Deep IV in Law
Towards Automated Impact Analysis of Judicial Precedents

Daniel L. Chen

December 2019
US Federal Courts as Natural Laboratory

- Random assignment of judges
  - Judge characteristics predict decisions
- Binding precedent within circuit
  - 98% of decisions are final
US Federal Courts as Natural Laboratory

- **Random** assignment of judges
  - Judge characteristics predict decisions
- Binding precedent within circuit
  - 98% of decisions are final
US Federal Courts as Natural Laboratory

- **Random** assignment of judges
  - Judge characteristics **predict** decisions
- Binding precedent **within** circuit
  - 98% of decisions are final
Random assignment of judges

- Judge characteristics predict decisions

Binding precedent within circuit

- 98% of decisions are final
Random assignment of judges

- Judge characteristics predict decisions

Binding precedent within circuit

- 98% of decisions are final
High-stakes common-law space

Introduce theories:

- **Contract** duty posits a general obligation to keep promises vs. a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law:** duty of care assigned to minimize costs (“least cost avoider” theory)

Shift in standards or thresholds:

- Shift from reasonable person standard to reasonable woman standard for what constitutes sexual harassment.

- Waive need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit allowed Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

What are causal effects of rulings like these?
High-stakes common-law space

Introduce theories:

- **Contract** duty posits a general obligation to keep promises vs. a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law:** duty of care assigned to minimize costs ("least cost avoider" theory)

Shift in standards or thresholds:

- Shift from reasonable person standard to reasonable woman standard for what constitutes sexual harassment.

- Waive need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit allowed Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

What are causal effects of rulings like these?
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.
  
a party should be allowed to breach a contract and pay damages, if it’s more **economically efficient** than performing (i.e., **efficient breach theory**) (Posner 7th Cir. 1985)

- **Tort law**: **duty of care** assigned to minimize costs (**“least cost avoider” theory**)

Shift in standards or thresholds:

- **Shift** from **reasonable person** standard to **reasonable woman** standard for what constitutes sexual harassment.

- **Waive** need to prove **emotional harm** in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law **requiring abortion clinics to meet building standards of ambulatory surgery centers**. (would reduce to < 10 clinics)

*What are causal effects of rulings like these?*
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.
  a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law: duty of care** assigned to minimize costs (“least cost avoider’ theory)

Shift in standards or thresholds:

- Shift from **reasonable person** standard to **reasonable woman** standard for what constitutes sexual harassment.

- **Waive** need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

*What are causal effects of rulings like these?*
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.

- A party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law: duty of care** assigned to minimize costs ("least cost avoider" theory)

Shift in standards or thresholds:

- Shift from reasonable person standard to reasonable woman standard for what constitutes sexual harassment.

- Waive need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

**What are causal effects of rulings like these?**
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.
  a party should be allowed to breach a contract and pay damages, if it’s more 
  economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law: duty of care** assigned to minimize costs (“least cost avoider’ theory)

Shift in standards or thresholds:

- Shift from *reasonable person* standard to *reasonable woman* standard for what 
  constitutes sexual harassment.

- Waive need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law requiring abortion clinics to meet building standards 
  of ambulatory surgery centers. (would reduce to < 10 clinics)

**What are causal effects of rulings like these?**
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.
  - a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law: duty of care** assigned to minimize costs (“least cost avoider” theory)

Shift in standards or thresholds:

- **Shift** from reasonable person standard to reasonable woman standard for what constitutes sexual harassment.

- **Waive** need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

**What are causal effects of rulings like these?**
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.
  - a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law: duty of care** assigned to minimize costs (“least cost avoider” theory)

Shift in standards or thresholds:

- Shift from **reasonable person** standard to **reasonable woman** standard for what constitutes sexual harassment.

- **Waive** need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

What are causal effects of rulings like these?
High-stakes common-law space

Introduce theories:

- Contract duty posits a general obligation to keep promises vs. a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- Tort law: duty of care assigned to minimize costs (‘least cost avoider’ theory)

Shift in standards or thresholds:

- Shift from reasonable person standard to reasonable woman standard for what constitutes sexual harassment.

- Waive need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit allowed Texas law requiring abortion clinics to meet building standards of ambulatory surgery centers. (would reduce to < 10 clinics)

What are causal effects of rulings like these?
High-stakes common-law space

Introduce theories:

- **Contract duty posits a general obligation to keep promises** vs.
  a party should be allowed to breach a contract and pay damages, if it’s more economically efficient than performing (i.e., efficient breach theory) (Posner 7th Cir. 1985)

- **Tort law: duty of care** assigned to minimize costs ("least cost avoider" theory)

Shift in standards or thresholds:

- Shift from **reasonable person** standard to **reasonable woman** standard for what constitutes sexual harassment.

- **Waive** need to prove emotional harm in court by plaintiff (to a jury).

Rule on states’ laws:

- 5th Circuit **allowed** Texas law **requiring abortion clinics to meet building standards of ambulatory surgery centers**. (would reduce to < 10 clinics)

**What are causal effects of rulings like these?**
Graphical Intuition of “coin flip”

Figure 3: Judicial Composition and Random Assignment, 1971-2004

- **Expected # Democratic appointees per seat**
- **Actual # Democrat appointees per seat**

![Graph showing judicial composition and random assignment from 1965 to 2010](image-url)
Biographies Predict Votes

What Matters, Chen, Cui, Shang, Zheng, NeurIPS-MLaw 2016

Minority religion judges prefer separate church and state

\[
\begin{align*}
\text{Law}_{ict} &= \alpha_{ict} + \phi Z_{ct} + \gamma_1 X_{ict} + \gamma_2 W_{ct} + \eta_{ict} \quad \text{(machine learning step)} \\
\gamma_{ict} &= \alpha_{ict} + \rho \text{Law}_{ct} + \beta_1 X_{ict} + \beta_2 W_{ct} + \epsilon_{ict} \quad \text{(causal inference step)}
\end{align*}
\]

Sparse Models and Methods for Optimal Instruments, Belloni, Chen, Chernozhukov, Hansen, ECMA 2012
Biographies Predict Votes

What Matters, Chen, Cui, Shang, Zheng, NeurIPS-MLaw 2016

Pro Religious Establishment Clause and Composition of Judicial Panels

% of Establishment Clause Cases favoring Church-State Separation

Minority religion judges prefer separate church and state

\[
\begin{align*}
L_{aw_{ct}} &= \alpha_{ict} + \phi Z_{ct} + \gamma_1 X_{ict} + \gamma_2 W_{ct} + \eta_{ict} \quad \text{(machine learning step)} \\
Y_{ict} &= \alpha_{ict} + \rho L_{aw_{ct}} + \beta_1 X_{ict} + \beta_2 W_{ct} + \varepsilon_{ict} \quad \text{(causal inference step)}
\end{align*}
\]

Sparse Models and Methods for Optimal Instruments, Belloni, Chen, Chernozhukov, Hansen, ECMA 2012
Biographies Predict Votes

What Matters, Chen, Cui, Shang, Zheng, NeurIPS-MLaw 2016

Minority religion judges prefer separate church and state

\[
\begin{align*}
\text{Law}_{ct} &= \alpha_{ict} + \phi Z_{ct} + \gamma_1 X_{ict} + \gamma_2 W_{ct} + \eta_{ict} \quad \text{(machine learning step)} \\
\text{Y}_{ict} &= \alpha_{ict} + \rho \text{Law}_{ct} + \beta_1 X_{ict} + \beta_2 W_{ct} + \varepsilon_{ict} \quad \text{(causal inference step)}
\end{align*}
\]

Sparse Models and Methods for Optimal Instruments, Belloni, Chen, Chernozhukov, Hansen, ECMA 2012
Impact of Environmental Decisions on Pollution

Calibration plot for cross-validated prediction

Rulings in favor of EPA regulations reduce air pollution
Graphical Model

- District Cases →
- District Judge Bio →
- **Circuit Case Appeal** →
- **Circuit Judge Characteristics** (\(z\)) →
- **Circuit Case Decision** (\(w\)) →
- Precedential Effects (e.g., State Laws) →
- Promulgation (e.g., News) →
- Outcomes (\(y\))

Data: All 380K cases, 1,150K judge votes, 94 topics, from 1870s-700M tokens, 2B 8-grams, 5M citation edges across cases
- 250 biographical features (D/R, law school, age)
- 5% sample, 400 hand-coded features; 6K cases hand-coded for meaning in 25 legal areas
- Link 145K cases out of 2.5M District Court opinions
Graphical Model

- District Cases →
- District Judge Bio →
- **Circuit Case Appeal** →
- **Circuit Judge Characteristics (z)** →
- **Circuit Case Decision (w)** →
- Precedential Effects (e.g., State Laws) →
- Promulgation (e.g., News) →
- Outcomes (y)

Data: All 380K cases, 1,150K judge votes, 94 topics, from 1870s-
  - 700M tokens, 2B 8-grams, 5M citation edges across cases
  - 250 biographical features (D/R, law school, age)
  - 5% sample, 400 hand-coded features; 6K cases hand-coded for meaning in 25 legal areas
  - Link 145K cases out of 2.5M District Court opinions
Graphical Model

- District Cases →
- District Judge Bio →
- Circuit Case Appeal →
- **Circuit Judge Characteristics** ($z$) →
- **Circuit Case Decision** ($w$) →
- Precedential Effects (e.g., State Laws) →
- Promulgation (e.g., News) →
- Outcomes ($y$)

**Data:** All 380K cases, 1,150K judge votes, 94 topics, from 1870s-
- 700M tokens, 2B 8-grams, 5M citation edges across cases
- 250 biographical features (D/R, law school, age)
- 5% sample, 400 hand-coded features; 6K cases hand-coded for meaning in 25 legal areas
- Link 145K cases out of 2.5M District Court opinions
Graphical Model

- District Cases →
- District Judge Bio →
- Circuit Case Appeal →
- Circuit Judge Characteristics (z) →
- Circuit Case Decision (w) →
- Precedential Effects (e.g., State Laws) →
- Promulgation (e.g., News) →
- Outcomes (y)

Data: All 380K cases, 1,150K judge votes, 94 topics, from 1870s-
- 700M tokens, 2B 8-grams, 5M citation edges across cases
- 250 biographical features (D/R, law school, age)
- 5% sample, 400 hand-coded features; 6K cases hand-coded for meaning in 25 legal areas
- Link 145K cases out of 2.5M District Court opinions
Deep IV  (Hartford et al. PMLR 2017)

- $z$: randomly assigned circuit judge characteristics
- $w$: decisions and opinions (7388 cases)
- $y$: $\Delta$sentencing (3 months before vs. after) in lower district courts
- $x$: covariates (legal topic)
  - demeaned by fixed effects for circuit and year
- $e$: confounders correlated with $w$ & $y$, but not with $z$

Deep OLS $F(y|w)$, Deep Reduced Form $F(y|z)$, Deep 2SLS
Deep IV  (Hartford et al. PMLR 2017)

- **z**: randomly assigned circuit judge characteristics
- **w**: decisions and opinions (7388 cases)
- **y**: $\Delta$sentencing (3 months before vs. after) in lower district courts
- **x**: covariates (legal topic)
  - demeaned by fixed effects for circuit and year
- **e**: confounders correlated with w & y, but not with z

Deep OLS $F(y|w)$, Deep Reduced Form $F(y|z)$, Deep 2SLS
Deep Reduced Form

**Figure**: Predicted vs. Actual Sentence Change
Deep 2SLS

- **First stage** $F(w|z)$: Predict $w$ using $z$
  - Represented text using 2 methods
    1. N-Gram Frequencies (TF-IDF adjusted) with PCA
    2. Document Embeddings
  - Random forest regressor to predict $(w)$

- **Second stage** $G(y|\hat{w})$: Use the predicted $\hat{w}$ to predict $y$
  - Neural Network
First stage $F(w|z)$: Predict $w$ using $z$
- Represented text using 2 methods
  1. N-Gram Frequencies (TF-IDF adjusted) with PCA
  2. Document Embeddings
- Random forest regressor to predict $(w)$

Second stage $G(y|\hat{w})$: Use the predicted $\hat{w}$ to predict $y$
- Neural Network
Deep First Stage

**Figure:** Feature Importance for Affirm or Reverse

Democrats more lenient: reversing lower-court criminal decisions.

- Predict 25-dimensional doc embedding; judges as input.
  - random forest
Deep First Stage

Figure: Feature Importance for Affirm or Reverse

Democrats more lenient: reversing lower-court criminal decisions.

- Predict 25-dimensional doc embedding; judges as input.
  - random forest
Deep 2SLS

- Input into neural network is output of first stage
  - Using the same architecture as DeepOLS, MSE outperforms DeepOLS.
    - Consistent with Deep Reduced Form - judge assignment having signal value
- Reversing a criminal’s appeal is weakly correlated with ↑ sentencing
  - Mechanism:
    - Essentially no effect on deviation from sentencing guidelines
    - Prosecutors may adjust charging decisions in response to appellate ruling
Deep 2SLS

- Input into neural network is output of first stage
  - Using the same architecture as DeepOLS, MSE outperforms DeepOLS.
    - Consistent with Deep Reduced Form - judge assignment having signal value
- Reversing a criminal’s appeal is weakly correlated with ↑ sentencing
  - Mechanism:
    - Essentially no effect on deviation from sentencing guidelines
    - Prosecutors may adjust charging decisions in response to appellate ruling
Deep 2SLS

- Input into neural network is output of first stage
  - Using the same architecture as DeepOLS, MSE outperforms DeepOLS.
    - Consistent with Deep Reduced Form - judge assignment having signal value
- Reversing a criminal’s appeal is weakly correlated with ↑ sentencing
  - Mechanism:
    - Essentially no effect on deviation from sentencing guidelines
    - Prosecutors may adjust charging decisions in response to appellate ruling
Automated **Prospective** Impact Analysis of Legal Precedents?

- **District Cases** →
- **District Judge Bio** ($z_1$) →
- **Circuit Case Appeal** ($w_1$) →
- **Circuit Judge Bio** ($z_2$) →
- **Circuit Case Decision** ($w_2$) →
- **Precedential Effects** ($y_1$) →
- **Promulgation** ($y_2$) →
- **Outcomes** ($y_3$)
  - 1. **$w$:** Identifying nearest cases \textit{Learning Policy Levers, Ash, Chen, Degaldo, Fierro, Lin}
  - 2. **$w$:** Fast decision classification \textit{Automated Fact-Value Distinction, Cao, Ash, Chen}
  - 3. **$w$:** Document embedding \textit{Case Vectors, Ash, Chen}
  - 4. **$z$:** Judge embedding using own corpora
Automated **Prospective** Impact Analysis of Legal Precedents?

- **District Cases** →
- District Judge Bio ($z_1$) →
- Circuit Case Appeal ($w_1$) →
- Circuit Judge Bio ($z_2$) →
- **Circuit Case Decision** ($w_2$) →
- Precedential Effects ($y_1$) →
- Promulgation ($y_2$) →
- Outcomes ($y_3$)

- 1. $w$: Identifying nearest cases *Learning Policy Levers, Ash, Chen, Degaldo, Fierro, Lin*
- 2. $w$: Fast decision classification *Automated Fact-Value Distinction, Cao, Ash, Chen*
- 3. $w$: Document embedding *Case Vectors, Ash, Chen*
- 4. $z$: Judge embedding using own corpora
Automated **Prospective** Impact Analysis of Legal Precedents?

- **District Cases** →
- District Judge Bio \((z_1)\) →
- Circuit Case Appeal \((w_1)\) →
- Circuit Judge Bio \((z_2)\) →
- **Circuit Case Decision** \((w_2)\) →
- Precedential Effects \((y_1)\) →
- Promulagation \((y_2)\) →
- Outcomes \((y_3)\)
  
  1. **w**: Identifying nearest cases *Learning Policy Levers, Ash, Chen, Degaldo, Fierro, Lin*

  2. **w**: Fast decision classification *Automated Fact-Value Distinction, Cao, Ash, Chen*

  3. **w**: Document embedding *Case Vectors, Ash, Chen*

  4. **z**: Judge embedding using own corpora
Automated Prospective Impact Analysis of Legal Precedents?

- **District Cases** →
- **District Judge Bio** \( (z_1) \) →
- **Circuit Case Appeal** \( (w_1) \) →
- **Circuit Judge Bio** \( (z_2) \) →
- **Circuit Case Decision** \( (w_2) \) →
- **Precedential Effects** \( (y_1) \) →
- **Promulgation** \( (y_2) \) →
- **Outcomes** \( (y_3) \)
  1. **w**: Identifying nearest cases *Learning Policy Levers, Ash, Chen, Degaldo, Fierro, Lin*
  2. **w**: Fast decision classification *Automated Fact-Value Distinction, Cao, Ash, Chen*
  3. **w**: Document embedding *Case Vectors, Ash, Chen*
  4. **z**: Judge embedding using own corpora ...
Automated Prospective Impact Analysis of Legal Precedents?

- **District Cases** →
- **District Judge Bio** ($z_1$) →
- **Circuit Case Appeal** ($w_1$) →
- **Circuit Judge Bio** ($z_2$) →
- **Circuit Case Decision** ($w_2$) →
- **Precedential Effects** ($y_1$) →
- **Promulgation** ($y_2$) →
- **Outcomes** ($y_3$)
  - 1. $w$: Identifying nearest cases. *Learning Policy Levers, Ash, Chen, Degaldo, Fierro, Lin*
  - 2. $w$: Fast decision classification. *Automated Fact-Value Distinction, Cao, Ash, Chen*
  - 3. $w$: Document embedding. *Case Vectors, Ash, Chen*
  - 4. $z$: Judge embedding using own corpora...
Automated **Prospective** Impact Analysis of Legal Precedents?

- District Cases →
- District Judge Bio \((z_1)\) →
- Circuit Case Appeal \((w_1)\) →
- Circuit Judge Bio \((z_2)\) →
- **Circuit Case Decision** \((w_2)\) →
- Precedential Effects \((y_1)\) →
- Promulgation \((y_2)\) →
- Outcomes \((y_3)\)
  - 1. \(w\): Identifying nearest cases *Learning Policy Levers, Ash, Chen, Degaldo, Fierro, Lin*
  - 2. \(w\): Fast decision classification *Automated Fact-Value Distinction, Cao, Ash, Chen*
  - 3. \(w\): Document embedding *Case Vectors, Ash, Chen*
  - 4. \(z\): Judge embedding using own corpora
Visual Structure of Case Vectors by Circuit

Figure 1: Centered by Topic-Year, Averaged by Judge, Labeled by Court
Visual Structure of Case Vectors by Decade

Figure 2: Centered by Court-Topic, Averaged by Court-Year, Labeled by Decade
Visual Structure of Case Vectors by Topic

Figure 3: Centered by Judge-Year, Averaged by Topic-Year, Labeled by Topic
Visual Structure of Case Vectors by Birth Cohort

Figure 5: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Judge Birth Cohort
Visual Structure of Case Vectors by Party

Figure 4: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Political Party
Visual Structure of Case Vectors by Law School

Figure 6: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Law School Attended
Lexical slant matters in the judiciary (proxy for IAT using judge’s own corpora)

Lexical Gender Slant by Judge Gender

Heterogeneous Effects by Case Topic

Campaign Finance
Sexual Harassment
Age Discrimination
Takings
Sex Discrimination
Americans with Disabilities Act
Piercing Corporate Veil
Abortion
Capital Punishment
Title VII
Affirmative Action
Federalism
EPA
Contract Clause

.. and we still have a long to do list!
Lexical slant matters in the judiciary (proxy for IAT using judge’s own corpora)

... and we still have a long to do list!