Forecasting the use, costs and benefits of HSR in the years ahead

Samer Madanat

UC Berkeley

Outline

- Demand models and ridership forecasts
- Errors in demand models and consequences
- Case study: the CA HSR
- Conclusion

The importance of ridership forecasts

- Evaluation of the feasibility of new infrastructure projects require forecasting costs and benefits
- Construction costs are typically easier to estimate than operations costs, which depend on ridership
- Benefits (reduction of congestion on freeways or at airports, reduction in GHG emissions and criteria pollutants, etc.) depend on ridership forecasts
- Revenues, and thus the financial soundness of new systems are primarily a function of ridership forecasts

Development of demand models

- Forecasting demand for existing transportation systems is performed by developing demand models and then applying them:
  1. Collect observations about current users’ mode choices and socio-economic characteristics; also collect observations of the attributes of the available transport modes (explanatory variables)
  2. Use data and statistical methods to estimate the parameters of demand models
  3. Use demand models with estimated parameters to forecast future usage, by using forecasts of explanatory variables as inputs

Errors in model forecasts

- All demand model forecasts have errors
- Sources of errors:
  - Errors in forecasting explanatory variables (income, prices, travel times, etc… unavoidable but can be reduced)
  - Modeling errors (missing explanatory variables, measurement errors… unavoidable but can be reduced)
  - Parameter estimation errors (difference between true parameters and estimated parameters… unavoidable because sample is a subset of population: statistical variance of estimate)

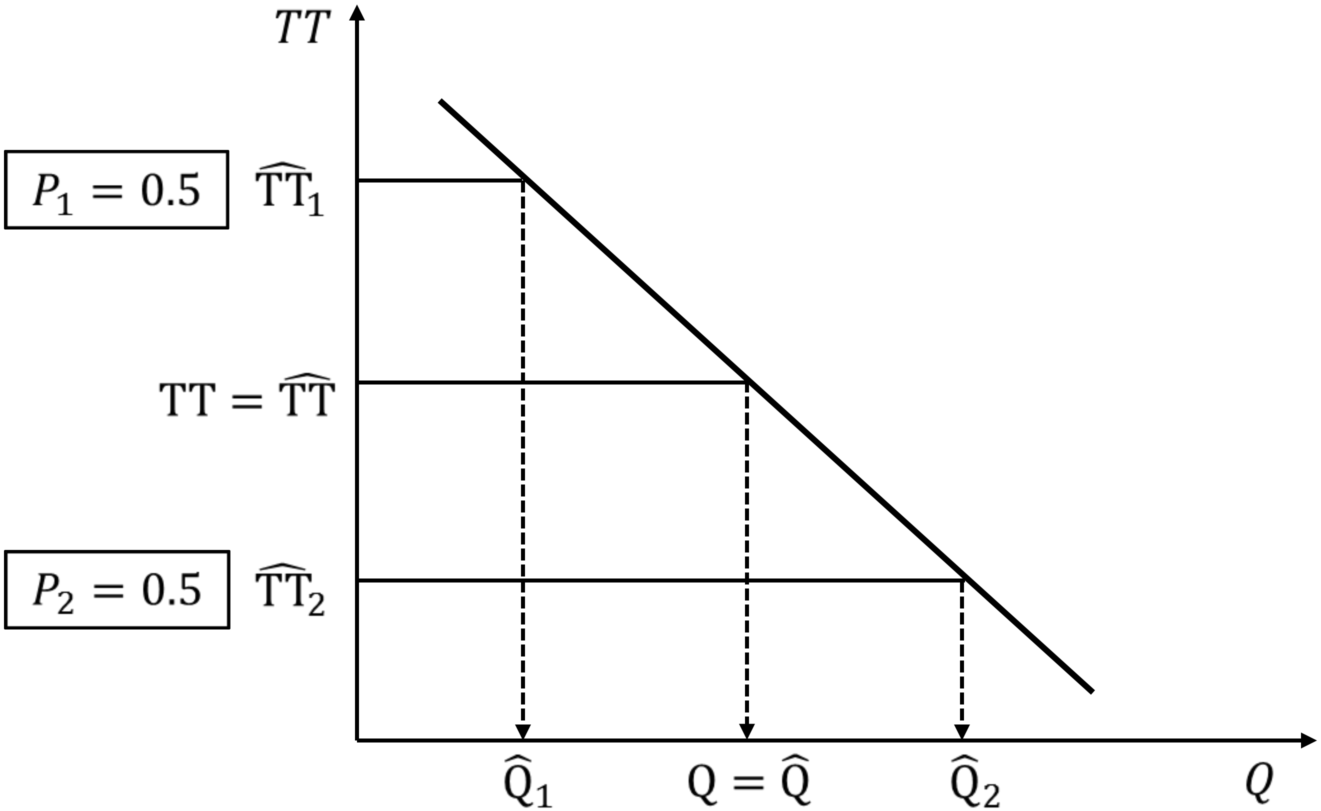
Forecast bias and variance

- Bias: the difference between the realized ridership and the <u>mean model forecast</u>
  - The mean forecast: if we repeated the model development exercise N (very large number) times, obtained N models, producing N forecasts, then took their average
- Variance: the difference between <u>our forecast</u> and the <u>mean model forecast</u>
  - Variance can <u>never</u> be eliminated because modeling error, estimation error and error in forecasting explanatory variables cannot be eliminated

An ideal world

- The explanatory-variable-forecast errors are symmetric with zero means
  - we are as likely to over-estimate the explanatory variables as we are to under-estimate them
- The modeling error is symmetric and has zero mean:
  - Specification error, missing explanatory variables, measurement errors cancel out
- The estimation error is symmetric and has zero mean:
  - Consistent parameter estimates
- Then, mean forecast of ridership is unbiased: on average, equal to realized ridership

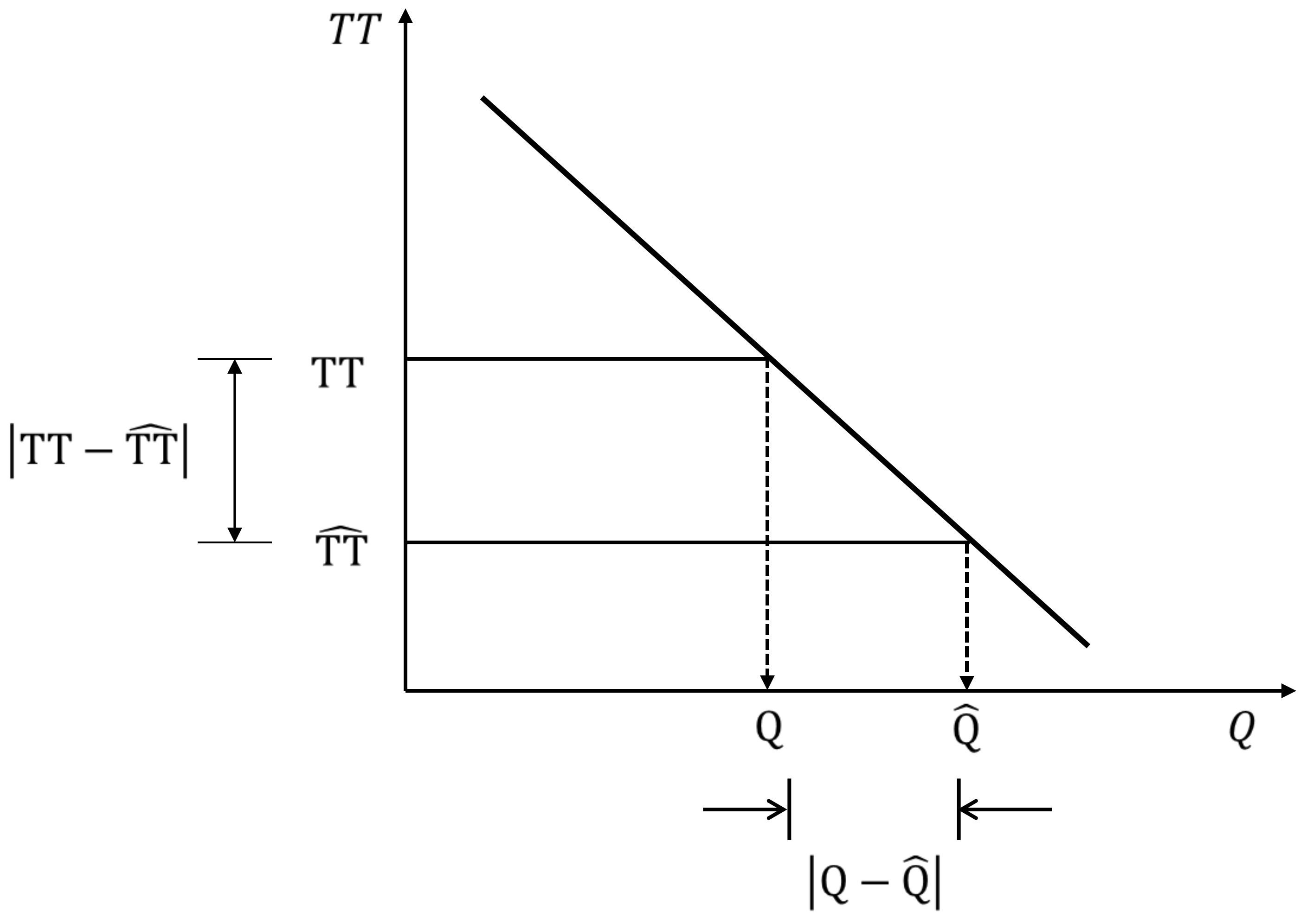
Graphically: An ideal world



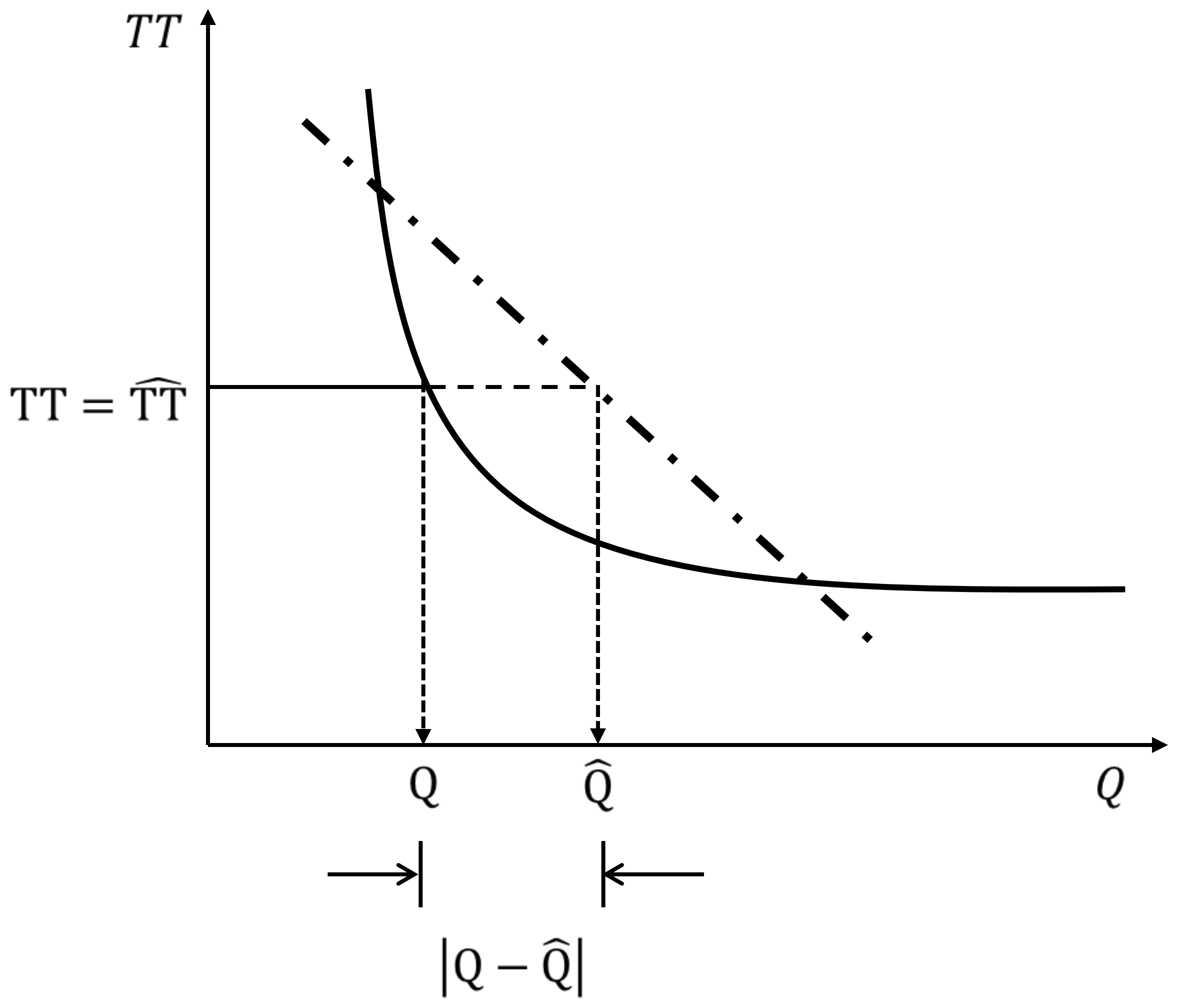
How things can go wrong…

- The explanatory-variable-forecast errors may be asymmetric with non-zero means
- The modeling error may be asymmetric (e.g., the model is mispecified)
- The estimation error may have non-zero mean (some parameter estimates are inconsistent)

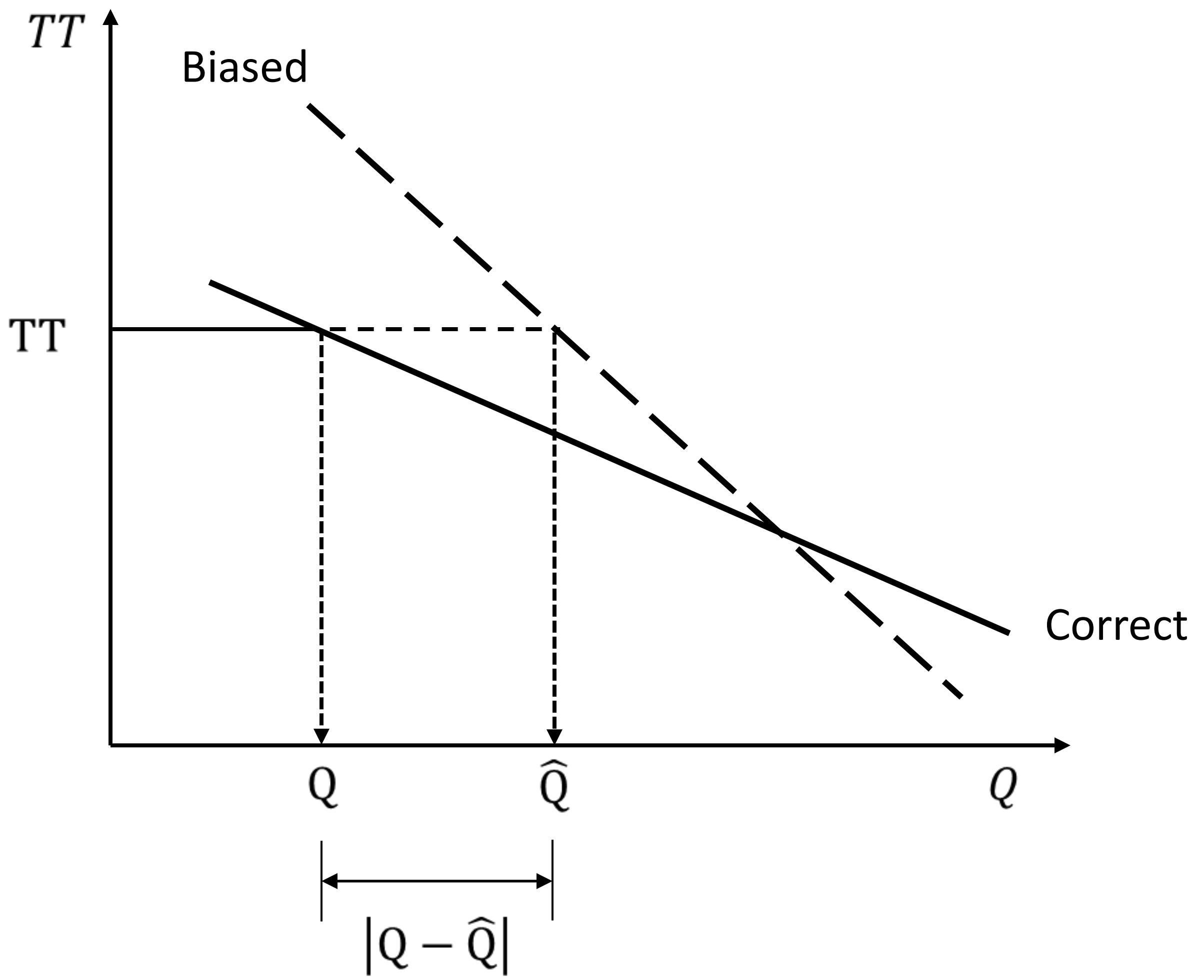
Asymmetric explanatory-variable-forecast errors



Modeling error (mispecified model)



Inconsistent parameter estimates



Case Study: the HSR travel demand models

- Objectives:
  - To apply concepts we have just covered to a real life case study
  - To review some features of the CA HSR travel demand models in light of these concepts

- Review was conducted by request of California Senate Committee on Transportation and Housing, and funded by the CA HSR Authority.

Sampling of respondents

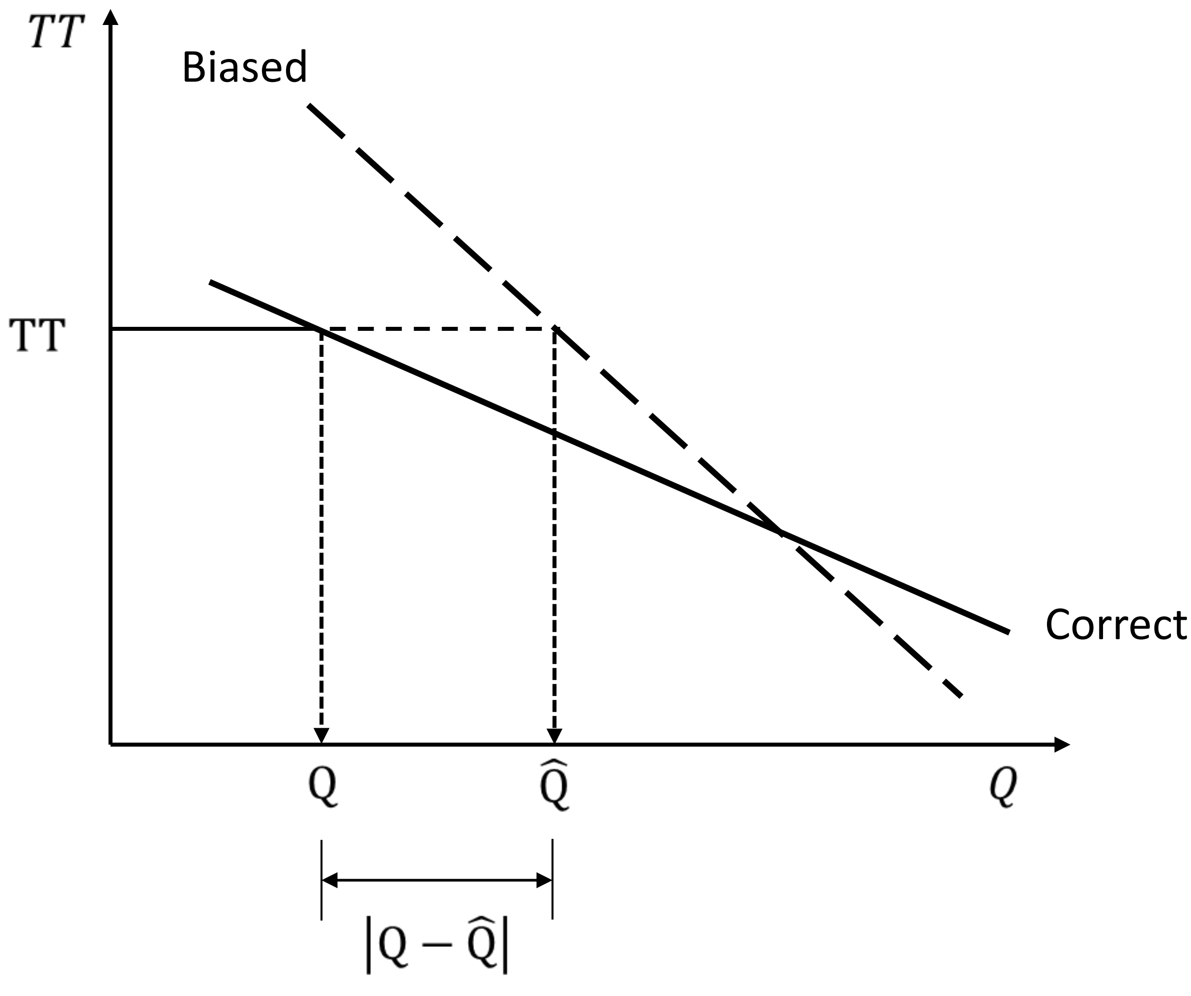
- Ideally, obtain a random sample, which is representative of the population
- But a random sample may not contain a sufficient number of users of low-share modes (e.g., rail)
- In such cases, a choice-based sample is utilized: users of low-share modes are <u>oversampled</u>
- A choice-based sample is not representative of the population, and this leads to inconsistent estimates of the parameters
- This inconsistency must be accounted for after model parameter estimation by applying a correction

Correcting for choice-based sampling bias

- For some simple models (MNL), the inconsistency is limited to a subset of parameters (the constants)
- The constants can be adjusted by calibrating the model against observed demand (from a previous year), thus eliminating the inconsistency
- For complex models (NL), all parameters (not only the constants) are inconsistent
- Here, the parameter inconsistency is not eliminated by adjusting the model parameters through "calibration"

- Effect: <u>introduce Bias in the forecast</u>

Inconsistent parameter estimates



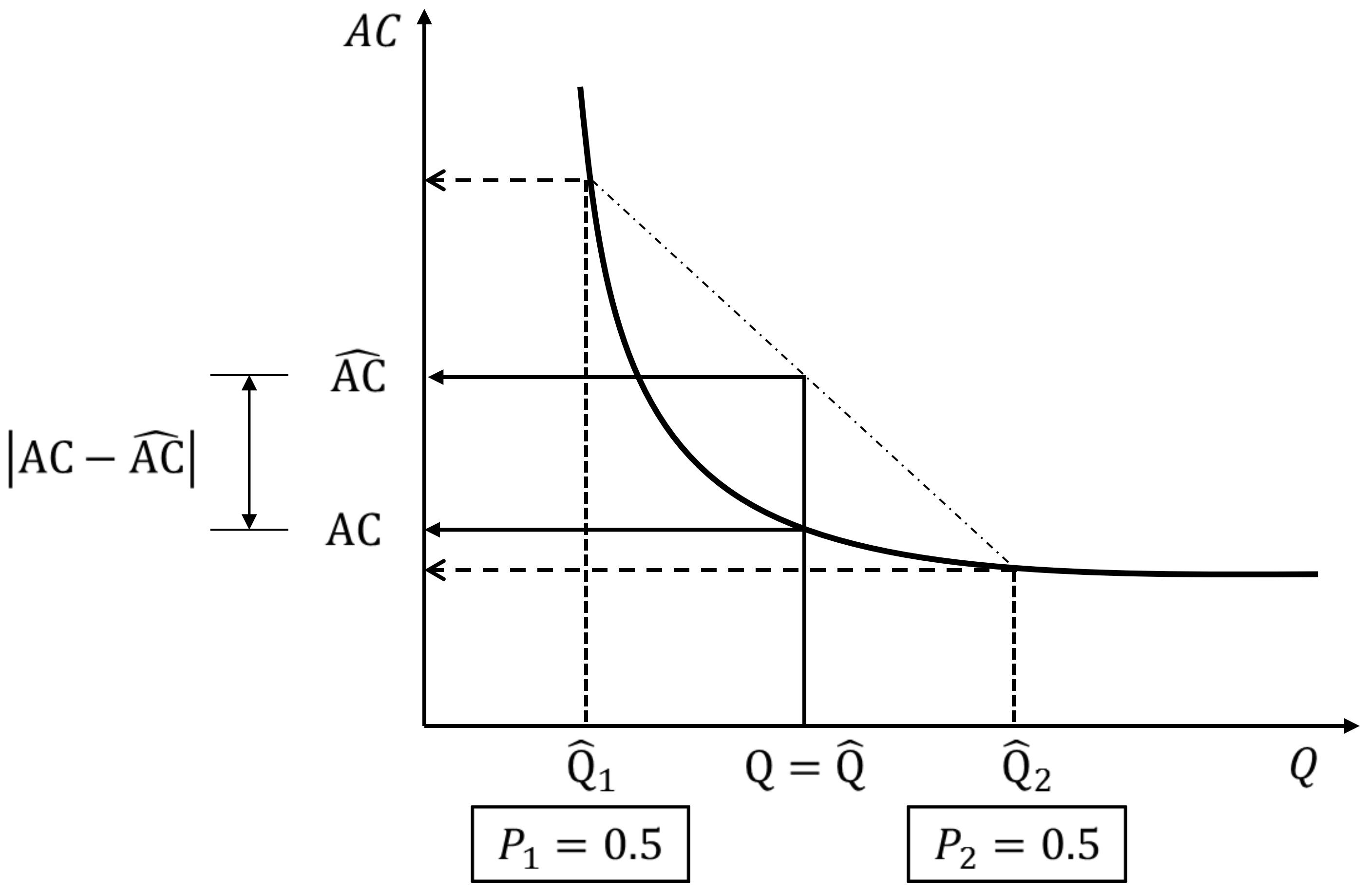
Adjustment of model parameters

- As seen, adjusting the <u>constants</u> by "calibrating" the model forecasts against observed demand is legitimate for some models (MNL).
- Not legitimate: adjusting <u>other parameters</u> from values obtained by statistical estimation to:
  - Match the a-priori expectations of modelers
  - Provide better match between the model "back-cast" and observed demand (from a previous year)
- Effect: <u>increase the Variance of the forecast</u>

So what?

- The <u>forecast variance is important</u> because benefits and costs don't necessarily vary linearly with ridership
- Example: HSR (and other public transport modes) exhibits economies of scale
- Thus, average cost decreases (non-linearly) with ridership
- Large variance leads to overestimation of average cost

Effect of Ridership Forecast Variance on Average Cost Forecast



Conclusion

- Parameters of CA HSR demand models are inconsistent, which leads to biased ridership forecasts
- Variance of ridership forecasts likely large; leads to biased forecasts of some costs and benefits (those that vary nonlinearly with ridership)
- Models should be revised to minimize bias and reduce variance of forecasts