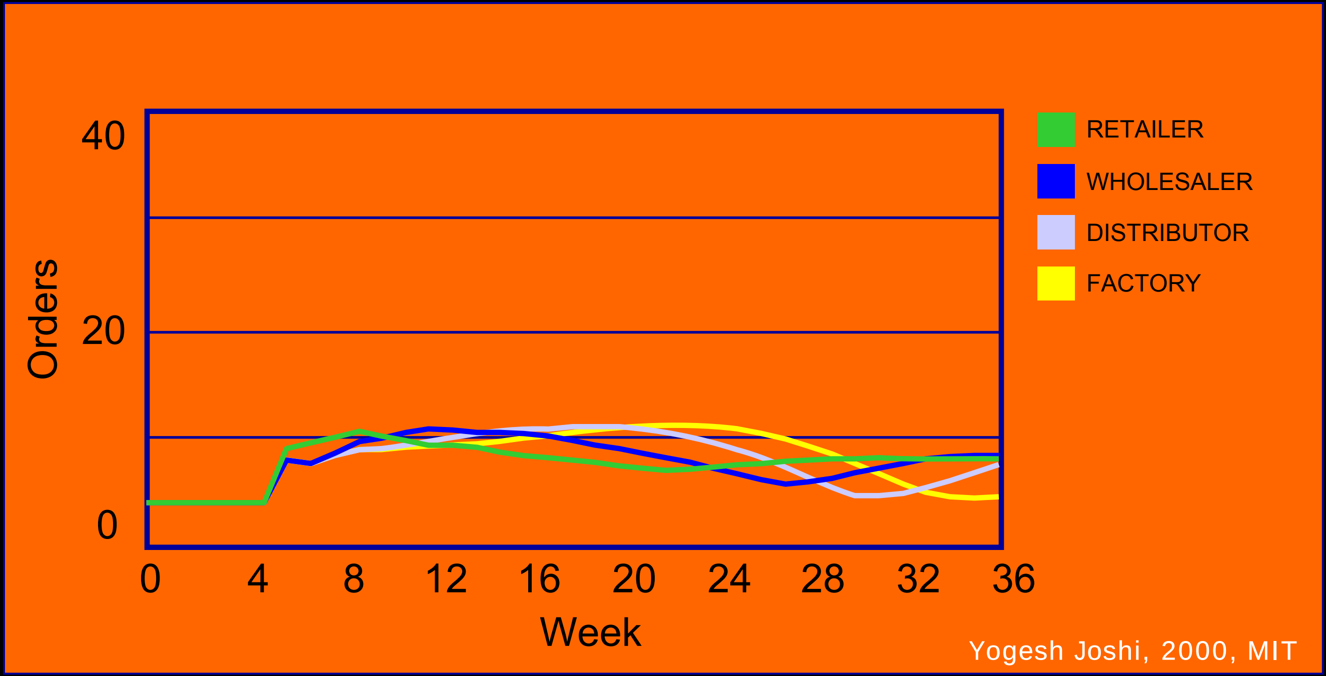
# Uncertainty



# Cannot Ignore Errors



# Can real-time data reduce transaction costs?
# Can macro-economics predict business cycles?
# Can Game Theory strategies reduce volatility?




## Transaction Cost Economics

Nobel Prize in Economics 1991 >> Ronald Coase

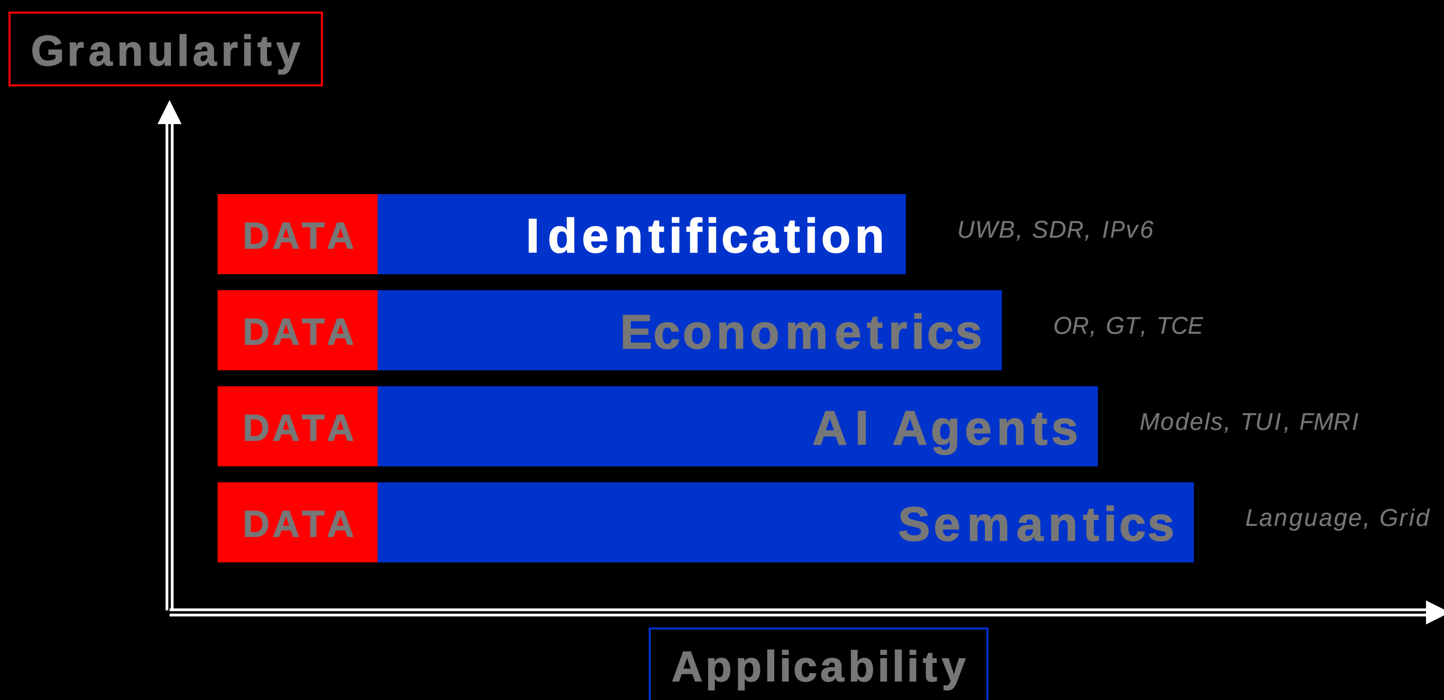
## Macroeconomics in Business Cycles

Nobel Prize in Economics 2004 >> Finn Kydland & Edward Prescott

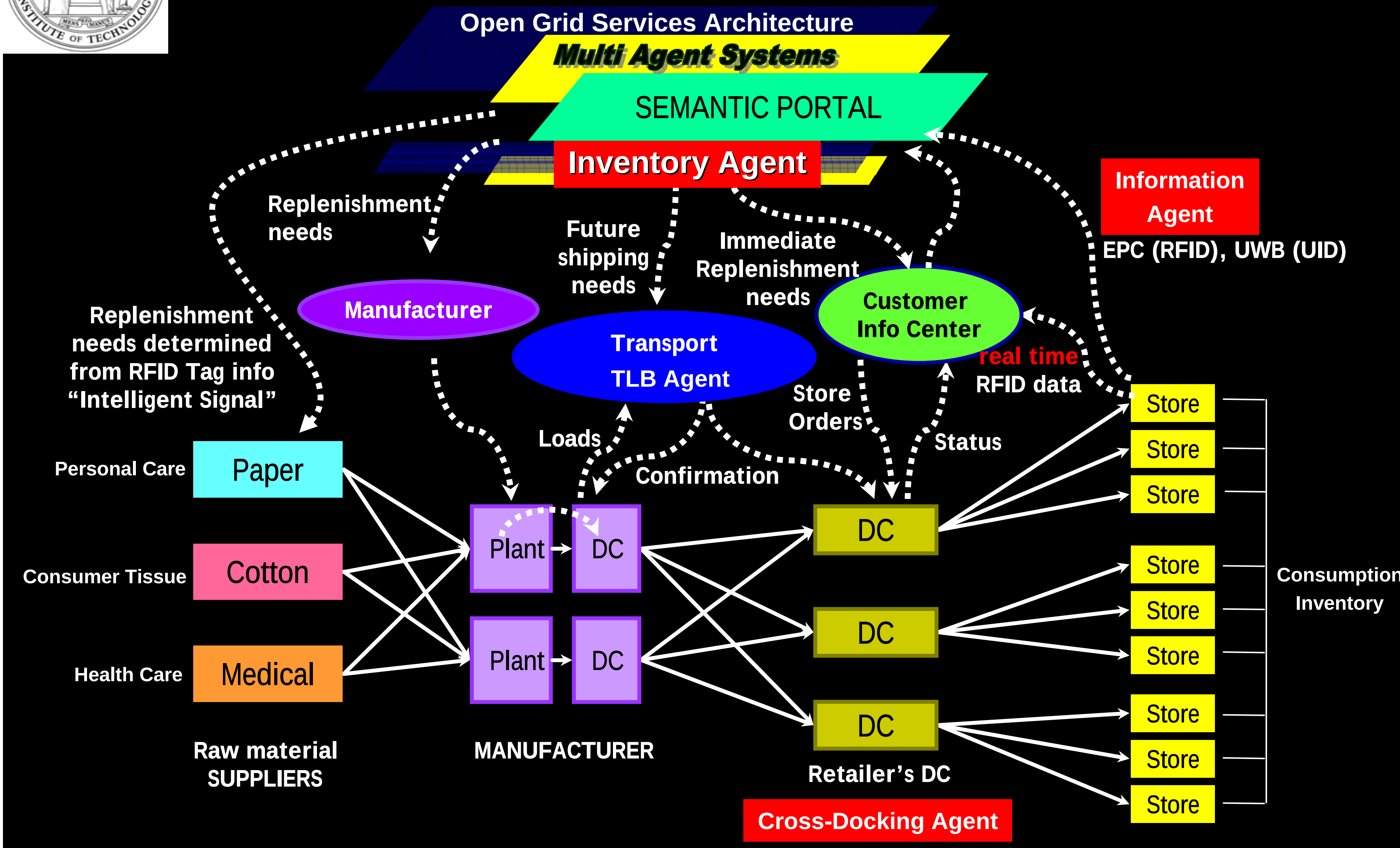
## Game Theory Strategies in Cooperation

Nobel Prize in Economics 2005 >> Robert Aumann & Thomas Schelling

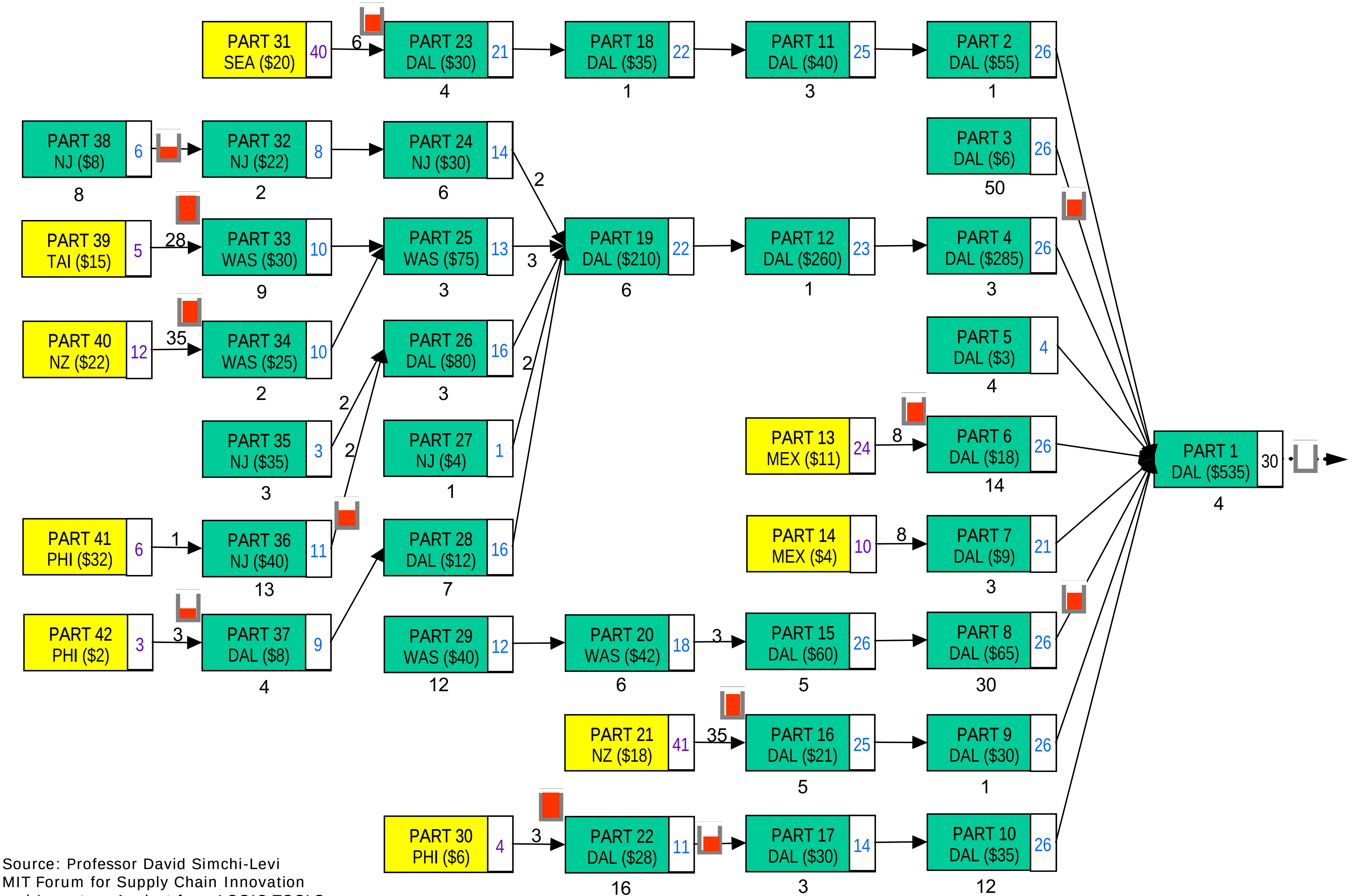
# Tools


Granularity
DATA Identification
UWB, SDR, IPv6
DATA Econometrics
OR, GT, TCE
DATA AI Agents
Models, TUI, FMRI
DATA Semantics
Language, Grid
Applicability

# Network


Open Grid Services Architecture
Multi Agent Systems
SEMANTIC PORTAL
Inventory Agent
Information Agent
EPC (RFID), UWB (UID)
Replenishment needs
Future shipping needs
Immediate Replenishment needs
Manufacturer
Transport TLB Agent
Customer Info Center
real time RFID data
Replenishment needs determined from RFID Tag info "Intelligent Signal"
Store Orders
Status
Loads
Confirmation
Personal Care
Paper
Consumer Tissue
Cotton
Health Care
Medical
Plant
DC
Plant
DC
DC
DC
DC
Store
Store
Store
Store
Store
Store
Store
Store
Store
Consumption Inventory
Raw material SUPPLIERS
MANUFACTURER
Retailer's DC
Cross-Docking Agent

# Adaptability in Supply Chain Inventory Optimization: Local and Global Multi-parametric Inter-dependencies




Source: Professor David Simchi-Levi
MIT Forum for Supply Chain Innovation
and Inventory Analyst from LOGIC TOOLS

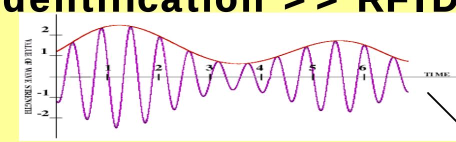
# Auto ID: Data Acquisition

**Manual Identification >> Barcode**

Then

Repository

**Automatic Identification >> RFID**



*Paradigm Shift ?*

**01.0203D2A.916E8B.0719BAE03C**

Now

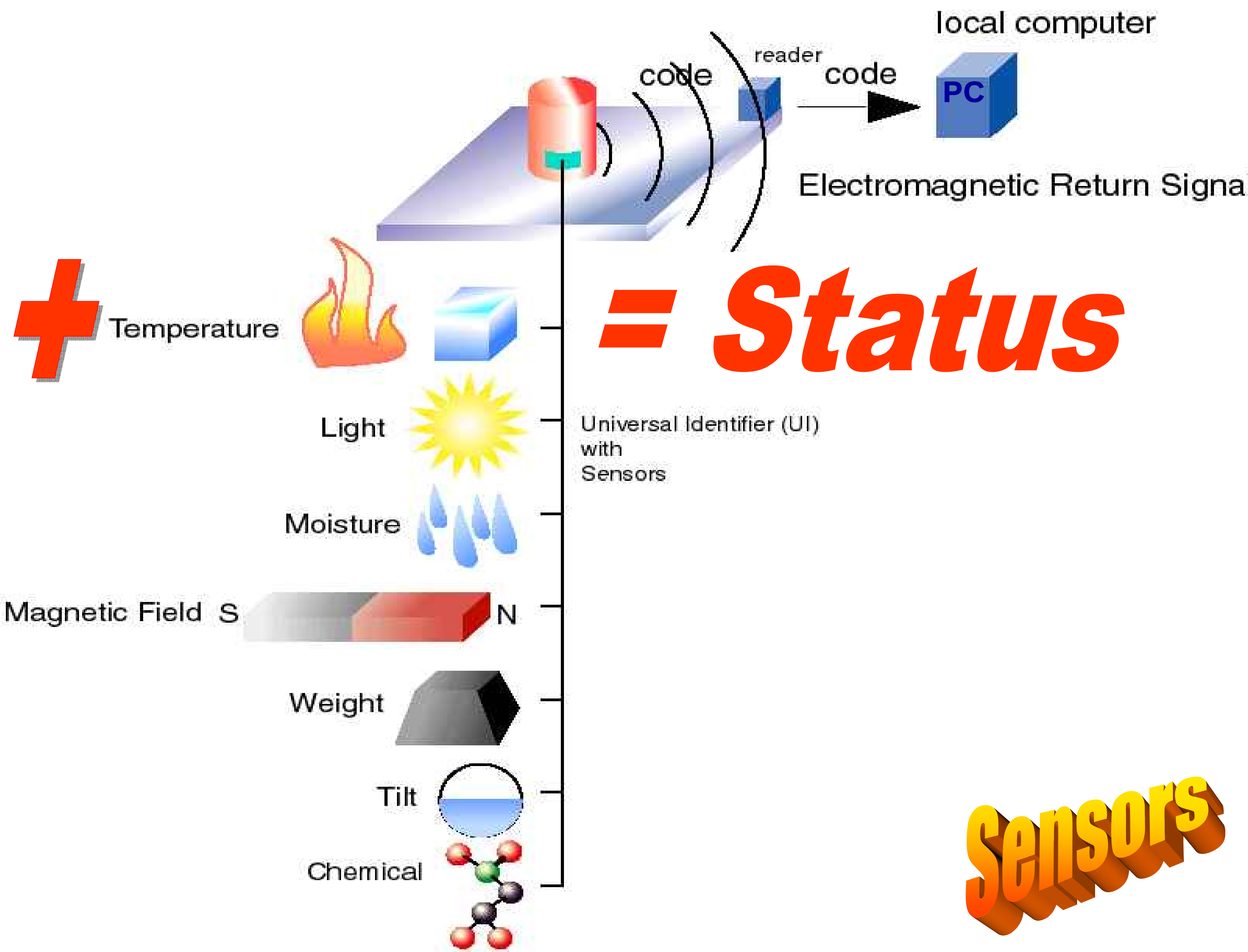
Repository

***Dynamic Systems Adaptability***

# DCA to BOS

**151.193.204.72**

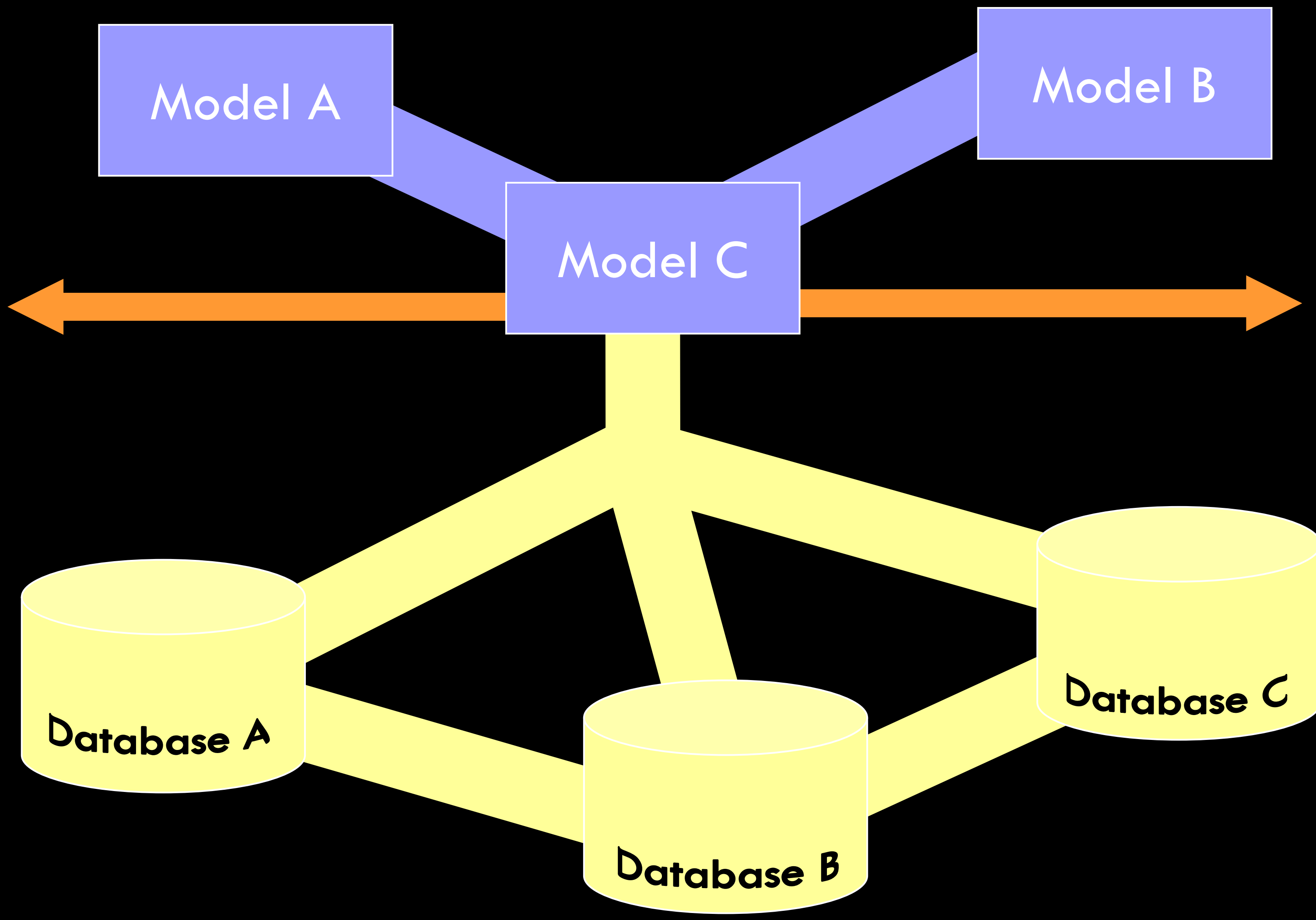
# Data is not Information

ID +  = Status

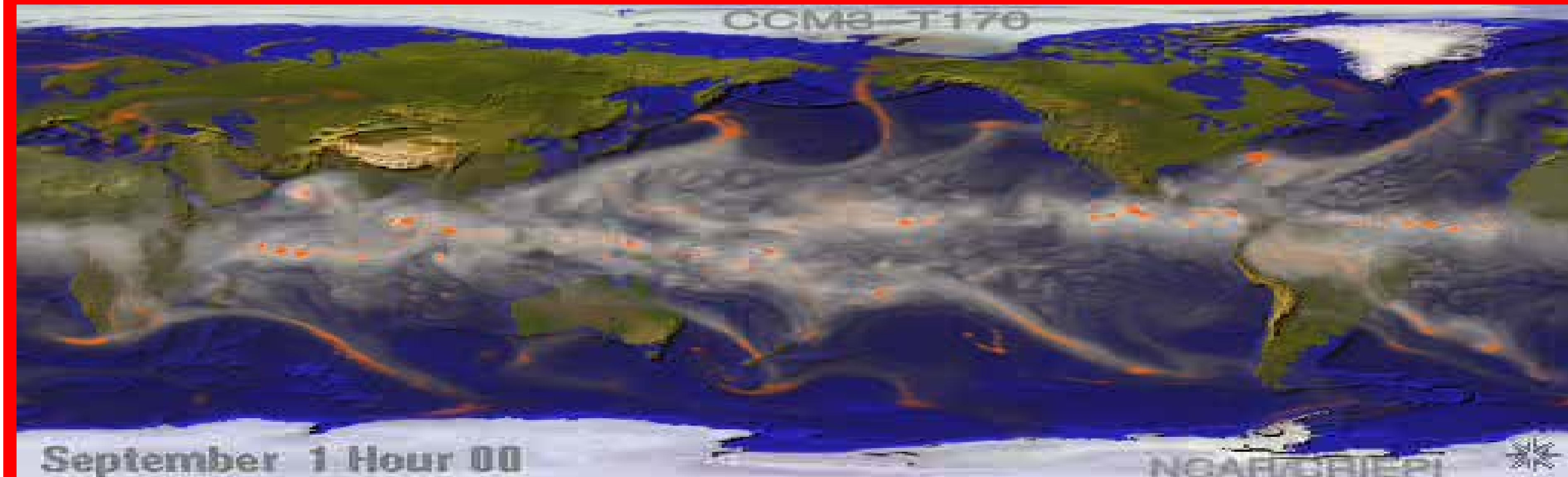


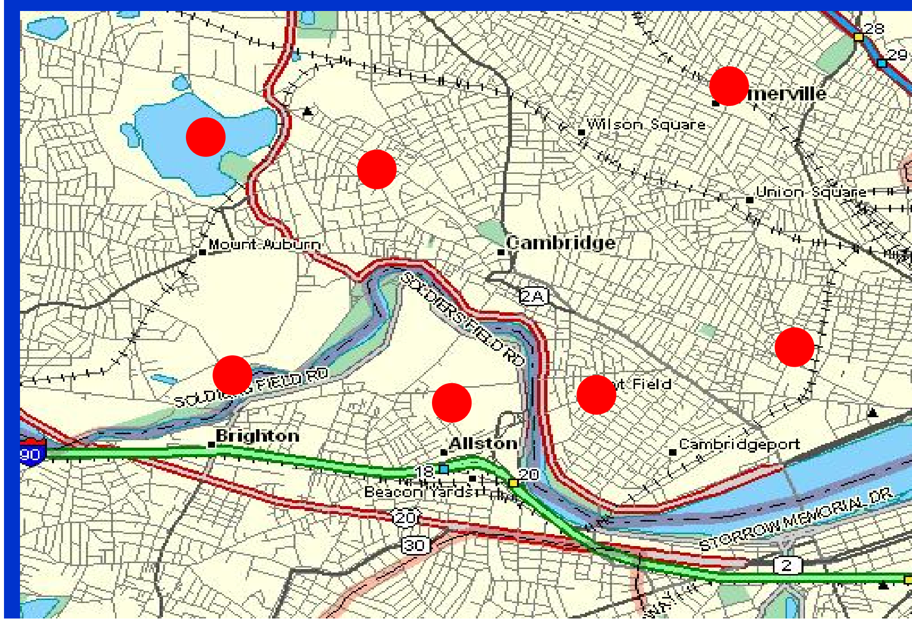
Sensors

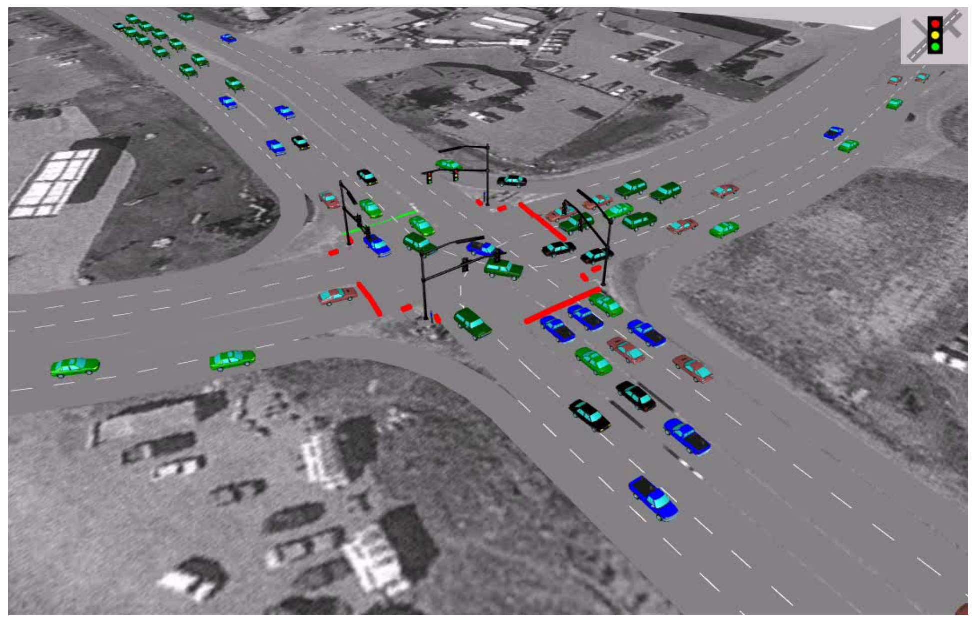
# Data and Information Asymmetry: Isolated Systems



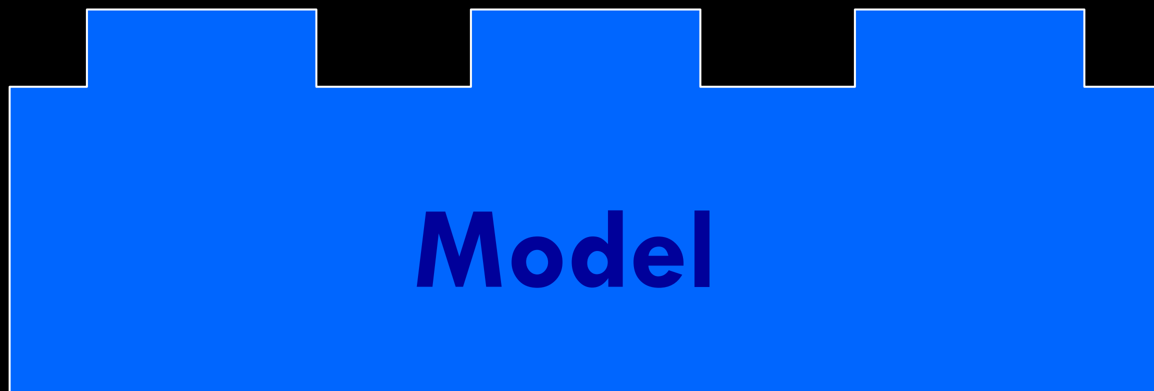
# International Logistics : Isolated and un-Integrated



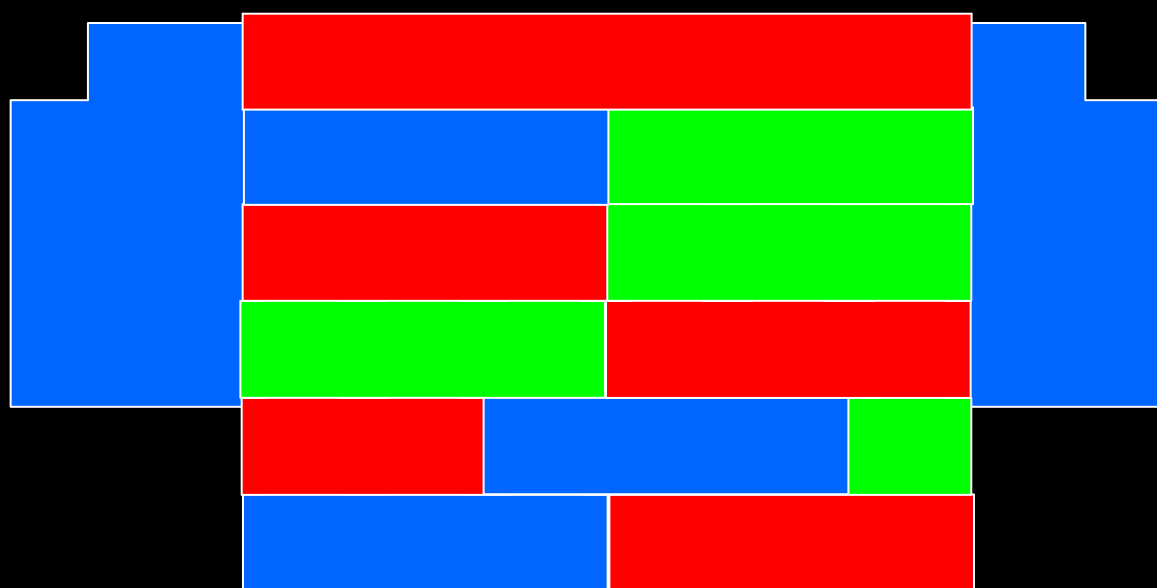



# Vision: Functional Data Integration & Interoperability



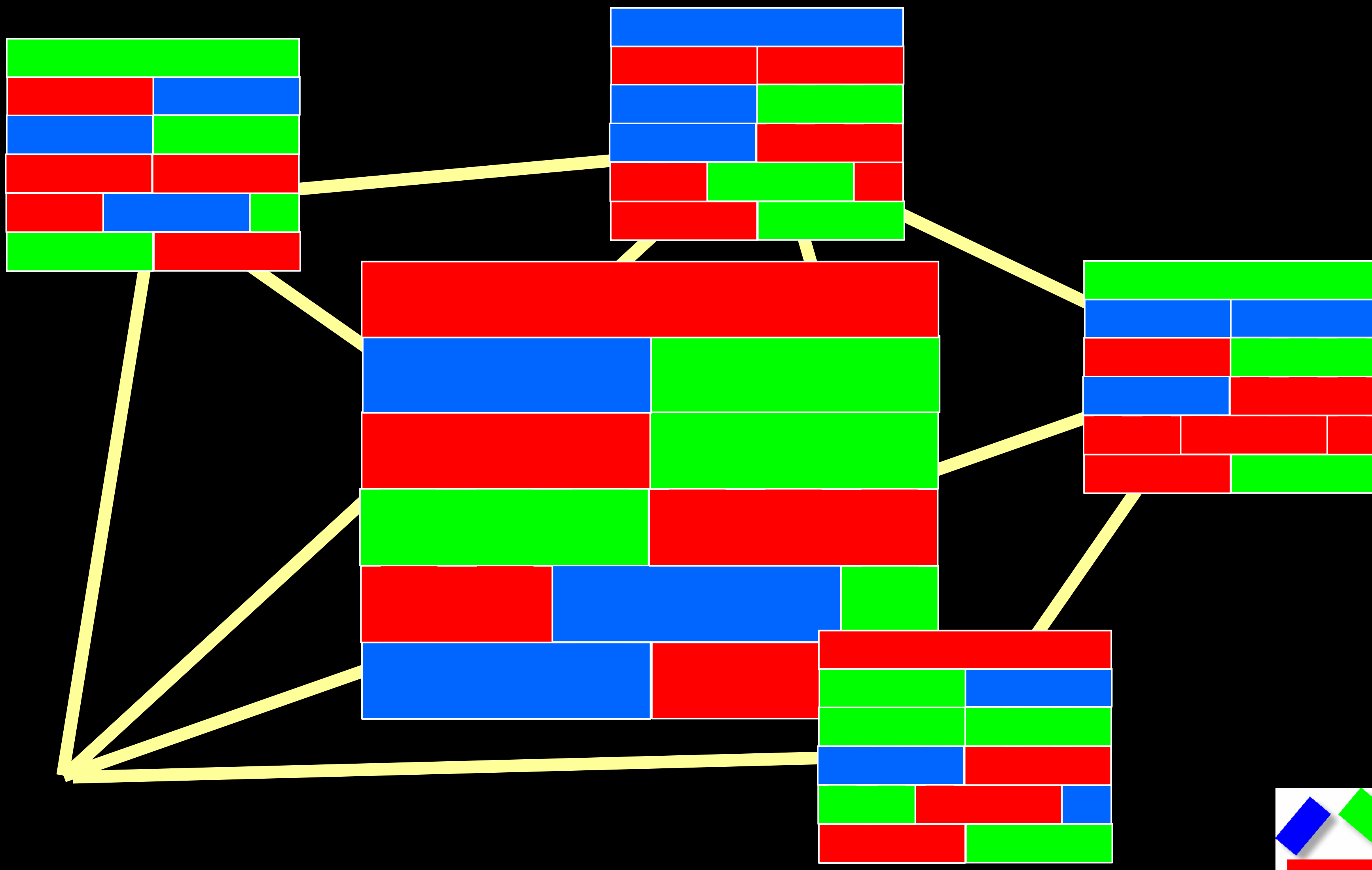

Dr Shoumen Datta, MIT <shoumen@mit.edu>

# Vision: Functional Data Integration & Interoperability



Dr Shoumen Datta, MIT <shoumen@mit.edu>

# Network Data Integration & Interoperability



Dr Shoumen Datta, MIT <shoumen@mit.edu>

# Modern Game Theory (1950)

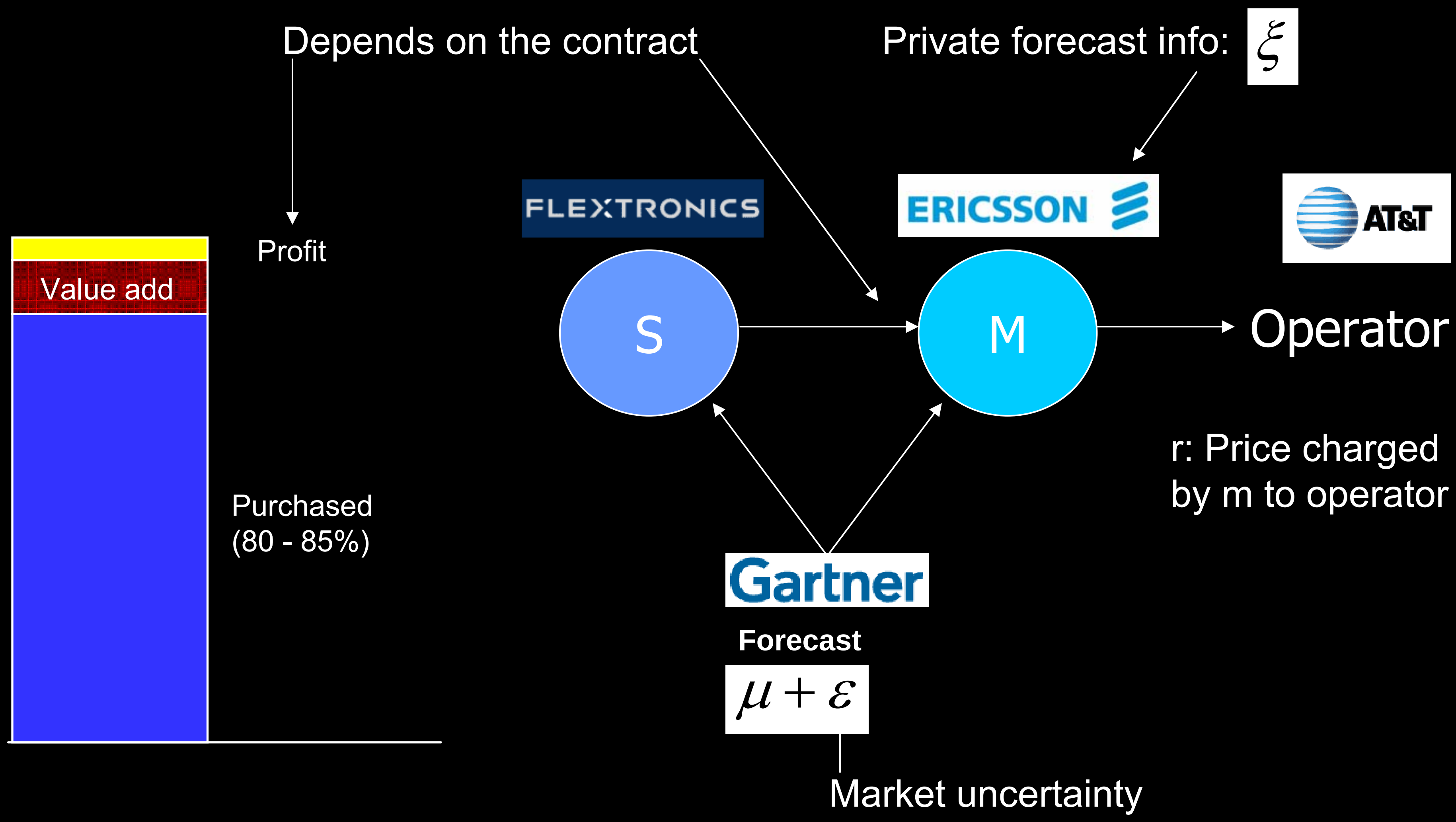
**Jon Von Neumann**

**Oskar Morgenstern**

- Classical Game Theory (circa 1800, Bertrand and Cournot)
- Commercial use in pricing optimization  (Stackelberg Equation)
- 1994 Nobel Prize for Economics (John Forbes Nash)
- 2005 Nobel Prize for Economics (Aumann and Schelling)

Source: Özalp Özer and Wei Wei

# Telecom Supply Chain Case



# Objective : Reduce Information Asymmetry

- Achieve credible information sharing
- Eliminating sources of inefficiency

# Capacity Planning Problem

- Short product lifecycle (clockspeed)
- Demand is uncertain prior to capacity decision

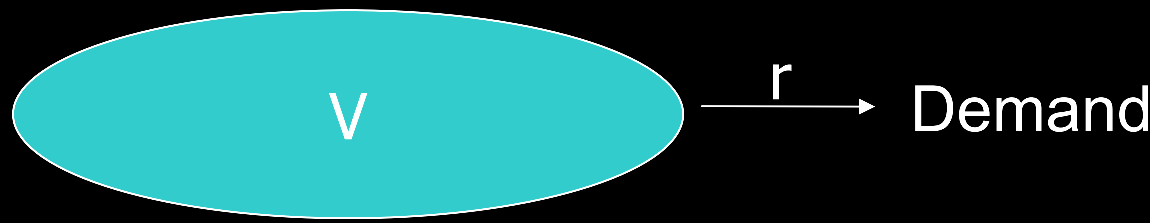
$$D = \mu + \xi + \varepsilon$$

Market uncertainty $\varepsilon \sim \mathrm{G}(\cdot)$

Manufacturer's private forecast update

Supplier's prior belief $\xi \in [\underline{\xi}, \overline{\xi}]$ $\xi \sim F(\cdot),$

# Centralized System




Optimization Problem

$$\max_{K \geq 0} E\left[ (r-c)\min\left( K, \mu + \xi + \underbrace{\varepsilon}_{random} \right) \right] - c_k K$$

Optimal Capacity

$$K^{cs} = G^{-1}\left( \frac{r - c - c_k}{r - c} \right) + \mu + \xi$$

## Decentralized System Wholesale Contract with **Symmetric Information**

- **Manufacturer's profit:**

$$(r-w)E\left[\min\left(K, \mu+\xi+\underbrace{\varepsilon}_{random}\right)\right]$$

- **Supplier's optimization problem:**

$$\max_{K\geq 0}(w-c)E\left[\min\left(K, \mu+\xi+\underbrace{\varepsilon}_{random}\right)\right]-c_k K$$

- **Optimal capacity: $K^{ws} < K^{cs}$**

$$K^{ws}=\mu+\xi+G^{-1}\left(\frac{w-c-c_k}{w-c}\right)$$

## Decentralized System Wholesale Contract with **Asymmetric Information**

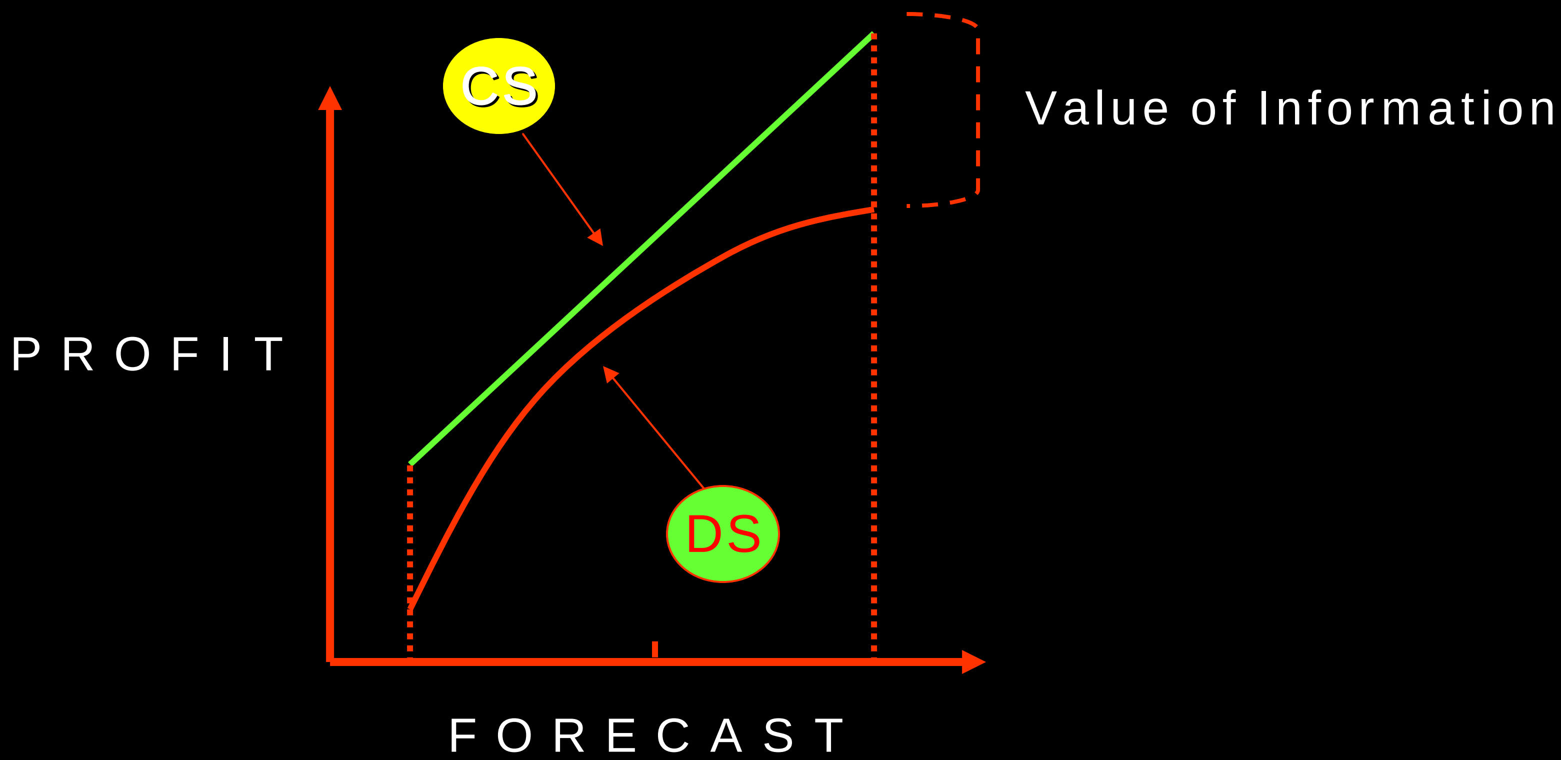
- **Supplier's Optimization Problem:**

$$\max_{K\geq 0} E\left[(w-c)\min\left(K, \mu + \underbrace{\xi + \varepsilon}_{random}\right)\right] - c_k K$$

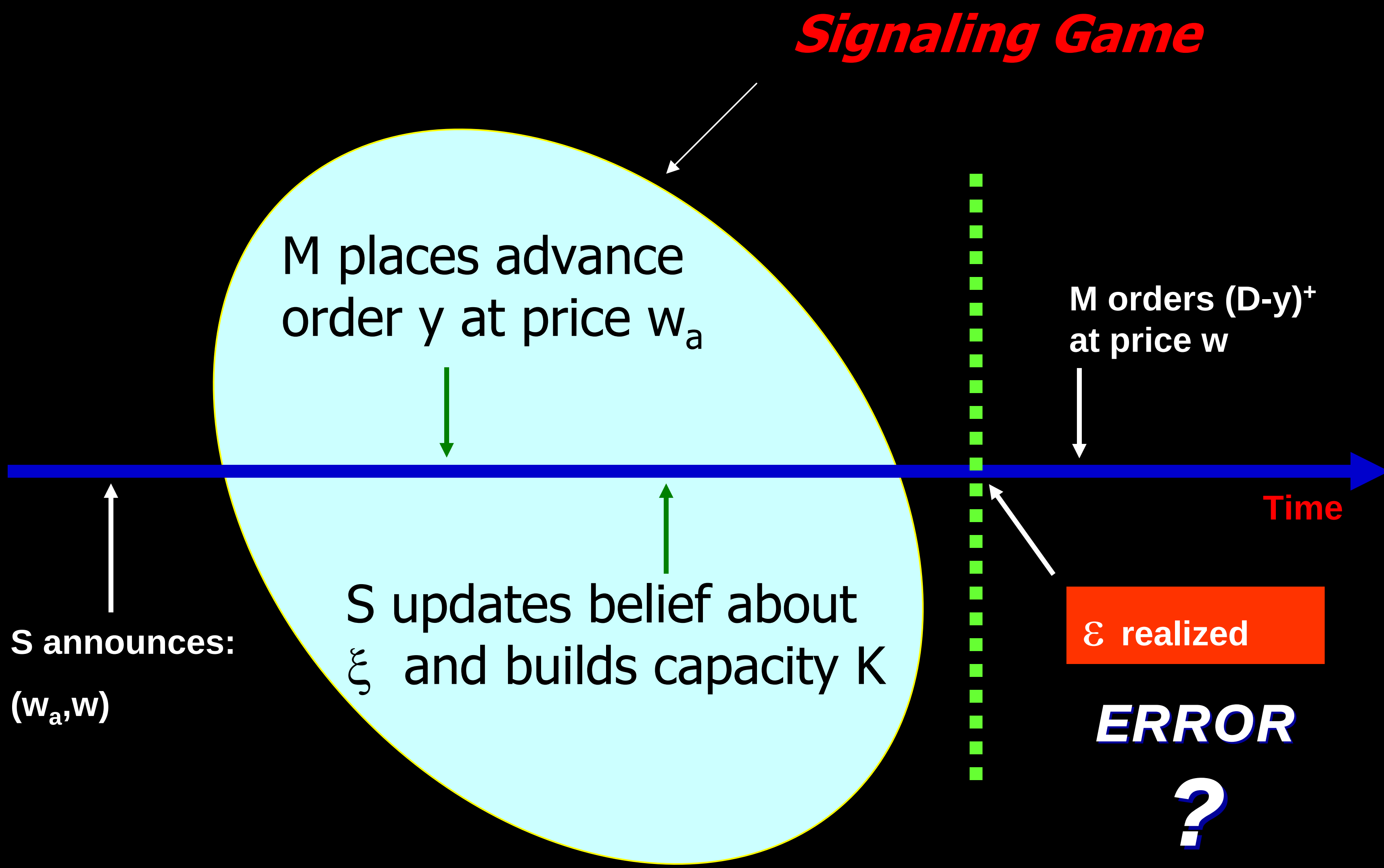
- **Optimal Capacity (in reality):**

$$K^{ds} = (F \circ G)^{-1}\left(\frac{w-c-c_k}{w-c}\right) + \mu$$

# Value of Information



# Advance Purchase Contract: Sequence of Events



# Forecasting Models

***Uncertainty – error terms are assumed to be a distribution***

$$D = \mu + \xi + \varepsilon$$

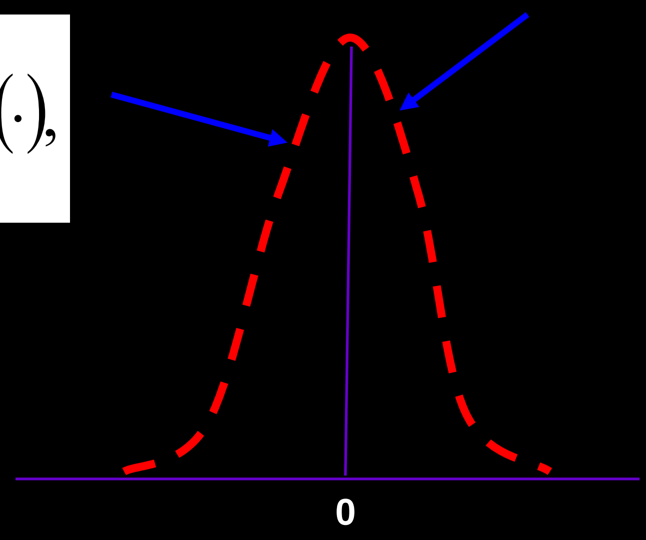
Market uncertainty $\varepsilon \sim G(\cdot)$

Manufacturer's private forecast update

Supplier's prior belief

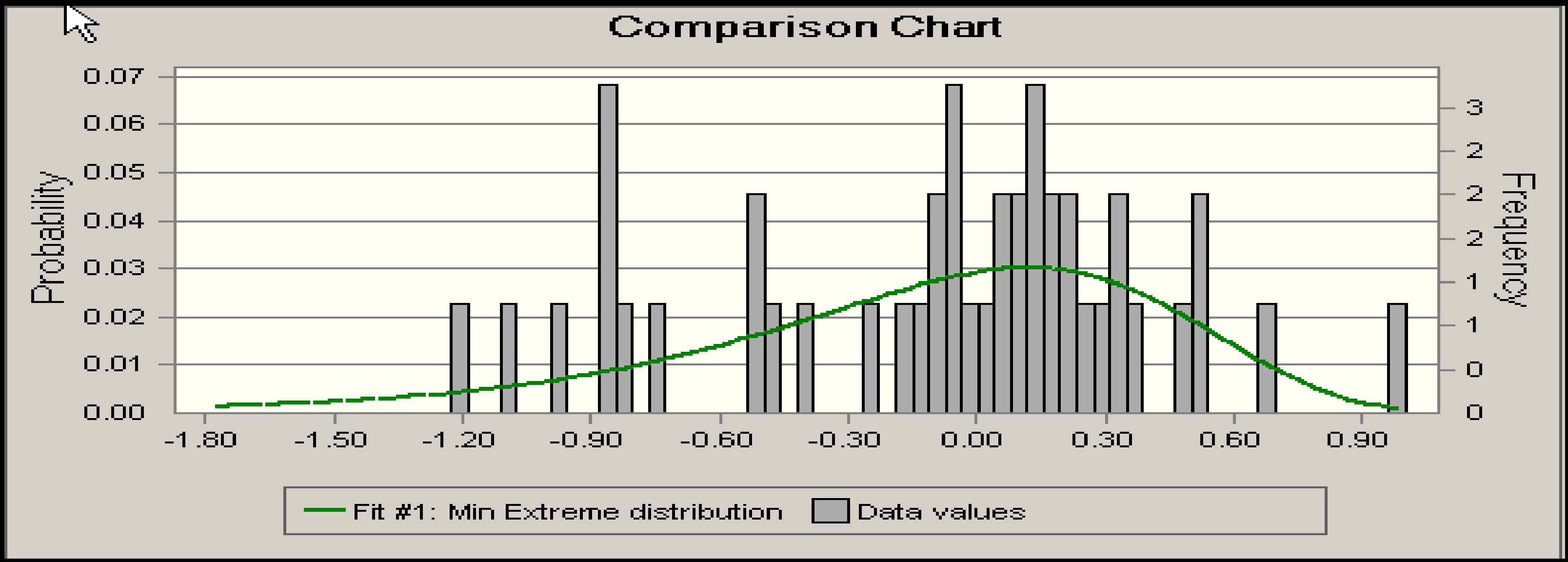
$$\xi \in [\underline{\xi}, \overline{\xi}]$$

$$\xi \sim F(\cdot),$$



# Analysis: The Crystal Ball ??

- No consistent best fit distribution of forecast error
- Normal distribution maybe a good fit

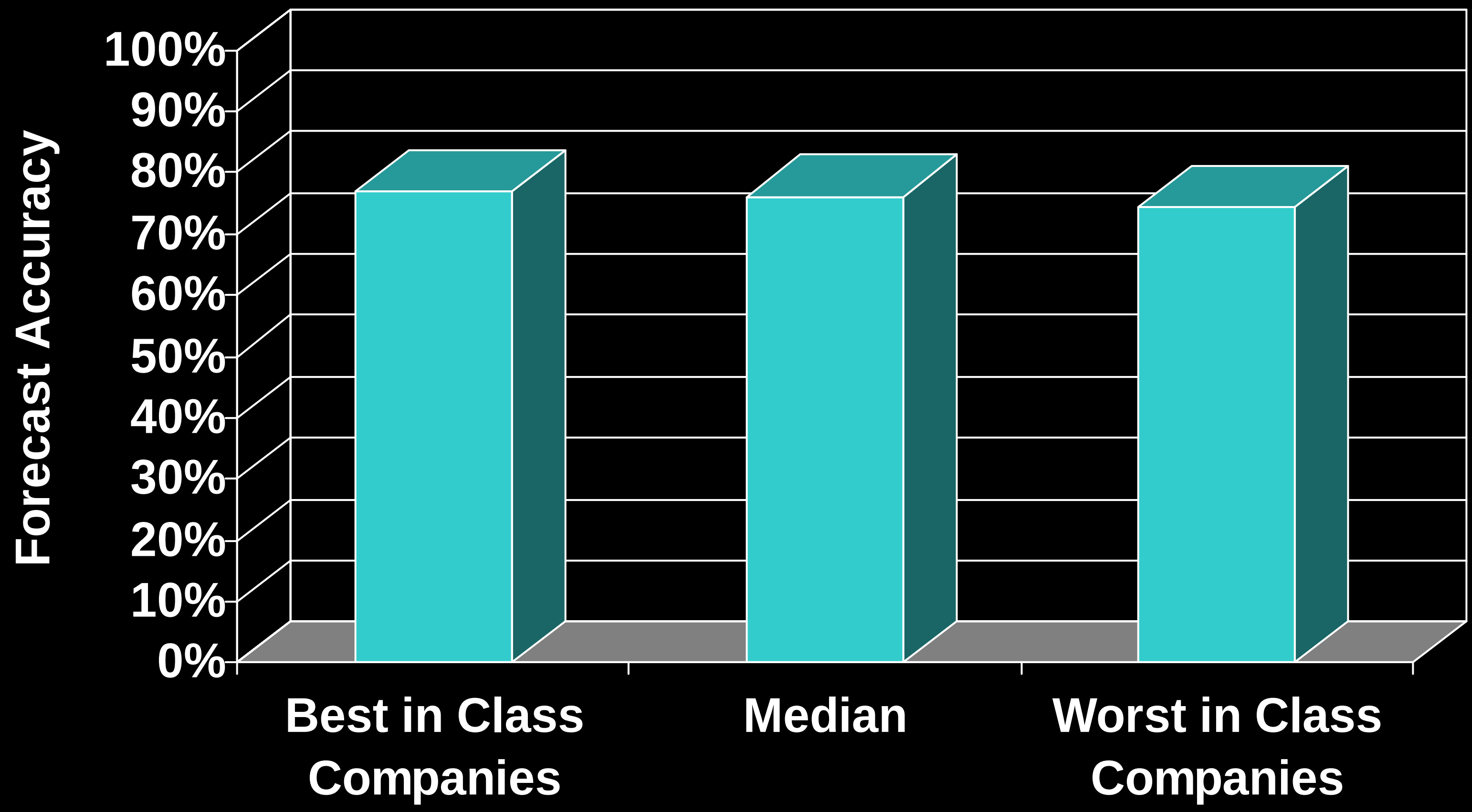



**Task**: Demand distribution analysis on SKU plus warehouse combinations

**Results**: Some distributions Gamma, some Exponential, others Normal

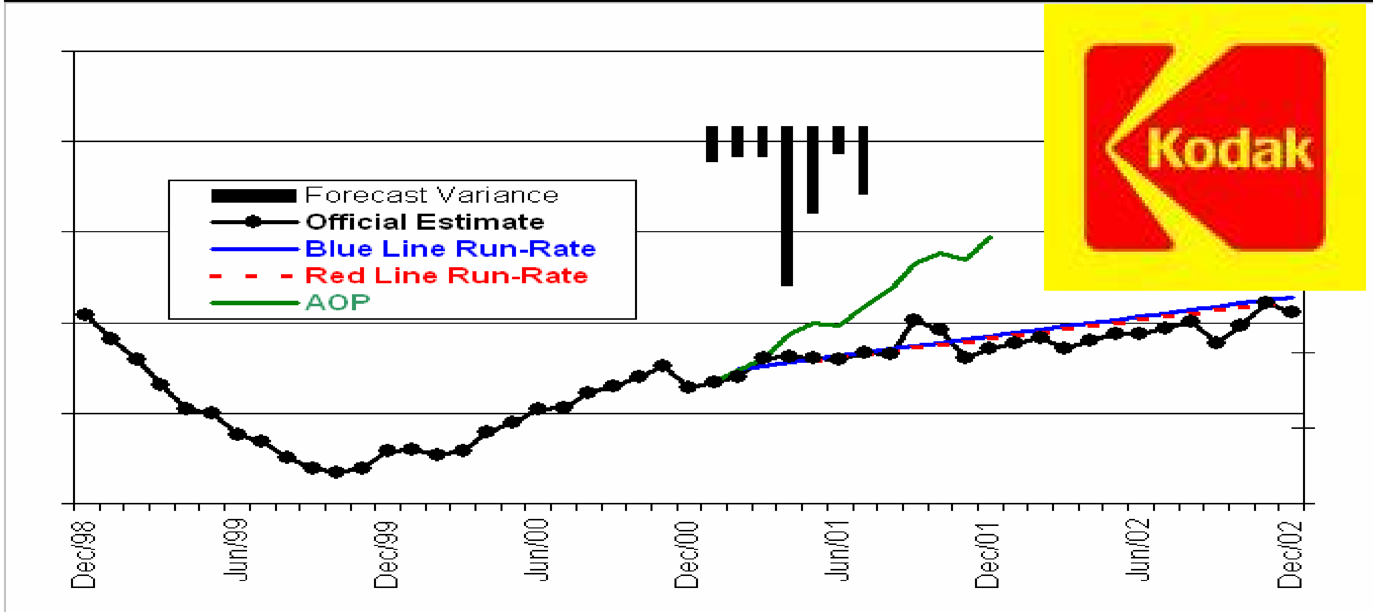
**Conclude**: Assumption of normal distribution generates best results

# Is Ability to Forecast Accurately a Non-issue?




Source: PRTM (06/04/2004)

# Sales Forecasting



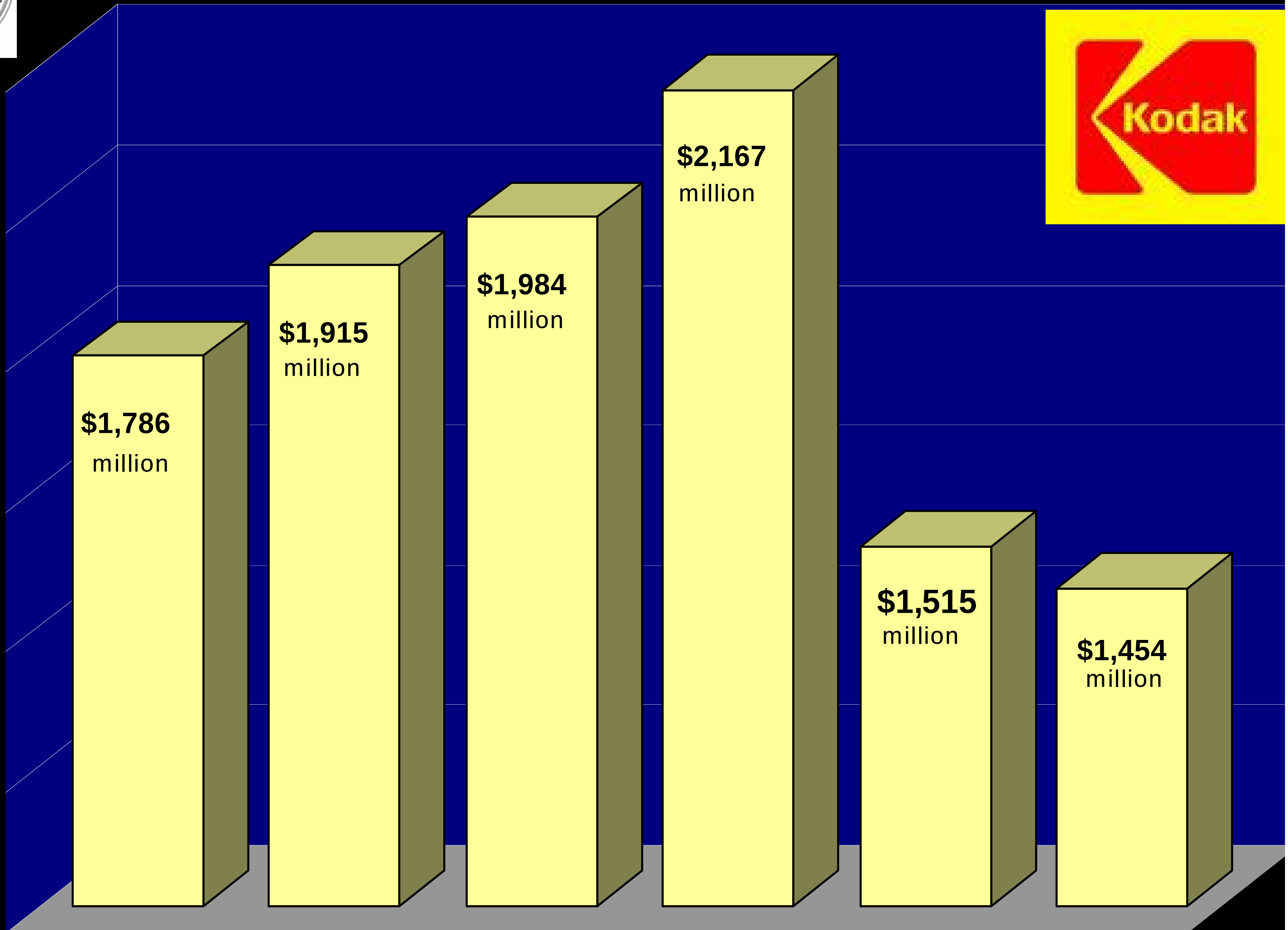

**Justificiation for Official Estimate being Different than the Run-Rate (Blue 50/50):**

>

| | | |
|---|---|---|
| Prior Blue Line | | 5482 |
| Prior Red Line | | 5247 |

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Qtr1 | Qtr2 | Qtr3 | Qtr4 | Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998 | 344 | 408 | 488 | 397 | 396 | 555 | 392 | 398 | 338 | 591 | 705 | 526 | 1240 | 1348 | 1128 | 1822 | 5538 |
| 1999 | 215 | 296 | 346 | 266 | 374 | 430 | 359 | 307 | 281 | 563 | 732 | 617 | 857 | 1070 | 947 | 1913 | 4787 |
| 2000 | 227 | 265 | 367 | 374 | 424 | 506 | 362 | 390 | 321 | 616 | 791 | 499 | 859 | 1304 | 1073 | 1906 | 5142 |
| *2000 vs 1999* | *6%* | *-10%* | *6%* | *41%* | *13%* | *18%* | *1%* | *27%* | *14%* | *9%* | *8%* | *-19%* | *0%* | *22%* | *13%* | *0%* | *7%* |
| *2001* | 257 | 293 | 472 | 377 | 417 | 498 | 405 | | | | | | 1021 | 1292 | 405 | 0 | 2718 |
| 2001 Forecast Variance | (47) | (39) | (40) | (210) | (115) | (35) | (89) | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| 2001 | | | | | | | | | | | | | | | | | |
| Blue Line (50/50) | 257 | 293 | 472 | 377 | 417 | 498 | 405 | 408 | 339 | 634 | 809 | 517 | 1021 | 1292 | 1152 | 1960 | 5425 |
| Red Line (Warning) | 257 | 293 | 472 | 377 | 417 | 498 | 405 | 399 | 338 | 633 | 808 | 516 | 1021 | 1292 | 1142 | 1958 | 5413 |
| Black Line (Official Estimate) | **257** | **294** | **472** | **377** | **417** | **498** | **405** | **376** | **511** | **564** | **636** | **546** | 1022 | 1292 | 1292 | 1746 | 5353 |
| *2001 vs 2000* | *13%* | *11%* | *29%* | *1%* | *-2%* | *-1%* | *12%* | *-4%* | *59%* | *-8%* | *-20%* | *9%* | *19%* | *-1%* | *20%* | *-8%* | *4%* |
| Upside Demand Potential | **257** | **290** | **477** | **392** | **459** | **600** | **435** | **449** | **592** | **665** | **763** | **656** | 1025 | 1450 | 1476 | 2085 | 6035 |
| Downside Demand Potential | **257** | **293** | **477** | **349** | **423** | **534** | **398** | **339** | **346** | **511** | **600** | **503** | 1027 | 1306 | 1082 | 1614 | 5030 |
| AOP | **282** | **306** | **432** | **510** | **481** | **496** | **468** | **480** | **464** | **679** | **752** | **625** | 1020 | 1487 | 1412 | 2056 | 5974 |
| Prior Official Estimate | **258** | **297** | **472** | **376** | **417** | **500** | **415** | **371** | **415** | **571** | **708** | **516** | 1026 | 1293 | 1201 | 1795 | 5316 |
| | | | | | | | | | | | | | | | | | |
| 2002 | | | | | | | | | | | | | | | | | |
| Blue Line (50/50) | 275 | 311 | 490 | 395 | 435 | 516 | 423 | 426 | 357 | 652 | 827 | 535 | 1075 | 1346 | 1206 | 2014 | 5641 |
| Red Line (Warning) | 274 | 310 | 489 | 394 | 435 | 515 | 422 | 416 | 356 | 651 | 825 | 534 | 1073 | 1344 | 1194 | 2010 | 5620 |
| Black Line (Official Estimate) | **289** | **322** | **414** | **425** | **453** | **495** | **440** | **407** | **395** | **664** | **760** | **494** | 1025 | 1373 | 1242 | 1918 | 5558 |
| 2002 vs 2001 | 13% | 10% | -12% | 13% | 9% | -1% | 9% | 8% | -23% | 18% | 20% | -9% | 0% | 6% | -4% | 10% | 4% |
| Upside Demand Potential | **277** | **297** | **366** | **375** | **395** | **422** | **385** | **344** | **310** | **479** | **590** | **466** | 941 | 1193 | 1039 | 1536 | 4708 |
| Downside Demand Potential | **246** | **267** | **333** | **336** | **363** | **390** | **356** | **313** | **277** | **437** | **538** | **421** | 847 | 1090 | 946 | 1396 | 4278 |
| Prior Official Estimate | **285** | **362** | **464** | **472** | **536** | **574** | **547** | **456** | **447** | **736** | **832** | **569** | 1111 | 1582 | 1451 | 2137 | 6281 |

# Projected cash flow from inventory reduction



# ODD-VAR-GARCH

ODD - Object Data Dependent

VAR - Vector AutoRegression

GARCH - Generalized AutoRegressive Conditional Heteroskedasticity

MGARCH – Multivariate GARCH

**Clive Granger and Robert Engle**
**Nobel Prize in Economics 2003**

# Forecasting Models

***Uncertainty – error terms are assumed to be a distribution***

$$D = \mu + \xi + \varepsilon$$

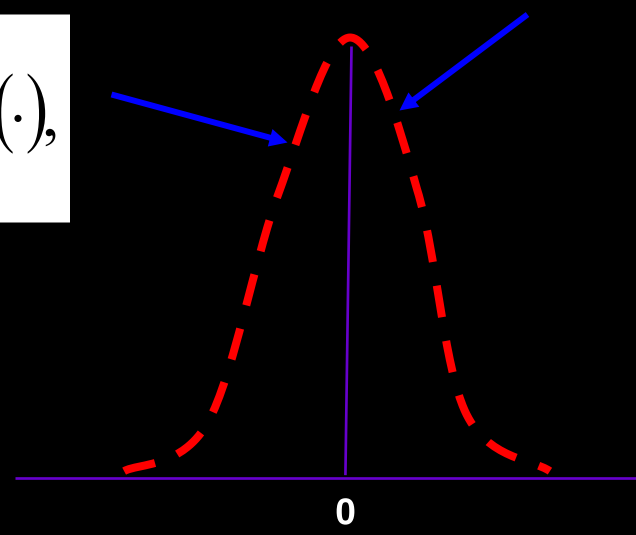
Market uncertainty $\varepsilon \sim G(\cdot)$

Manufacturer's private forecast update

$\xi \sim F(\cdot),$

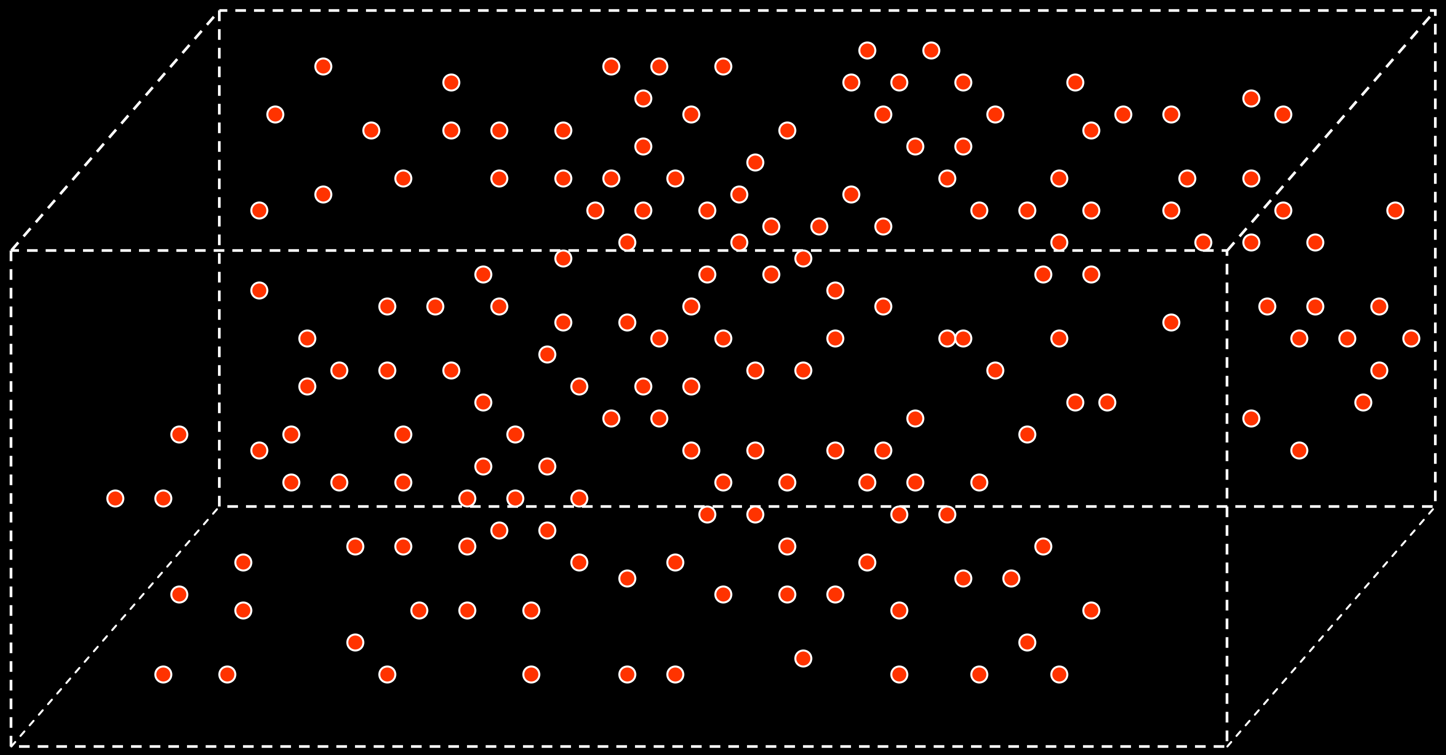
Supplier's prior belief

$\xi \in [\underline{\xi}, \overline{\xi}]$



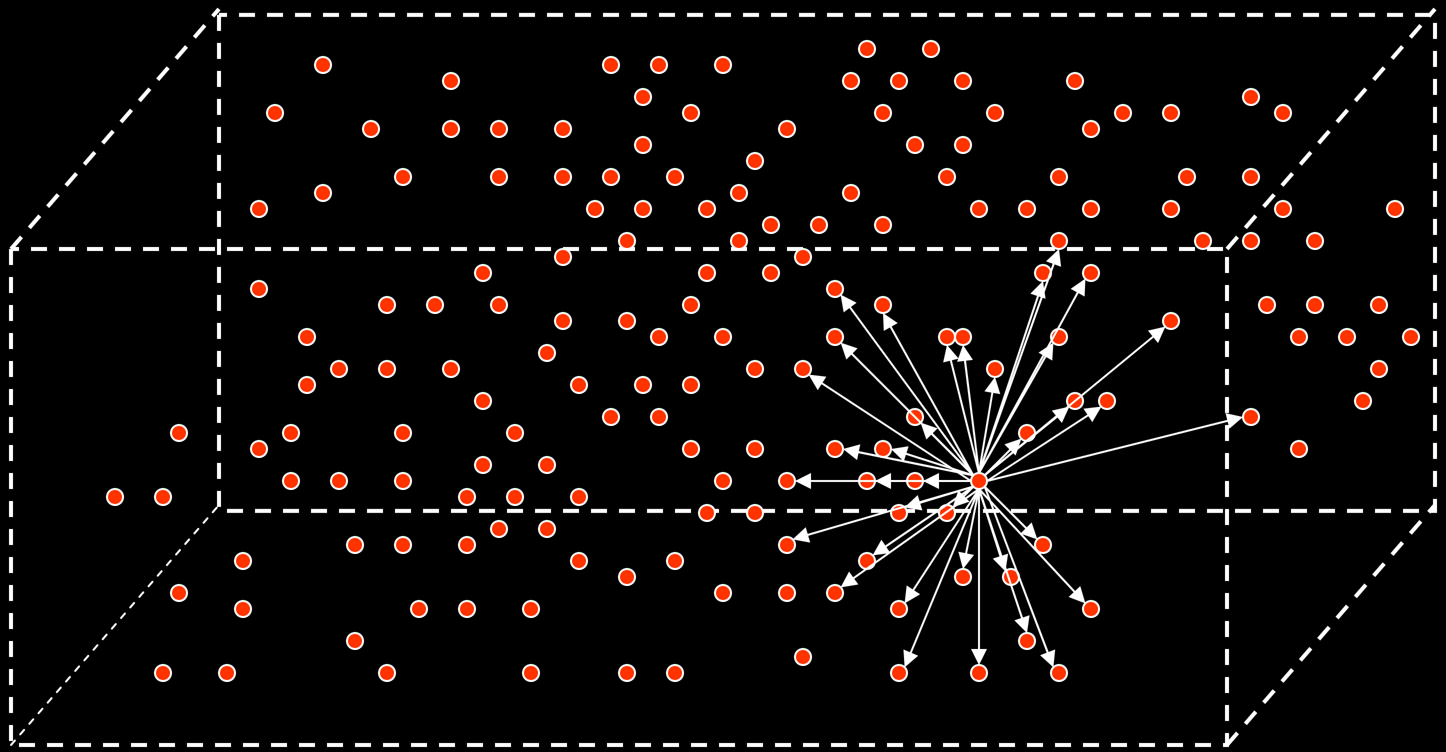
# Clustering: Classification by Reduction

**Patterns plotted in 'n' dimensional space. Each point (pattern) can represent multiple (n) pieces of information (dimensions).**



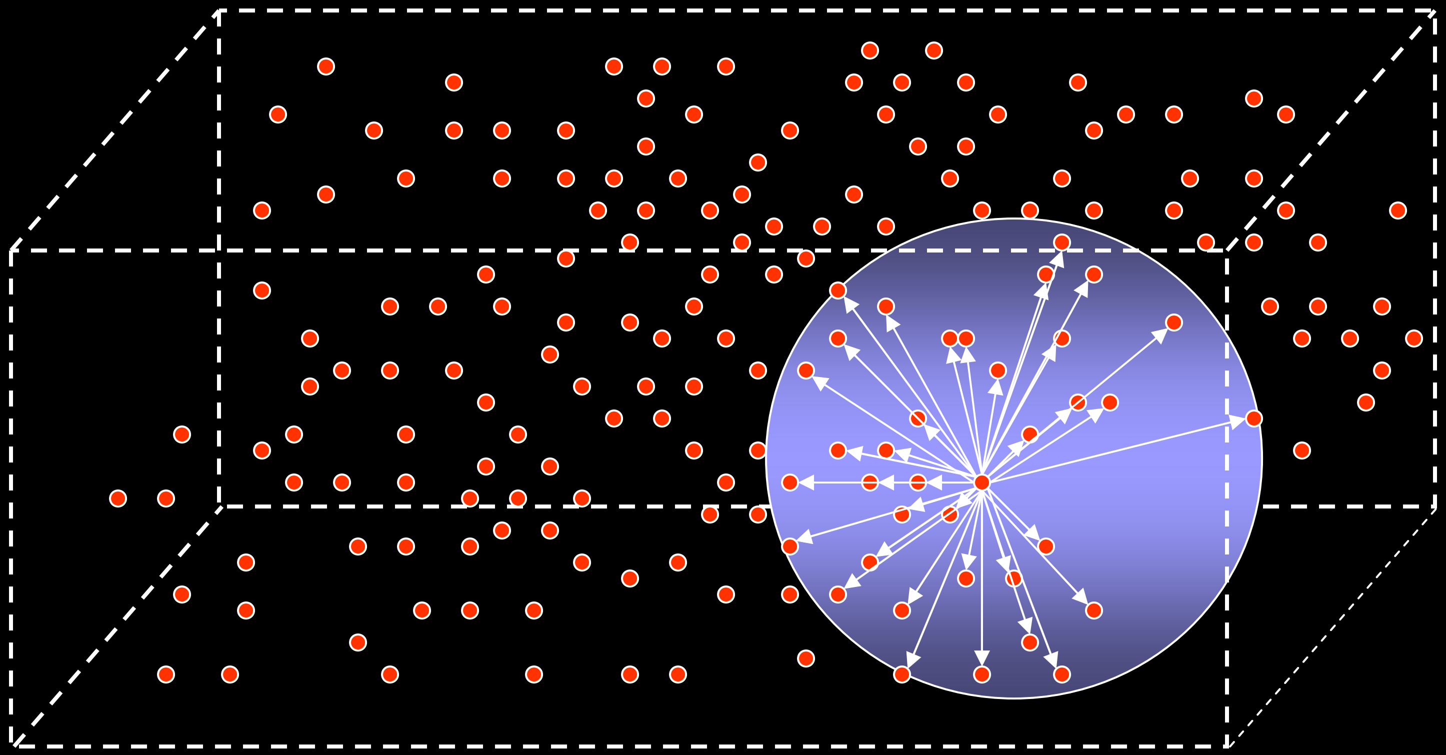
# Clusters

Distances are calculated to determine similarity.

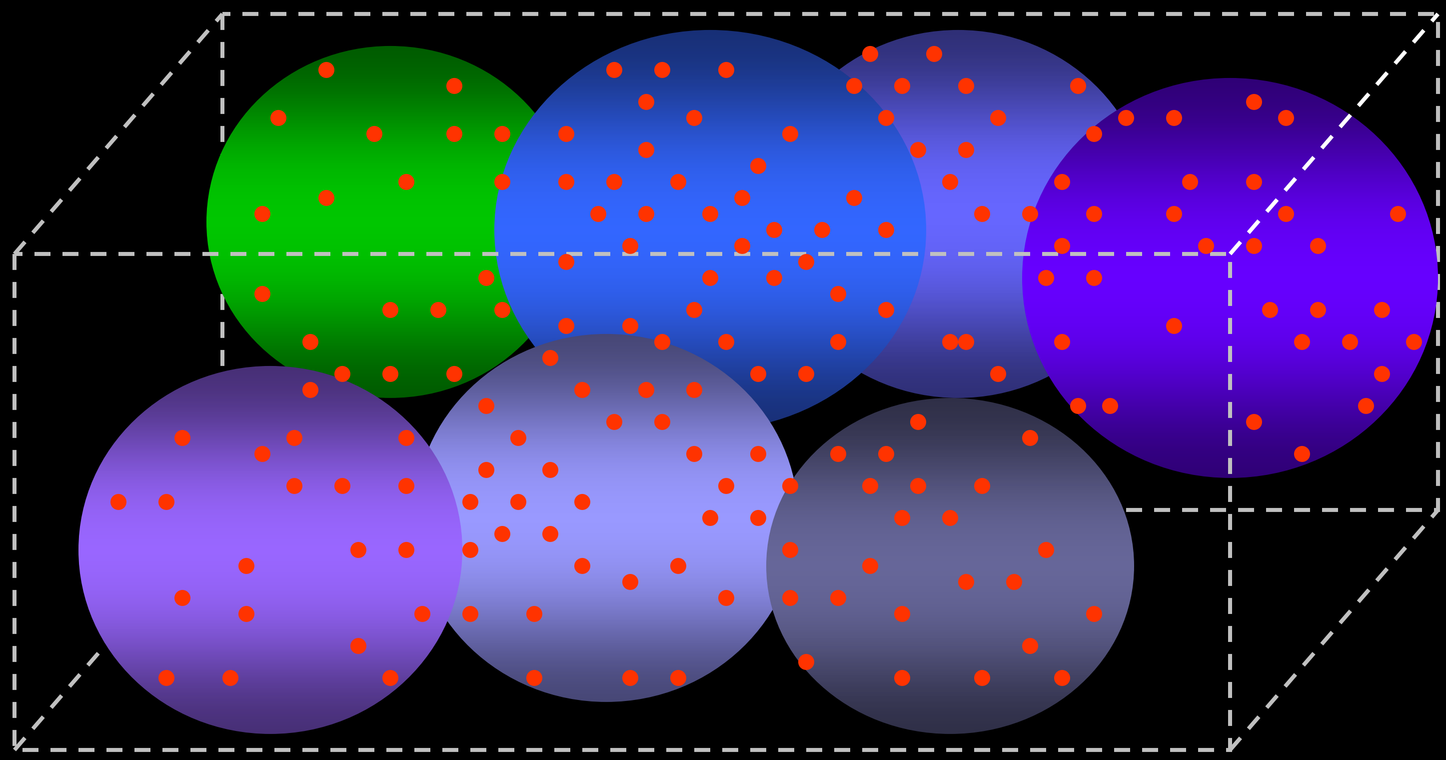


# Clustering

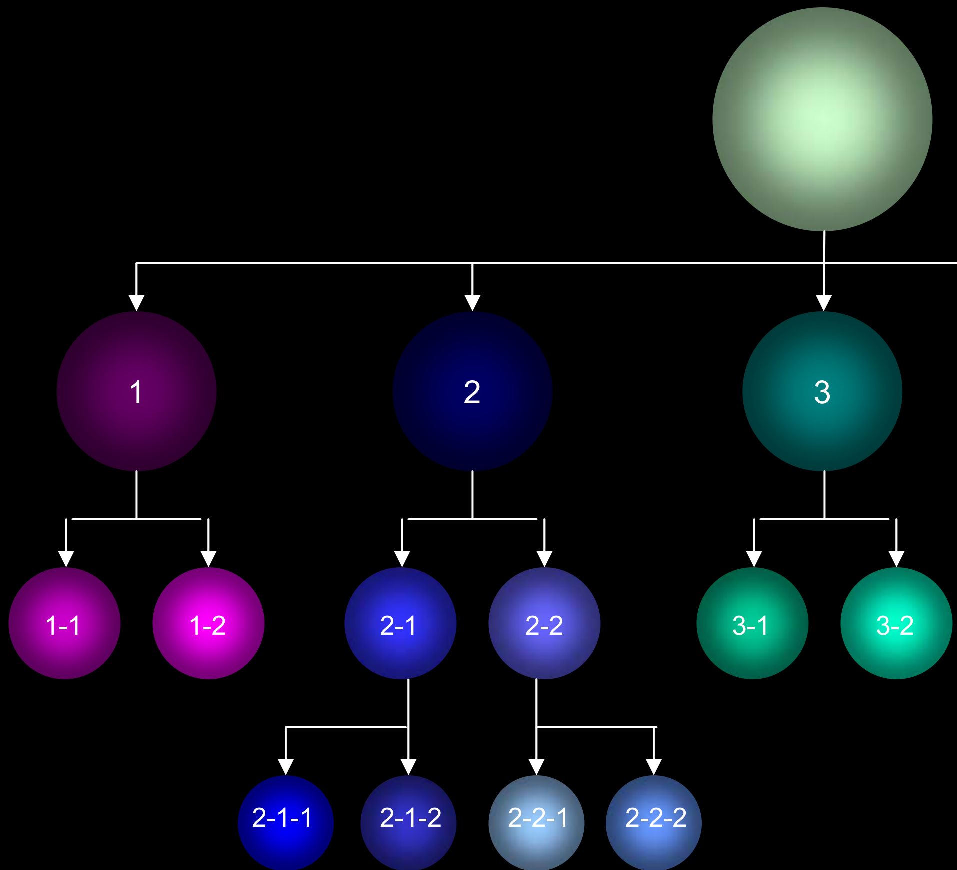
**Clusters of ‘similar’ patterns are grouped.**



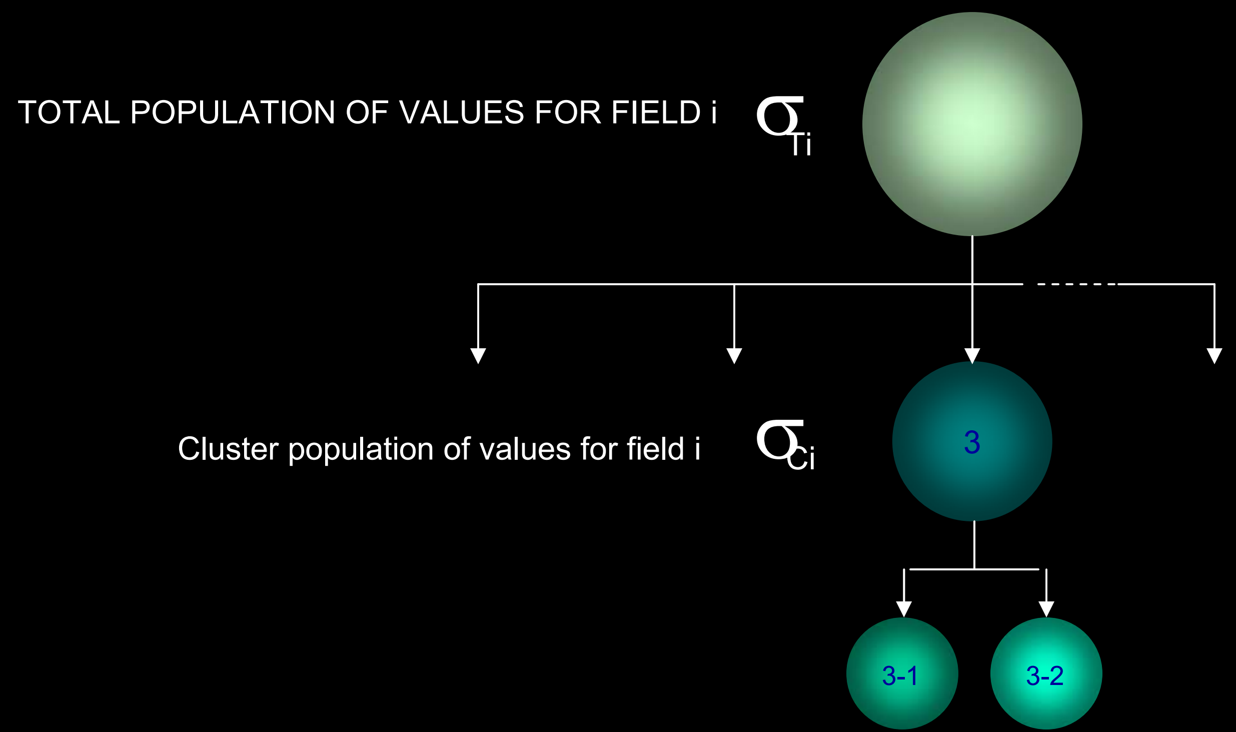
# Patterns contained within family of clusters.



# Hierarchical Clustering :
# Determines granularity of the clustering

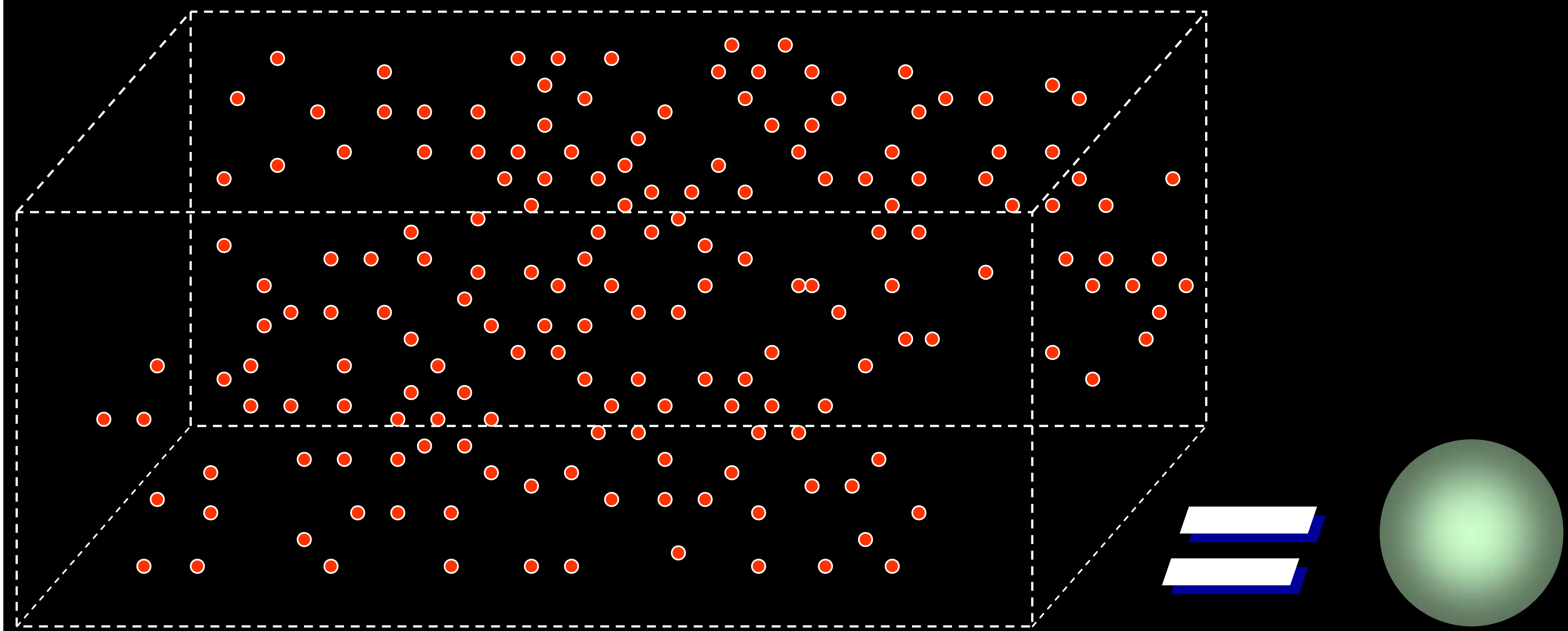


# Characterizing Clusters

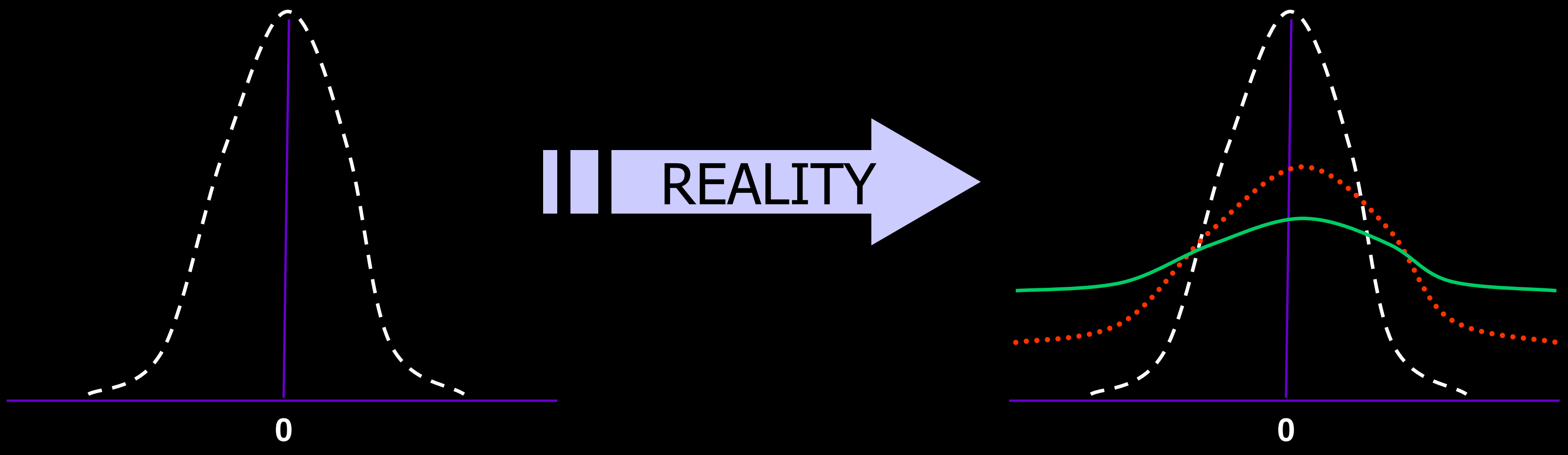



**Standard Deviation Ratio:** Identifies how much a field in a particular cluster varies in comparison to all clusters. The *standard deviation ratio for field i* is calculated by dividing $\sigma_{Ci}$ by $\sigma_{Ti}$. If the standard deviation ratio for a field is small, the field may partly characterize this cluster.

# Reductionist Approach

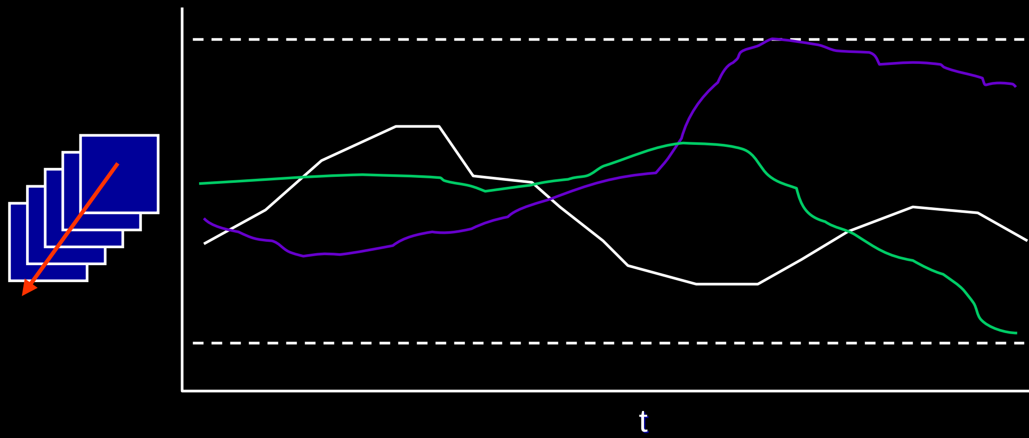


TOTAL POPULATION
OF VALUES FOR FIELD i

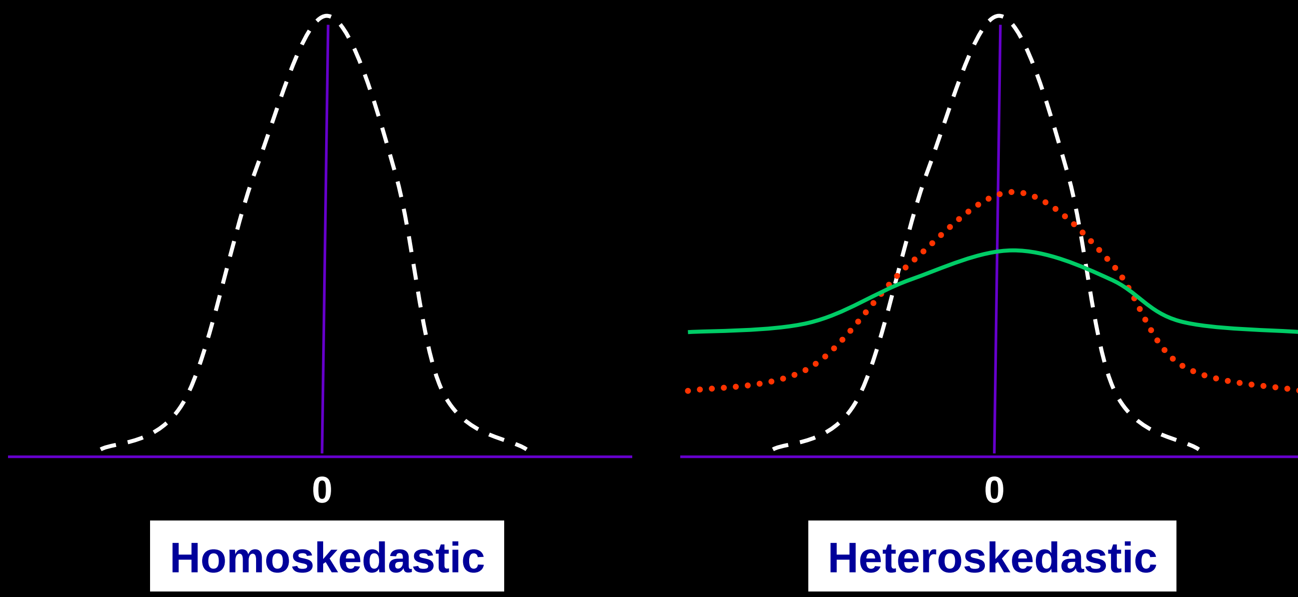


REALITY
0
0

# *Looking "across" recent history of same SKU*

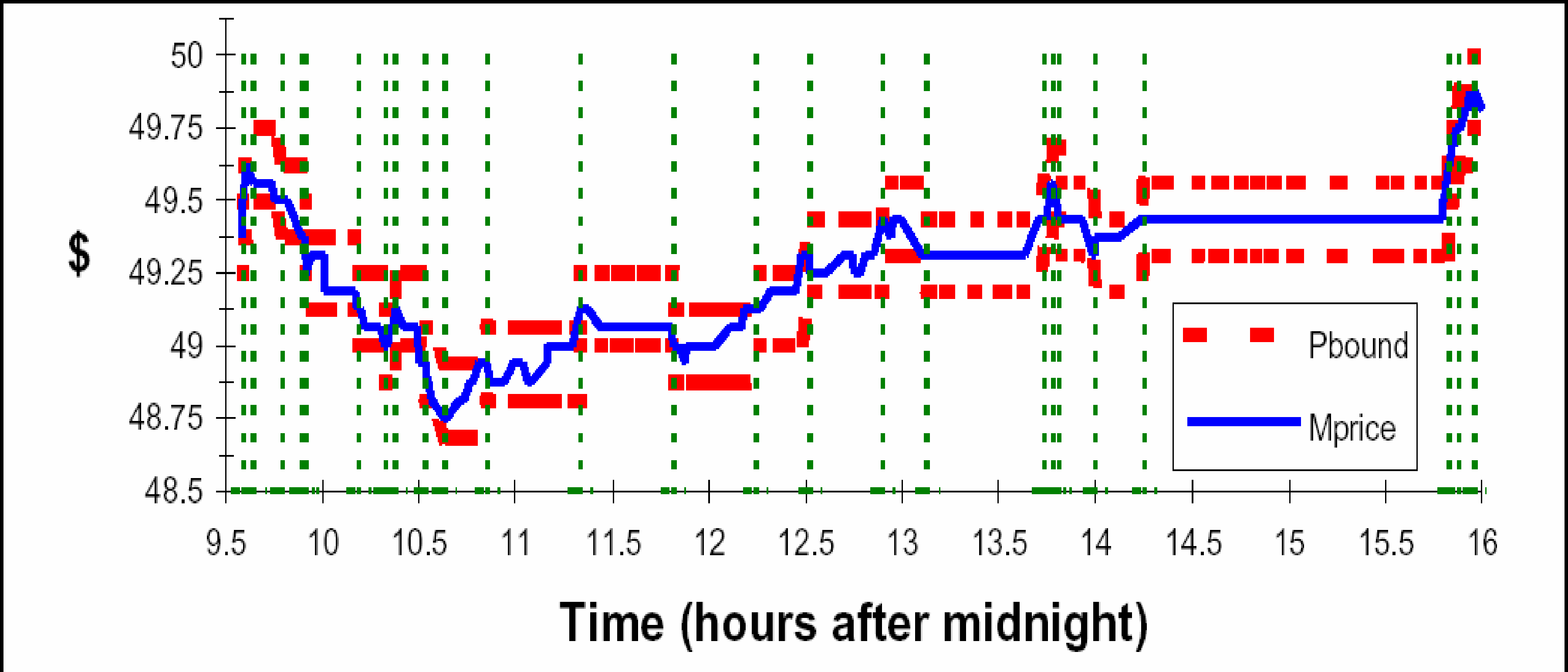


# Volatility




| | |
|---|---|
| homo = | same |
| skedasticity = | variance |

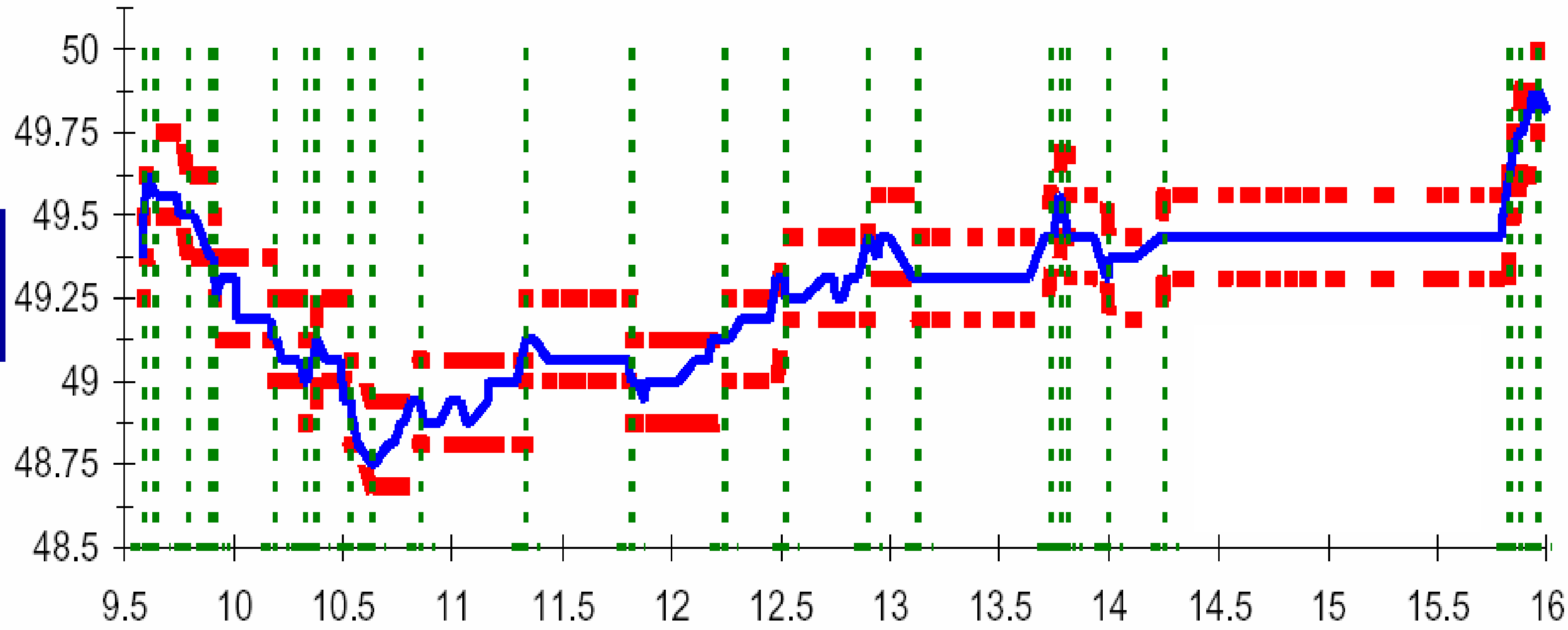
# NYSE quotes for Exxon on 01 November 2001




Source: New York Stock Exchange and Robert Engle, New York University

# RFID Data ? Sensor Network Data ?

Number of Gillette Razors on Store Shelf




RFID EPC data since store opening at 9:30 AM

## Basic CLRM

**Error Term Assumption**
**Normal distribution**

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

Dependent Variable

Example: Sales of Aspirin

Explanatory Variable

Example: In-store inventory of Aspirin

Example: Sales of Aspirin

Dependent Variable

Error Term

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_K x_{Kt} + \varepsilon_t$$

Explanatory Variables

Example:
[1] In-store inventory of Aspirin
[2] Price
.
.
.
[K] Expiration date

# " What if "

What happens to sales of aspirin if competing brand cuts price by 10% ?

Example: Sales of Aspirin

Dependent Variable

Error Term

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_K x_{Kt} + \varepsilon_t$$

Explanatory Variables

Example:
[1] In-store inventory of Aspirin
[2] Price
.
.
.
[K] Expiration date

## Estimating Coefficients: Do we need high volume data ?

**K = 10 (ten factors that impact sales)**

**N = 10 (past 10 hours, past 10 days, past 10 weeks)**

$$x_{1t} = \alpha_{01} + \alpha_{11}x_{1t-1} + \alpha_{12}x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}}x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

$$x_{2t} = \alpha_{02} + \alpha_{21}x_{2t-1} + \alpha_{22}x_{2t-2} + \ldots + \alpha_{2N_{x_{1t}}}x_{2t-N_{x_{1t}}} + u_{x_{1t}}$$

$$x_{3t} = \alpha_{01} + \alpha_{11}x_{1t-1} + \alpha_{12}x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}}x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

$$x_{4t} = \alpha_{01} + \alpha_{11}x_{1t-1} + \alpha_{12}x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}}x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

$$x_{5t} = \alpha_{01} + \alpha_{11}x_{1t-1} + \alpha_{12}x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}}x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

$$x_{6t} = \alpha_{01} + \alpha_{11}x_{1t-1} + \alpha_{12}x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}}x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

$$\vdots$$

$$x_{Kt} = \alpha_{01} + \alpha_{11}x_{1t-1} + \alpha_{12}x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}}x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

Coefficients to estimate = NK = 100 ***(excluding constants)***

***Enough degrees of freedom ?***

## ***Forecasting*** explanatory variables to determine sales of Aspirin

$$y_t = \beta_0 + \sum_{i=1}^{N_{x_1}} \alpha_{1i} x_{1t-i} + \ldots + \sum_{i=1}^{N_{x_{kt}}} \alpha_{Ki} x_{Kt-i} + \varepsilon_t$$

$$x_{1t} = \alpha_{01} + \alpha_{11} x_{1t-1} + \alpha_{12} x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}} x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

⋮

$$x_{Kt} = \alpha_{01} + \alpha_{11} x_{1t-1} + \alpha_{12} x_{1t-2} + \ldots + \alpha_{1N_{x_{1t}}} x_{1t-N_{x_{1t}}} + u_{x_{1t}}$$

---

$$y_t = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

***How many "assumed" normal distribution error terms are aggregated into this random error term ?***

# *Prior sales may help predict future sales*

## *Using lagged values of dependent variable*

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

Lagged values of dependent variable, y, say, sales of aspirin:
Lagged values of sales – last hour, yesterday, last week, etc.

# *Prior sales may help predict future sales*

## *How many lagged values of dependent variable ?*

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

**Assume, t-1 through t-10 or 10 lagged values:**
**φ Coefficients = 10 (excluding constants)**

**Coefficients = 100**

***110***
Coefficients

# *Products with Short Life Cycle*

*j = 1000 (lagged values of y for past 1000 hours)*

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

**Therefore:**

**φ Coefficients = 1,000 (excluding constants)**

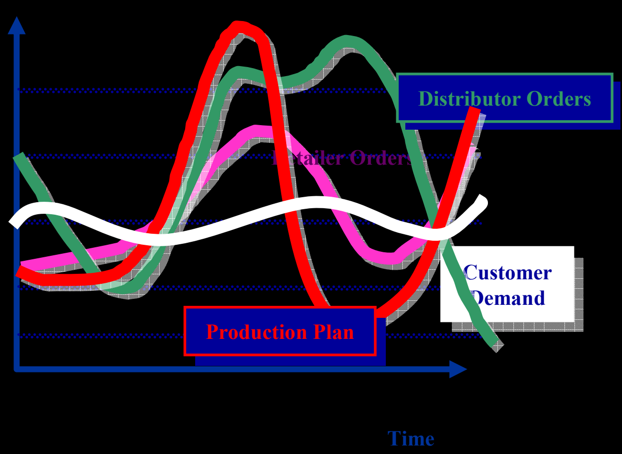
Coefficients = 10,000

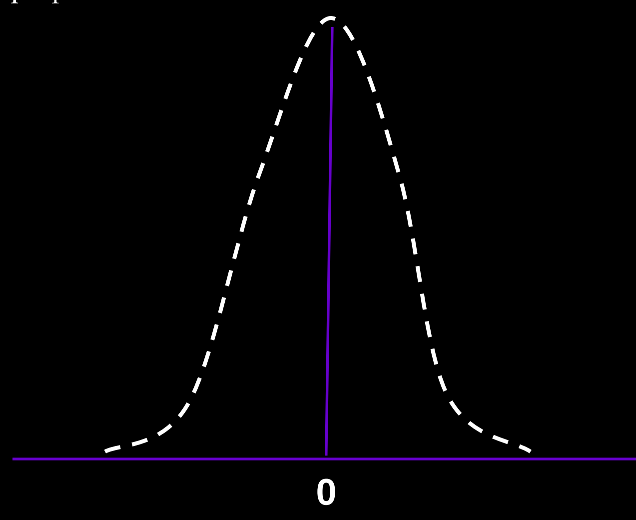
***11,000***
coefficients

# ERROR TERM

***Uncertainty – lumped as a non-variant (homo-skedastic) constant***

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki}\, x_{kt-i} + \varepsilon_t$$



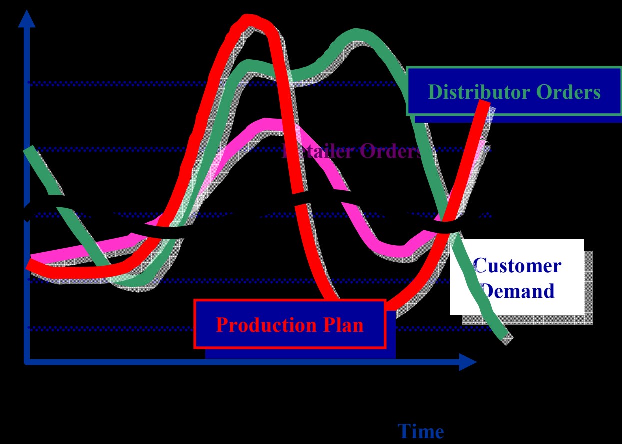
# ARCH

**Robert Engle captures the time varying volatility of the random error term**

AutoRegressive Conditional Heteroskedasticity (Robert Engle, Nobel Prize in Economics, 2003)

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \boxed{\varepsilon_t}$$

?




Variance of random error term over time depends on previous lagged errors (t-1, t-2, …., t-q)

$$\sigma_t^2 = \theta_0 + \theta_1 \varepsilon^2_{t-1} + \theta_2 \varepsilon^2_{t-2} + \ldots + \theta_q \varepsilon^2_{t-q}$$

# GARCH Generalized AutoRegressive Conditional Heteroskedasticity

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

$$\sigma_t^2 = \theta_0 + \theta_1 \varepsilon^2_{t-1} + \theta_2 \varepsilon^2_{t-2} + .... + \theta_q \varepsilon^2_{t-q}$$

**Variance of the random error term DEPENDS NOT ONLY on previous lagged errors (t-1, t-2, …., t-q) but also on LAGGED VALES OF THE VARIANCE (t-1, t-2, …, t-p)**

$\sigma_{t-1}^2$

$\sigma_{t-2}^2$

•
•
•

$\sigma_{t-p}^2$

# GARCH

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

$$\sigma_t^2 = \theta_0 + \theta_1 \varepsilon^2_{t-1} + \theta_2 \varepsilon^2_{t-2} + .... + \theta_q \varepsilon^2_{t-q}$$

**Variance of the random error term DEPENDS NOT ONLY on previous lagged errors (t-1, t-2, ...., t-q) but also on LAGGED VALES OF THE VARIANCE (t-1, t-2, ..., t-p)**

$$y_t = \beta_0 + \sum_{j=1}^{N_y} \varphi_j y_{t-j} + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varepsilon_t$$

$$\sigma_t^2 = \theta_0 + \sum_{i=1}^{q} \theta_i \varepsilon^2_{t-i} + \sum_{j=1}^{p} \tau_j \sigma^2_{t-j}$$

# VAR-GARCH

n = 2, p = 1
2 locations
1 lag period
Single SKU

$$y_{1t} = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varphi_{11} y_{1t-1} + \varphi_{12} y_{2t-1} + \varepsilon_{1t}$$

$$y_{2t} = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varphi_{21} y_{1t-1} + \varphi_{22} y_{2t-1} + \varepsilon_{2t}$$

$$\sigma_{1t}^2 = \theta_0 + \sum_{i=1}^{q} \theta_i \varepsilon^2_{1t-i} + \sum_{j=1}^{p} \tau_j \sigma^2_{1t-j}$$

$$\sigma_{2t}^2 = \theta_0 + \sum_{i=1}^{q} \theta_i \varepsilon^2_{2t-i} + \sum_{j=1}^{p} \tau_j \sigma^2_{2t-j}$$

# VAR-GARCH

**ROI from very high volume auto id data**

**n = 10; p = 1,000**
**10 locations**
**1,000 lags**

**Estimate Coefficients:**

**10,000 φ**
**+**
**10,000 for x's**
**=**
**20,000 per stage**
**or**
**200,000 for n=10**
***(excluding constants and error coefficients)***

$$y_{1t} = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varphi_{11} y_{1t-1} + \varphi_{12} y_{2t-1} + \varepsilon_{1t}$$

$$y_{2t} = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varphi_{21} y_{1t-1} + \varphi_{22} y_{2t-1} + \varepsilon_{2t}$$

$$\sigma_{1t}^2 = \theta_0 + \sum_{i=1}^{q} \theta_i \varepsilon^2_{1t-i} + \sum_{j=1}^{p} \tau_j \sigma^2_{1t-j}$$

$$\sigma_{2t}^2 = \theta_0 + \sum_{i=1}^{q} \theta_i \varepsilon^2_{2t-i} + \sum_{j=1}^{p} \tau_j \sigma^2_{2t-j}$$

# VAR-GARCH

*Auto id nodes in Supply Network Planning*

n = 10; p = 1,000

**10 locations**

$$y_{1t} = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varphi_{11} y_{1t-1} + \varphi_{12} y_{2t-1} + \varepsilon_{1t}$$

$$y_{2t} = \beta_0 + \sum_{k=1}^{K} \sum_{i=1}^{N_{x_{kt}}} \alpha_{ki} x_{kt-i} + \varphi_{21} y_{1t-1} + \varphi_{22} y_{2t-1} + \varepsilon_{2t}$$

$y_{10t}$

*Will ROI increase if business process is optimized before tech investment?*

*Will precision of forecasting depend on an optimized supply network planning?*

# Transforming EBM to ABM

## Cross-docking Variables: Decouple ‘Chains’ to Include/Exclude Local Effects

Traditional EBM (CLRM example): Sales of Aspirin and Variables that Impact Sales

**EBM** - Explanatory Variables

Example of X:
[1] Inventory
[2] Price
[K] Expiration

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_K x_{Kt} + \varepsilon_t$$

Inventory Agent

Price Agent

Expiration Agent

**ABM** - Agents in EBM DSS

Transformed EBM plus ABM within CLRM construct: Sales of Aspirin

## Decoupling Equation-Based Models (EBM) prevalent in SCM

*Agent-integrated business models rapidly respond to changes in value network partners and incorporate local changes for global optimization.*

MODEL 1

Inventory, Price, Expiration

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_K x_{Kt} + \varepsilon_t$$

MODEL 2

$$y_t = \beta_0 + \beta_1 x_{1t} + \boxed{\phantom{\beta_2 x_{2t}}} + \dots + \beta_K x_{Kt} + \varepsilon_t$$

3

$$y_t = \beta_0 + \beta_1 x_{1t} + a_2 z_{2t} + \dots + \beta_K x_{Kt} + \varepsilon_t$$

*CROSS-DOCKING VARIABLES*

# VAR-GARCH (+ ABM) : Real World Behaviour

Real world outcomes are influenced by events or interactions between decision domains (supply chain or value network partners). Coefficient $\varphi_{ij}$ refers to changes in $y_i$ with respect to $y_j$ (hence, importance of SNP).

For example, if $y_1$ represents Michelin tire sales at Sears retail store and $y_2$ represents Michelin tire sales at the distributor, Merisel, then parameter $\varphi_{12}$ refers to changes in sales at retail store ($y_1$) with respect to sales at the distributor ($y_2$).

Random error term ($\varepsilon_{1t}$ and $\varepsilon_{2t}$) volatility will impact both dependent variables ($y_1$ and $y_2$). Uncertainty in the sales at retail store impact sales at the distributor.

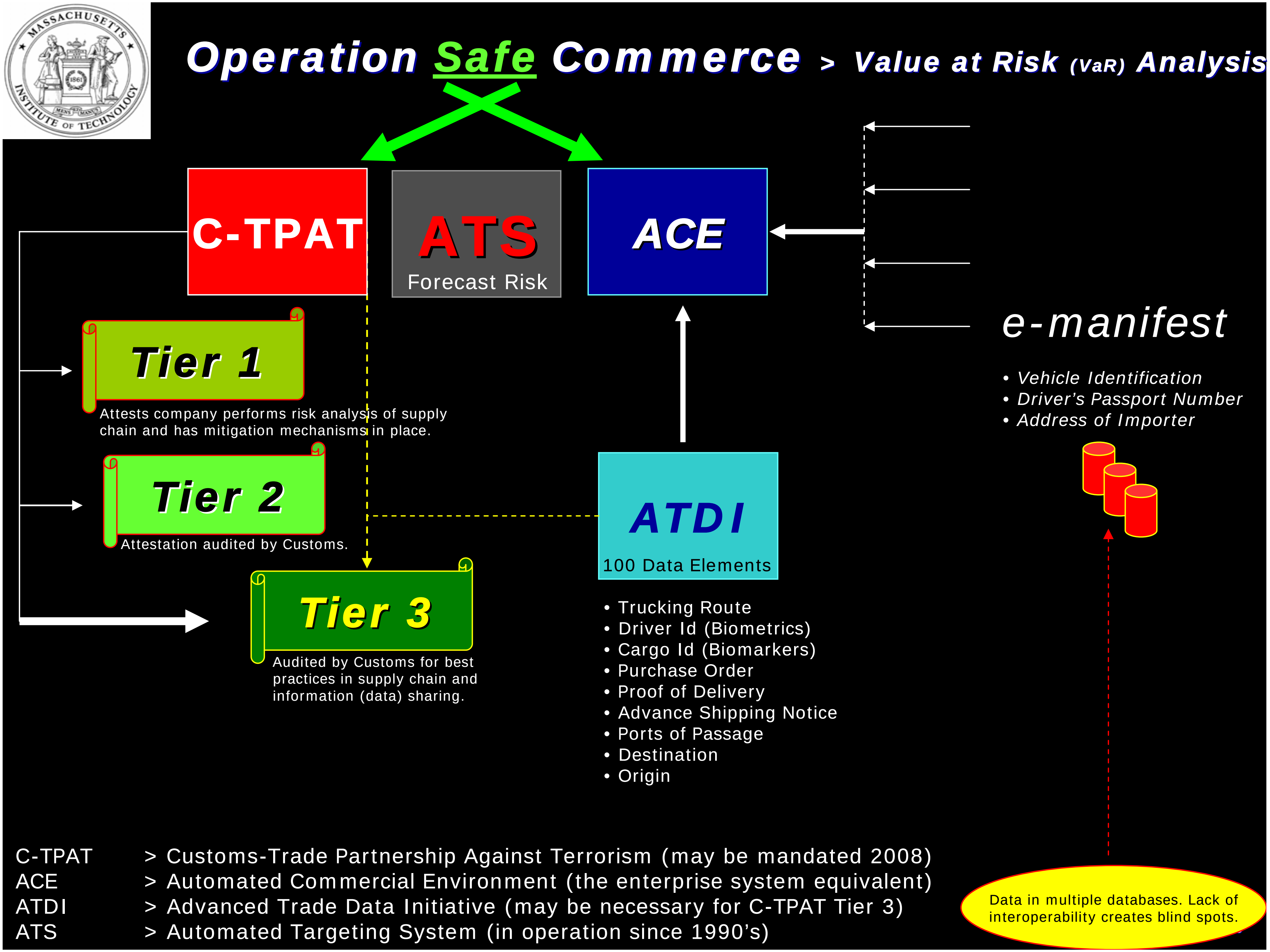
If $\varepsilon_{1t}$ changes, it will change $y_{1t}$ and $y_{2t}$ since $y_{1t}$ appears as one of the regressors (explanatory variable) for $y_{2t}$ **(thus, volatility or uncertainty of 1 error term impacts all dependent variables). This impact was completely ignored thus far in all models, tools and forecasts.**

## Interoperability Management ?
## Supply Chain Management ?
## Information Management ?
## Data Flow Management ?

- Data
- Status
- Analytics
- Information
- Collaboration
- Decisionable Information
- Intelligent Contextual Response

***CONTEXT***

| | |
|---|---|
| Data: | Driver or Facilitator? |
| Process: | Optimize or Adaptive? |
| Information: | Static or Decisionable? |


Operation Safe Commerce > Value at Risk (VaR) Analysis
C-TPAT
ATS
Forecast Risk
ACE
Tier 1
Attests company performs risk analysis of supply chain and has mitigation mechanisms in place.
Tier 2
Attestation audited by Customs.
Tier 3
Audited by Customs for best practices in supply chain and information (data) sharing.
ATDI
100 Data Elements
• Trucking Route
• Driver Id (Biometrics)
• Cargo Id (Biomarkers)
• Purchase Order
• Proof of Delivery
• Advance Shipping Notice
• Ports of Passage
• Destination
• Origin
e-manifest
• Vehicle Identification
• Driver's Passport Number
• Address of Importer
Data in multiple databases. Lack of interoperability creates blind spots.
C-TPAT > Customs-Trade Partnership Against Terrorism (may be mandated 2008)
ACE > Automated Commercial Environment (the enterprise system equivalent)
ATDI > Advanced Trade Data Initiative (may be necessary for C-TPAT Tier 3)
ATS > Automated Targeting System (in operation since 1990's)

# Ubiquitous Data Infrastructure: Information Life Cycle Management