

• Using entire dataset for estimation risks overfitting

- Using entire dataset for estimation risks overfitting
- Matching targeted moments can also overfit

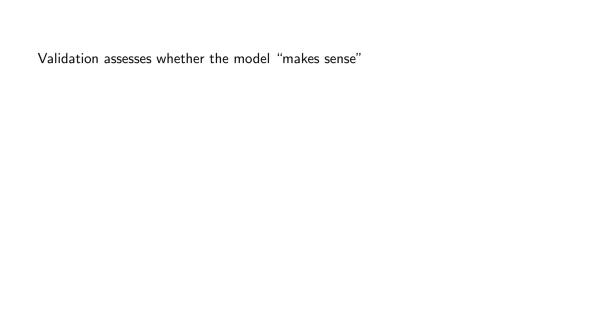
- Using entire dataset for estimation risks overfitting
- Matching targeted moments can also overfit
- Model validation demonstrates model generalizes beyond data used for estimation

- Using entire dataset for estimation risks overfitting
- Matching targeted moments can also overfit
- Model validation demonstrates model generalizes beyond data used for estimation
- i.e. for counterfactual prediction

- Using entire dataset for estimation risks overfitting
- Matching targeted moments can also overfit
- Model validation demonstrates model generalizes beyond data used for estimation
- i.e. for counterfactual prediction
- A valid model:

- Using entire dataset for estimation risks overfitting
- Matching targeted moments can also overfit
- Model validation demonstrates model generalizes beyond data used for estimation
- i.e. for counterfactual prediction
- A valid model:
  - Reproduces patterns in a "holdout sample" excluded from estimation

- Using entire dataset for estimation risks overfitting
- Matching targeted moments can also overfit
- Model validation demonstrates model generalizes beyond data used for estimation
- i.e. for counterfactual prediction
- A valid model:
  - Reproduces patterns in a "holdout sample" excluded from estimation
  - Or matches moments of estimation sample not explicitly targeted



• Qualitative: Do parameters conform to theory?

• Qualitative: Do parameters conform to theory?

• Quantitative: Is prediction error sufficiently low?

- Qualitative: Do parameters conform to theory?
- Quantitative: Is prediction error sufficiently low?
- Machine learning uses cross-validation to determine optimal complexity

- Qualitative: Do parameters conform to theory?
- Quantitative: Is prediction error sufficiently low?
- Machine learning uses cross-validation to determine optimal complexity
- Cross-validation enables automated specification search

- Qualitative: Do parameters conform to theory?
- Quantitative: Is prediction error sufficiently low?
- Machine learning uses cross-validation to determine optimal complexity
- Cross-validation enables automated specification search
- Structural estimation rarely uses automated specification search

Key question: How much unobserved heterogeneity should we specify?

• Cannot determine optimal amount ex ante

- Cannot determine optimal amount ex ante
- Typical approach: Add "enough" for good in-sample fit

- Cannot determine optimal amount ex ante
- Typical approach: Add "enough" for good in-sample fit
- Authors search for optimal number of types

- Cannot determine optimal amount ex ante
- Typical approach: Add "enough" for good in-sample fit
- Authors search for optimal number of types
- Use 20% holdout sample for out-of-sample assessment

- Cannot determine optimal amount ex ante
- Typical approach: Add "enough" for good in-sample fit
- Authors search for optimal number of types
- Use 20% holdout sample for out-of-sample assessment
- Formally compare model fit between estimation and holdout data

- Cannot determine optimal amount ex ante
- Typical approach: Add "enough" for good in-sample fit
- Authors search for optimal number of types
- Use 20% holdout sample for out-of-sample assessment
- Formally compare model fit between estimation and holdout data
- Take-away: Visual model fit often insufficient

Alternative validation: Predict treatment effects in control group

• Compare model predictions across states of the world

- Compare model predictions across states of the world
- New state: Schooling without costs

- Compare model predictions across states of the world
- New state: Schooling without costs
- Elicited preferences under both cost regimes

- Compare model predictions across states of the world
- New state: Schooling without costs
- Elicited preferences under both cost regimes
- Estimated model using only one regime

- Compare model predictions across states of the world
- New state: Schooling without costs
- Elicited preferences under both cost regimes
- Estimated model using only one regime
- Tests whether model predicts behavior in counterfactual regime

• Analyze school choice vouchers in India

- Analyze school choice vouchers in India
- Use a field experiment to validate structural model

- Analyze school choice vouchers in India
- Use a field experiment to validate structural model
- Estimate model on control villages, predict behavior in treatment villages

- Analyze school choice vouchers in India
- Use a field experiment to validate structural model
- Estimate model on control villages, predict behavior in treatment villages
- Initial model: underpredicts take-up, overpredicts school quality

- Analyze school choice vouchers in India
- Use a field experiment to validate structural model
- Estimate model on control villages, predict behavior in treatment villages
- Initial model: underpredicts take-up, overpredicts school quality
- Add new mechanisms based on prediction failures

- Analyze school choice vouchers in India
- Use a field experiment to validate structural model
- Estimate model on control villages, predict behavior in treatment villages
- Initial model: underpredicts take-up, overpredicts school quality
- Add new mechanisms based on prediction failures
- Enriched model predicts treatment outcomes much better

- Analyze school choice vouchers in India
- Use a field experiment to validate structural model
- Estimate model on control villages, predict behavior in treatment villages
- Initial model: underpredicts take-up, overpredicts school quality
- Add new mechanisms based on prediction failures
- Enriched model predicts treatment outcomes much better
- Final policy experiments use enriched model