LECTURE 13: MODEL RISK

Yeganeh Hayeri
SYS 660
Spring 2019

MODEL RISK

Overview of Model Risk

Model Uncertainty Framework

Uncertainty Quantification (UQ)

Issues with UQ for Complex Systems

Multi-modeling of Complex Systems

MODEL RISK

An often overlooked risk in decision making is that of model risk

That is the risk that the model you are using for decision support leads to a faulty decision

The concern with regard to policy making is obvious

However, there are actually substantial risks with regard to engineering decision making as well

- What happens if an incorrect engineering model results in a flawed product design that is then reproduced in thousands or millions of units?

Understanding model risk involves understanding the interplay between uncertainty, prediction, and validation

- How do we know that what a model says is the truth?

SOME EXAMPLES

How do we handle the fact that the probability distribution that we elicit from an expert may be incorrect?

When we assume a probability distribution to describe a risk such as probability of a large flood, how do we know that we have the correct probability distribution or that it won't change in the future?

What happens if we left out an important factor from our decision support system? Will we make a bad decision?

What if we assume that utility independence exists in order to construct a MAUT model, but it is not actually true?

What if we are using an aeronautical simulation to make design decisions for a new airliner, but there is an unknown error in the code?

MODEL VALIDATION

Many of these questions relate to the issue of model validation

Trying to answer these questions can lead you very quickly to questions involving the philosophy of science and epistemology

In fact, there is an ongoing discussion as to whether or not simulation constitutes a new source of knowledge beyond traditional experimentation

The big issue is this:

- When most people think of model validation, they think about comparing a model to experimental or observational data
- However, we often want to use simulation to extrapolate to conditions for which we either will not or cannot collect data or conduct experiments

MODEL VALIDATION

Zeigler lists three forms of model validation:

- Replicative validity – the model can reproduce the data that was used to develop it
- Predictive validity – the model can correctly predict data that was not used to develop it
- Structural validity – the structure of the model itself is comports with known theory. Components of the structure themselves have been validated

People seem to neglect the last category, but it may be essential when we extrapolate to areas where we don't have data

I would modify structural validity to incorporate an idea by Winsberg: the model conforms to accepted methods of construction

MODEL VALIDATION

Even if we “validate” our model using all three forms, there is still a risk that our model could be wrong

Really, this is just another form of the old problem of induction

Just because we used a model successfully in the past doesn’t mean that it will continue to work in the future

The problem of induction has never been solved…

So we try to do the best that we can with what we know

There is nothing that we can do about the unknown unknowns

TAXONOMY OF UNCERTAINTY

Aleatory Uncertainty

- Irreducible uncertainty in that no amount of new information will remove it
- Example: coin flip
- Can be modeled with a probability distribution

Epistemic Uncertainty

- Reducible through the acquisition of new information
- Example: a measurable parameter such as system mass

Error

- A deficiency in the model implementation process
- Example: applying inconsistent units of measure in simulation code
- Example: numerical errors when discretizing a continuous model

ISN'T THIS ALL JUST PROBABILITY?

There is a long running philosophical argument as to whether or not there is a difference between aleatory and epistemic uncertainty

- If you flip a coin, does it really matter whether you have not flipped the coin yet or if you have but haven't looked at the result?

From a modeling perspective, the distinction is practical

- Aleatory uncertainties are uncertainties we have given up on
- Epistemic uncertainties are those that we could reduce by conducting additional measurements, tests, analysis, etc. We are trying to decide whether or not it is worth it
- Errors are uncertainties that we mange through our development processes

The challenge is that it can be difficult to assess a probability distribution for each of these categories

Uncertainty Quantification

UNCERTAINTY QUANTIFICATION

Approaches to capture model uncertainty are known as uncertainty quantification

One popular approach breaks the problem up into one of handling aleatory and epistemic uncertainty:

- We start with the aleatory uncertainties and these are captured as probability distributions
- If we want to capture the uncertainty in the model prediction we could run a Monte Carlo simulation (or related technique) to get a probability distribution on the prediction
- But what if we don’t know the probability distribution exactly?
- We could vary the distribution parameters as well

The results is a double loop evaluation of the model, and it is very computationally expensive

UNCERTAINTY QUANTIFICATION

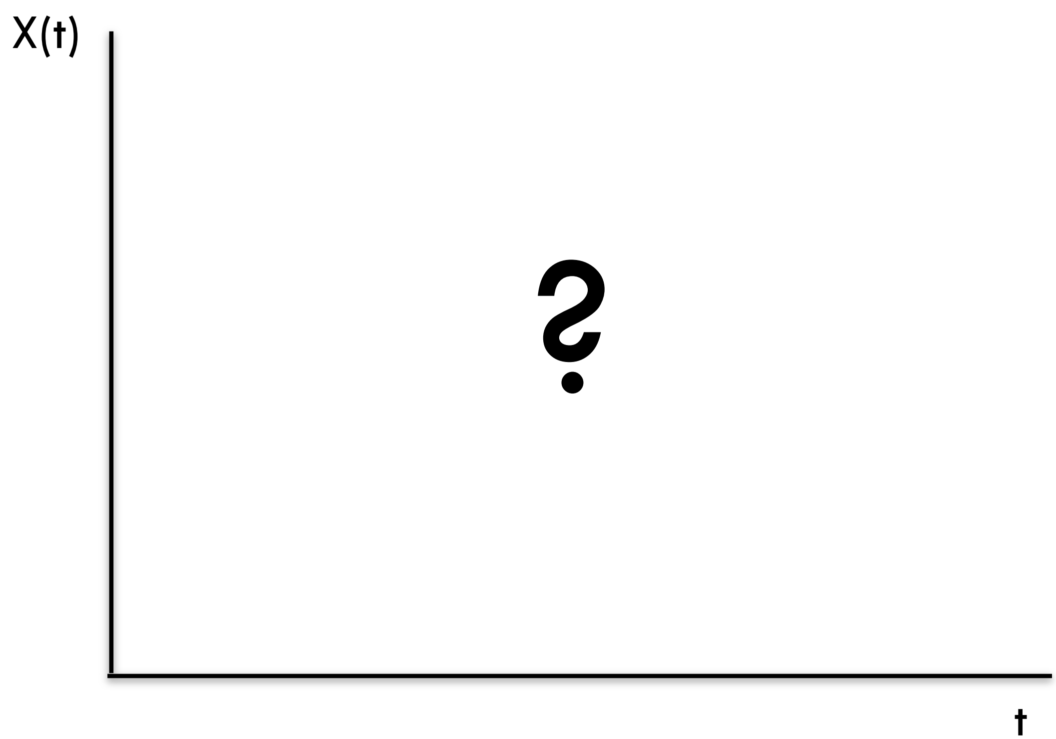
Much of the research in uncertainty quantification revolves around how to make the propagation of uncertainties tractable when the system model gets complicated

This is critical for systems where tests are difficult, expensive, or impossible

Before we get into the details, let's look conceptually at why we would do this

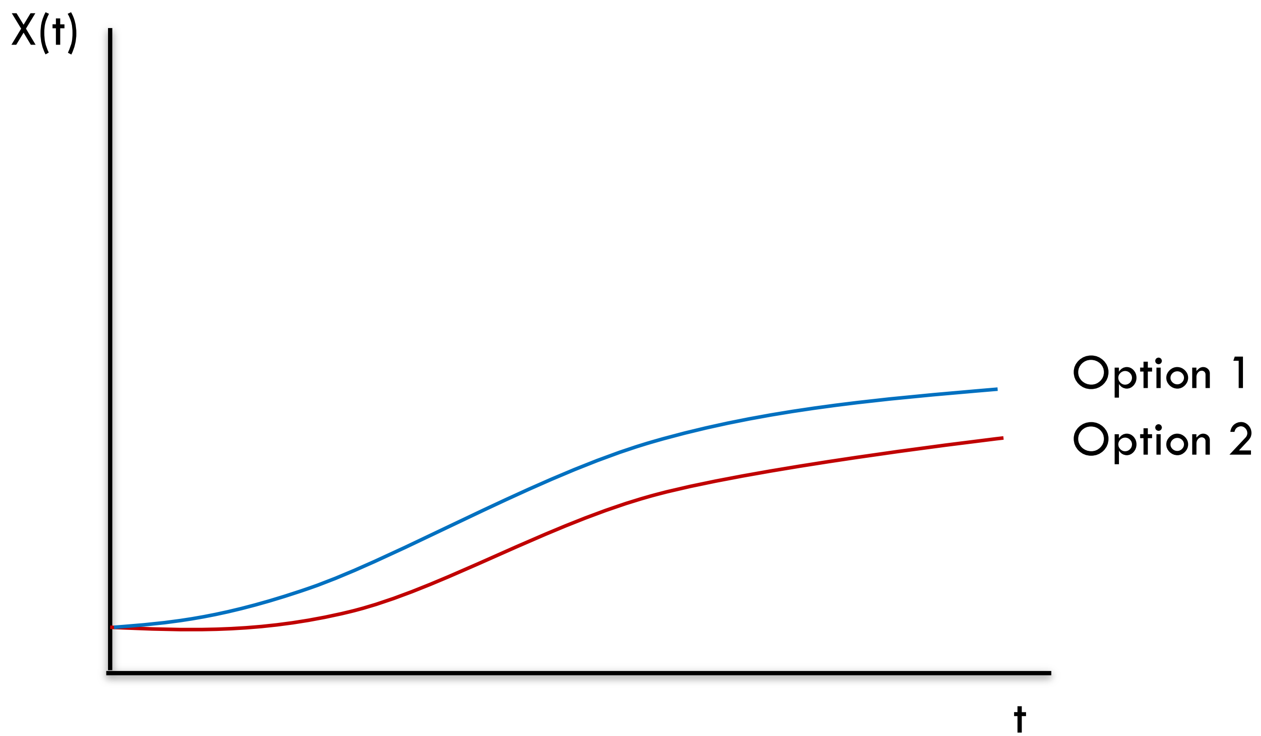
THE PROBLEM OF PREDICTION

We want to predict the behavior of a system in response to design decisions



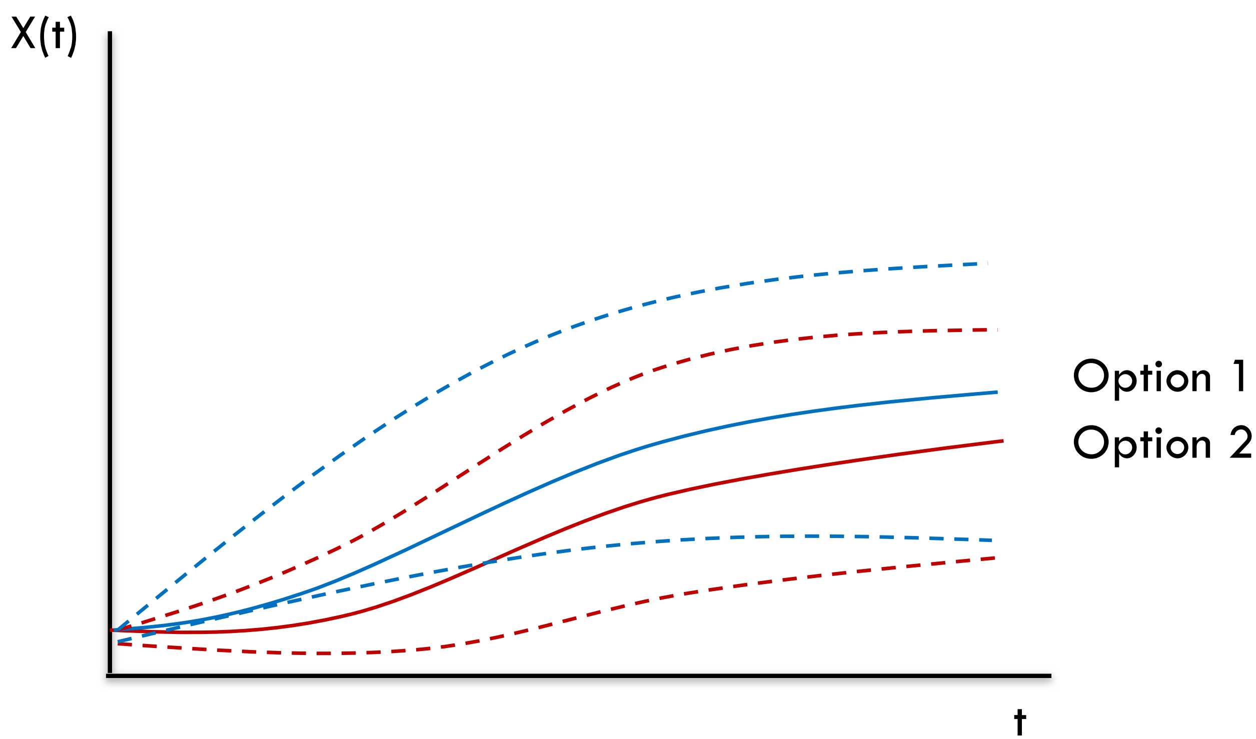
THE PROBLEM OF PREDICTION

If we can accurately model the of dynamics of a system, we can make a clear decision



THE PROBLEM OF PREDICTION

Unfortunately, during the design phase of the system, there are a large number of uncertainties that can complicate this choice



THE PROBLEM OF PREDICTION

Characterizing the uncertainty associated with the forecasted dynamics of systems is critical to determining whether or not we can discriminate among design choices

The classical way to do this is with probability distributions, but what if we are not sure what the probability distributions should be?

- We have hit upon this problem repeatedly in this class

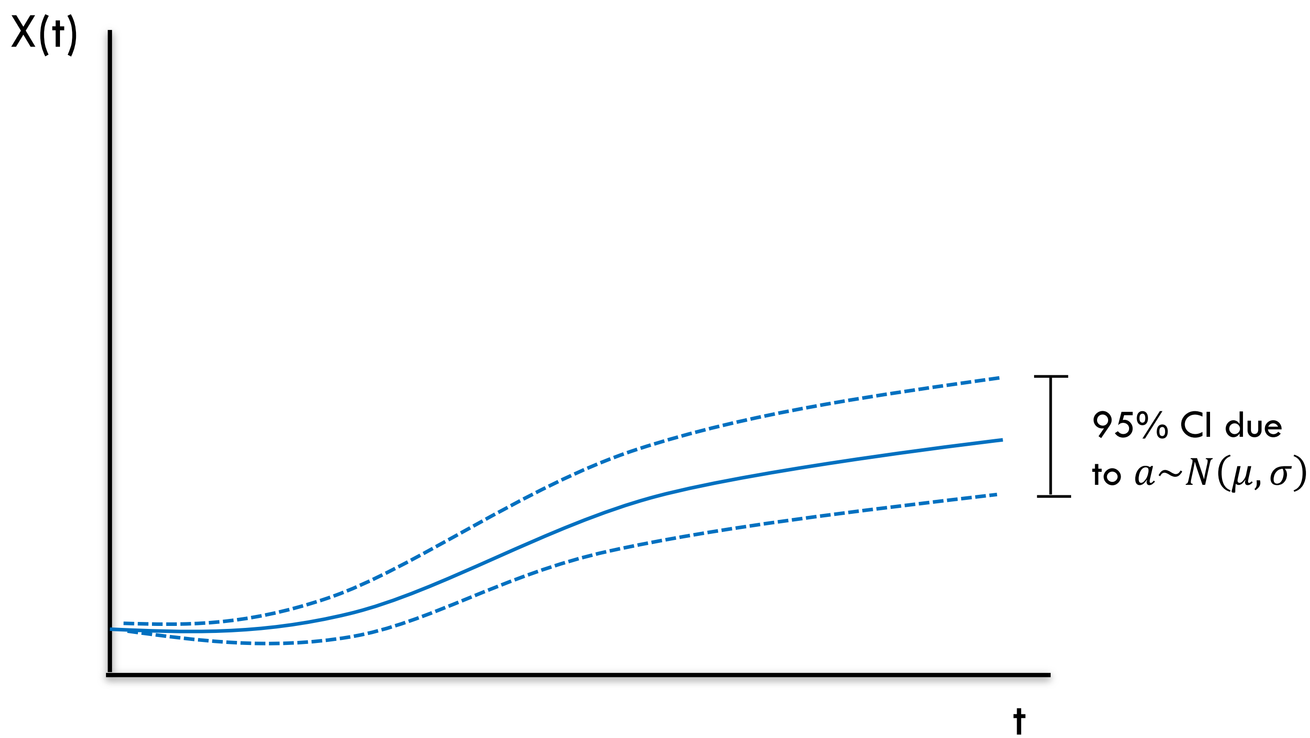
One Uncertainty Quantification (UQ) approach involves layering different sources of parametric uncertainty onto the trajectory through various means

- The approach I am about to show you has been heavily criticized
- Think about why that might be

CLASSICAL APPROACH

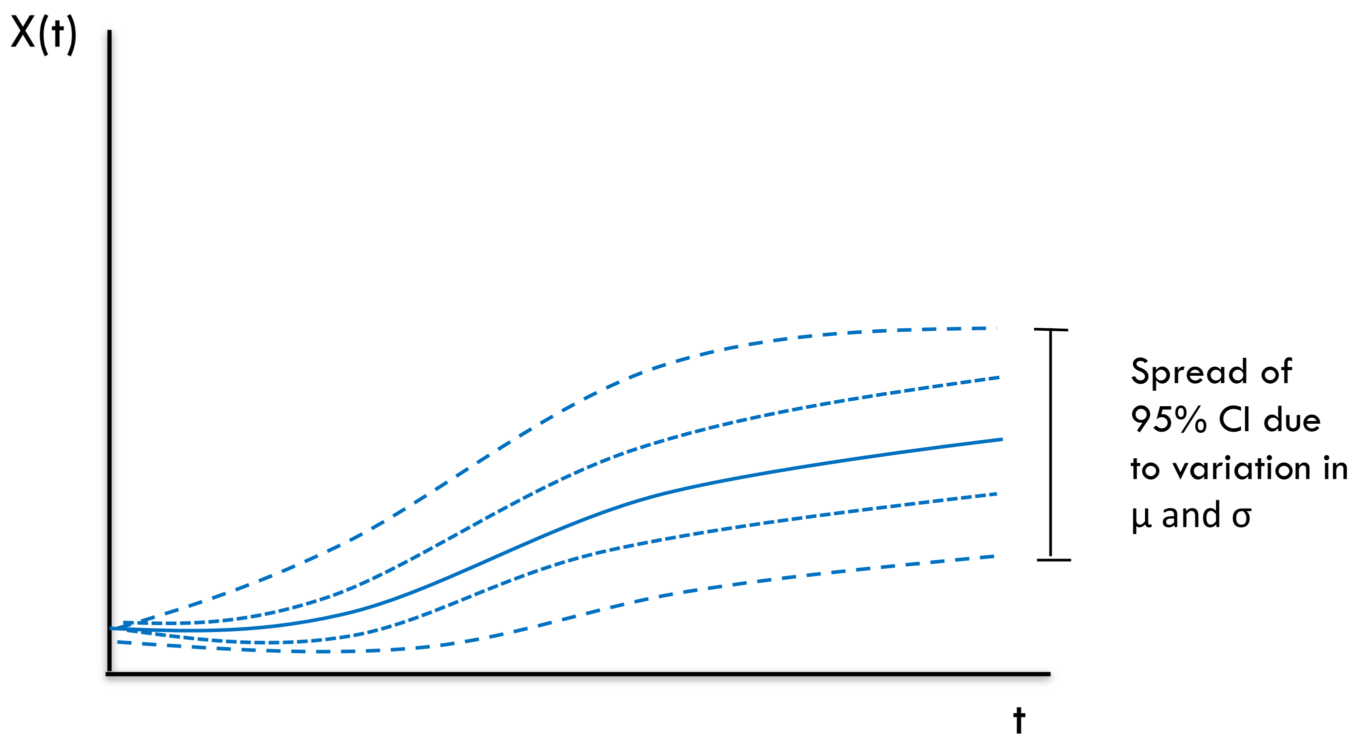
Uncertain parameter is captured via a probability distribution

- E.g., $a \sim N(\mu, \sigma)$



UQ APPROACH

We may not be certain of μ and σ, so we could vary those over intervals of possible values



DECISION RULES

The decision rule for the classical approach is to set a confidence level and then only consider options where the whole confidence interval exceeds the requirement

- E.g., A system design with a top speed of 101 +/- 10 exceeds the design requirement of 90

For this UQ approach, the decision rule is similar, but the bar is raised because we have to account for the parametric uncertainty

- E.g., if the CI for top speed could shift up or down by 5 due to parametric uncertainty, the requirement would no longer be met with confidence

UQ IMPLEMENTATION

UQ methods are often implemented via a double loop approach

The inner loop explores the impact of aleatory uncertainties

Aleatory uncertainties are modeled via probability distribution

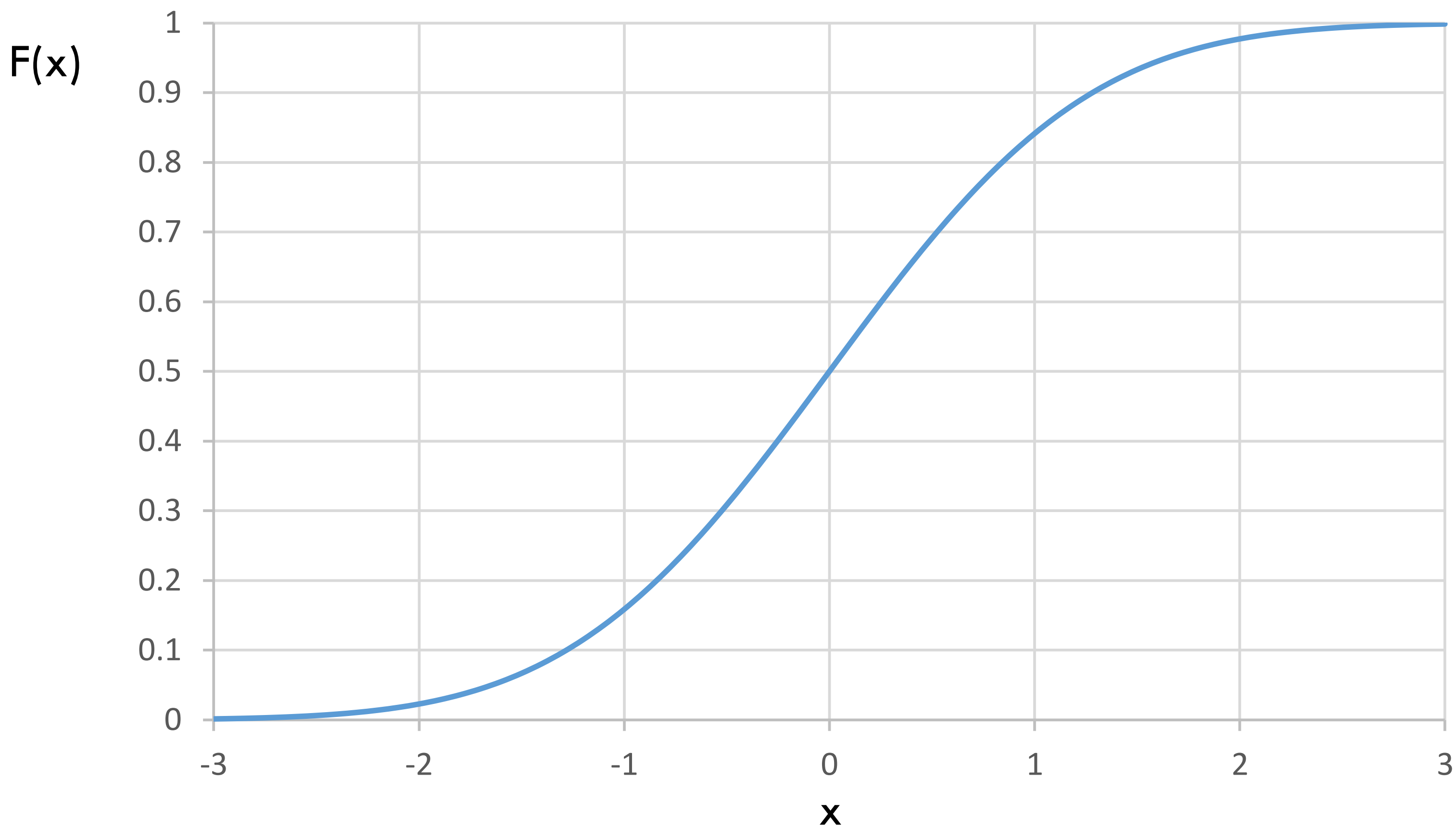
The most direct way to execute the inner loop is via Monte Carlo simulation

- Stochastic expansion methods are sometimes used instead

Recall that the output of a Monte Carlo simulation is a probability distribution

Thus, the inner loop results in a CDF for each output value of interest

CUMULATIVE DISTRIBUTION FUNCTION



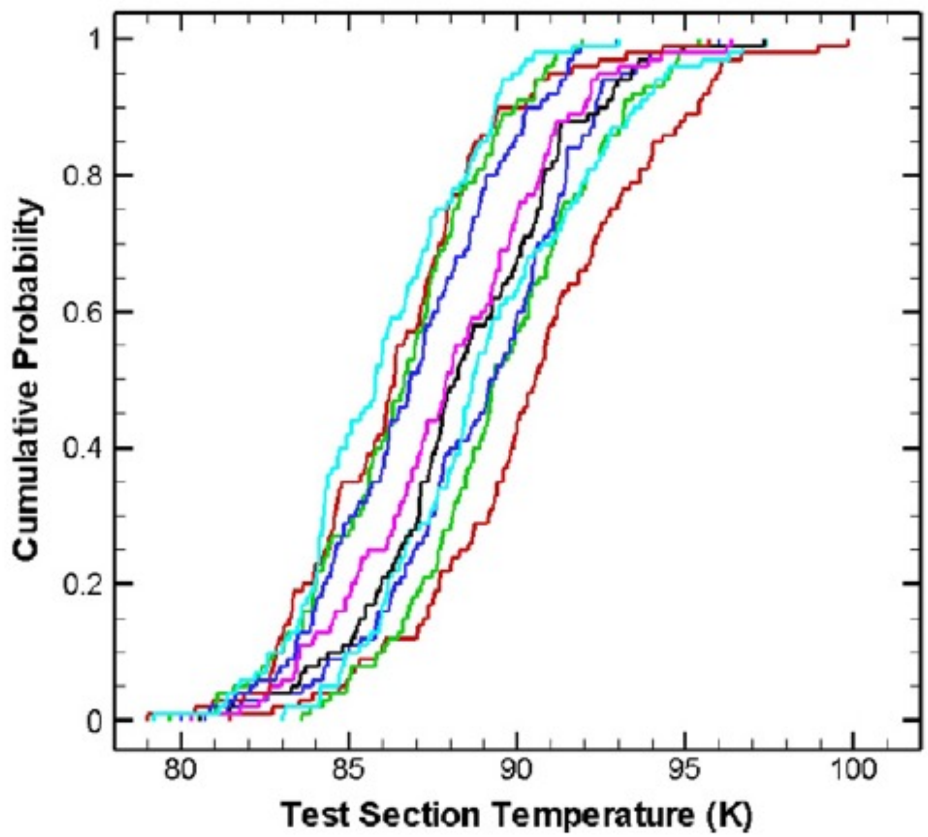
UQ IMPLEMENTATION

The outer loop varies the epistemically uncertain parameters using one of several techniques including:

- Interval analysis
- Bayesian probabilities
- Dempster-Shafer
- Possibility theory
- Latin Hypercube Sampling

If interval analysis is used, the result is a set of CDFs that cover the spread of possibilities resulting from the epistemic uncertainties

EXAMPLE OF SPREAD OF CDFS FROM UQ



Source: Roy, C. J., & Oberkampf, W. L. (2011). A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, *200*(25), 2131-2144.

MODEL FORM AND NUMERICAL UNCERTAINTY

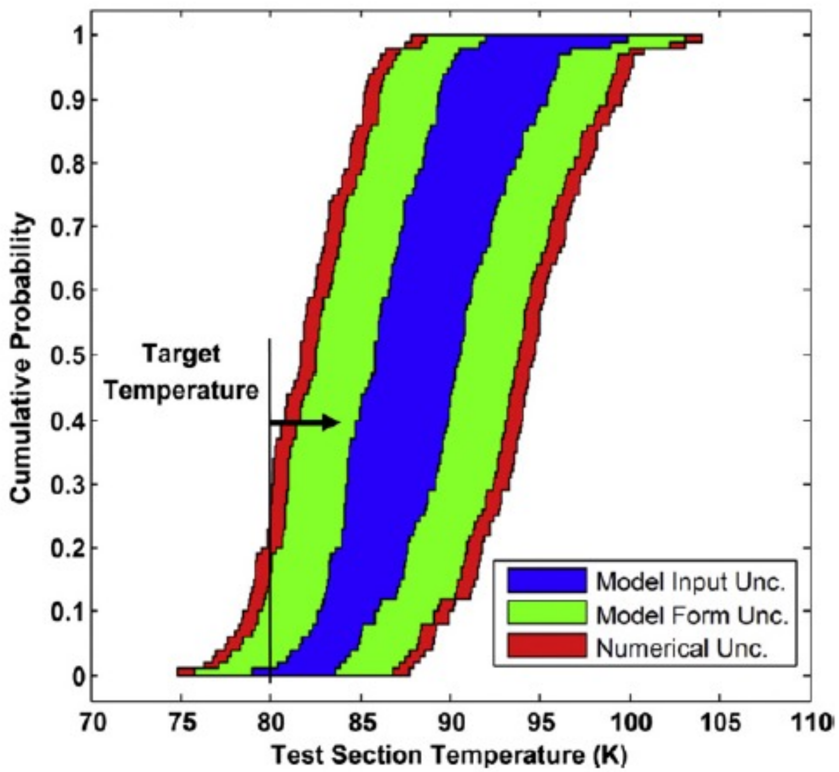
In addition to parametric uncertainties, Roy and Oberkampf (2011) also consider uncertainty in the form of the model form and numerical precision

Model form uncertainty is captured via the prediction interval that results from fitting the model to data

Numerical precision issues can sometimes be estimated based on the integration technique employed

Both of these can be added to the spread of CDFs

EXAMPLE OF UNCERTAINTY FROM MULTIPLE SOURCES



Source: Roy, C. J., & Oberkampf, W. L. (2011). A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, *200*(25), 2131-2144.

ISSUES AND QUESTIONS

What is the problem with using the prediction interval for model uncertainty?

Do you have any concerns about using a sensitivity analysis over a probability distribution?

What if we shifted from the “frequentist” view of probability to a “Bayesian” view of probability?

A COUNTER EXAMPLE

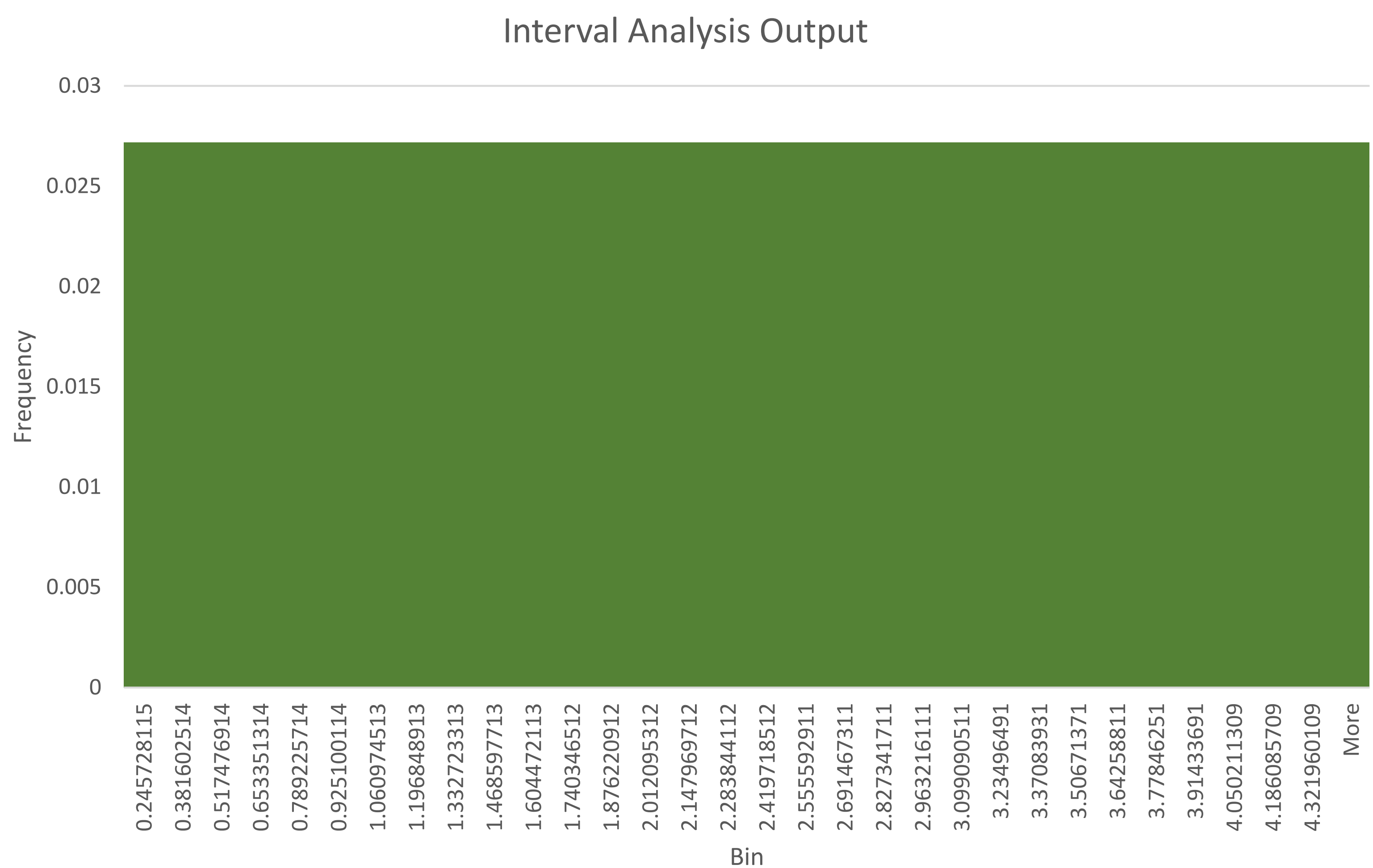
Let's take a very simple model and compare the interval approach to the probabilistic approach

- $y = x_1 + x_2 + x_3 + x_4 + x_5$

For each x_i any value between 0 and 1 is equally likely

Using the interval approach, we would characterize y as taking any value between 0 and 5

FREQUENCY PLOT



PROBABILISTIC APPROACH

What if we used a probabilistic approach instead?

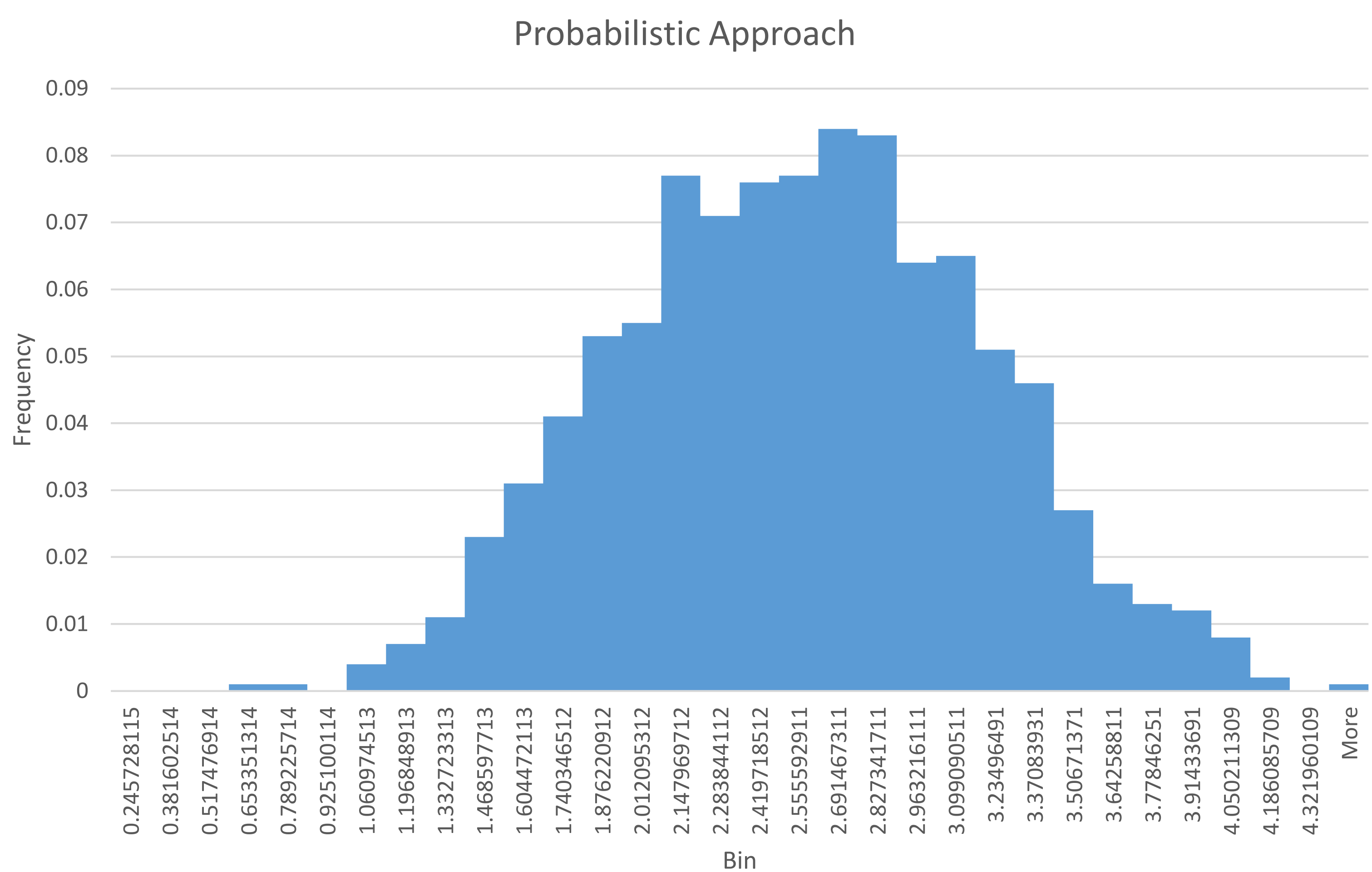
We could assign a uniform probability distribution to each variable

- $x_i \sim U(0,1)$

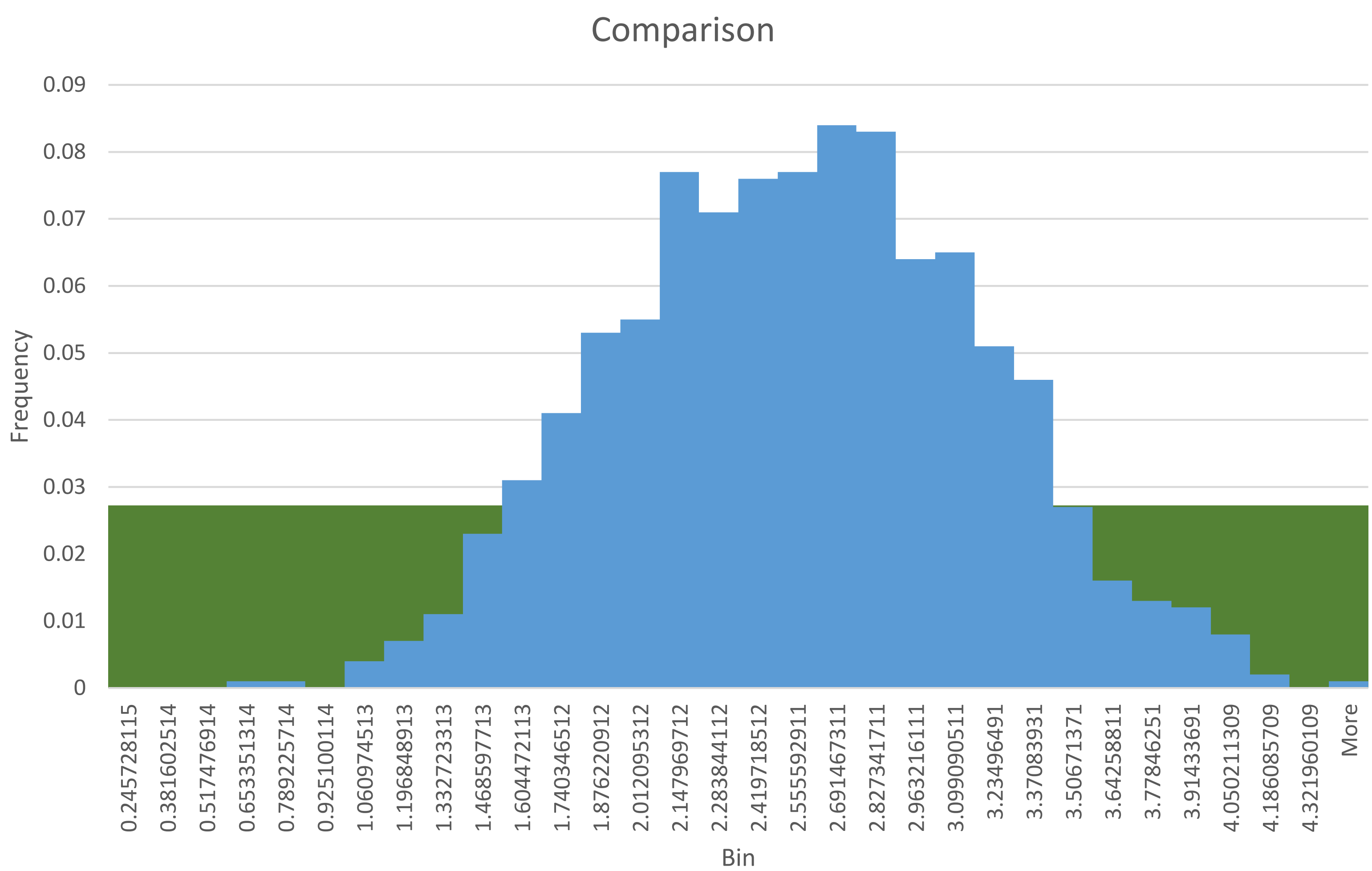
Then we could run a Monte Carlo simulation over the model and plot the results

Remember the Central Limit Theorem?

FREQUENCY PLOT



METHOD COMPARISON



IMPLICATIONS

The interval or “p-box” approach can implicitly overweight extremely unlikely outcomes and underweight likely outcomes

As you add sources of uncertainty, which there are many, the problem gets worse

For critical analyses, it is better use a Bayesian approach

- Though there are still substantial computational challenges for complicated models

APPLICATIONS OF UNCERTAINTY QUANTIFICATION

Once one can quantify all of the sources of uncertainty for a simulation, one would like to use it for design trades

The problem is that this now adds optimization to an already challenging analysis

We want to conduct a form of robust optimization that balances the expected value and spread of an output variable of interest

This substantially increases the computational challenges

Reducing the computational burden of these types of design analyses is an active field of research

Issues with UQ for Complex Systems

A MOTIVATING EXAMPLE

Let's imagine that you are advising the government of an imaginary country

This government is concerned with the major economic strides made by a rival country. Their fear is that if this trend continues, they will be surpassed both economically and militarily

They have asked you to develop a projection the size of their rival's economy over time to support policy making

CLASSICAL APPROACH

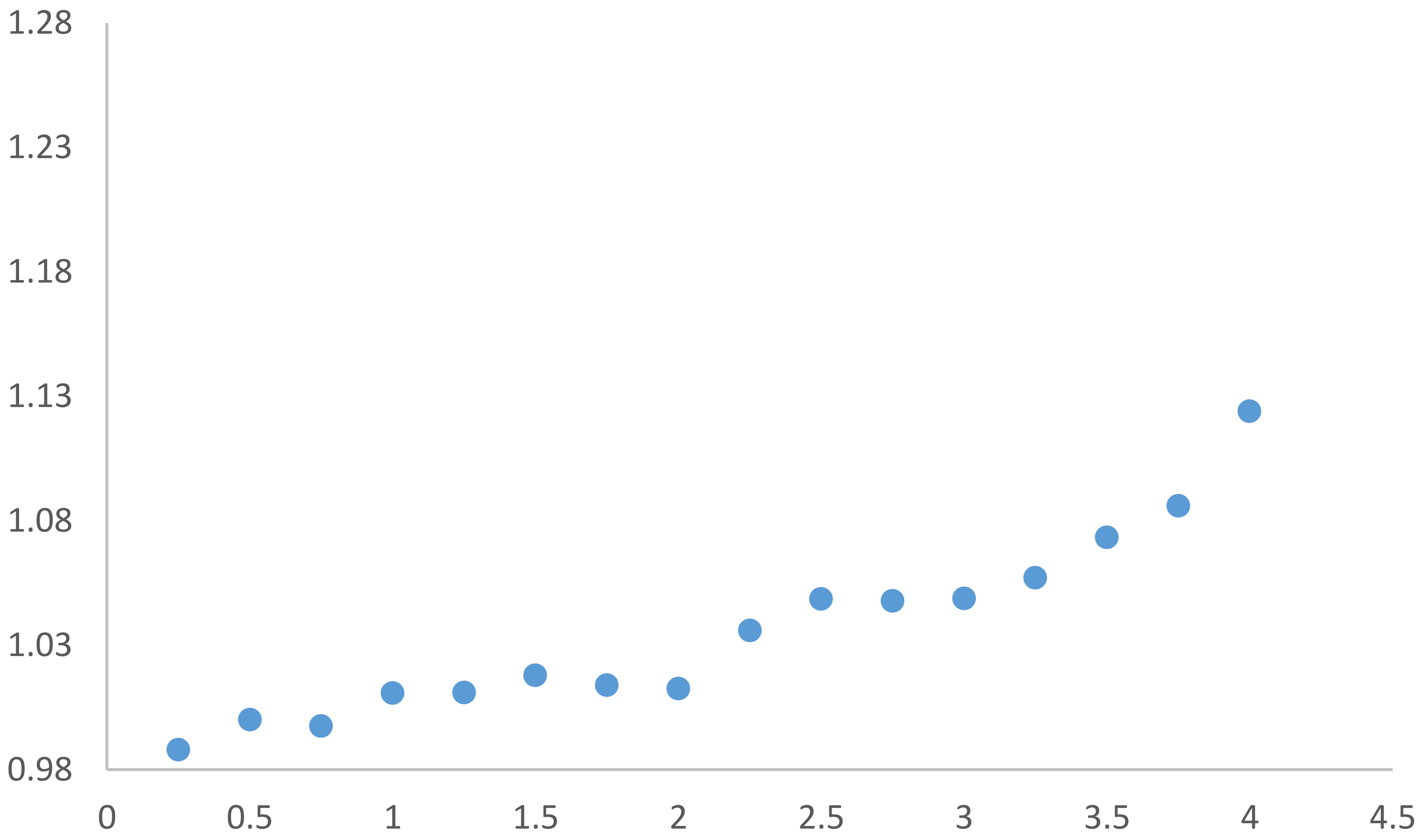
Let’s consider how we would approach this problem by via the nominal approach to model building

We will then subsequently argue that an expert wouldn’t really do it this way

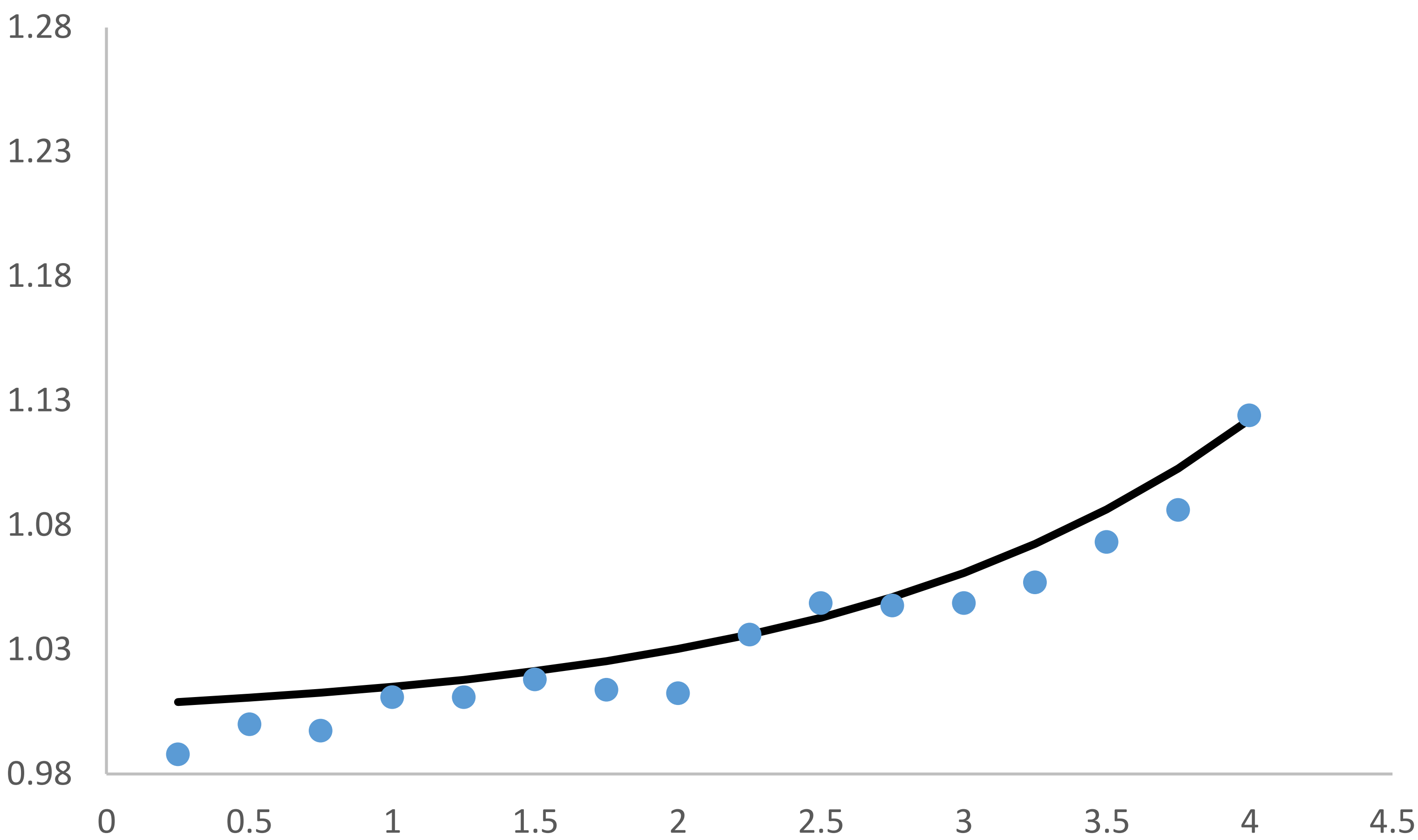
The first thing we want to do is split the data into training and test sets to avoid over-fit

Next we would want to find the simplest model that explains the training set and then evaluate it against the test set

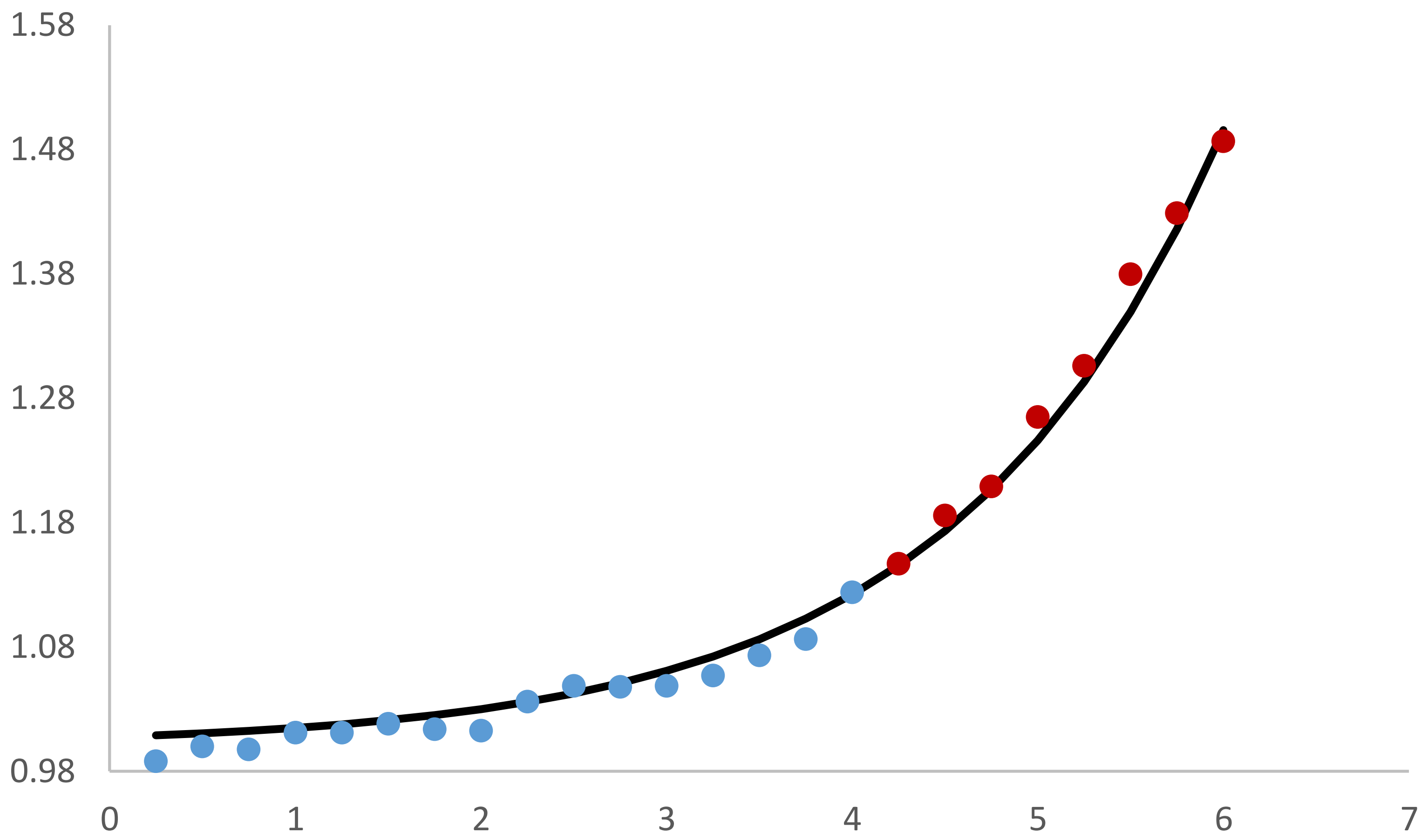
TRAINING SET



YOU THINK THE DATA MIGHT BE EXPONENTIALLY INCREASING



NOW YOU COMPARE THE MODEL TO THE TEST SET



OUTCOME OF ANALYSIS

The fit looks okay

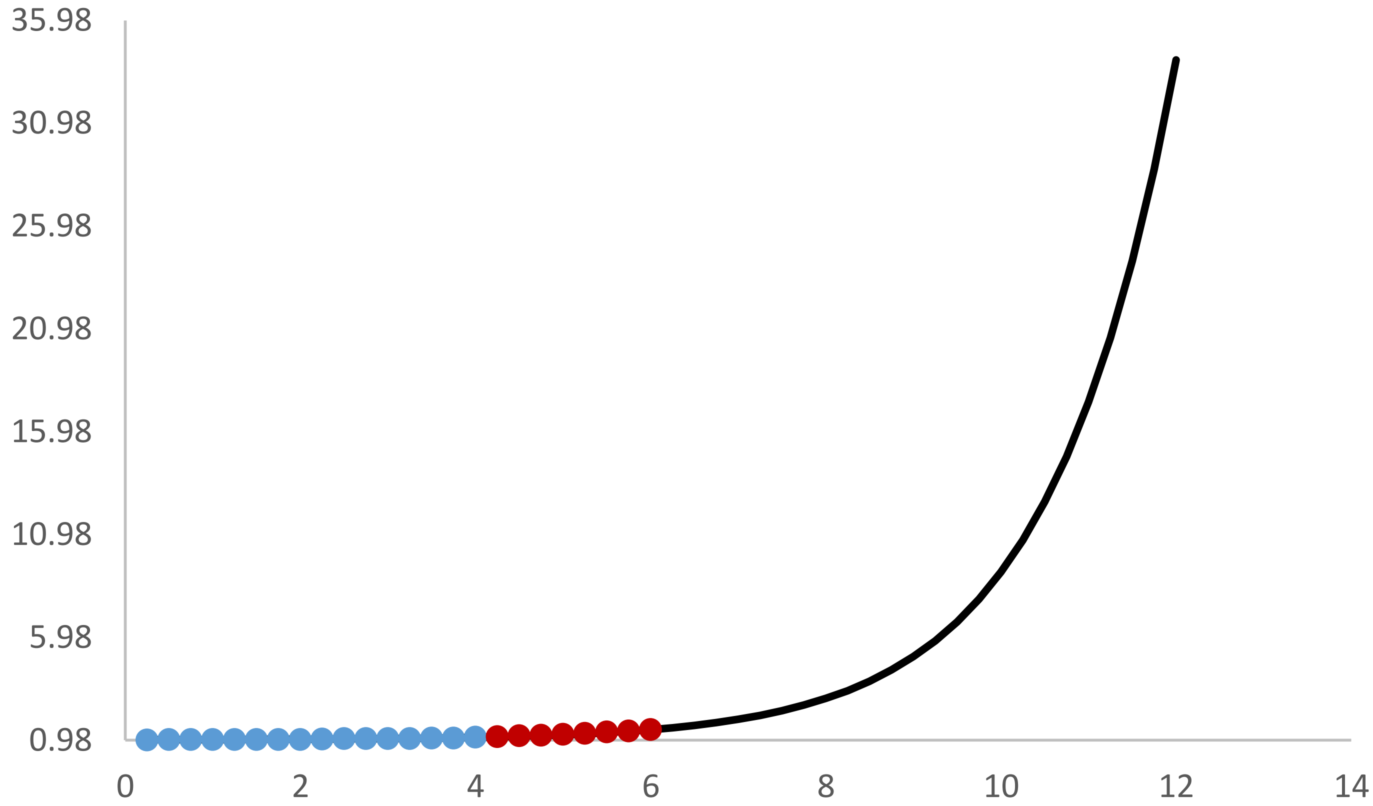
By conventional standards, you are doing well

- Since you fit the training set you have replicative validity
- Since you fit the test set you have predictive validity

Now you should be able to forecast the growth of the rival country's economy

You extrapolate the model forward

THE RESULTING PROJECTION



INCORPORATE UNCERTAINTY

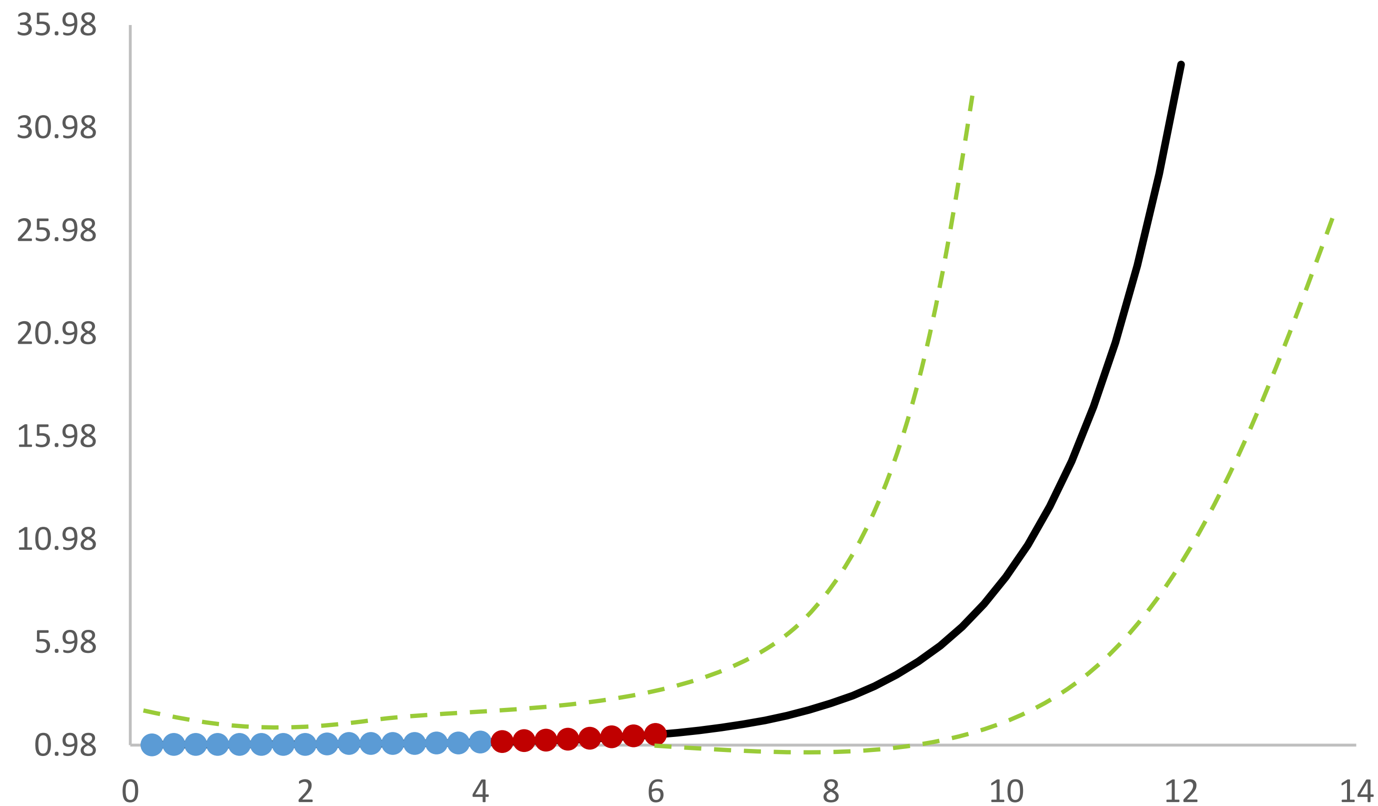
You realize that you are not 100% certain of you model fit

You decide to add some measure of uncertainty

The traditional approach is to add a confidence or prediction interval

This accounts for the impact of the "noise" in the model fit

MODEL PLUS PREDICTION INTERVAL (NOTIONAL)



WOULD YOU USE THIS MODEL?

Of course, no true economic expert would ever use this model for prediction

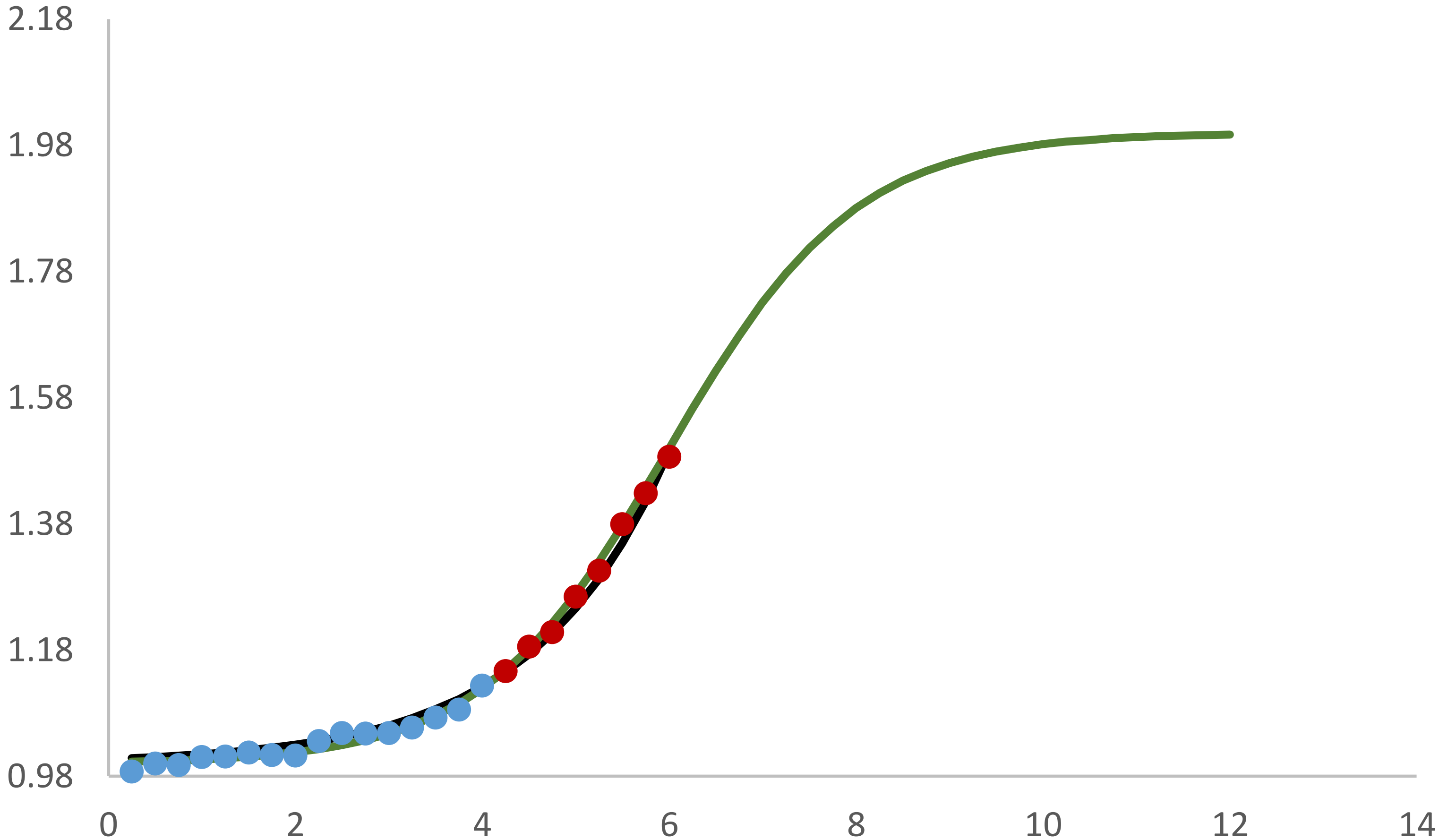
Why? We followed all the steps for model validation

The issue is that we actually know more than just what is in the data set

Obviously, considering the data and avoiding over-fit is incredibly important

But under-fit is also dangerous

HERE IS THE ACTUAL SOURCE FUNCTION



THE ACTUAL MODEL

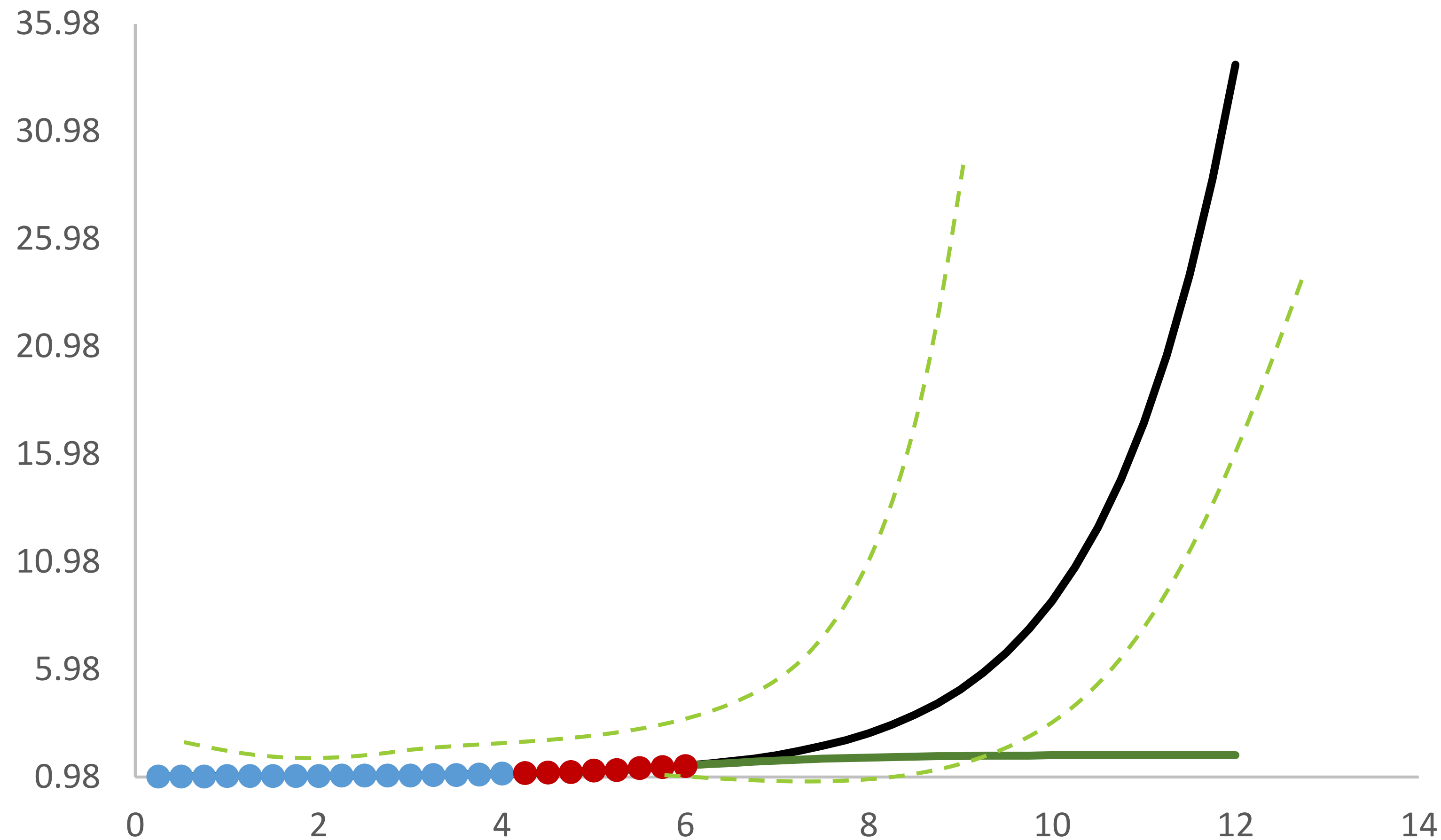
The data were actually generated using a logistic function

- A very common functional form in economics

In the real world, super-normal growth cannot continue indefinitely due to capacity and other constraints

From an economic perspective, a rapidly growing economy will eventually hit the same growth limits as everyone else

OF COURSE IF YOU HAD CONTINUED TO USE YOUR EXPONENTIAL MODEL...



STRUCTURAL UNCERTAINTY

The issue here is one of structural uncertainty

- The prediction intervals on the exponential model would not have helped in this case
- The two different models bifurcate relative to each other

Perhaps more importantly from a policy making standpoint, the two models would likely trigger two radically different policy approaches

UNDERREPRESENTING COMPLICATED AND COMPLEX SYSTEMS

Some challenges of modeling complicated and complex systems are:

- The data set under-represents the problem
- We know more than just what is in the data set
  - E.g., we have centuries of accumulated economic knowledge
- The problem is that either the knowledge exists and the model developer doesn't know it or there are multiple, potentially applicable theories and approaches

Thus, the structural uncertainty is epistemic in nature

ISSUES WITH UQ

Question: Would the Uncertainty Quantification approach described have handled the issues raised in the motivating example?

The methods discussed don't really account for uncertainty about which theory to apply when there are multiple possibilities

- Sometimes ensemble modeling with Bayesian updating is used to address such situations
- But how would this help when we need to make a decision now?

Is capturing the spread of possibilities the right way to go?

PHYSICAL SYSTEMS VS. SOCIO-TECHNICAL SYSTEMS

For physical engineered systems, the previously described UQ approach may be usable

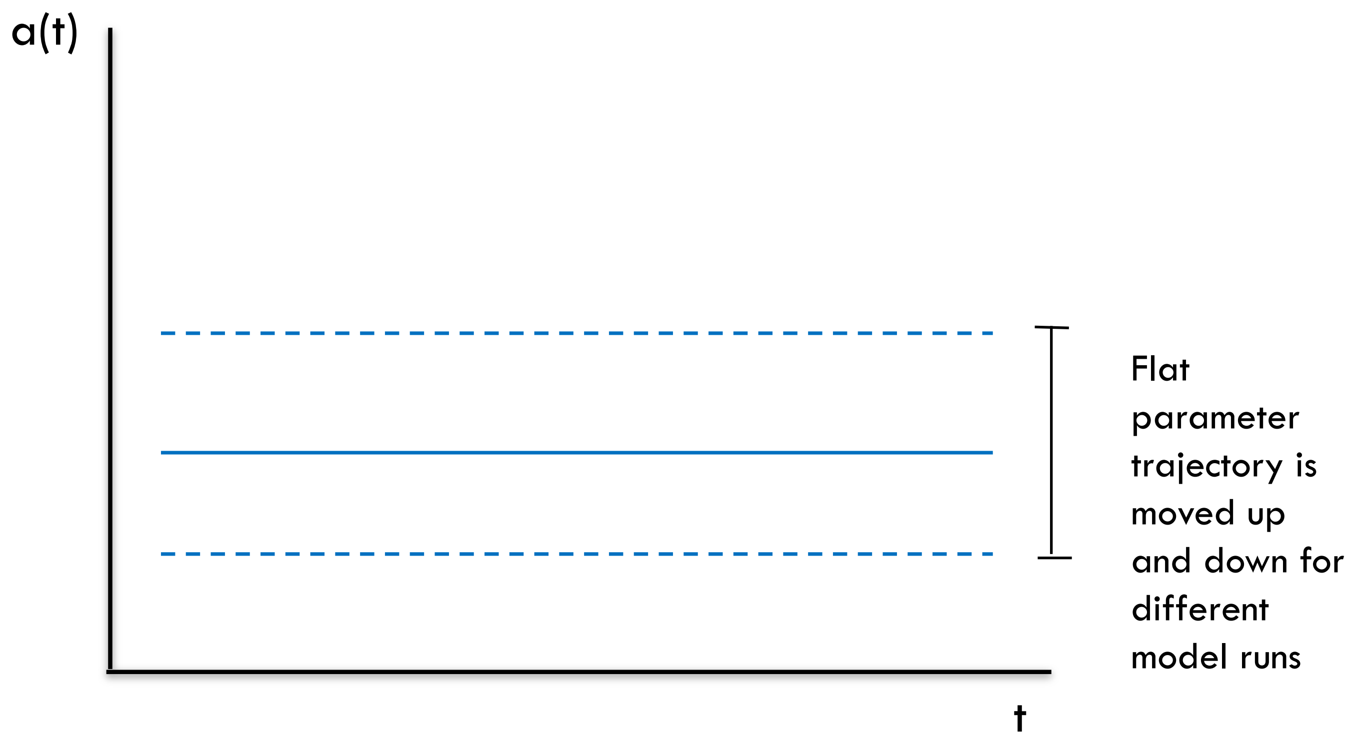
- The dynamics are often well understood
- The parametric uncertainties are truly parametric uncertainties
  - e.g., uncertainty of the final value for the system mass

For "engineering" complex, socio-technical systems, this approach will not be sufficient

- The dynamics of the system are not well understood
- The parametric uncertainties are really model uncertainties masquerading a parametric uncertainties
  - E.g., uncertainty in the future value of an interest rate is governed by its own dynamics
- Will require some sort of Bayesian approach

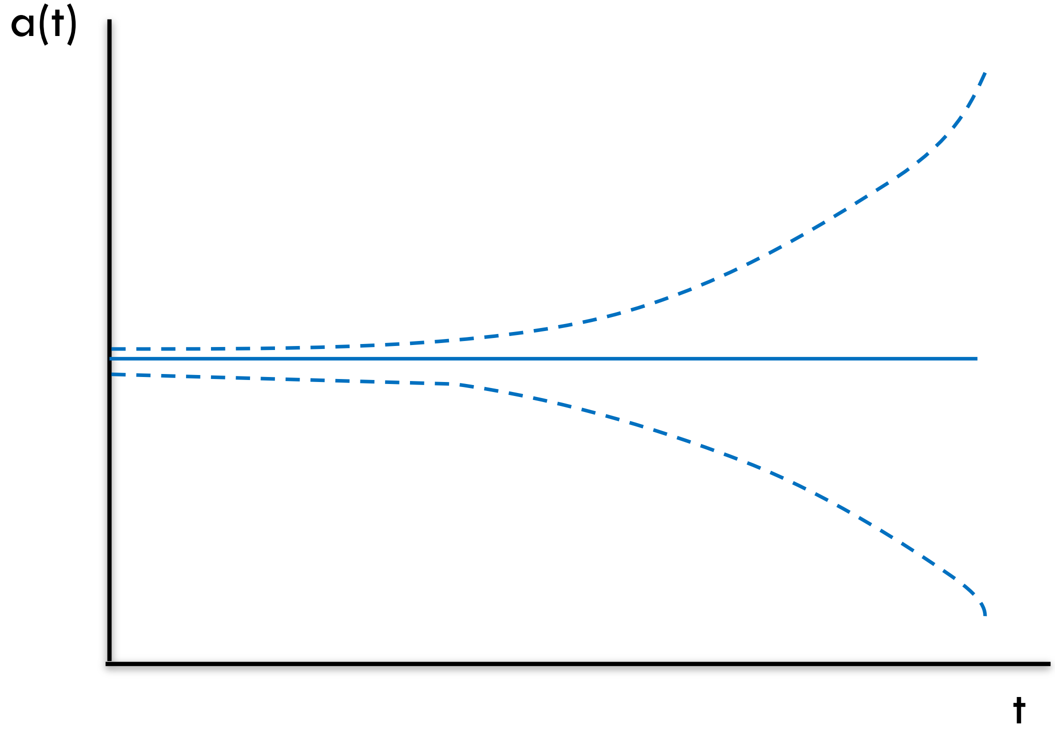
PARAMETRIC UNCERTAINTY

Varying a parameter in a dynamic model over multiple “runs” is equivalent to varying a level trajectory over the parameter



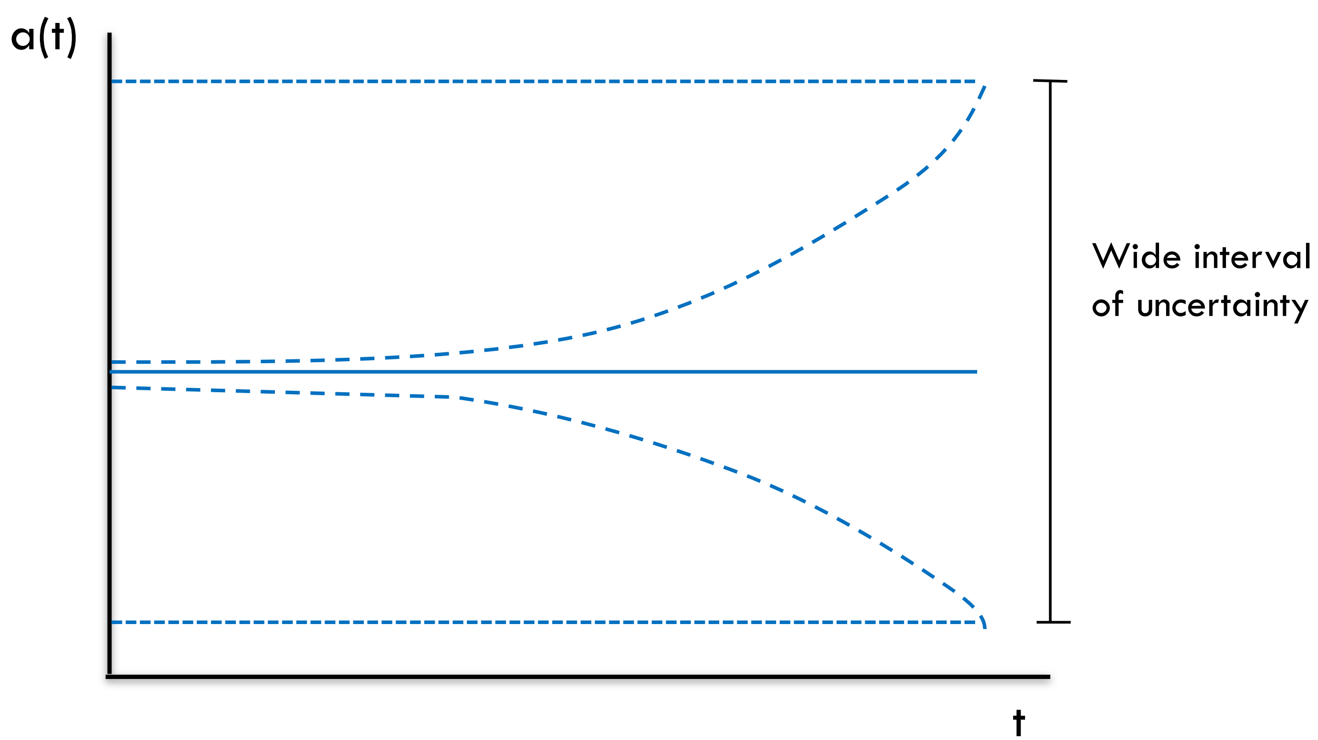
PARAMETRIC UNCERTAINTY

A parameter in a model of socio-technical system may have its own dynamics resulting in a spread that increases dramatically over time



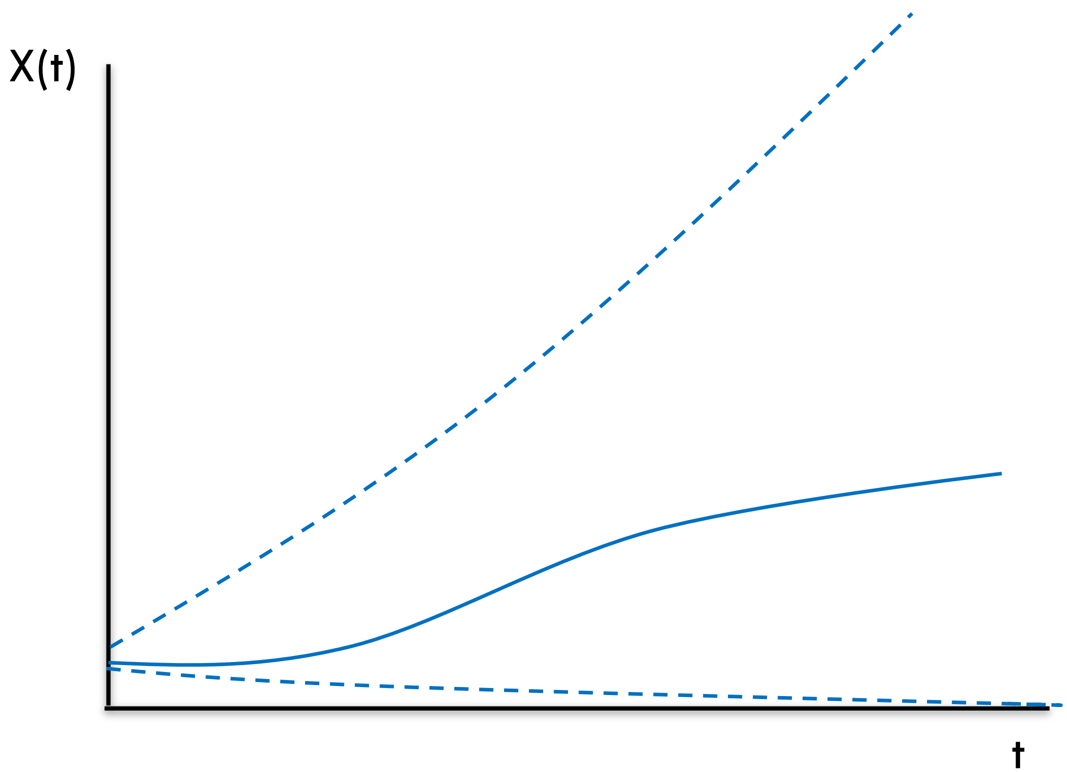
PARAMETRIC UNCERTAINTY

Given the potentially dramatic uncertainty due the parameter's own dynamics, using a interval approach could result in a very large range for parametric uncertainty



UQ APPROACH

If the interval-based UQ approach is applied to a model of a socio-technical system, the result is almost useless once we account for the variation of all of the uncertain parameters



AN ALTERNATIVE APPROACH

When the spread of possibilities is huge, it is almost meaningless

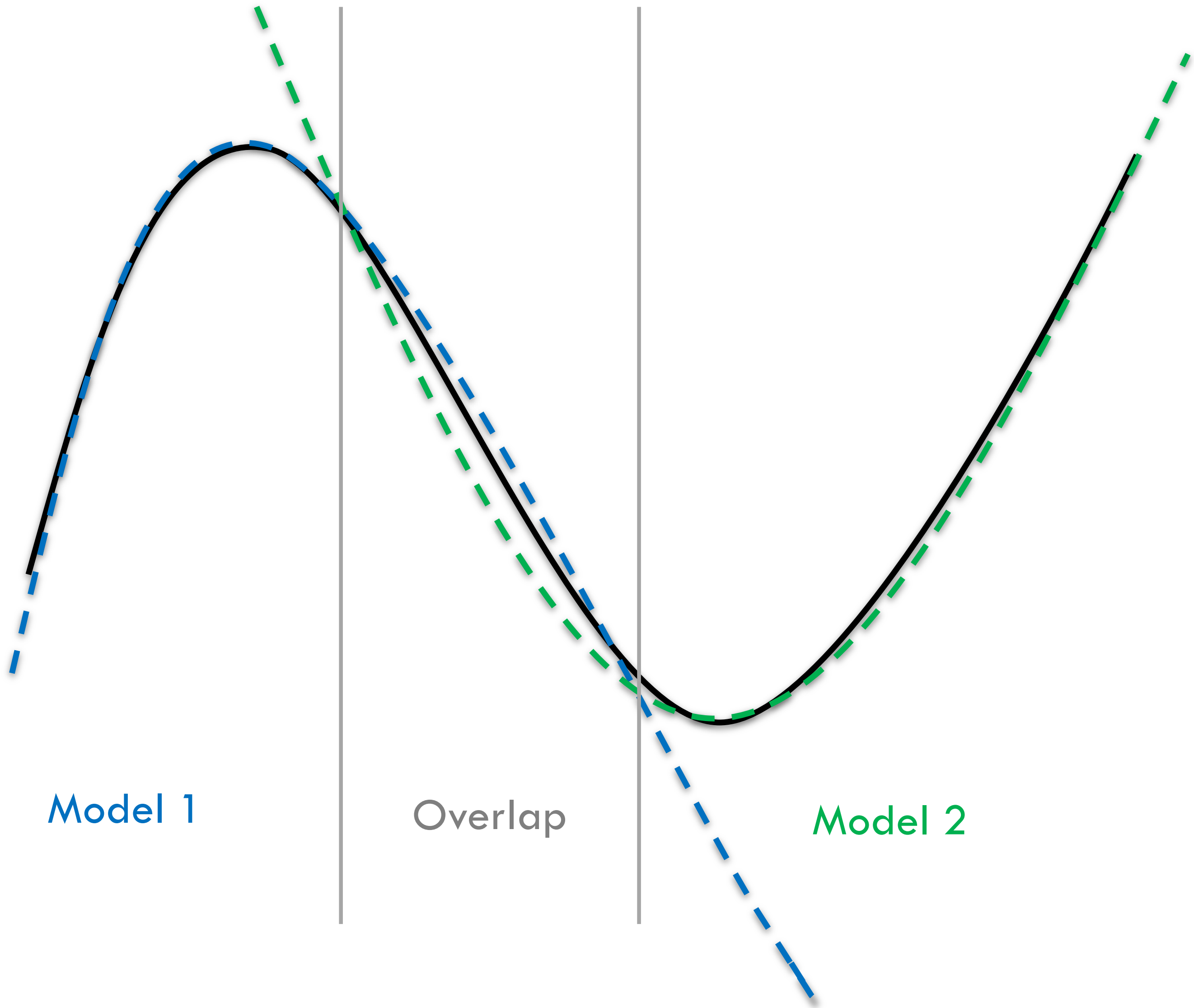
Consider the motivating example. The spread of possible economic trajectories is substantial as we look further forward in time

Yet, I suspect that you would make a very different policy choice if you expected exponential growth as opposed to capacity constrained growth

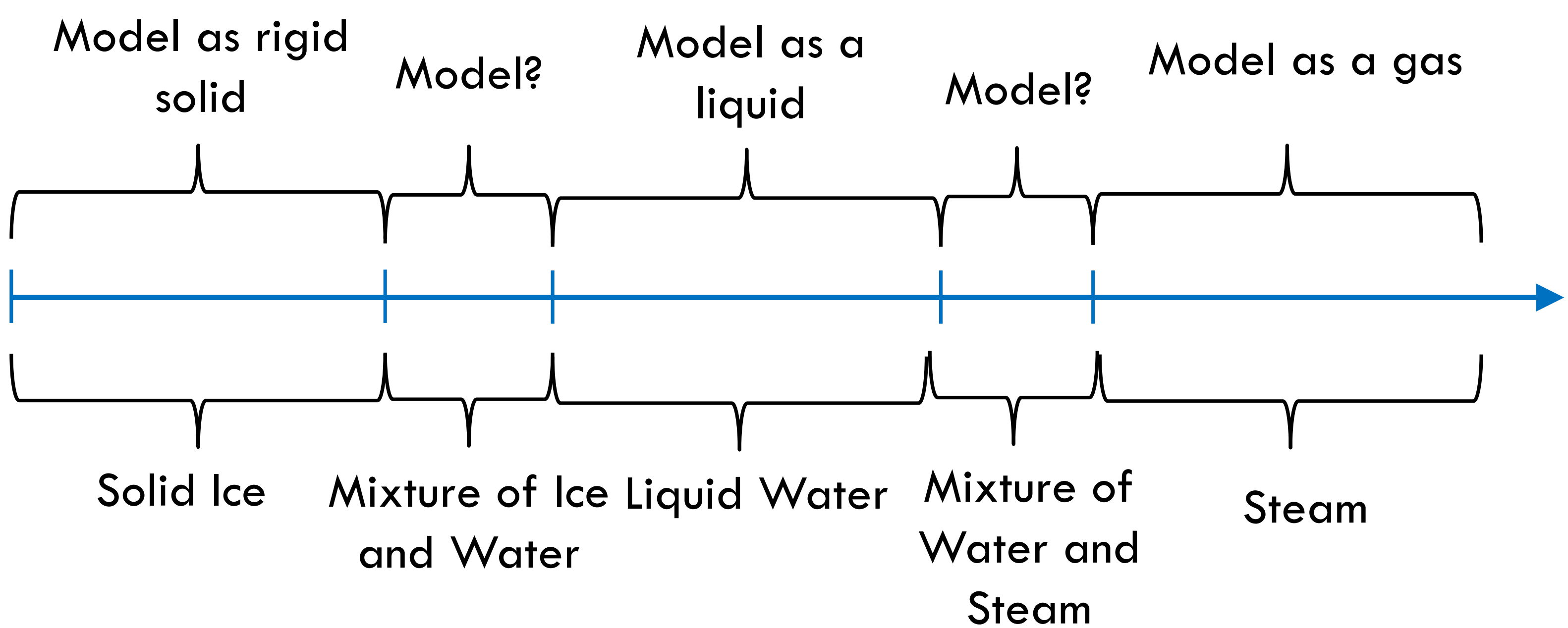
As a decision maker, it is probably more important to identify potential policy tipping points or counterintuitive results than to quantify the prediction interval

- We could characterize these as phase transitions or bifurcation points

MODEL BIFURCATIONS

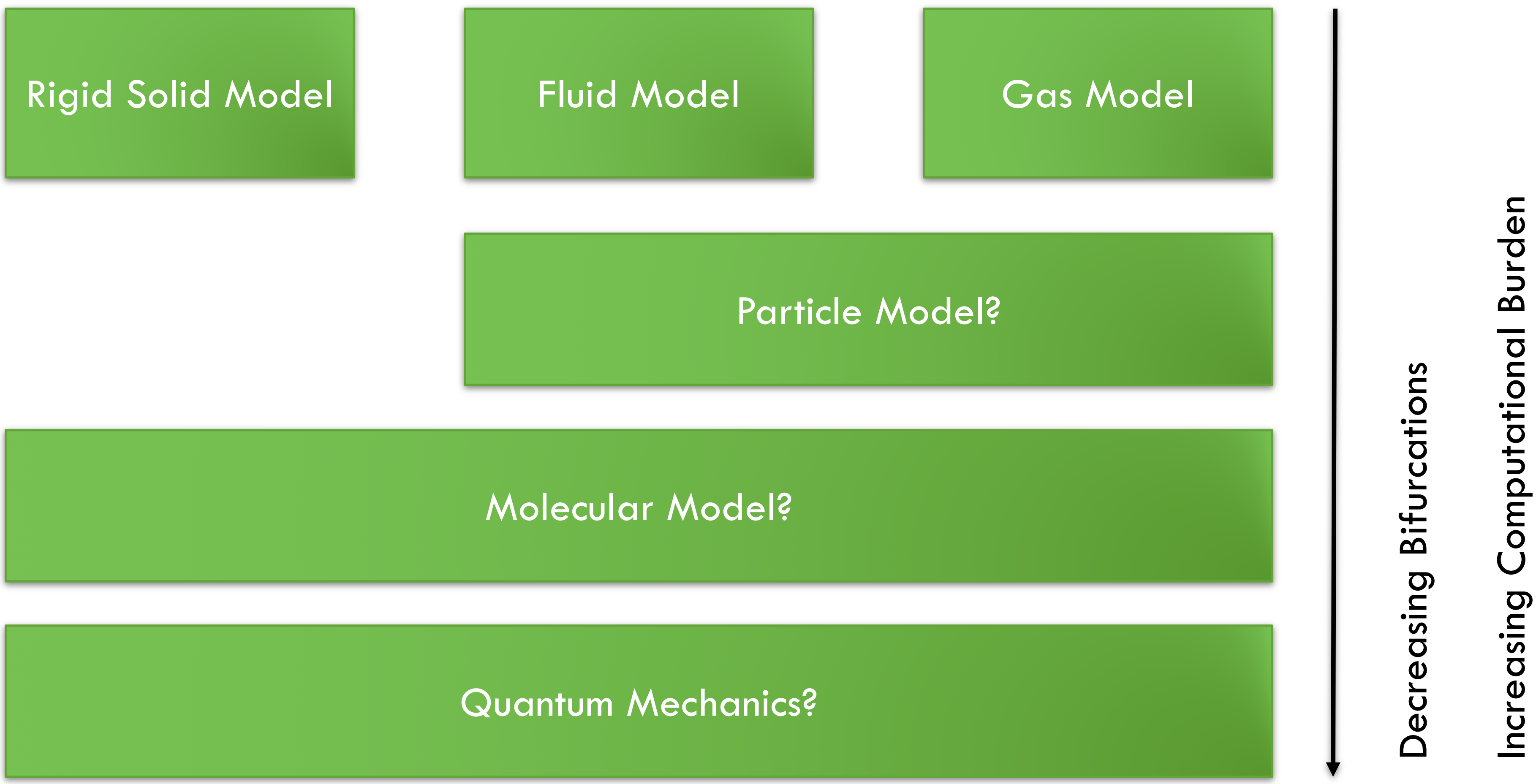


EXAMPLE: HEATING ICE



* Assume Constant Pressure

EXAMPLE: HEATING ICE



TAXONOMY OF UNCERTAINTY

Uncertainty	Model Structure	What Varies	Example
Deterministic	$y = ax + b$	N/A	Newton's Laws
Aleatory	$y = aX + b$ $X \sim \mathrm{N}(\mu, \sigma)$	X	Monte Carlo Financial Modeling
Epistemic - UQ	$y = aX + b$ $X \sim \mathrm{N}(\mu, \sigma)$	a, b, μ, σ	Robust Design Optimization
Epistemic – Phase shift	$y = \begin{cases} ax + b & t < t' \\ ax^2 + b & t \geq t' \end{cases}$	t'	Phase transition
Epistemic – Model Structure	$ax + b < y < ax^2 + b$	Model	Weather Modeling
Epistemic – Model Ontology	$y \approx ax + b$ $y \approx ct - du$	Ontology	Traffic Modeling

DECISION STRATEGY VS TYPE OF UNCERTAINTY

Strategy is driven by level of epistemic uncertainty

When epistemic uncertainty is low (the first three rows) the best strategy is to optimize

When epistemic uncertainty is high (the second three rows) it is best to adapt or hedge

- This uncertainty is due to system bifurcations

Multi-Modeling of Complex Systems

MULTI-MODELING OF COMPLEX SYSTEMS

The hallmark of complex systems is that they are difficult to capture with a single model

- Each view of the system reveals different aspects of their behavior
- Think of each model as a projection
- E.g., projecting a 3D globe onto to 2D maps
- Each projection has its pros and cons

In principle, we could rectify the different projections with the "true" model

But what makes complex systems complex is that we don't have a "true" model, we only have the projections

- Projections can vary in levels of abstraction, aggregation, and ontology

EXAMPLE: MERCATOR PROJECTION

Source: Wikipedia (https://en.wikipedia.org/wiki/Mercator_projection)

MULTI-MODELING OF COMPLEX SYSTEMS

Each of these different “projections” or views of the system was created to capture specific phenomena

These tend to be well understood

The interesting or unexpected results occur when these views intersect

We could view inconsistencies at the intersections as model bifurcations

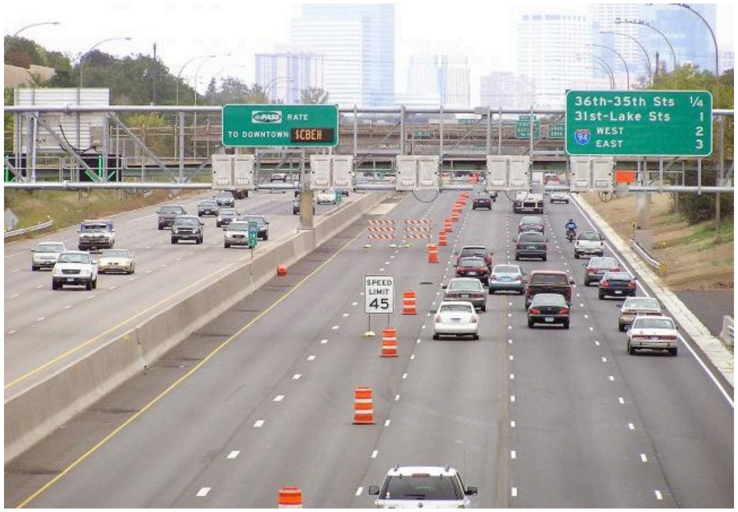
These bifurcations are potential sources of counterintuitive behaviors and policy tipping points

So one way we could use multi-models is to try to find these points as opposed to making accurate predictions

TOLLING FOR CONTROL

A congestion pricing scheme was implemented in Minneapolis in 2005

- High Occupancy Toll (HOT) lanes with dynamically changing tolls run along side non-tolled lanes
- The toll rate is based on HOT lane traffic density; adjusted upward when density increases and downward when density decreases



TOLLING FOR CONTROL

In late 2012 and early 2013, several experiments were performed to test driver response to tolling (Janson and Levinson, 2014)

Tolls were allowed to increase at lower than normal thresholds, effectively increasing toll prices; this was done without the public's knowledge

Contrary to expectations, HOT lane density increased with the higher tolls

The intuition behind the tolling strategy is reasonable

But the empirical evidence is counter-intuitive

How could this behavior been predicted *a priori*?

COMBINING TWO VIEWS

Dynamic tolling could be viewed as combining a traffic flow view from civil engineering with a driver decision model from economics

The rationale behind dynamically tolled roads seems to be the classical view

- The higher the price of the good, the more demand for that good will drop

However there are also non-classical views that capture behavior that seems to defy classical views

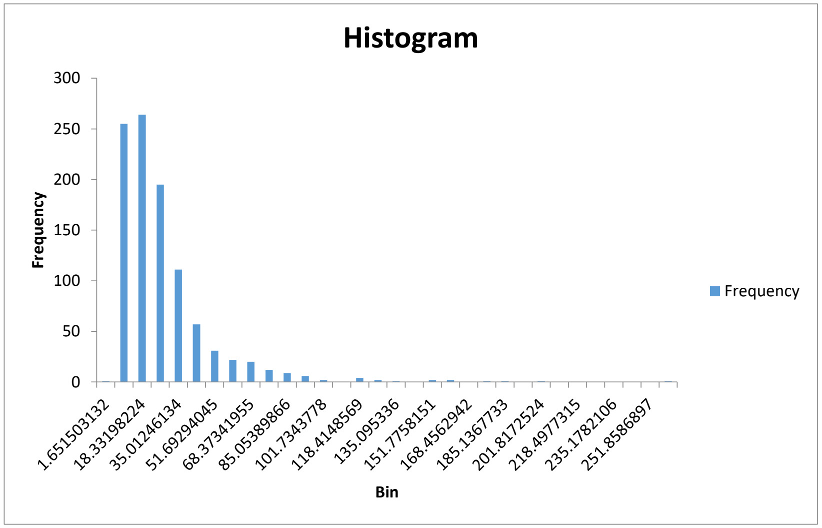
What if we tried substituting in these different views

- Prospect theory
- Information economics

MONTE CARLO SIMULATION

I created a simple Monte Carlo simulation of driver decision making

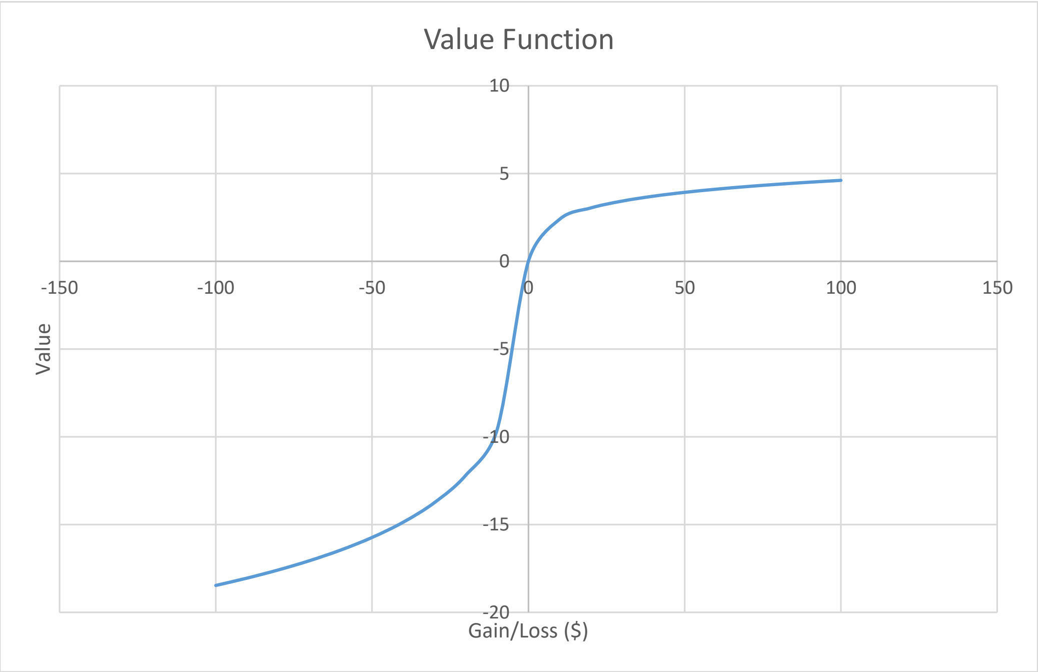
- Population of 1000 drivers
- Each was randomly assigned an hourly wage as a stand in for the time value of money
- The wage was drawn from a lognormal distribution parameterized with data from the US Bureau of Labor Statistics



ENTRY DECISION MODEL

A prospect theory value function was used to determine if each driver decided to enter the toll road for a given toll level

- The value function was intentionally exaggerated to capture any effects due to differences in gain/loss perception



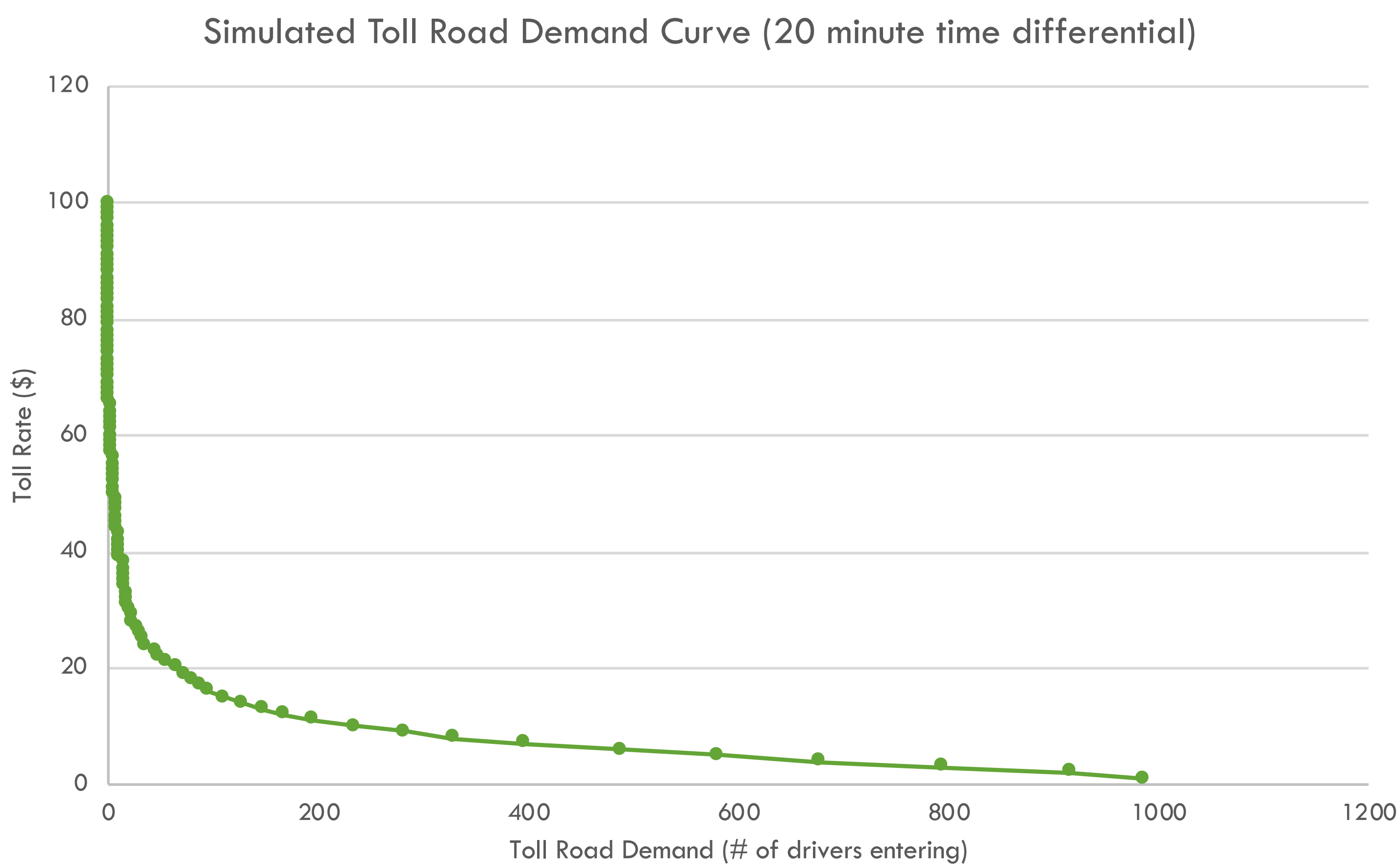
PROSPECT THEORY APPROACH

For the prospect theory view, I evaluated the entry decision for two different cases for each toll level

- Using the toll road is viewed as a gain of time
- Using the toll road is viewed as avoiding a loss of time

By varying the toll and tallying up each simulated driver's choice, I generated aggregate demand curves

PROSPECT THEORY APPROACH



INFORMATION ECONOMICS APPROACH

In retrospect, the results shouldn't be surprising

- The values generated in the two scenarios are very different, but the order of preference never changed

An alternative way is to view the toll level as a signal

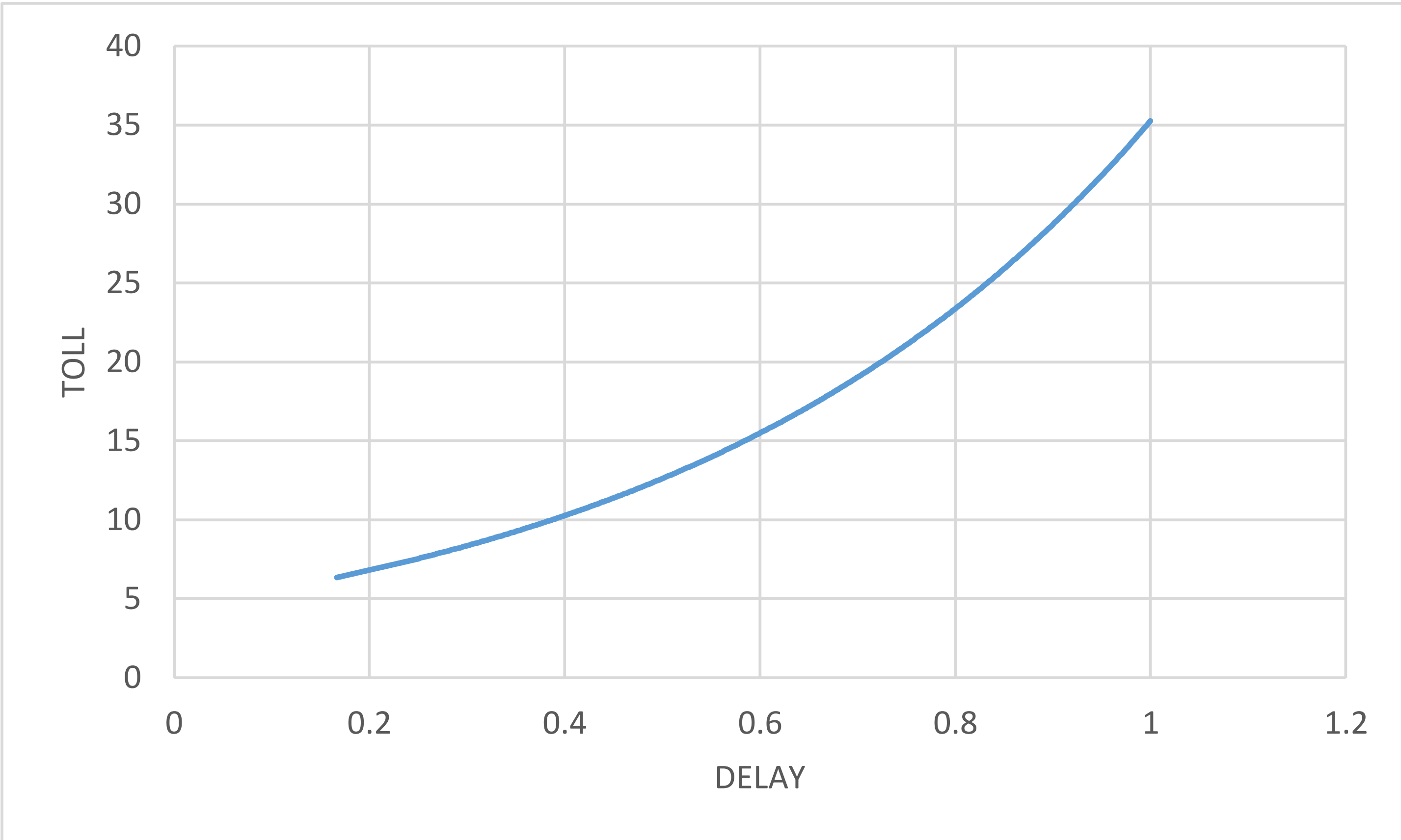
- The drivers do not know the level of congestion, but they can use the toll level as a signal
- They know the toll goes up as the usage increases

This situation falls under information economics

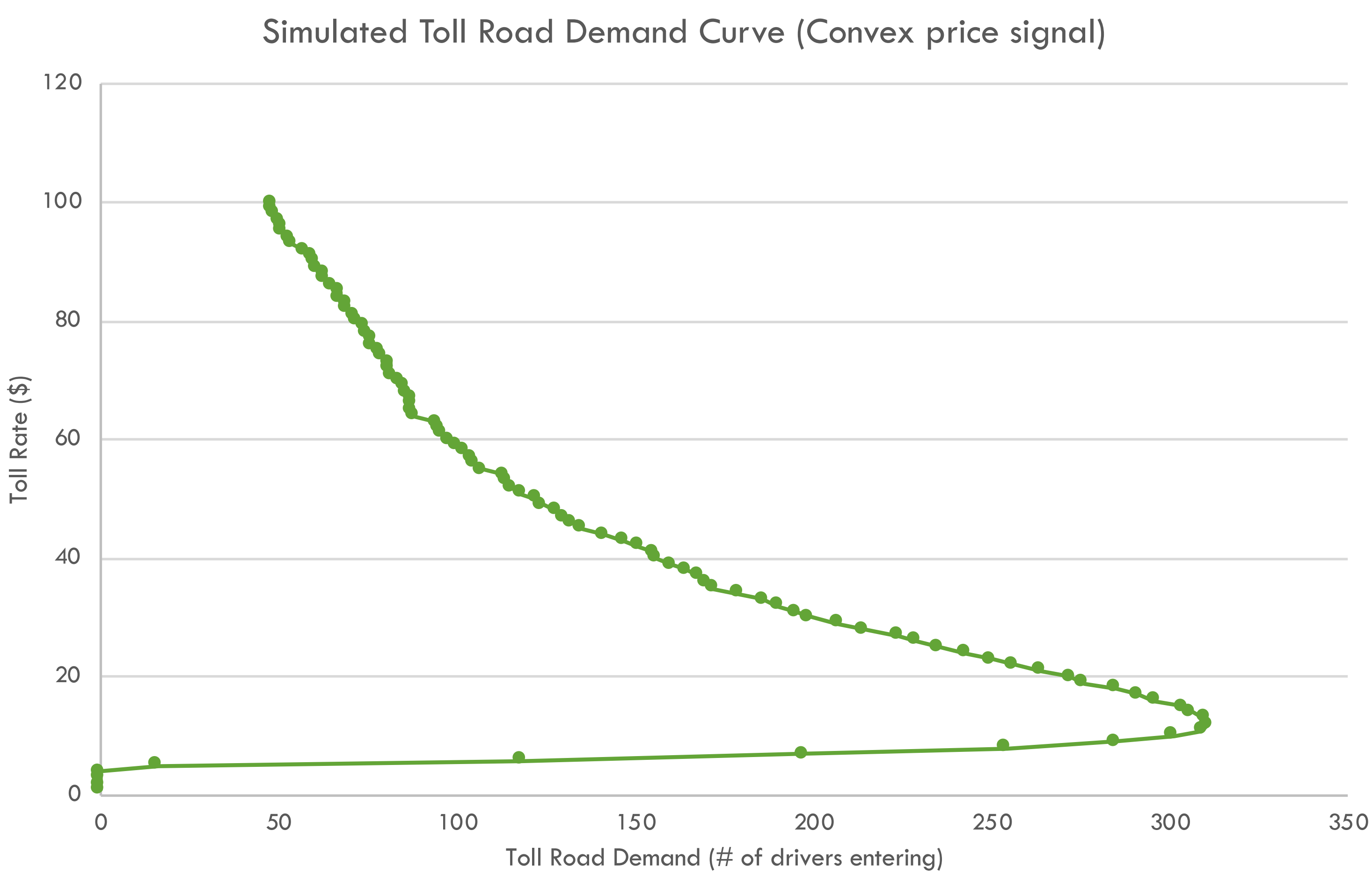
- Information economics studies market behavior under information uncertainty

I modified the model to translate the current toll level into a perceived time delay

TOLL LEVEL VS. PERCEIVED DELAY



INFORMATION ECONOMICS APPROACH



HOW DO YOU VIEW DYNAMIC TOLL ROADS NOW?

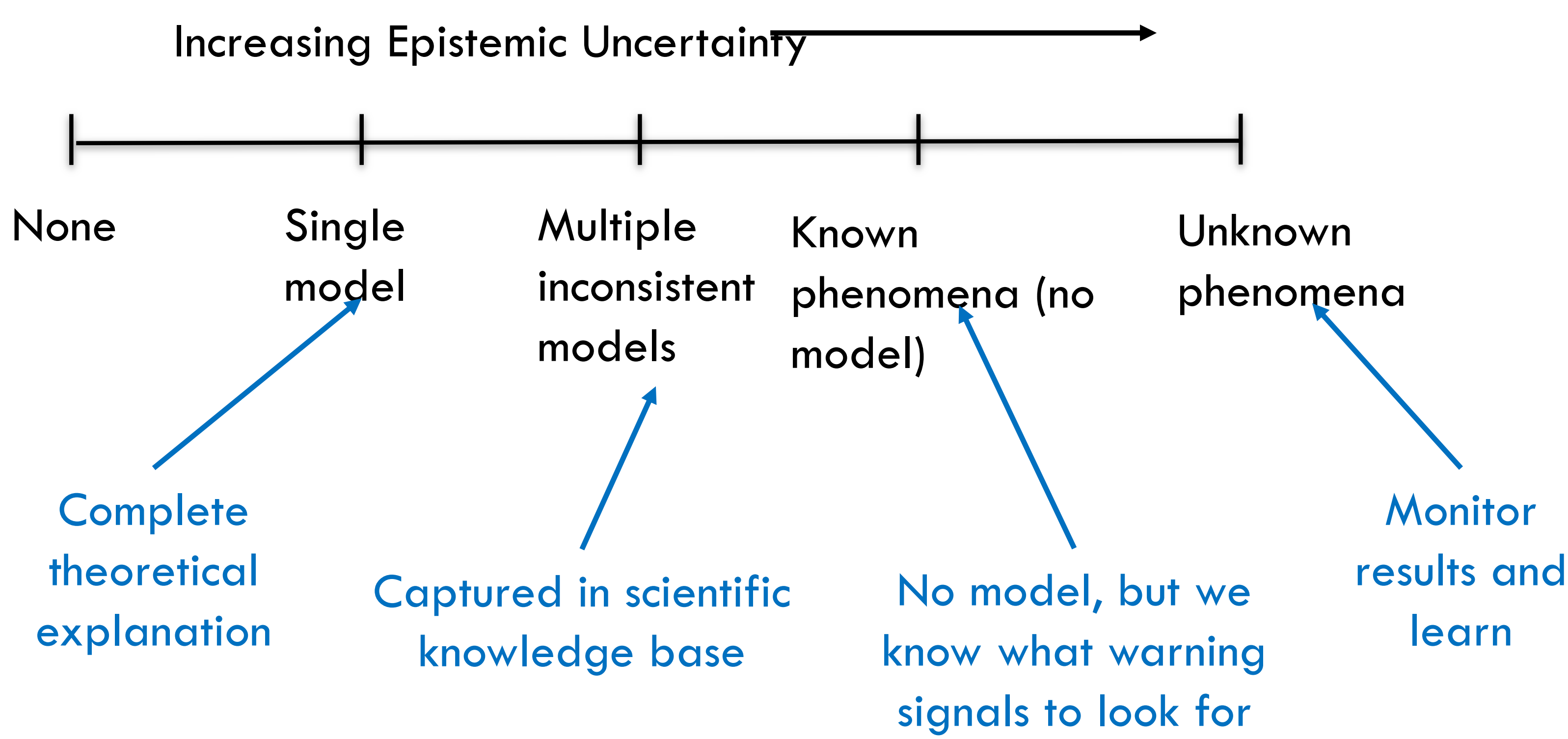
Do the results of this analysis change your views of dynamic toll roads?

What might you do differently because of this result?

Note that we did not make any specific, accurate predictions

It is the policy tipping points and the counterintuitive nature of the results that are of interest

SPECTRUM OF BIFURCATIONS



WHERE DO WE STAND?

We often worry about uncertainty out in the world, but neglect uncertainty in the models we use for engineering analysis and decision making

One simple way to account for model uncertainty is with sensitivity analysis

- We have discussed sensitivity analysis repeatedly in this class

Basic sensitivity analysis has some limitations in that it essentially treats every parameter equally and doesn't account for model structural and ontological uncertainty

Some of these shortcomings are addressed by uncertainty quantification techniques

WHERE DO WE STAND?

If you are using detailed physics-based modeling for engineering design, UQ methods should serve you well

Unfortunately, current UQ methods may not help when the level of epistemic uncertainty is extremely high due to complex and/or evolving system structure

- We talked about the risks entailed by such systems in the last lecture

When the uncertainty interval is too big, identifying policy tipping points and counterintuitive behaviors become the drivers

- These can be viewed as bifurcation points in the system behavior
- We would like to be able to buy options to hedge these bifurcation points

WHERE DO WE STAND?

Some bifurcation points can be found with a single model

- Recall the Ising model from the last lecture

But for many complex systems, the plethora of available representations are themselves indicative of bifurcation points

Exploring the interactions of different views of the system at points of intersection are another way we can identify potential bifurcation points

Once we know the existence of the these points we can consider the policy implications and adapt, hedge, or accept as appropriate