Model validation addresses skepticism about model fit

Model validation addresses skepticism about model fit

- Using entire dataset for estimation risks overfitting

Model validation addresses skepticism about model fit

- Using entire dataset for estimation risks overfitting

- Matching targeted moments can also overfit

Model validation addresses skepticism about model fit

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

Model validation addresses skepticism about model fit

- 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

Model validation addresses skepticism about model fit

- 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:

Model validation addresses skepticism about model fit

- 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

Model validation addresses skepticism about model fit

- 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

Validation assesses whether the model "makes sense"

Validation assesses whether the model “makes sense”

- Qualitative: Do parameters conform to theory?

Validation assesses whether the model “makes sense”

- Qualitative: Do parameters conform to theory?
- Quantitative: Is prediction error sufficiently low?

Validation assesses whether the model "makes sense"

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

Validation assesses whether the model "makes sense"

- 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

Validation assesses whether the model “makes sense”

- 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

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

- Cannot determine optimal amount *ex ante*

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

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

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

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

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

- 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

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

- 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

Lang & Palacios (2018, NBER WP)

Key question: How much unobserved heterogeneity should we specify?

- 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

Delavande & Zafar (2019, JPE)

Alternative validation: Predict treatment effects in control group

Delavande & Zafar (2019, JPE)

Alternative validation: Predict treatment effects in control group

- Compare model predictions across states of the world

Delavande & Zafar (2019, JPE)

Alternative validation: Predict treatment effects in control group

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

Delavande & Zafar (2019, JPE)

Alternative validation: Predict treatment effects in control group

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

Delavande & Zafar (2019, JPE)

Alternative validation: Predict treatment effects in control group

- 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

Delavande & Zafar (2019, JPE)

Alternative validation: Predict treatment effects in control group

- 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

Arcidiacono et al. (2025, NBER WP)

Arcidiacono et al. (2025, NBER WP)

- Analyze school choice vouchers in India

Arcidiacono et al. (2025, NBER WP)

- Analyze school choice vouchers in India

- Use a field experiment to validate structural model

Arcidiacono et al. (2025, NBER WP)

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

Arcidiacono et al. (2025, NBER WP)

- 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

Arcidiacono et al. (2025, NBER WP)

- 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

Arcidiacono et al. (2025, NBER WP)

- 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

Arcidiacono et al. (2025, NBER WP)

- 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