Uncertainty in Statistical Inference – Part 1

Frank Schorfheide
University of Pennsylvania

November 4, 2024

Uncertainty in Statistical Inference

- Much of what you have seen in econometrics classess is about attaching standard errors (s.e.) to estimates.
- They can be used to construct coverage intervals, e.g.,

$$\text{point estimate } \pm 2 \times s.e. \tag{1}$$

- The s.e. are supposed to summarize uncertainty associated with your estimates.
- **Question for next two lectures:** Do they? In what sense?
- Along the way, we will cover:
  - Bayesian vs. frequentist inference
  - Interval forecast evaluation
  - Meta analysis and Bayesian hierarchical modeling.

A November 2024 JPE Paper

Inferring Trade-Offs in University Admissions: Evidence from Cambridge

Debopam Bhattacharya

University of Cambridge

Julia Shvets

University of Cambridge

How do elite universities balance diversity and academics during admissions? We develop a theory-based empirical framework to identify and quantify this potential trade-off, using postentry outcomes. We apply this to matched admission and exam-performance data from Cambridge. Comparing first- versus second-round admits from different demographic groups yields bounds on the trade-off magnitude, which (*a*) hold irrespective of whether we observe all applicant characteristics and (*b*) require no information on rejected applicants. We find robust evidence of trade-off between gender balance and future performance in math-intensive subjects but not for state/private school students or for gender in competitive nonmathematical disciplines.

TABLE 5
IDENTIFYING AND MEASURING THE TRADE-OFF: GENDER

	Pooled Males (g) (1)	DA Females (h) (2)	Lower Bound of β (3)	DA Males (g) (4)	Pooled Females (h) (5)	Upper Bound of β (6)
Non-MI subjects:						
Mean	−.04	−.07	.02	.17	−.23	.39
Standard error	.07	.03	.08	.03	.05	.06
MI subjects:						
Mean	−.06	−.30	.25	.14	−.45	.59
Standard error	.04	.04	.06	.02	.06	.07

NOTE.—"Pooled Males (g)/Pooled Females (h)" gives the average standardized performance in year-1 exams of students admitted from the pool from group g (males)/h (females). "DA Males (g)/DA Females (h)" gives the average standardized performance in year-1 exams of students directly admitted (DA) from group g (males)/group h (females). "Lower Bound of β" indicates the lower bound of the weight on diversity in the university objective function (pooled g − DA h). "Upper Bound of β" indicates the upper bound of the weight on diversity in the university objective function (DA g − pooled h). MI subjects are economics, engineering, mathematics, and physical sciences. Non-MI subjects are biological sciences, medicine, and law. Standard errors are equal to the estimate of the standard deviation of performance divided by the square root of the number of observations, for each group.

How Are S.E. Constructed in the First Place?

There is a generative model...

Ex 1. Generate N draws from a superpopulation:

$$Y_i|\theta \overset{iid}{\sim} \mathcal{N}(\theta, 1), \quad i = 1, \ldots, N \tag{2}$$

Estimate θ based on $Y_1, \ldots, Y_N$.

Ex 1. Potential outcomes, finite population:

- Potential outcomes for unit i:

$$\begin{aligned} \text{Not treated} &: \quad Y_i(0) \\ \text{Treated} &: \quad Y_i(1) \end{aligned} \tag{3}$$

- Treatment D_i is randomly assigned, e.g.,

$$D_i = \begin{cases} 0 & \text{with prob. } 1/2 \\ 1 & \text{with prob. } 1/2 \end{cases}, \quad \text{observe } Y_i(D_i). \tag{4}$$

- Estimate average treatment effect $\frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))$ based on $Y_1(D_1), \ldots, Y_N(D_N)$.

How Are S.E. Constructed in the First Place?

- Frequentist Inference:
  - pre-experimental perspective;
  - condition on "true" but unknown θ_0;
  - treat data Y as random;
  - study behavior of estimators and decision rules under repeated sampling.
- Bayesian Inference:
  - post-experimental perspective;
  - condition on observed sample Y;
  - treat parameter θ as unknown and random;
  - derive estimators and decision rules that minimize expected loss (averaging over θ) conditional on observed Y.

Frequentist Inference in Example 1

- Postulate estimator: e.g., sample average, OLS, maximum likelihood (MLE), generalized method of moments (GMM), ... Here:

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} Y_i. \tag{5}$$

- Derive sampling distribution: sums of Normals are Normal; expectation of a sum is sum of expectation; variance of a sum is sum of variances plus covariance terms

$$\hat{\theta}|\theta \sim \mathcal{N}\left(\theta, \frac{1}{N}\right). \tag{6}$$

- Thus, frequentist s.e.$=1/\sqrt{N}$, 95% freq confidence interval is

$$CI_F(Y_{1:N}) = \left[\hat{\theta} - 1.96/\sqrt{N},\ \hat{\theta} + 1.96/\sqrt{N}\right]. \tag{7}$$

Bayesian Inference

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}. \tag{8}$$

$$\mathbb{P}(\theta|\text{data}) = \frac{\mathbb{P}(\text{data}|\theta)\mathbb{P}(\theta)}{\mathbb{P}(\text{data})}. \tag{9}$$

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}, \quad p(Y) = \int p(Y|\theta)p(\theta)d\theta, \quad p(\theta|Y) \propto p(Y|\theta)p(\theta) \tag{10}$$

$\propto$ means proportionality

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{marginal data density}} \tag{11}$$

Some Terminology

- Prior Distribution: summarizes the knowledge/beliefs a researcher has about θ **before observing the sample Y.**

- Likelihood function: is the density of the data Y conditional on the parameter θ, interpreted as a function of θ, holding the data fixed. Also used in frequentist approaches to construct maximum likelihood estimators.

- Posterior distribution: is obtained after updating the prior distribution using the information contained in the sample Y. Summarizes the knowledge/beliefs a researcher has about θ **after observing the sample Y.**

Bayesian Inference for Example 1

- Prior:

$$\theta \sim U[-M, M], \quad p(\theta) = \frac{1}{2M}\mathbb{I}\left\{|\theta| < M\right\}. \tag{12}$$

- Likelihood for N observations $Y_i|\theta \stackrel{iid}{\sim} \mathcal{N}(\theta, 1)$:

$$\begin{aligned} p(Y_{1:N}|\theta) = \prod_{i=1}^{N} p(y_i|\theta) &= (2\pi)^{-N/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}(y_i - \theta)^2\right\} \\ &\propto \exp\left\{-\frac{N}{2}\left(\theta^2 - 2\hat{\theta}\theta\right)\right\}, \quad \hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} Y_i. \end{aligned} \tag{13}$$

- Posterior distribution:

$$p(\theta|Y_{1:N}) \quad \propto \quad \exp\left\{-\frac{N}{2}(\theta - \hat{\theta})^2\right\}\mathbb{I}\{-M \leq \theta \leq M\} \tag{14}$$

$$\theta|Y_{1:N} \quad \stackrel{\text{approx}}{\sim} \quad \mathcal{N}\left(\hat{\theta}, \frac{1}{N}\right). \tag{15}$$

Bayesian Inference for Example 1

- Thus, Bayesian s.e.$=1/\sqrt{N}$, 95% Bayesian credible interval is

$$CI_B(Y_{1:N}) \approx \left[\hat{\theta} - 1.96/\sqrt{N},\ \hat{\theta} + 1.96/\sqrt{N}\right]. \tag{16}$$

- **Remarks:**
  - We have two s.e. now. In this stylized example they are numerically identical, but they have a different probabilistic interpretation.
  - Also two coverage intervals...
  - **Frequentist confidence interval:**

$$\mathbb{P}^Y_\theta\{\theta \in CI_F(Y)\} = 95\%. \tag{17}$$

  - **Bayesian credible interval:**

$$\mathbb{P}^\theta_Y\{\theta \in CI_B(Y)\} = 95\%. \tag{18}$$

  - Frequentist and Bayesian intervals are not always numerically identical, but often if the model is identified and the sample information dominates the prior information.

Where Does That Leave Us?

- One can do similar calculations over and over again for different generative models.
- That's what much of the econometrics literature does...
- But, we never really find out whether our probability statements are consistent with empirical frequencies.
- **Because we never see the "true" θs.**
- We can make promises but we never really have to deliver...

Let's Look at a Different Problem First: Forecasting

- In a forecasting setting, our promises can be verified *ex post* because the actual observation becomes available eventually.

- Suppose that we want to generate forecasts for a cross-section of households or firms:

$$y_{it} = \lambda_i + \rho y_{it-1} + u_{it}, \quad u_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_i^2), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T \tag{19}$$

- Estimation periods: $t = 1, \ldots, T$.

- Forecast origin $t = T$.

- Forecast target $t = T + 1$.

Let's Look at a Different Problem First: Forecasting

- Express y_{iT+1} as function of y_{iT} and future shock(s):

$$y_{iT+1} = \lambda_i + \rho y_{iT} + u_{iT+1}, \quad u_{iT+1} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_i^2). \tag{20}$$

- Ignoring the fact that we don't know the parameters, a one-step ahead interval forecast is

$$\left[\lambda_i + \rho y_{iT} - 1.96\sigma_i,\ \lambda_i + \rho y_{iT} + 1.96\sigma_i\right]. \tag{21}$$

- Advantage of Bayesian approach: uncertainty about parameters $(\lambda_i, \rho, \sigma_i^2)$ and future shock u_{iT+1} are conceptually the same.

- For now, we skip the details of how to construct the Bayesian interval forecasts. We will look at an application.

Let's Look at a Different Problem First: Forecasting

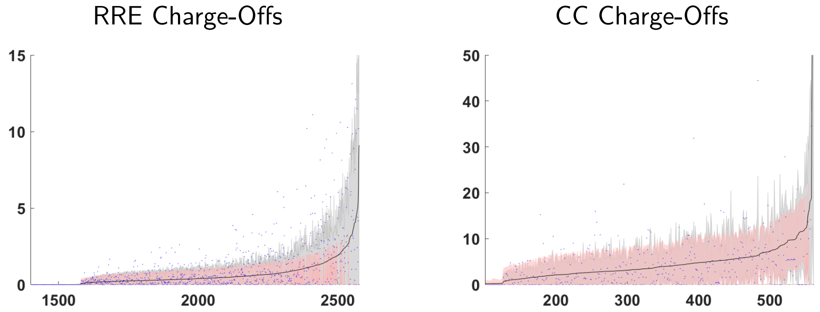
- **Reference:** Liu, Moon, and Schorfheide (2023): "Forecasting with a Panel Tobit Model," *Quantitative Economics*.
- Forecast loan charge off rates for a panel of small banks (assets ¡ 1b); many banks N, few time periods T.
- There is some censoring at zero (charge off rates are weakly positive):

$$y_{it} = \max\{y_{it}^*, 0\}, \quad y_{it}^* = \lambda_i + \rho y_{it}^* + \dots \tag{22}$$

- It can be shown, that if the interval forecasts are valid $1-\alpha$ credible sets, then the empirical coverage frequency across $i = 1, \dots, N$ should be close to $1-\alpha$.
- $\Longrightarrow$ **One can verify if promise is kept!**

Set Forecast Performance

		Cover.	AveLen	Fraction of Sets w Form $\{0\}$	$[0, b]$	$\{0\} \cup [a, b]$
RRE	Target Avg Cover.	0.88	0.31	0.68	0.28	0.04
	Target Ptwise Cover.	0.94	0.75	0.61	0.36	0.03
CC	Target Avg Cover.	0.91	6.48	0.02	0.81	0.17
	Target Ptwise Cover.	0.91	7.74	0.19	0.56	0.25

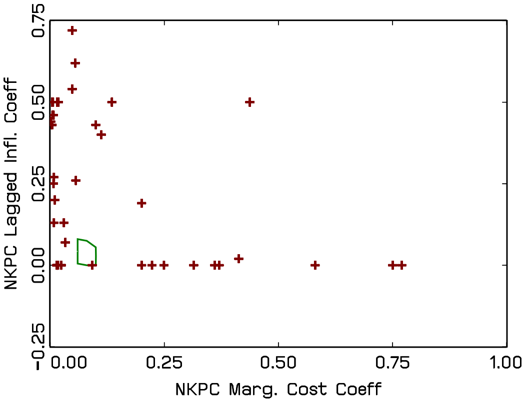


Solid black line: posterior mean forecasts. Blue dots: actuals. Grey/pink set: targeting pointwise/average coverage. Nominal coverage probability: 90%.

Can We Use a Similar Approach to Assess S.E. of Parameter Estimates?

- In economics and other social sciences we often re-estimate parameters that have been estimated before.
- Does the uncertainty measure of the first study provide a reliable prediction for the estimate obtained in a subsequent study? (External validity)
- Reasons for why or why not:
  - parameters tend not to exist outside of models;
  - subsequent studies use a different model;
  - subsequent studies uses different data / variable definitions;
  - (...)
- We need a model that describes how parameters are connected across studies.
  $\Longrightarrow$ Related to the field of meta analysis.

NK Phillips Curve Parameters (Slope and Degree of Backward Looking)



Reference: Parameter estimates are from Schorfheide (2008): "DSGE Model-Based Estimation of the New Keynesian Phillips Curve," *FRB Richmond Quarterly*; figure is from Schorfheide (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," *Advances in Economics and Econometrics*.

What Is Meta Analysis?

Meta Analysis:

- Synthesize parameter estimates across various studies.
- Key ingredient: a set of assumptions that link the parameters estimated in different studies to a meta parameter that is then inferred from estimates reported in the literature.
- Widely used outside of economics.

Some References:

- Sutton and Abrams (2001): "Bayesian Methods in Meta-Analysis and Evidence Synthesis," *Statistical Methods in Medical Research*.
- Meager (2019): "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics.*
- (...)

Grouping Empirical Studies

- $N \cdot (J+1)$ studies that try to estimate a similar parameter.
- Use ij subscripts.
- $i = 1, \ldots, N$ groups of studies.
- Baseline studies (for group i): $j = 1, \ldots, J$.
- Validation study (for group i): $j = J + 1$.
- Notation similar to panel forecasting problem: previously i was bank, t was time period, T periods to for estimation, forecast outcome in $T + 1$.

Linking Parameters Across Studies

- **Bayesian** hierarchical modeling assumption

$$\theta_{ij}|(\tau_i, \nu_i) \overset{iid}{\sim} \mathcal{N}(\tau_i, \nu). \tag{23}$$

- Parameters are highly correlated but not identical.
- Parameters in each group share common mean τ_i.
- If $\nu = 0$ parameters in each group are identical.
- **Note:** parameters θ_{ij} are random variables $\Longrightarrow$ Bayesian approach.

Parameter Estimates and S.E.

- Study ij reports point estimate $\hat{\theta}^{ij}$ and s.e. σ_{ij}.
- Give estimate a frequentist interpretation to obtain quasi likelihood:

$$\hat{\theta}_{ij}|\theta_{ij} \sim \mathcal{N}\big(\theta_{ij}, \sigma_{ij}^2\big). \tag{24}$$

- Some hand-waving: we typically only have a consistent estimator, $\hat{\sigma}_{ij}^2$, of σ_{ij}^2.

Next Steps

- Construct a posterior distribution for τ_i.
- Need to deal with the unknown ν.
- Generate interval prediction for $\hat{\theta}_{iJ+1}$.

Likelihood Function for $\hat{\theta}_{i,1:J}$

- Suppose $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, X, Y are independent, and $Z = X + Y$. Then

$$
\begin{aligned}
p(z|y, \mu_x, \sigma_x^2) &= p_{\mathcal{N}}(z - y; \mu_x, \sigma_x^2), && (25) \\
p(z|\mu_x, \sigma_x^2, \mu_y, \sigma_y^2) &= \int p_{\mathcal{N}}(z - y; \mu_x, \sigma_x^2) p_{\mathcal{N}}(y; \mu_y, \sigma_y^2) dy. && (26)
\end{aligned}
$$

- Looks complicated: but folks before us have shown that the sum of Normals is Normally distributed:

$$
Z \sim \mathcal{N}\big(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\big). \quad (27)
$$

- **Likelihood function:**

$$
\begin{aligned}
\hat{\theta}_{ij}|(\tau_i, \nu_i) &\stackrel{iid}{\sim} \mathcal{N}\big(\tau_i, \nu + \sigma_{ij}^2\big) && (28) \\
p\big(\hat{\theta}_{i,1:J}|\tau_i, \nu_i\big) &\propto \prod_{j=1}^{J} (\nu_i + \sigma_{ij}^2)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{J} \frac{(\hat{\theta}_{ij} - \tau_i)^2}{\nu_i + \sigma_{ij}^2} \right\}. && (29)
\end{aligned}
$$

Outlook

- We will continue this in the next lecture...
- Thus far, we discussed
  - Bayesian vs. frequentist inference
  - Interval forecast evaluation
  - Meta analysis and Bayesian hierarchical modeling.