

A comment on Vorada Limjaroenrat's  
**Courtroom Surveillance:**  
Evidence from a State Intervention in the Courts

Thanee Chaiwat  
Faculty of Economics  
Chulalongkorn University

# Main Concept of the Paper

- Hierarchical monitoring shapes judicial behavior and case outcomes.
- The key question is whether it improves consistency or enables political control.
- The 2017 reform required senior review of draft rulings in selected serious and politically sensitive cases.
- The reform increased sentence severity and reduced inter-judge variation.
- Effects were heterogeneous: harsher penalties in narcotics and social-stability cases, but more favorable outcomes for politicians and state officials.

# Some Remarks

- The paper examines how state-led hierarchical monitoring reshapes judicial power under authoritarian politics.
  - The relevant mechanism is hierarchical, not external monitoring.
- Empirically, the design exploits suboffense-by-region legal variation and large-scale case-level data.
- The effects emerge mainly after enforcement in 2019, not immediately after the 2017 reform.
- The key comparison isolates ex ante monitoring from ex post reporting.
- The central rule-of-law finding is asymmetric treatment across case types.
- Sentences became harsher in severe drug and social-stability cases, but more favorable in cases involving politicians and public officials.
- Greater uniformity need not imply greater fairness.

# Comments

# 1. Separate whether the effect comes from the 2017 reform or the 2019 enforcement.

- **Objective:** clarify whether the paper estimates the effect of the **legal rule** or **its actual enforcement**.
  - 2017 = Legal reform / rule change / constitution 2017
    - Certain cases were required to submit draft rulings for senior review.
    - Behavioral effects on judges were limited at first.
  - 2019 = Enforcement shock / effective monitoring / general election
    - Monitoring intensified through extra enforcement, electronic tracking, and upward reporting.
    - The main effects appear stronger after 2019 than after 2017.
- **Methodological concerns**
  - A single treatment may conflate the 2017 reform with the 2019 enforcement phase.
  - The 2019 period may also capture a post-election political shock.

# Result I: Sentence lengths

	Sentence lengths(months)				
	(1)	(2)	(3)	(4)	(5)
$Treated_{gr}^1$	121.387*** (18.515)	129.920*** (22.203)	130.067*** (21.858)	128.927*** (21.949)	127.890*** (21.737)
$Treated_{gr}^1 \times Post_t$	39.172*** (7.597)	37.545*** (7.417)	37.366*** (7.540)	38.566*** (7.360)	
$Treated_{gr}^1 \times Post_t^{2017}$					14.660* (8.801)
$Treated_{gr}^1 \times Post_t^{2019}$					69.555*** (9.817)
Observations	72309	72190	72190	72190	72190
R-square	0.382	0.388	0.389	0.395	0.399
Mean DV.	36	36	36	36	36
Offense fe.	Yes	Yes	Yes	Yes	Yes
Severity bin fe.	Yes	Yes	Yes	Yes	Yes
Quarterly fe.	Yes	Yes	Yes	Yes	Yes
Region fe.	No	No	Yes	Yes	Yes
Region trend	No	No	Yes	Yes	Yes
Judge fe.	No	No	No	Yes	Yes
Control vars.	No	Yes	Yes	Yes	Yes

# 1. Separate whether the effect comes from the 2017 reform or the 2019 enforcement.

- Technical suggestion

- Frame or interpret it as a **two-stage intervention**.

- Stage 1: 2017 legal mandate.
- Stage 2: 2019 enforcement / salience shock.

- Estimate  $Treated \times Post2017$  and  $Treated \times Post2019$  as the main effects.

- The core specification is  $Sentence_{igrt} = \beta_1(Treated_{gr} \times Post2017_t) + \beta_2(Treated_{gr} \times Post2019_t) + controls + FE + \epsilon_{igrt}$
- $\beta_1$  = effect of the 2017 rule change before strict enforcement.
- $\beta_2$  = additional effect after 2019, when monitoring became more binding.

## 2. Make Treated and Control More Comparable

- **Objective:** rule out the concern that the estimated effect reflects treated cases being more serious or politically sensitive, rather than monitoring itself.
- **Main concern**
  - Treated cases are not randomly assigned at the case level.
  - Treatment is determined by charge severity, offense category, and region.
  - Treated and control cases therefore differ systematically from the control.

## 2. Make Treated and Control More Comparable

- Potential problem in the dataset

Dimension	Control cases	Treated cases	Interpretation
processing time	22.3 days	52.2 days	Treated cases take longer to process, suggesting greater complexity.
defendant confess	94.6%	86.6%	Treated cases have a lower share of guilty pleas.
major drugs	10.3%	63.6%	Treated cases are concentrated in serious drug offenses.
minor drugs	68.2%	14.2%	Control cases are concentrated in less serious drug offenses.
political stability	0.1%	6.5%	Treated cases more often involve politically or stability-sensitive charges.

## 2. Make Treated and Control More Comparable

- **Potential problems out of the dataset**
  - Treated and control cases may differ along unobserved dimensions. (Prosecutorial strategy / Political sensitivity / Regional or court-level practices)
- **Technical issues**
  - Treatment variation is defined at the **suboffense-by-region** level.
  - The estimated effect may mix between
    - The impact of monitoring.
    - Differences in offense composition.
    - Regional implementation or local sentencing trends.
- **Suggestions to strengthen the design**
  - Compare only cases close to the 10-/20-year sentencing thresholds.
  - Estimate effects within offense categories.
  - Use matching or reweighting to make treated and control cases more comparable.

### 3. Consider Monitoring as Behavioral Pressure, Not Only Outcome

- **Objective**

- Shift the contribution from “monitoring increases punishment” to “hierarchical monitoring changes judges’ behavioral responses.”

- **Main point**

- The paper models monitoring well through a **principal-agent framework**.
- But in hierarchical courts, monitoring may also operate through **behavioral pressure**.
- Judges may respond not only to the probability of review, but also to **the perceived gaze of senior judges**.

### 3. Consider Monitoring as Behavioral Pressure, Not Only Outcome

- **Possible mechanisms**

- **Salience Effect** senior judges become the relevant audience in judges' decision-making.
- **Loss aversion** judges may fear appearing too lenient.

- **How to test this**

- Estimate heterogeneous effects by seniority, career stage, or years before rotation.
- Test for spillovers to non-treated cases handled by the same judge.
- If non-treated cases also become harsher, the reform may have shifted perceived norms, not just case-specific incentives.

## 4. Move to Defensive Judging and Legalistic Reasoning

- **Objective**
  - Distinguish whether the reform makes judges “**decide better**” or primarily makes them “**decide more defensively.**”
- **Main point**
  - The fact that lenient judges become harsher need not mean they “learn” from senior judges.
  - It may instead reflect **defensive judging.**
  - Judges may choose safe sentences and write opinions defensively under monitoring.
- **Possible mechanisms**
  - Judges may pick sentences that look “tough enough” to avoid appearing lenient.
  - Judgments may become longer, more formal, and more heavily referenced to statute.
  - The language may stress seriousness, social order, deterrence, or public order more strongly.
  - Increased uniformity may come from defensive standardization rather than improved justice.

## 4. Move to Defensive Judging and Legalistic Reasoning

- How to test this
  - (1) Examine **bunching in sentences** after the reform.
    - Check whether sentences cluster at round numbers or “safe sentencing points” (for example, 60 months, 120 months, near the upper bound, or a high share at the maximum).
  - (2) Use **text analysis of judgment texts**.
    - Track judgment length (word counts) and the number of statutory citations.
    - Test whether the language of judgments becomes more similar across cases and differs from the pre-monitoring period.

## 5. Shift focus from heterogeneity to the contribution of “redistribution of bias”

- **Objective**

- Recast the paper from “**monitoring increases punishment**” to “**monitoring reshapes the allocation of justice.**”

- **Main points**

- On average, monitoring raises sentence severity.
- Heterogeneous effects are more revealing, monitoring is not distributionally neutral.
  - Severe drug and social-stability cases receive harsher punishment.
  - Cases involving politicians and public officials receive more favorable treatment.
  - Effects are stronger in autocrat-aligned areas.

## 5. Shift focus from heterogeneity to the contribution of “redistribution (/ reshape / reallocation) of bias”

- **Political economy interpretation**
  - The reform does not **reduce** bias in the justice system.
  - Instead, it may **reallocate / reshape** bias along political priorities.
  - Move from an average “monitoring effect” toward **selective enforcement / selective leniency**.
- **Methodological concern**
  - Aggregate parallel trends pass the test, but main effects are driven by specific offense groups.
  - Pre-trends should therefore be checked separately by offense category.
  - Treated and control cases within each offense group may have different pre-reform trends.

## 5. Shift focus from heterogeneity to the contribution of “redistribution of bias”

- How to test this

- (1) Offense-specific event studies

- Test whether each offense category exhibits similar pre-trends.
    - Run separate event studies for: severe drugs, social/political stability, politician/public officer, and minor drugs.

- (2) Triple-difference model (DDD)

- Include direct interactions between treatment and offense type, for example:

$$\textit{Sentence} = \beta_1 \textit{Treated} \times \textit{Post} \times \textit{SeriousDrugs} + \beta_2 \textit{Treated} \times \textit{Post} \times \textit{Officer} + \beta_3 \textit{Treated} \times \textit{Post} \times \textit{Stability} + FE + \textit{controls}$$

- (3) Leave-one-court-out checks

- Re-estimate the results excluding one court at a time.
    - Drop Court 1 and re-run, then Court 2, and so on for all courts.
      - If severe-drug effects remain positive and officer/politician effects remain negative in all specs, results are not driven by a single court.
      - If excluding one court makes the effects disappear, the pattern may reflect a local court effect rather than system-wide selective justice.

## 6. Add Dimensions and Philosophical Arguments in Law and Economics

- Justice can be understood as a service that delivers **credence goods** rather than fully observable outcomes. **Trust** is a key.
- In that sense, justice is shaped partly (mostly) by **perception**, not purely by **objective reality**.
- May **judicial discretion** either improve fairness through contextual judgment or weaken fairness by reducing consistency across decision makers?
  - Or is the relevant standard simply “justice”?
- Monitoring can function as a **second opinion**: it may reduce perceived judicial independence while increasing or decreasing **perceived legitimacy**?